Observer Variability in Metameric Color Matches using Color Reproduction Media

Richard L. Alfvin,* Mark D. Fairchild

Munsell Color Science Laboratory, Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, New York 14623-5604

Received 27 January 1996; accepted 16 November 1996

Abstract: Standard color-matching functions are designed to represent the mean color-matching response of the population of human observers with normal color vision. When using these functions, two questions arise. Are they an accurate representation of the population? And what is the uncertainty in color-match predictions? To address these questions in the dual context of human visual performance and cross-media reproduction, a color-matching experiment was undertaken in which twenty observers made matches between seven different colors presented in reflective and transmissive color reproduction media and a CRT display viewed through an optical apparatus that produced a simple split-field stimulus. In addition, a single observer repeated the experiment 20 times to estimate intra-observer variability. The results were used to evaluate the accuracy of three sets of color-matching functions, to quantify the magnitude of observer variability, and to compare intra- and inter-observer variability in color-matching. These results are compared with various techniques designed to predict the range of color mismatches. The magnitude of observer variability in this experiment also provides a quantitative estimate of the limit of cross-media color reproduction accuracy that need not be exceeded. On average, the differences between matches made by two different observers was approximately 2.5 CIELAB units. © *1997 John Wiley & Sons, Inc. Col Res Appl, 22, 174–188, 1997*

Key words: observer variability; observer metamerism; color-matching functions; color reproduction

INTRODUCTION

The primary objective in color reproduction systems is often to achieve faithful color reproduction, i.e., color matching. In colorimetric terms, color matches are defined by multiple stimuli with identical tristimulus values. Since tristimulus values are a function of the interaction between a physical stimulus and the human visual system, it is necessary to consider two distinctly different types of color-matches, spectral and metameric. Spectral color matches, identified by physical stimuli with identical spectral radiant power distributions, are perceived as color-matches by all observers. Spectral color-matches in color reproduction are only possible when a single medium is used for both the original and the reproduction. Metamers or metameric color stimuli are spectrally different color stimuli that have the same tristimulus values.[1] Metameric color matches are perceived as color-matches for a given observer.

Due to differences in the spectral characteristics of the primary colorants used in cross-media color reproduction, all cross-media color-matches are metameric. Since metameric matches are dependent upon the characteristic spectral responsivities of the human visual system, variations in color-matching functions will result in variations in metameric matches. A metameric color-match perceived by one observer may appear to be a significant mismatch to another observer. This phenomenon is caused by observer metamerism.

Pobboravsky[2] studied the effect of observer metamerism on color-matches between CRT displays and printed matter. He performed a set of theoretical calculations based on the color-matching data used in the CIE recom-

* Correspondence to: R. L. Alfvin, Eastman Kodak Company, Building 65, Rochester, NY 14650-1829

Contract grant sponsor: NSF-NYS/IURC

Contract grant sponsor: NYSSTF-CAT Center for Electronic Imaging Systems

Contract grant sponsor: Munsell Color Science Laboratory

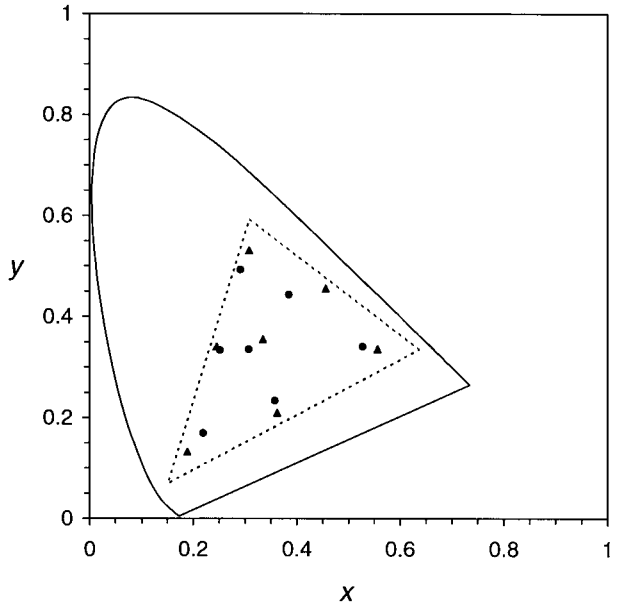

FIG. 1. CIE x, y chromaticity diagram (1931 CIE 2° Standard Colorimetric Observer, D50 fluorescent daylight simulator) showing the coordinates of the fourteen hard-copy stimuli with respect to the chromaticity gamut of the CRT display and the spectrum locus. Color prints (open circles); color transparencies (solid triangles); CRT chromaticity gamut (dashed triangular outline); and spectrum locus (solid line).

mended technique for assessing observer metamerism.[3] His conclusion was that observer metamerism was not a significant problem in these types of matches. He verified this conclusion with qualitative visual judgments of acceptability. These data, while useful, might well underestimate the importance of observer metamerism and they lack the quantitative nature required for application to digitally based cross-media imaging systems. Thus, the research described in this article was undertaken to better quantify observer variability for such matches.

The following section describes an experiment designed to explore metameric color-matching of color reproduction media by quantifying the precision and accuracy of three sets of color-matching functions and the magnitude of observer variability found in hard-copy to CRT color matches. The effects for both intra- and inter-observer color matches are quantified and compared. In addition, the results are compared with the CIE recommendations on observer metamerism[3] and uncertainties derived by Nimeroff, Rosenblatt, and Dannemiller[4] for the 1964 CIE 10° Supplemental Standard Colorimetric Observer color-matching data.

Although the effects of observer variability are present in any color reproduction system, color prints and transparencies were selected for use in this research as typical media commonly used in CRT to hard-copy imaging systems. Observer variability in cross-media color-matching is composed of multiple variance components related to spectral differences, image characteristics, and viewing conditions. In order to assess better the observer metamerism related to spectral differences in this application, the experimental cross-media stimuli were designed to minimize nonspectral related components of observer variability associated with the task of cross-media color-matching. The use of simple homogeneous stimuli, devoid of texture and other spatial characteristics, viewed in contiguous fields eliminates the potential for experimental error associated with cross-media color-matching.

EXPERIMENTAL METHOD

A visual experiment was designed to permit observers to make critical color-matches between color prints or transparencies and a CRT display. Seven color prints and seven color transparencies were prepared as fixed, solid-color, matching stimuli. The seven colors included red, green, blue, gray, cyan, magenta, and yellow. The color print samples were produced with a Fujix Pictrography 3000 color printer. The Fujix printer is a hybrid photographic/thermal-transfer continuous-tone digital printer. The color transparencies were imaged with an MGI Solitaire 8_{xp} film recorder using 4×5 Ektachrome 100 Plus Professional film. The chromaticities of the fixed hard-copy samples illuminated with a fluorescent D50 simulator were designed to sample effectively the color gamut of the Sony Trinitron CRT display model PVM-1942Q used to generate the soft-copy color matches. The CIE x, y chromaticity coordinates of the hard-copy stimuli and the chromaticity gamut of the CRT display are presented in Fig. 1.

With the use of a simple optical apparatus, consisting of an equilateral glass prism mounted on an optical bench, the observers were able to view simultaneously both the soft- and hard-copy matching stimuli. As shown in Fig. 2, a CRT display and a combination lightbooth/light box

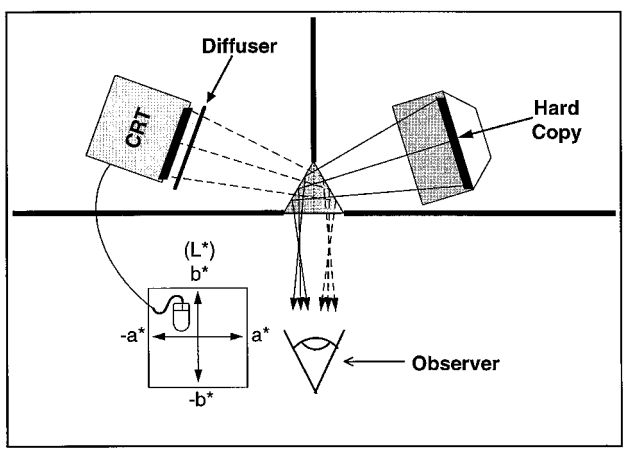

FIG. 2. Schematic overhead view of experimental set-up used for color-matching.

were aligned with the optical prism and shielded from the observer. A diffuser was placed in front of the CRT display to eliminate the appearance of scan lines. The fixed hard-copy stimulus and the adjustable soft-copy stimulus were presented in a vertical symmetric bipartite field. The color-matching stimuli were presented as solid colors appearing self-luminous in a completely darkened surround. It is important to note that the observers were completely unaware of the origins of the stimuli—they appeared as self-luminous unrelated colors that could have been produced using any sort of traditional color-matching apparatus.

The hard-copy stimuli were illuminated by a GTI Soft View D50 fluorescent simulator light booth designed for viewing both reflective and transmissive materials. The light booth was adjusted to yield equal luminance from the reflective and transmissive gray hard-copy stimuli. The absolute luminance of the matching stimuli was about 50 cd/m^2 for all colors. The color appearance of the soft-copy stimulus produced with the CRT display was adjusted by the observer to match the color appearance of the fixed hard-copy stimulus. With the use of a computer mouse, the observer was able to adjust independently the color-appearance attributes of the soft-copy image along a CIE L^* vector, and an $a^* - b^*$ plane. The CRT display was controlled with a Pixar II image computer with 10-bits per RGB color channel resolution (i.e., 1024 levels of R, G, and B luminance). Thus, the color resolution of the display system exceeded human visual color discrimination capabilities. The visual experiment was no different than the classical color-matching experiments that are the basis of modern colorimetry with the one exception that the particular metamers used are also of practical importance.

Twelve male and eight female observers between the ages of 21 and 56 participated in the color-matching experiment to assess inter-observer variability. The ages of the observers were uniformly distributed; there were five observers in each of four groups, 20–29, 30–39, 40–49, and 50–59 years of age. Each of the twenty observers successfully passed a screening for congenital color vision deficiencies when tested with a set of Ishihara Pseudoisochromatic Plates.[5] The Farnsworth–Munsell 100-Hue Test[6] was administered to each of the observers to assess their color discrimination abilities. Seventy-five percent of the twenty observers were rated as having ''superior'' color discrimination, while the remaining 25% achieved an ''average'' color discrimination rating.

A 22-year-old male observer performed the color-matching experiment twenty times to assess intra-observer variability. The single observer tested color normal with the set of Ishihara Pseudoisochromatic Plates, and received a ''superior'' color discrimination rating with the Farnsworth–Munsell 100-Hue Test.

The color-matching experiments performed by each of the observers were divided into two 1-hour sessions consisting of seven color matches. The observers participated in only one session per day. Each of the observers completed their two sessions within a one-week time period. All of the experiments were performed within a 60-day time period.

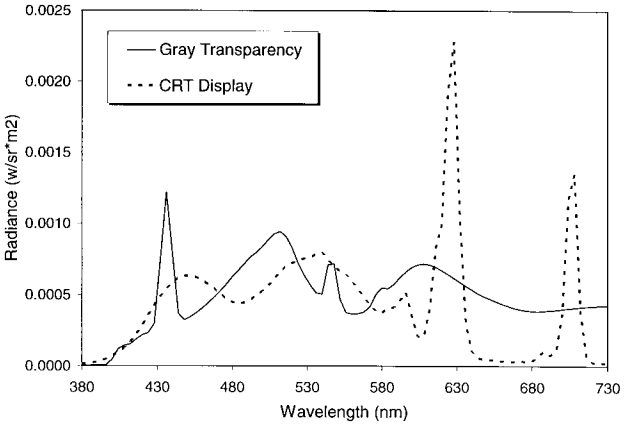

FIG. 3. The spectral power distributions of a gray transparency and a CRT display adjusted to yield a metameric match for the 1931 CIE 2° Standard Colorimetric Observer.

The observers were seated 1 meter away from the matching stimuli. The 5 cm × 5 cm matching field subtended a visual angle of 2.9°. The observers were asked to adjust the color appearance of the soft-copy stimulus to create an exact color-match for each of the fourteen different hard-copy stimuli. A color match was achieved when the two matching stimuli appeared as a single homogeneous stimulus. After a match was attained, a Photo Research PR-650 telespectroradiometer was used to measure the spectral radiance ($W/sr*m^2$) of both the hard- and soft-copy stimuli from the observers point of view. While the optical apparatus was designed to eliminate all visible stray light in the matching fields, this was not critical, because the visually matching stimuli were measured immediately following each observer's setting of the match point. The order of color matches was randomized for each observer. Figure 3 shows an example of the disparate spectral power distributions required to produce a metameric match between the gray transparency sample and the CRT display, calculated for the 1931 CIE 2° Standard Colorimetric Observer.

The precision and accuracy of the Photo Research PR-650 telespectroradiometer used to record experimental data for this research were examined and found to be within acceptable limits as defined by the manufacturer. The photometric accuracy of the telespectroradiometer was tested with a standardized tungsten source. The recorded luminance was within 2.1% of the expected value, well within the 4% variation of the NIST standard used for calibration. The spectral wavelength accuracy of the instrument was satisfactorily tested by correctly measuring the expected spectral peaks of a mercury-cadmium source. The precision of the telespectroradiometer was examined by making twenty repeated spectral radiance

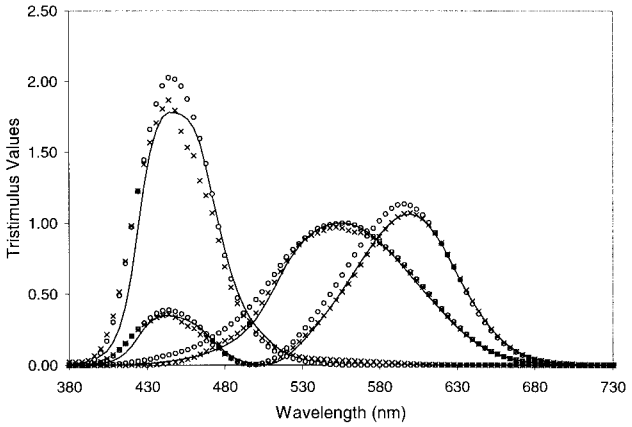

FIG. 4. Color-matching functions for Stiles–Burch 2° mean observer (**x**) transformed to approximate the CIE 2° Standard Colorimetric Observer (——), and CIE 10° Supplemental Standard Colorimetric Observer (○).

measurements without replacement of the gray print and gray transparency hard-copy samples. The mean color difference from the mean was approximately 0.07 ΔE^*_{ab} for both of the hard-copy samples. The standard deviation was less than 0.10 CIELAB units in each CIELAB vector direction for both samples. The results of the precision and accuracy tests indicate that the systematic and random error associated with the instrument are minimal and acceptable for the purpose of this research.

RESULTS AND DISCUSSION

Three different sets of color-matching functions were used in conjunction with the spectral radiant power distributions, recorded from both the hard- and soft-copy stimuli, to calculate tristimulus values for each of the observer color-matches. Figure 4 shows the three sets of color-matching functions, the 1931 CIE 2° Standard Colorimetric Observer, the 1964 CIE 10° Supplemental Standard Colorimetric Observer, and the 1955 Stiles–Burch 10° mean observer, used for tristimulus calculations. The 1955 Stiles–Burch 10° $\bar{r}_{sb}$, $\bar{g}_{sb}$, and $\bar{b}_{sb}$ color-matching functions were transformed to CIE 2°-based $\bar{x}_{sb}$, $\bar{y}_{sb}$, and $\bar{z}_{sb}$ color-matching functions by a linear combination of the 1955 Stiles–Burch 10° $\bar{r}_{sb}$, $\bar{g}_{sb}$, and $\bar{b}_{sb}$ color-matching functions.[7]

$$\begin{bmatrix} \bar{x}_{sb}(\lambda) \\ \bar{y}_{sb}(\lambda) \\ \bar{z}_{sb}(\lambda) \end{bmatrix} = \begin{bmatrix} 0.35811 & 0.14600 & 0.35820 \\ 0.13993 & 0.82873 & 0.01423 \\ 0 & 0.06250 & 1.87194 \end{bmatrix} \begin{bmatrix} \bar{r}_{sb}(\lambda) \\ \bar{g}_{sb}(\lambda) \\ \bar{b}_{sb}(\lambda) \end{bmatrix}. \quad (1)$$

The transformation matrix was determined from a best fit to the 1931 CIE 2° Standard Colorimetric Observer functions by least-squares multiple linear regression. This transformation was performed such that the data analyses could be completed in the familiar and approximately visually uniform CIELAB color space.

Reference-white tristimulus values for both hard- and soft-copy stimuli were calculated for the D50 fluorescent simulator using each of the three sets of color-matching functions. To use the CIELAB color space most effectively, the reference-white *Yn* value was adjusted so that L^* value for the gray stimuli would be approximately equal to 50. The selection of *Yn* values will have a scaling effect on CIELAB L^*, color-difference, and variance metrics. Intuitively, a reduction in the *Yn* value will cause an increase, while an increase in the *Yn* value will result in an apparent decrease in the value of these metrics. Since this was a single stimuli in a dark surround experiment, it would be expected that the L^* values of the achromatic stimuli would be relatively large due to the effects of adaptation. Therefore, selecting a *Yn* value for the reference white stimulus that results in an L^* value of 50 for the gray stimuli is, in effect, a very conservative choice, causing the experimental data to reflect color-difference and variance values smaller than expected. CIELAB coordinates were calculated according to CIE methods,[8] using each of the three sets of color-matching functions, for both hard- and soft-copy stimuli constituting observer-determined metameric pairs. CIE ΔL^*, Δa^*, and Δb^* values were calculated for all metameric pairs.

For all sample colors, the maximum CIELAB unit differences were as large as 19.7 units for inter-observer matches, and 11.4 units for intra-observer matches, indicating a relatively large spread of color-matches. As shown in Table I, the mean CIELAB deviations ranged from −0.81–0.67 units for inter-observer matches, and between −1.63–0.78 units for intra-observer matches. The relatively small mean CIELAB deviations for the inter-observer data indicate that the average of all experimental color-matches correlates well with the theoretical color-matches of the standard observers.

As a measure of intra- and inter-observer variability, the mean color difference from the mean color match (MCDM) was calculated using each of the three sets of color-matching functions. The CIE ΔE^*_{ab} color-difference equation was used to compute color differences between matching stimuli. Table II shows the MCDM[9] statistics tabulated according to observers, media, and color-matching functions. The MCDMs for inter-observer matches were approximately twice as large as the MCDMs for intra-observer matches. The three sets of color-matching functions used to calculate MCDM values yielded similar results, indicating a lack of dependence on color-matching function.

The sample covariance matrix, $\mathbf{S}_{Lab}$, defined by the sample covariances and variances of ΔL^*, Δa^*, and Δb^* values was calculated for each combination of observer,

TABLE I. Minimum, maximum, and mean CIELAB unit deviations of color matches tabulated by observer, and color-matching function.

	Intra-observer			Inter-observers		
	CIE 2°	CIE 10°	Stiles–Burch 2°	CIE 2°	CIE 10°	Stiles–Burch 2°
ΔL^* minimum	−4.14	−3.99	−4.12	−14.08	−14.62	−14.25
ΔL^* maximum	4.41	4.73	4.53	13.70	13.91	13.75
ΔL^* mean	0.18	0.17	0.13	0.11	0.07	0.06
Δa^* minimum	−6.98	−8.76	−7.15	−15.71	−9.94	−14.35
Δa^* maximum	6.97	8.70	7.61	9.76	10.75	10.32
Δa^* mean	0.36	0.78	0.60	0.15	0.67	0.37
Δb^* minimum	−11.23	−11.36	−10.96	−19.69	−19.50	−18.89
Δb^* maximum	8.29	9.12	8.19	18.51	16.49	18.00
Δb^* mean	−1.55	−1.63	−1.09	−0.70	−0.81	−0.22

color-matching functions, color, and medium, color-matching data. These data are presented in Table III as a summary of the experimental results. Assuming a multivariate normal distribution, the inverse of the sample covariance matrix can be used to construct a 95% confidence region for the sample distribution[10] of the CIE ΔL^*, Δa^*, and Δb^* multivariate dataset. Such an ellipsoidal confidence region can be thought of as enclosing 95% of the observer responses for this stimulus.

A multidimensional confidence region for the mean of a sample distribution with a given probability can also be described by including the reciprocal of the sample size. As the number of samples increases, the uncertainty with which the mean is known decreases and the confidence region for the mean shrinks.

The 95% confidence regions for the sample distribution and sample mean of the inter-observer cyan-transparency color-matches calculated with the 1931 CIE 2° Standard Colorimetric Observer are plotted together in Fig. 6(a). The larger of the two ellipses encompasses the distribution of the sample population of color matches, while the smaller ellipse describes the uncertainty of the mean color match of the sample population.

In addition to the $\Delta a^* - \Delta b^*$ relationship examined in Fig. 6(a), it is important to consider the $\Delta a^* - \Delta L^*$, and $\Delta b^* - \Delta L^*$ planes, because the experimental color-matches involved adjustments to each of the three independent variables defining the CIELAB color space. Figures 6(b) and 6(c) show the 95% confidence regions for the sample distribution and sample means of inter-observer cyan-transparency color-matches calculated with the 1931 CIE 2° Standard Colorimetric Observer for the $\Delta a^* - \Delta L^*$, and $\Delta b^* - \Delta L^*$ planes, respectively.

It is apparent from Figs. 6(a)–(c), that, at a 95% confidence level, the mean color-match of the twenty observers for the cyan transparency is not significantly different than the predicted color-match for the 1931 CIE 2° Standard Colorimetric Observer. If any one of the three confidence regions defining the sample means of the $\Delta a^* - \Delta b^*$, $\Delta a^* - \Delta L^*$, or $\Delta b^* - \Delta L^*$ planes do not contain the theoretical mean match for a given standard observer, the mean color-matches are statistically significantly different. As shown in Fig. 7, the intra-observer mean color-matches for the cyan transparency are significantly different than the predicted color-match for the 1931 CIE 2° Standard Colorimetric Observer for all three CIELAB planes. It can be concluded that the 1931 CIE 2° Standard Colorimetric Observer match point for the cyan transparency is not included or represented in the population of color-matches made by the single observer.

The variability of different multidimensional populations can be compared with a multivariate analog to Bartlett's test for homogeneity of variance in a univariate population, called Bishop's test.[11] Bishop's test was used to compare the sample covariances of the three sets of color-matching functions for color-matches made by both intra- and inter-observers given the following hypothesis:

$$H_0: \mathbf{S}_{\text{CIE 2°}} = \mathbf{S}_{\text{CIE 10°}} = \mathbf{S}_{\text{Stiles-Burch 2°}} = \mathbf{S} \qquad (2)$$

$$H_1: \mathbf{S}_i \neq \mathbf{S} \text{ for at least one } i. \qquad (3)$$

There was no significant difference between the sample covariance matrices of the three sets of color-matching functions for all fourteen of the color-matches made by the group of twenty observers. In the case of the single

TABLE II. Mean color differences from the mean (MCDM) for all color matches tabulated by media, color-matching functions, and observers.

Prints	Inter-observers	Intra-observers	SDO
CIE 2°	3.03	1.15	0.70
CIE 10°	2.79	1.10	0.67
Stiles–Burch 2°	2.98	1.11	0.72
Transparencies			
CIE 2°	2.31	1.56	0.24
CIE 10°	2.45	1.53	0.22
Stiles–Burch 2°	2.45	1.59	0.24
Prints & Transparencies			
CIE 2°	2.67	1.35	0.47
CIE 10°	2.62	1.31	0.45
Stiles–Burch 2°	2.72	1.35	0.48

TABLE III. Experimentally established intra- and inter-observer color-matching covariance data for each of the fourteen hard-copy samples.

Intra-observer color-matching data

	Variance			Covariance		
	(ΔL^*)	(Δa^*)	(Δb^*)	$(\Delta L^*, \Delta a^*)$	$(\Delta L^*, \Delta b^*)$	$(\Delta a^*, \Delta b^*)$
Prints						
Red	3.726	3.695	7.437	3.195	4.049	3.542
Green	1.310	2.520	2.532	−0.853	1.018	−1.788
Blue	0.676	1.835	3.321	0.213	−0.730	−2.313
Gray	0.964	1.089	3.307	−0.217	−0.082	−1.498
Cyan	1.066	1.573	3.717	−0.211	0.047	−1.894
Magenta	2.684	1.300	3.958	1.202	1.157	−0.241
Yellow	2.480	1.171	6.711	−0.891	3.088	−1.677
Transparencies						
Red	1.634	2.940	10.247	1.904	2.855	3.784
Green	1.140	4.061	4.947	−1.580	1.209	−3.072
Blue	0.366	6.783	9.386	0.890	−1.192	−7.856
Gray	0.719	0.856	2.460	−0.198	0.700	−1.206
Cyan	1.543	1.384	1.280	−0.547	−0.363	−0.549
Magenta	1.842	4.656	8.924	1.965	−1.725	−5.587
Yellow	2.856	1.114	9.447	0.119	3.793	−0.567

Inter-observer color-matching data

	Variance			Covariance		
	(ΔL^*)	(Δa^*)	(Δb^*)	$(\Delta L^*, \Delta a^*)$	$(\Delta L^*, \Delta b^*)$	$(\Delta a^*, \Delta b^*)$
Prints						
Red	5.271	9.269	8.154	5.610	3.004	3.709
Green	8.268	12.471	17.712	−3.640	9.439	−1.163
Blue	12.983	11.152	17.595	−0.086	−2.961	−12.844
Gray	4.356	2.197	22.569	0.400	5.777	−0.652
Cyan	12.507	2.800	8.970	−1.165	0.544	−2.466
Magenta	4.573	11.637	9.120	3.134	0.842	−5.556
Yellow	13.120	5.050	19.472	−3.234	12.872	−1.570
Transparencies						
Red	8.583	20.082	17.443	11.857	10.680	15.456
Green	18.876	11.562	7.412	−8.791	4.191	−4.773
Blue	6.988	25.587	33.763	0.952	−7.342	−26.274
Gray	6.592	3.024	8.759	−0.361	−2.562	−3.363
Cyan	3.367	9.241	6.185	−1.328	1.633	−2.346
Magenta	16.026	12.680	11.423	11.560	−4.619	−7.674
Yellow	5.801	1.224	19.037	1.372	7.184	1.319

observer, a significant difference existed only between the sample covariance matrices of the three sets of color-matching functions for the color-matches of the blue transparency.

Bishop's test was also used to compare the sample covariances of intra- and inter-observer color-matches for each combination of color, medium, and color-matching functions, given the following hypothesis:

$$H_0: \mathbf{S}_{\text{Intra}} = \mathbf{S}_{\text{Inter}} \quad (4)$$

$$H_1: \mathbf{S}_{\text{Intra}} \neq \mathbf{S}_{\text{Inter}}. \quad (5)$$

A significant difference existed between all of the sample covariance matrices of the intra- and inter-observer color matches for each combination of color, medium, and color-matching functions, except for the yellow transparency matches.

The accuracy of the three sets of color-matching functions for predicting the mean color-matches produced by the intra- and inter-observers was examined with a multivariate test for means. Hotelling's T^2 test for means was used to compare the experimental mean color-matches with the theoretical color-matches calculated for each of the three sets of color-matching functions.

Table IV shows the results of Hotelling's T^2 test for means comparing the mean color-matches made by the group of twenty observers with the hypothetical means predicted using each of the three sets of color-matching functions. The checkmarks ($\surd$) indicate that the mean color-matches are not significantly different from the specified standard observer at the $\alpha = 0.05$ level. The Stiles–Burch 2° color-matching functions accurately predicted nine of the fourteen inter-observer mean color

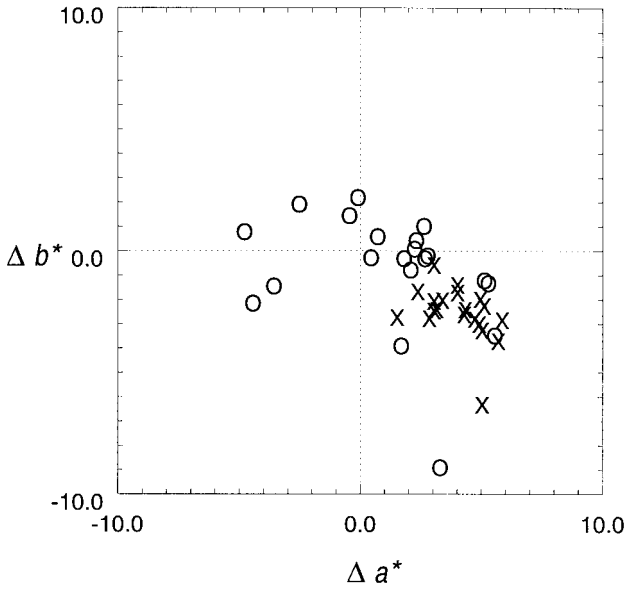

FIG. 5. Intra-observer (X), and inter-observer (○), cyan-transparency color matches relative to the 1931 CIE 2° Standard Colorimetric Observer matchpoint located at the origin of a CIE $\Delta a^* - \Delta b^*$ plane.

matches; the CIE 2° Standard Colorimetric Observer color-matching functions accurately predicted six of the fourteen inter-observer mean color-matches; and the CIE 10° color-matching functions accurately predicted only one of the fourteen inter-observer mean color-matches. The Stiles–Burch and CIE 2° color-matching functions outperformed the CIE 10° Supplemental Standard Colorimetric Observer color-matching functions in terms of predicting the mean color-match for a population of color normal observers. The relatively poor performance of the CIE 10° Supplemental Standard Colorimetric Observer color-matching function is not unexpected, considering the fact that the experimental color-matching stimulus was restricted to a 2° visual field. Although the standard observer color-matching functions were not able to predict the mean color-matches of the group of twenty observers for every sample, it is noteworthy that in every case the predicted color-matches were contained in the ellipsoids that defined the 95% confidence regions at the $\alpha = 0.05$ level for the sample distributions of inter-observer color-matches, as shown in Table V. Therefore, it can be said that the hypothetical color-matches determined with the standard observers are representative of a member of the population of inter-observer color-matches determined in this experiment.

Table VI shows the results of Hotelling's T^2 test for means comparing the mean color-matches made by the single observer with the hypothetical mean color-matches predicted with the three sets of color-matching functions. The checkmarks (✓) indicate that the mean color-matches are not significantly different from the specified standard observer at the $\alpha = 0.05$ level. The means were found to

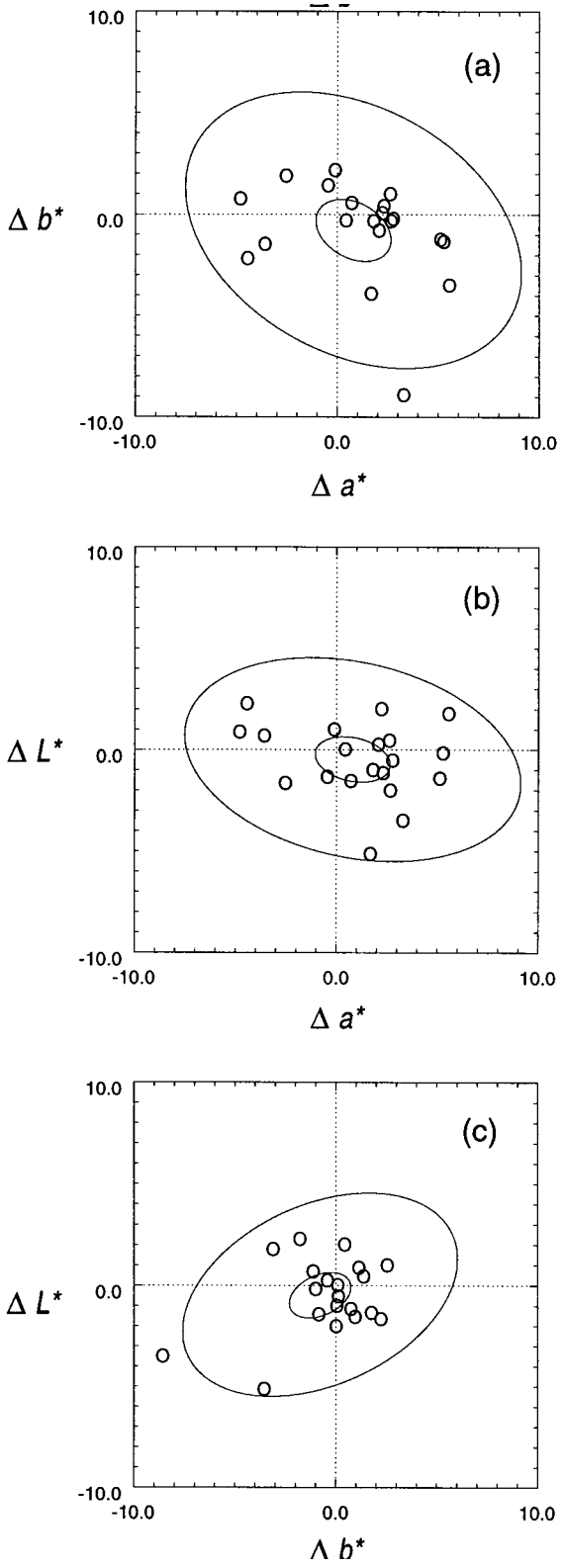

FIG. 6. Twenty inter-observer (○) cyan-transparency color matches bound by; (a) both 95% $\Delta a^* - \Delta b^*$ bivariate confidence ellipses of the sample distribution (outer ellipse), and the sample mean (inner ellipse). (b) 95% $\Delta a^* - \Delta L^*$ bivariate confidence ellipses of the sample distribution, and the sample mean; (c) 95% $\Delta b^* - \Delta L^*$ bivariate confidence ellipses of the sample distribution, and the sample mean. The 1931 CIE 2° Standard Colorimetric Observer matchpoint is located at the origin.

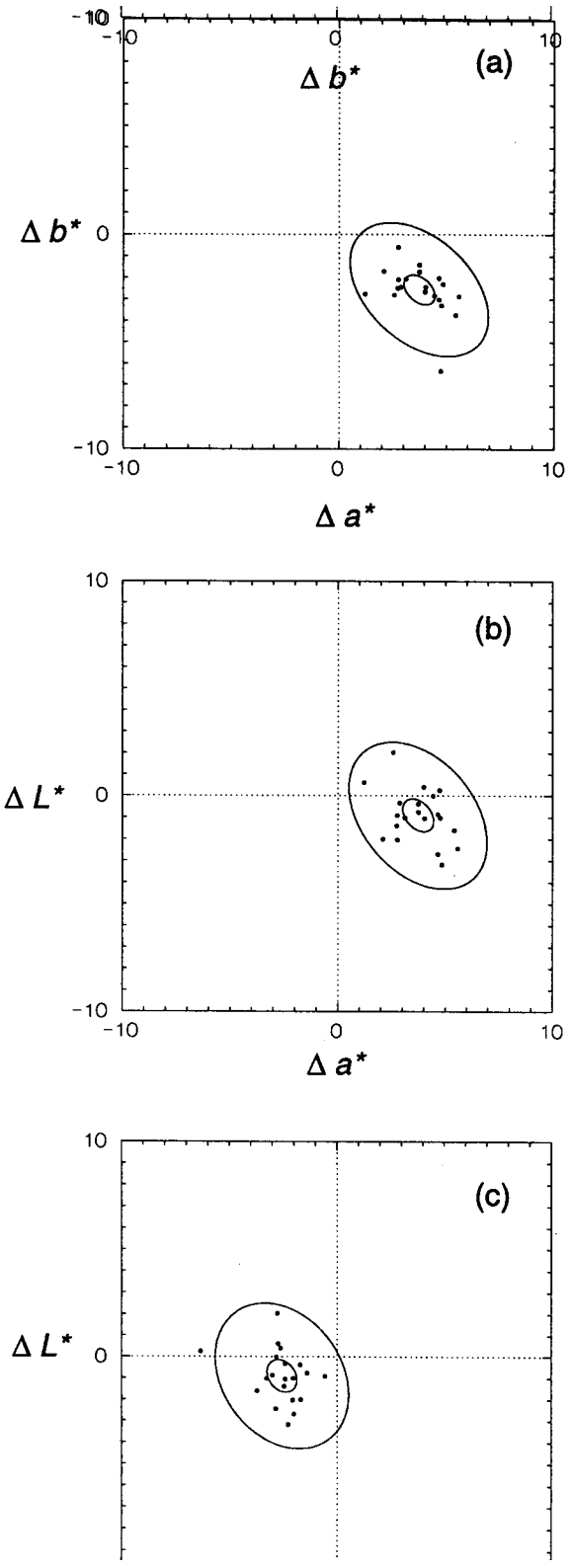

FIG. 7. Twenty intra-observer (○) cyan-transparency color matches bound by; (a) both 95% $\Delta a^* - \Delta b^*$ bivariate confidence ellipses of the sample distribution (outer ellipse), and the sample mean (inner ellipse). (b) 95% $\Delta a^* - \Delta L^*$ bivariate confidence ellipses of the sample distribution, and the sample mean; (c) 95% $\Delta b^* - \Delta L^*$ bivariate confidence ellipses of the sample distribution, and the sample mean. The 1931 CIE 2° Standard Colorimetric Observer matchpoint is located at the origin.

be significantly different for all color-matches except the green transparency and the 2° observers. As shown in Table VII, the checkmarks (✓) indicate that the predicted mean color-matches were contained in the ellipsoids that defined the 95% confidence regions at the $\alpha = 0.05$ level for the sample distributions of intra-observer color-matches in only 26 out of 42 cases. It is apparent from these results that the single observer is not characteristic of the standard observers examined in this experiment. This result is not unusual, as the standard observers are represented by sets of mean color-matching functions determined from older populations of observers, and it is not unreasonable to expect any single observer to deviate significantly from any given average or standard observer.

CIE STANDARD DEVIATE OBSERVER

The CIE Standard Deviate Observer, developed by CIE Committee TC 1-07,[3,12] was designed to represent the individual variations among color-normal observers. The Standard Deviate Observer was derived from singular value decomposition using 20 of Stiles' 49 observers (10°) from 20–60 years of age. The decomposition method gave four sets of deviation functions. The four deviation functions are used in combination to describe a range of color mismatch. The first set of deviation functions, shown in Fig. 8, in conjunction with a CIE Standard Colorimetric Observer (either 1931 or 1964), is used to define the color-matching functions of the standard deviate observer:

$$\begin{aligned} \bar{x}_{dev}(\lambda) &= \bar{x}(\lambda) + \Delta\bar{x}_l(\lambda) \\ \bar{y}_{dev}(\lambda) &= \bar{y}(\lambda) + \Delta\bar{y}_l(\lambda), \\ \bar{z}_{dev}(\lambda) &= \bar{z}(\lambda) + \Delta\bar{z}_l(\lambda), \end{aligned} \tag{6}$$

where $\bar{x}_{dev}(\lambda)$ is a Standard Deviate Observer color-matching function, $\Delta\bar{x}_l(\lambda)$ is a first deviation function, and $\bar{x}(\lambda)$ is a CIE Standard Colorimetric Observer color-matching function. The set of Standard Deviate Observer color-matching functions based on the 1931 CIE 2° Standard Colorimetric Observer is shown in Fig. 9. The individual color-matching functions, measured by Stiles–Burch and referenced by the CIE Pub. No. 80, were derived through a normalization process; whereby the intensities of the primary stimuli were divided by the intensity, measured in energy units of the spectral test stimulus for each experimental color-match.[13] In addition, Stiles normalized the color-matching functions to a value of 1.0 at each of the primary wavelengths.[13] The estimation of

TABLE IV. The results of Hotelling's T^2 test for means comparing the mean color matches made by the group of twenty observers with the hypothetical means predicted with three sets of color-matching functions. A (✓) indicates that the mean color matches are not significantly different from the specified standard observer at the $\alpha = 0.05$ level.

	Transparencies			Prints		
Colors	Stiles–Burch 2°	CIE 2°	CIE 10°	Stiles–Burch 2°	CIE 2°	CIE 10°
Red				✓		
Green	✓	✓		✓	✓	✓
Blue				✓		
Gray		✓		✓		
Cyan		✓		✓	✓	
Magenta	✓	✓		✓		
Yellow	✓					

inter-observer variability is fundamentally influenced by the normalization of the color-matching function data used to determine the inter-individual variation. The normalization procedures used by Stiles on the color-matching data result in a significant loss of luminance variance. The loss of luminance variance inherent in the dataset of color-matching functions experimentally measured by Stiles–Burch is ultimately propagated through the statistical analyses performed by the CIE. It is, therefore, expected that the observer variability predicted by this method will be artificially small. The reduction in observer variability will be most severe in the CIE ΔL^* direction when plotting predicted ranges of color mismatch in CIELAB space. The effects of normalization can be seen in the minimal deviation between the $\bar{y}_{dev}$ and the $\bar{y}_{cie\,2^\circ}$ functions depicted in Fig. 9, and the small magnitude of $\Delta\bar{y}_l$ in Fig. 8. These effects are also illustrated by the coincidental zero-crossings of the three deviation functions in Fig. 8 that suggest that there are wavelengths for which there is no inter-observer variability. Clearly this is physiologically implausible.

Much of the variability among observers is due to the eye-lens absorption and scattering.[14] The density of the lens has been shown to increase with age.[12,14] The first deviate was found to correlate with the lens density function. The CIE[3] proposed that the CIE standard deviate observer could be used to show the degree of color mismatch according to the observers age N. This dependence is modeled as a linear function of age modifying the first deviation function, where

$$L(N) = 0.064N - 2.31, \quad \text{and} \tag{7}$$

$$\begin{bmatrix} \Delta\bar{x}_l(\lambda, N) \\ \Delta\bar{y}_l(\lambda, N) \\ \Delta\bar{z}_l(\lambda, N) \end{bmatrix} = L(N) \begin{bmatrix} \Delta\bar{x}_l(\lambda) \\ \Delta\bar{y}_l(\lambda) \\ \Delta\bar{z}_l(\lambda) \end{bmatrix}. \tag{8}$$

This equation is effective only for ages 20–60 years, because the scope of Stiles' observer data was limited to this range. The method was used to predict soft-copy color-matches for the hard-copy samples used in this experiment, based on CIE standard deviate observers of 20, 30, 40, 50, and 60 years of age. The predicted CIE-standard-deviate-observer color-matches for the magenta transparency are plotted in Fig. 10, along with the mean color-matches of the twenty observers divided into four groups of five observers, averaging 25, 33, 44, and 54 years of age. As expected, CIE standard deviate observer color-matches predicted for observers between 20–60 years of age exhibit a linear trend. The mean color-matches for the four age groups of observers exhibit no clear trend with respect to age. The experimental data for

TABLE V. The checkmarks (✓) indicate that the mean color matches predicted by the respective sets of color-matching functions were contained in the ellipsoids, which defined the 95% confidence regions at the $\alpha = 0.05$ level of the sample distributions of inter-observer color matches.

	Transparencies			Prints		
Colors	Stiles–Burch 2°	CIE 2°	CIE 10°	Stiles–Burch 2°	CIE 2°	CIE 10°
Red	✓	✓	✓	✓	✓	✓
Green	✓	✓	✓	✓	✓	✓
Blue	✓	✓	✓	✓	✓	✓
Gray	✓	✓	✓	✓	✓	✓
Cyan	✓	✓	✓	✓	✓	✓
Magenta	✓	✓	✓	✓	✓	✓
Yellow	✓	✓	✓	✓	✓	✓

TABLE VI. The results of Hotelling's T^2 test for means comparing the mean color matches made by the single observer with the hypothetical means predicted with three sets of color-matching functions. A (✓) indicates that the mean color matches are not significantly different from the specified standard observer at the $\alpha = 0.05$ level.

	Transparencies			Prints		
Colors	Stiles–Burch 2°	CIE 2°	CIE 10°	Stiles–Burch 2°	CIE 2°	CIE 10°
Red						
Green	✓	✓				
Blue						
Gray						
Cyan						
Magenta						
Yellow						

this and all other colors suggest that the variability between observers in this experiment is not well described by a simple linear function of age.

The CIE[3] outlined a methodology for evaluating a range of color mismatches for metameric pairs by a statistical confidence ellipse in a $u'v'$ chromaticity diagram. Tristimulus value deviations, $\Delta^2 X_i$, $\Delta^2 Y_i$, $\Delta^2 Y_i$ are used to define a CIE XYZ covariance matrix. Tristimulus value deviations are calculated from the spectral differences of the metameric stimuli in conjunction with the four deviation functions as shown in Eqs. (9) and (10):

$$\begin{bmatrix} \Delta^2 X_i \\ \Delta^2 Y_i \\ \Delta^2 Z_i \end{bmatrix} = \sum_{\lambda} \Delta\rho(\lambda) S(\lambda) \begin{bmatrix} \Delta\bar{x}_i(\lambda) \\ \Delta\bar{y}_i(\lambda) \\ \Delta\bar{z}_i(\lambda) \end{bmatrix} \Delta\lambda \quad (9)$$

$$\Delta\rho(\lambda) = \rho^2(\lambda) - \rho^1(\lambda). \quad (10)$$

The CIE XYZ covariance matrix Σ_{xyz} is defined as

$$\Sigma_{XYZ} = \sum_{i=1}^{4} \begin{bmatrix} (\Delta^2 X_i)^2 & \Delta^2 X_i \cdot \Delta^2 Y_i & \Delta^2 X_i \cdot \Delta^2 Z_i \\ \Delta^2 Y_i \cdot \Delta^2 X_i & (\Delta^2 Y_i)^2 & \Delta^2 Y_i \cdot \Delta^2 Z_i \\ \Delta^2 Z_i \cdot \Delta^2 X_i & \Delta^2 Z_i \cdot \Delta^2 Y_i & (\Delta^2 Z_i)^2 \end{bmatrix}. \quad (11)$$

Instead of transforming the XYZ covariance matrix Σ_{xyz} to a $u'v'$ covariance matrix $\Sigma_{u'v'}$ as recommended by the CIE[3] for determining a confidence ellipse in the $u'v'$ chromaticity plane, the XYZ covariance matrix Σ_{xyz} was transformed to an $L^*a^*b^*$ covariance matrix in order to compare the ranges of the intra- and inter-observer mismatches from the experiment described in this article with the CIE predicted range of mismatch in the CIELAB space. The XYZ covariance matrix Σ_{xyz} is transformed to an $L^*a^*b^*$ covariance matrix Σ_{Lab} by pre- and post-multiplying the Σ_{xyz} covariance matrix with a 3×3 matrix and its transpose composed of partial derivatives of L^*, a^*, and b^* with respect to X, Y, and Z:

$$\Sigma_{Lab} = \mathbf{M}\Sigma_{XYZ}\mathbf{M}', \quad (12)$$

$$\mathbf{M} = \begin{bmatrix} 0 & \frac{\delta}{\delta Y} L^* & 0 \\ \frac{\delta}{\delta X} a^* & \frac{\delta}{\delta Y} a^* & 0 \\ 0 & \frac{\delta}{\delta Y} b^* & \frac{\delta}{\delta Z} b^* \end{bmatrix}. \quad (13)$$

The ellipsoidal 95% confidence region for the range of

TABLE VII. The checkmarks (✓) indicate that the mean color matches predicted by the respective sets of color-matching functions were contained in the ellipsoids, which defined the 95% confidence regions at the $\alpha = 0.05$ level of the sample distributions of intra-observer color matches.

	Transparencies			Prints		
Colors	Stiles–Burch 2°	CIE 2°	CIE 10°	Stiles–Burch 2°	CIE 2°	CIE 10°
Red	✓	✓		✓	✓	✓
Green	✓	✓	✓	✓	✓	✓
Blue						
Gray		✓		✓	✓	✓
Cyan				✓	✓	✓
Magenta	✓	✓	✓	✓	✓	
Yellow	✓	✓	✓			

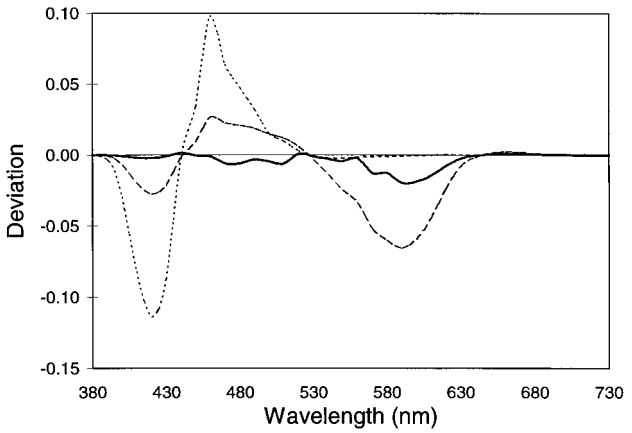

FIG. 8. The first deviation functions of the CIE Standard Deviate Observer; $\Delta\bar{x}_l$ (dashed line); $\Delta\bar{y}_l$ (solid line); $\Delta\bar{z}_l$ (dotted line).

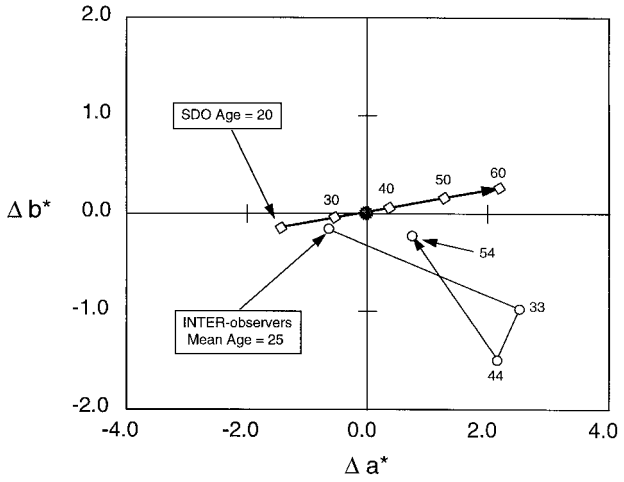

FIG. 10. Age effect on color mismatches. Magenta-transparency age-related color matches predicted by the CIE standard deviate observer; solid thick line. Magenta-transparency mean color matches for four mean age groups of observers. The matchpoint for the 1931 CIE Standard Colorimetric Observer is located at the origin.

color mismatches is defined using this covariance matrix. The center of the ellipsoid when $\Delta L^* = \Delta a^* = \Delta b^* = 0$ corresponds to the CIELAB coordinates of the reference stimuli.

Using the aforementioned CIE method, *XYZ* uncertainties were calculated from tristimulus value deviations, and bivariate 95% confidence ellipses containing the range of (ΔL^*, Δa^*, Δb^*) color mismatches were calculated for the cyan-transparency sample. The bivariate ellipses are plotted in Fig. 12 along with the experimentally determined bivariate 95% confidence ellipses of the sample distributions established for the intra- and inter-observer color-matches. It is clear that the CIE method substantially under-predicts the experimentally observed range of color mismatches.

There are several reasons for the discrepancies between the CIE predictions and the experimental results. One lies in the normalization of the spectral tristimulus data used to derive the CIE technique, as discussed above. Also, the CIE technique is designed to predict the range of color mismatches in a situation in which intra-observer variability has been eliminated, or at least minimized. This would require an experiment in which each observer made multiple matches and only the mean results were used to determine inter-observer variability. If such an

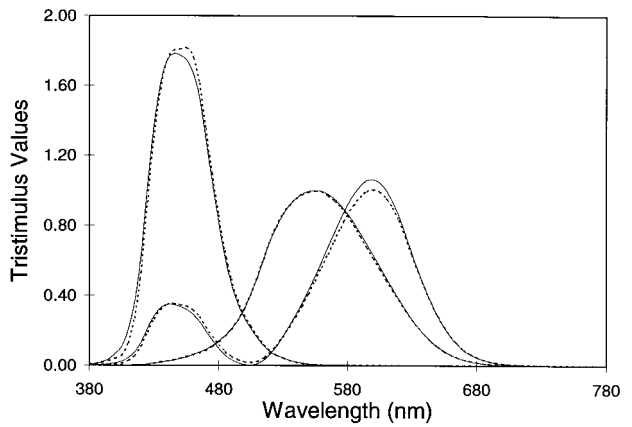

FIG. 9. CIE standard deviate observer (dashed lines), and 1931 CIE Standard Colorimetric Observer color-matching functions (solid lines).

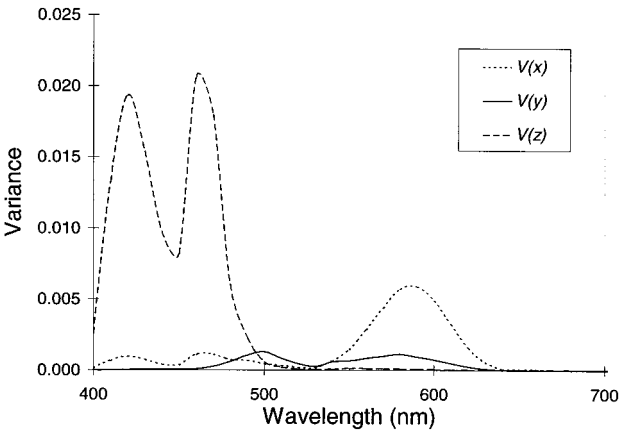

FIG. 11. Fundamental spectral tristimulus variances derived by Nimeroff, Rosenblatt, and Dannemiller for the 10°-field color-matching data of 53 Stiles–Burch observers, and 27 Speranskaya observers.

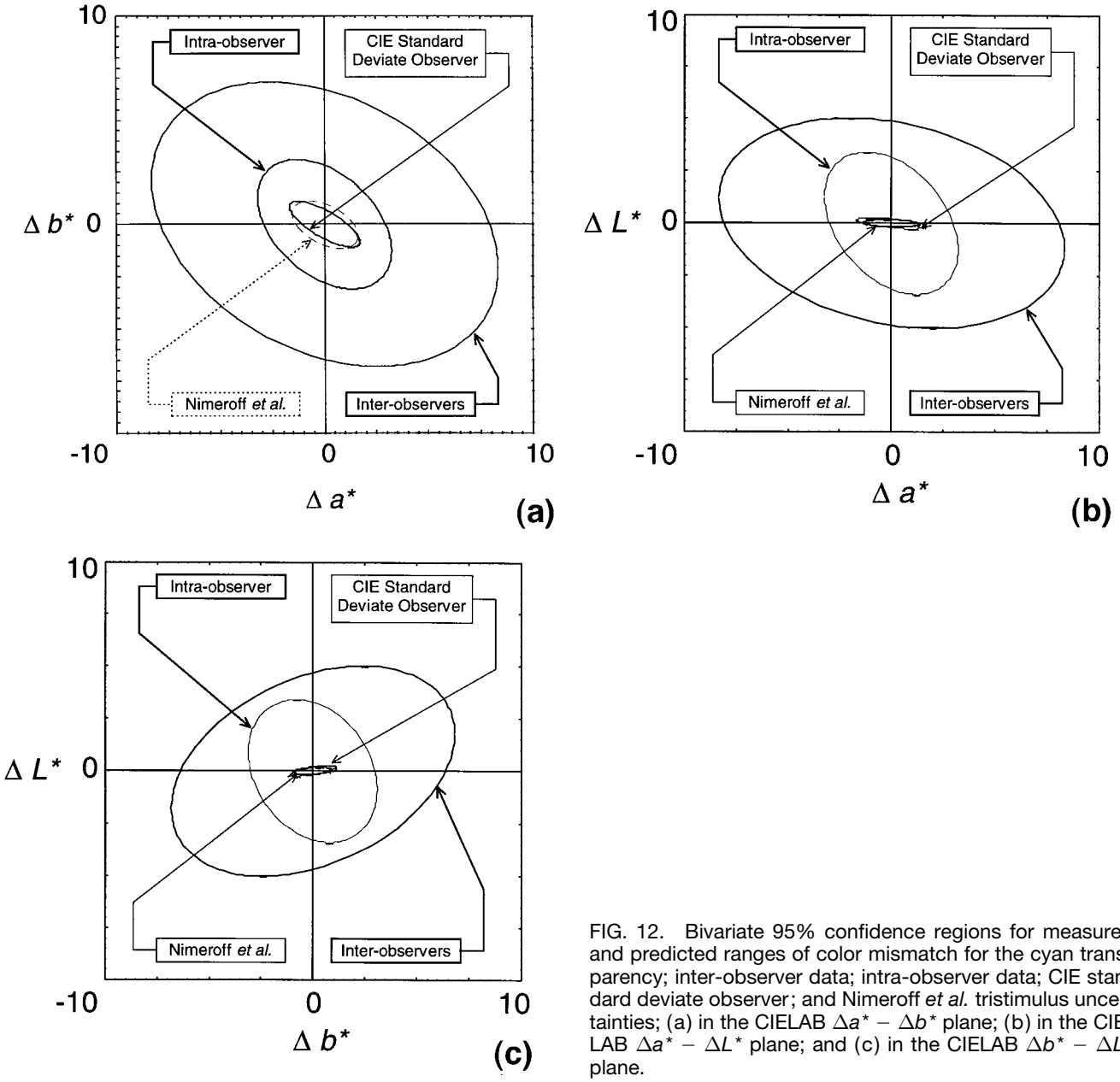

FIG. 12. Bivariate 95% confidence regions for measured and predicted ranges of color mismatch for the cyan transparency; inter-observer data; intra-observer data; CIE standard deviate observer; and Nimeroff *et al.* tristimulus uncertainties; (a) in the CIELAB $\Delta a^* - \Delta b^*$ plane; (b) in the CIELAB $\Delta a^* - \Delta L^*$ plane; and (c) in the CIELAB $\Delta b^* - \Delta L^*$ plane.

experiment were completed, the observed variability would be smaller. The results of the current experiment suggest that the intra-observer variability would account for approximately 50% of the observed inter-observer variability. While such data would be useful, they were not the objective of this research. The goal was to quantify the variability in individual matches made by a population of observers. This goal is similar to that discussed by Rich and Jalijali.[15] The experimental results of Rich and Jalijali,[15] obtained under similar viewing conditions, agree quite well with the results of this study.

In another study, Nayatani *et al.*[16] evaluated the performance of the CIE procedure in comparison to various psychophysical results. While Nayatani *et al.*[16] concluded that the CIE technique performed well when intra-observer variability was minimized, closer examination of the results provides some additional insight. Firstly, they only analyzed bivariate (chromaticity) data and did not examine variability in the luminance dimension. They were forced to normalize out the luminance variation using Eq. (2) in Ref. 16. Additionally, the results of their visual experiment are shown in Fig. 13 of Ref. 16, which illustrates that even when intra-observer variability is minimized through averaging, the match points for three out of 5 observers fall outside the predicted 95% confidence ellipse. Thus, in the Nayatani *et al.*[16] experiment, only 40% of the observed

matches fall inside predicted 95% ellipse. This result is in general agreement with the current results and those of Rich and Jalijali.[15] While the design of the CIE technique precludes any expectation that it could predict the observed experimental results of this study, *CIE Publication 80*[3] makes several claims that it can do so, the first of which appears in the summary. Clearly, *CIE Pub. 80* overstates the utility of the techniques described therein.

The current data do allow a test of the CIE recommendations that is free of any discrepancy introduced by intra-observer variability. The MCDM values reported in Table II provide measures of the average difference to be expected between individual matches and the mean of the population. If the intra-observer variability is assumed to be random across the 20 observers, then the MCDM should effectively average it out. This is clearly the case, as illustrated by the rather small MCDM values in comparison to the range of color-matches. The Standard Deviate Observer described in *CIE Pub. 80* is designed to predict the ''average of color mismatches for metameric pairs when test observers are substituted for a reference observer.''[3] These predictions are listed in the third column of Table II, labeled SDO. The SDO values significantly under predict, by factors ranging from $3.2X$ to $10.7X$, the observed MCDM values in every case. It would be of interest, if the SDO values were well correlated with the experimental results such that an appropriate scaling factor could be applied (or at least the SDO would provide appropriate relative values). This is examined in Fig. 13 in which the SDO ΔE^*_{ab} values are plotted as a function of the observed MCDM values for all 14 color stimuli using the CIE 1931 Standard Colorimetric Observer. Figure 13 shows no significant correlation between the SDO predictions and the MCDM data.

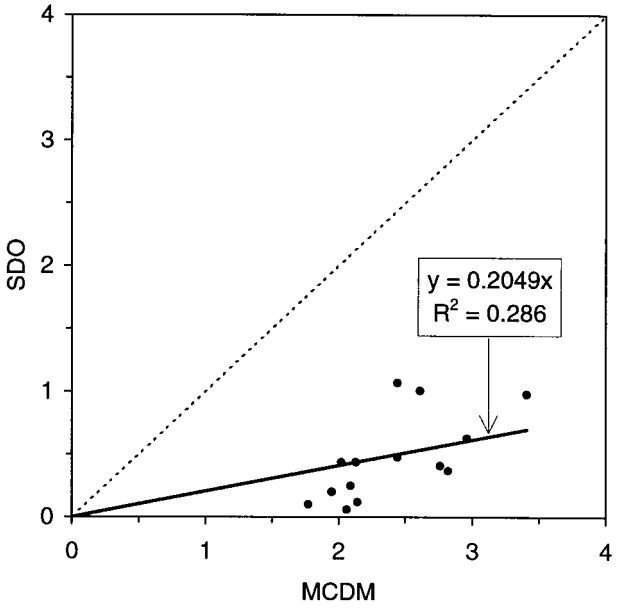

FIG. 13. SDO ΔE^*_{ab} values plotted as a function of the observed MCDM values for all 14 color stimuli using the CIE 1931 Standard Colorimetric Observer.

VARIABILITY OF SPECTRAL TRISTIMULUS VALUES

Nimeroff, Rosenblatt, and Dannemiller derived fundamental estimates of variances and covariances in spectral tristimulus values, based on color-matching for individual observers.[4] The inter-observer variances were derived for 10°-field color-matching data from 53 Stiles–Burch observers, and 27 Speranskaya observers. Inter-observer covariances were derived only using the color-matching data collected from the 53 Stiles–Burch observers. The color-matching functions measured by Stiles–Burch[13] were normalized. The normalization of color-matching functions has the effect of eliminating a significant component of luminance variance and can again be seen by the relatively small degree of variation associated with the $\bar{y}$ tristimulus function shown in Fig. 11.

A variation of the CIE methodology to define a range of color mismatch can be used to construct a 95% confidence region of color mismatch for the cyan transparency, based on the tristimulus value uncertainties derived by Nimeroff *et al.*[4] An XYZ covariance matrix is calculated from the variance and covariance components:

$$\begin{bmatrix} V(X_{10}) \\ V(Y_{10}) \\ V(Z_{10}) \\ \mathrm{Cov}(X_{10}, Y_{10}) \\ \mathrm{Cov}(X_{10}, Z_{10}) \\ \mathrm{Cov}(Y_{10}, Z_{10}) \end{bmatrix} = \sum_{\lambda} \Delta P(\lambda) \begin{bmatrix} V(\bar{x}_{10}(\lambda)) \\ V(\bar{y}_{10}(\lambda)) \\ V(\bar{z}_{10}(\lambda)) \\ \mathrm{Cov}(\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda)) \\ \mathrm{Cov}(\bar{x}_{10}(\lambda), \bar{z}_{10}(\lambda)) \\ \mathrm{Cov}(\bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)) \end{bmatrix} \Delta\lambda, \quad (14)$$

where ΔP is the difference in radiant spectral power between a pair of metameric stimuli. The XYZ covariance was used to construct a 95% confidence region containing the range of (ΔL^*, Δa^*, Δb^*) color mismatch predicted for the cyan-transparency sample, based on the uncertainties derived by Nimeroff *et al.* for the CIE 10° Supplemental Standard Colorimetric Observer.

Figures 12(a)–(c) each show four bivariate 95% confidence regions for measured and predicted ranges of color mismatch for the cyan transparency. The four regions are defined by the inter-observer sample data, the intra-observer sample data, the CIE standard deviate observer recommendations, and the Nimeroff *et al.* tristimu-

lus uncertainties. The intra- and inter-observer ellipses defined by experimental data are significantly larger than the predicted CIE, and Nimeroff *et al.* based ellipses. The loss of luminance variance associated with the normalization of the color-matching function data used by the CIE and Nimeroff *et al.* to assess observer variability is readily seen in the compressed L^* dimension of the CIE and Nimeroff *et al.* based ellipses predicted for the cyan transparency, shown in Figs. 12(b) and 12(c). The ΔL^* components of the CIE and Nimeroff *et al.* based confidence ellipses predicted for each of the thirteen other color-medium samples were found to be similarly compressed when compared with the CIE and Nimeroff *et al.* based confidence ellipse predicted for the cyan transparency. These ellipses also significantly under-estimated the experimentally observed variances in the $\Delta a^* - \Delta b^*$ dimensions for the other 13 colors. Again, intra-observer variability does enlarge the experimental inter-observer ellipses, but that would not be enough of a discrepancy to explain the poor predictions.

CONCLUSIONS

A visual experiment was designed and performed to measure the magnitude of uncertainties associated with both intra- and inter-observer variability in cross-media color matching. The range of color mismatch was found to be as large as 19 CIELAB units for hard- to soft-copy color matches. The results indicate that the variability of inter-observer color matches is approximately twice as large as the variability of one observer's intra-observer color matches for the metameric pairs of soft- and hard-copy media examined in this experiment. All metameric match points for the three standard observers were found to be inside the 95% confidence ellipses of the sample distributions of inter-observer color-matches. The majority of the match points for the 2° standard observers were contained within the 95% confidence ellipses of the sample means of inter-observer color-matches. These results suggest that the existing CIE Standard Colorimetric Observers are a reasonably good representation of the population of normal trichromats, and confirm the concepts regarding CIE colorimetry described by Fairchild.[17]

The experimentally established ranges of intra- and inter-observer color mismatch were found to be significantly larger than the ranges of color mismatch calculated from CIE recommendations on observer metamerism and the ranges of color mismatch calculated from the estimated spectral tristimulus uncertainties derived by Nimeroff *et al.* The CIE method and Nimeroff *et al.* uncertainties are based on normalized color-matching function data, which appear to have significantly diminished the variance component associated with luminance, resulting in regions of color mismatch unnaturally compressed in the L^* dimension. The experimental inter-observer color-matching data showed no correlation with age, as is predicted by the linear function of age modeled for the CIE standard deviate observer. While there is no doubt that age-related changes in color vision do exist, their magnitude is apparently overwhelmed by inter-observer variability at any given age. It is suggested that the CIE recommendations on observer metamerism be reviewed as proposed by the CIE,[18] and further research be performed to better quantify the variability associated with observer metamerism in practical cross-media color-matching, in addition to developing more accurate techniques for predicting the range of color mismatches.

Several approaches might be taken to derive an improved predictor of the total range of color mismatches (inter-observer plus intra-observer variability) in practical applications. Scaling of the predictions based on the techniques evaluated in this article was attempted. Such scaling did not produce satisfactory results, indicating that the discrepancy is in nature as well as magnitude. A more accurate technique could be derived by applying similar analyses to absolute color-matching functions (data that have never been normalized). Such data are extremely expensive and time-consuming to collect and it is not likely that they will become available for a large enough population of observers in the near future. Another approach is currently under development that will take advantage of an analytical model of color-matching functions[19] and knowledge of the physiological variations in color vision[20] in a Monte Carlo simulation aimed at producing large numbers of typical absolute color-matching functions.

ACKNOWLEDGMENTS

The authors thank Jason Gibson for his assistance in collecting the experimental data and participating as an observer for twenty trials, in addition to all the other observers who graciously volunteered their time.

1. *International Electrotechnical Vocabulary,* Publication 50(845-03-05), Central Bureau of the International Electrotechnical Commission, Geneva, 1987.
2. I. Pobboravsky, Effect of small color differences in color vision on the matching of soft and hard proofs. *TAGA Proceedings,* 62–79 (1988).
3. *Special Metamerism Index: Change in Observer,* CIE Publ. No. 80, Central Bureau of the CIE, Vienna, 1989.
4. I. Nimeroff, J. R. Rosenblatt, and M. C. Dannemiller, Variability of spectral tristimulus values. *J. Res. NBS A* **65,** 475–483 (1961).
5. S. Ishihara, *The Series of Plates Designed as a Test for Colour-Blindness,* Kanehara Shuppan Co., Tokyo.
6. D. Farnsworth, The Farnsworth–Munsell 100 Hue and dichotomous tests for color vision. *J. Opt. Soc. Am.* **33**, 568–578 (1943).
7. A. Stockman, D. I. A. MacLeod, and N. E. Johnson, Spectral sensitivities of the human cones. *J. Opt. Soc. Am. A* **10,** 2515–2517 (1993).
8. *Colorimetry,* 2nd Ed., CIE Publ. No. 15.2, Central Bureau of the CIE, Vienna, 1986.
9. F. W. Billmeyer, Jr. and P. J. Alessi, Assessment of color-measuring instruments. *Color Res. Appl.* **6,** 195–203 (1981).
10. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical*

Analysis Third Ed., Prentice Hall, Englewood Cliffs, 188–191, 1992.
11. J. E. Jackson, Some multivariate statistical techniques used in color-matching. *J. Opt. Soc. Am.* **49,** 585–592 (1959).
12. A. D. North and M. D. Fairchild, Measuring color-matching functions. Part II. New data for assessing observer metamerism. *Color Res. Appl.* **18,** 167–169 (1993).
13. W. S. Stiles and J. M. Burch, N.P.L. colour-matching investigation: Final report (1958). *Opt. Acta* **6,** 6–7 (1959).
14. J. Pokorny, V. C. Smith, and M. Lutze, Aging of the human lens. *Appl. Opt.* **26,** 1437–1440 (1987).
15. D. C. Rich and J. Jalijali, Effects of observer metamerism in the determination of human color-matching functions. *Color Res. Appl.* **20,** 29–35 (1995).
16. Y. Nayatani, T. Tanaka, H. Sobagaki, K. Takahama, and K. Hashimoto, Field trials for assessing the method of observer metamerism adopted by CIE. *Color Res. Appl.* **16,** 97–107 (1991).
17. M. D. Fairchild, The CIE 1931 Standard Colorimetric Observer: Mandatory retirement at age 65? *Color Res. Appl.* **18,** 129–133 (1993).
18. CIE, Proceedings of the CIE Symposium on Advanced Colorimetry, *Publ. CIE x007*, Central Bureau of the CIE, Vienna, 106, 1993.
19. A. D. North and M. D. Fairchild, Measuring color matching functions, Part I. *Color Res. Appl.* **18,** 155–162 (1993).
20. V. C. Smith and J. Pokorny, Chromatic-discrimination axes, CRT phosphor spectra, and individual variation in color vision. *J. Opt. Soc. Am. A* **12,** 27–35 (1995).