Observer variability experiment using a four-primary display and its relationship with physiological factors

Yuta Asano [1], Mark D. Fairchild [1], Laurent Blondé [2]
[1] Munsell Color Science Laboratory, Rochester Institute of Technology; Rochester, NY, USA
[2] Technicolor; Rennes, France

Abstract

There exist individual differences in color matching functions and the use of a single standard observer as a representative of a whole population often limits the accuracy of color reproduction, especially for narrowband stimuli. We conducted a paired comparison experiment for 58 color-normal people involving color difference judgments using four nearly metameric spectra pairs. The performance of observer functions: CIEPO06, Sarkar's observers, and the extended CIEPO06 incorporating peak-shift in L and M cones were investigated. Large observer variability was found in the obtained results, which is much larger than what CIEPO06 predicts. At least two different groups were found in the experimental results, which could be explained by eye-lens and macular pigment optical density variations. We estimated the individual cone fundamentals from another experiment where observers performed five color matching, and used them to predict the paired comparison results. They gave better, or at least comparable prediction to those of CIE 1964 observer and CIEPO06.

Introduction

The CIE 1931 standard colorimetric observer and the CIE 1964 standard colorimetric observer, also known as 2°and 10°(standard) observers, have satisfied industries for many decades. However, the advent of narrowband stimuli such as lasers and LEDs used in monitors, TVs, mobile displays, or cinema projection is changing this situation. The use of a standard observer is based on the assumption that a single observer or a single set of color matching functions (CMFs) can reasonably represent a whole population of people with normal color vision. If narrowband stimuli such as lasers and LEDs are observed, however, they magnify the individual differences in CMFs and this assumption often breaks [1, 2, 3]. Researchers have tackled this problem and proposed methods to evaluate observer metamerism, which are summarized below.

Related Work

In mid-20th century, devices to evaluate observer and illuminant metamerism were invented. They include Glenn Color Rule [4], Davidson and Hemmendinger Color Rule (D&H Color Rule) [5], and Macbeth Matchpoint. These devices have a series of color patches arranged in pairs. The observer's task is to find a pair of patches that appear the same according to their perception. Thus, using different illuminations and a single observer allows us to evaluate illuminant metamerism and using different observers and a single illumination allows us to evaluate observer metamerism. A number of researches have been performed using D&H Color Rule with respect to observer metamerism evaluation [5, 6, 7, 8, 9, 10, 11].

In 1989, the CIE proposed a "Standard Deviate Observer", a method to evaluate observer metamerism [12, 13]. However, it is reported that the standard deviate observer significantly underestimates observer variability [14, 15].

In 2006, the CIE proposed a model [16] (CIEPO06) that computes cone fundamentals by specifying an observer's age and a field size. It allows us to generate a number of theoretical observers and use them to evaluate observer metamerism.

In 2011, Sarkar [17, 18] proposed observer categories by analyzing 47 individual CMFs of Stiles and Burch data, and 61 CIE 2006 CMFs using cluster analysis. In a test experiment, forty-nine human observers were classified into nine categories (including CIE 1964 observer as one category) using an observer calibrator prototype [19], which displays two different spectra side by side in a bipartite field. The performance of these eight observer categories is not widely verified and needs more investigation.

Research Objective

Most of the observer variability experiments in the past involve colors appearing as unrelated colors, comparison of adjoining stimuli, a single bipartite field, and/or monocular view, which might be different from practical viewing conditions.

Our goal in this study was to conduct an experiment under more practical viewing conditions, to see how much observer variability we would obtain, and to see if observer functions such as CIEPO06 and Sarkar's observers cover the variability of human observers. Note that it is not our purpose to seek for better standard observer functions or to investigate the accuracy of standard observer functions.

Experiment and Analysis

Overview

To achieve our goal, we designed a paired comparison experiment using SHARP Quattron display. Quattron is a display having four primaries, R, G, B, and Y (for Yellow) with 10-bit assignment that enables us to produce metameric spectra on a single display. The spectral power distributions (SPDs) of the four primaries are shown in Figure 1. Four nearly metameric spectra pairs are generated and used in the experiment. One metameric pair consists of two color patches: One patch is made from red, green, and blue primaries. The other patch is made from blue and yellow primaries. Figure 2 illustrates the experiment workflow.

Paired comparison is the method that several stimulus intensity levels are chosen (four color difference levels in this case) and compared pair-wise repeatedly. Unlike traditional color matching

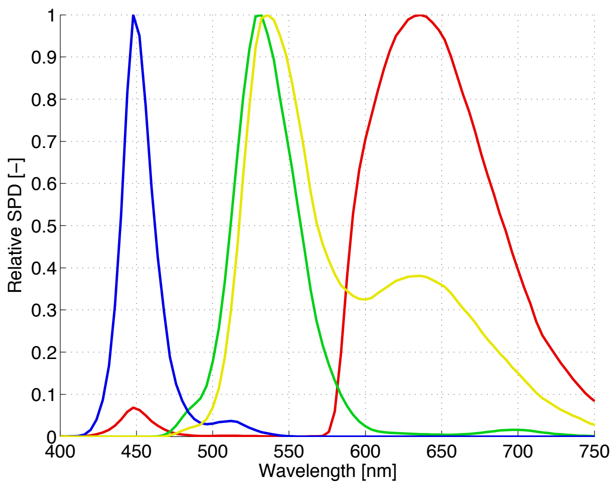

Figure 1. *Spectral Power Distributions of Quattron display primaries.*

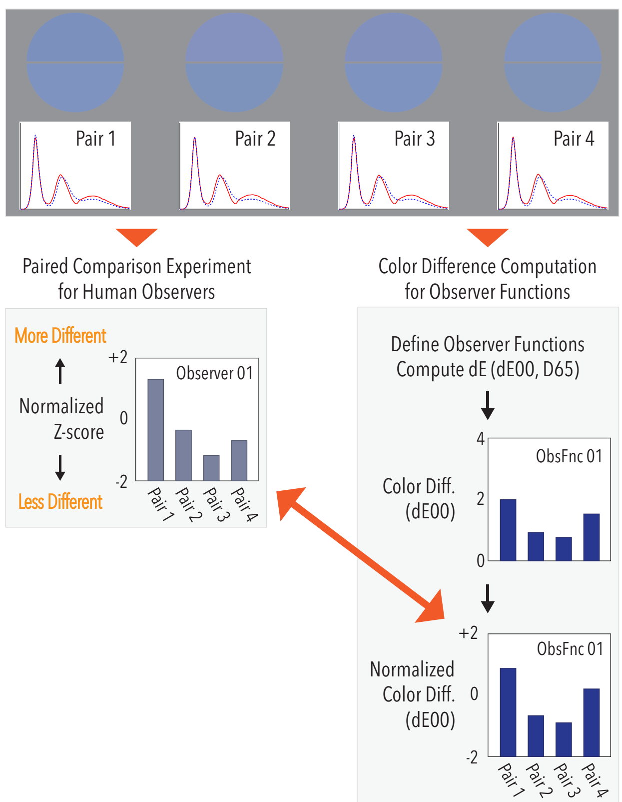

Figure 2. *Experiment workflow.*

experiments, paired comparison is very simple and easy for naive subjects, allowing us to collect a number of observers easily. The number of comparisons (N_{cmp}) without repetition is expressed as Equation 1.

$$N_{cmp} = \frac{n \cdot (n-1)}{2} \tag{1}$$

where n is the number of stimuli. The number of repetition is empirically determined to 16 times in a preliminary experiment. The decision is made such that the experiment time is as short as possible but still retains reliability. Therefore, the number of total judgments per one subject is 96 ($= 4 \cdot (4-1)/2 \times 16$).

By applying Case V of Thurstone's law of comparative judgments, it is possible to extract the perceptual scale [20]. How to obtain the perceptual scale is briefly explained below. First, the frequency was recorded in a matrix where the number of rows and columns corresponded to the number of stimuli; A value of 1 was added to the matrix if the color difference of j-th pair was judged greater than that of i-th pair. The frequency possibly varied from 0 to 16. Next, the frequencies were converted to probability simply dividing by the maximum (=16). The diagonal line was filled with 0.5 assuming the comparison of same stimuli would produce 50% probability. In case we have probability of 0 and 1, which produces the z-score of $\pm\infty$ and makes further analysis impossible, the probability of 0 and 1 was replaced by 10^{-3} and $1-10^{-3}$. Then, the probability was converted to the perceptual scale, z-score using a psychometric function expressed as Equation 2.

$$S = \log \frac{p}{1-p} \tag{2}$$

where S is z-score and p is the probability ($0 < p < 1$). The z-scores in the matrix were averaged for each column, which gave us a z-score for each of the four SPD pairs. Finally, the four z-scores were normalized such that they have the mean value of zero and the standard deviation of unity. The obtained results would tell us which pair appears more/less different in a perceptual order for each human observer.

On the other hand, from measurements, we can compute color differences of the four SPD pairs for a given observer function. The computation steps are summarized below. First, an observer function (CMFs) was defined. Second, the CMFs were transformed into CIE 1964 observer space using a matrix $M_{3\times3}$, expressed as Equation 3. The $M_{3\times3}$ could be estimated by linear regression expressed as Equation 4.

$$T_{trans} = M_{3\times3} \cdot T_{target} \tag{3}$$

$$M_{3\times3} = T_{CIE1964} \cdot (T_{target}^T T_{target})^{-1} T_{target} \tag{4}$$

Here, T_{target} is a set of CMFs for the defined observer and $T_{CIE1964}$ is a set of CMFs for CIE 1964 observer. All the matrices T are $3 \times n$ matrices where n is the number of wavelength sampling. The expression $(T_{target}^T T_{target})^{-1} T_{target}$ is a pseudo-inverse of T_{target}. This transformation was preferred to use the color difference formula as uniformly as possible. Note that T_{target} can be any CMFs including the cone fundamentals from CIEPO06 and *rgb* CMFs, because they are all in a linear relationship with each other. Third, using the transformed CMFs, $pseudo-XYZ_{10^\circ}$ were computed for the four SPDs and the reference white spectrum (D65). Forth, CIELAB values were computed, and then the color differences were computed using CIE DE2000 (ΔE_{00}). Finally, the same normalization as z-scores was performed to the color differences in order to bring color differences and z-scores to the same space and compare them directly. In this way, we can investigate which observer function correlates with a given human observer.

Experiment Conditions

The image that subjects observe during the experiment is illustrated in Figure 3. Two SPD pairs (left and right circles) are shown side by side on a gray background. The distance between two circles is about 2°and each circle subtends about 4.5°in visual angle. The letters and the cross at the center are written in white, which is considered as reference white. Both the background and the reference white are made from all the four primaries, and approximately have D65 white point. The background has L^* of around 50, the luminance of the reference white is 318.7 $[cd/m^2]$. The calculations here are done using CIE 1964 observer.

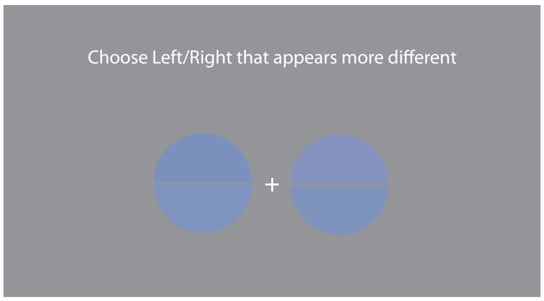

Figure 3. *Experiment view.*

There are eight stimuli (two patches for four pairs) used in this experiment. The eight stimuli are all very similar colors (purplish blue) but have slightly different spectra, and appear differently for different observers. Their differences with respect to CIELAB values for CIE 1931 and 1964 observers are shown in Table 1.

Table 1. CIELAB values of four SPD pairs for CIE 1931, 1964 standard observers

		Pair 1		Pair 2		Pair 3		Pair 4	
		RGB_{mix}	BY_{mix}	RGB_{mix}	BY_{mix}	RGB_{mix}	BY_{mix}	RGB_{mix}	BY_{mix}
	L^*	60.3	60.8	60.7	61.1	60.4	61.2	60.9	61.1
CIE	a^*	0.1	0.3	2.5	-1.1	2.4	-0.2	0.6	-1.4
1931	b^*	-24.8	-25.1	-24.6	-23.1	-24.7	-24.5	-24.5	-22.4
	ΔE_{00}	0.5		3.0		2.4		1.7	
	L^*	61.1	61.3	61.4	61.6	61.1	61.7	61.6	61.5
CIE	a^*	-2.3	-0.1	-0.2	-1.2	-0.3	-0.5	-1.9	-1.4
1964	b^*	-23.7	-24.7	-23.6	-22.8	-23.7	-24.1	-23.5	-22.1
	ΔE_{00}	1.9		0.8		0.6		0.9	

The stimuli are presented in a random order. The locations of the presentations are also randomized so that each stimulus is assigned to the four locations uniformly.

Regarding the stability of the display, the temporal change in colorimetric values between 1-hour warm-up and 9-hour warm-up is about 0.1% on average. The spatial change in SPDs for the four different locations is about 0.3% on average. The changes are overall very small for the purpose of this study. In addition, the changes would be further discounted by randomly presenting stimuli.

Experiment Procedure and Subjects

Subjects sat on a chair and looked at the display. Before the experiment started, they were asked to adjust the chair such that their eyes were at about the same height as the center of display and about 1m away from the display. Then, they were asked to judge which color difference appears greater than the other by pressing L/R button on keyboard. Subjects had an option to go back and modify the previous judgment.

Subjects were instructed to fixate their view to the center (the cross in Figure 3) during judgments. This fixation was meant to avoid the macular pigment intense region affecting the results, and to simulate 10°viewing condition.

58 color-normal subjects participated in this experiment. Subjects were screened using Ishihara's Tests for Colour-Blindness. The youngest observer was 18 year-old and the oldest was 69 year-old. The number of males and females is 42 and 16, respectively. The number of naive observers and expert observers is 28 and 30, respectively.

Results and Discussion

Results obtained from Human Observers

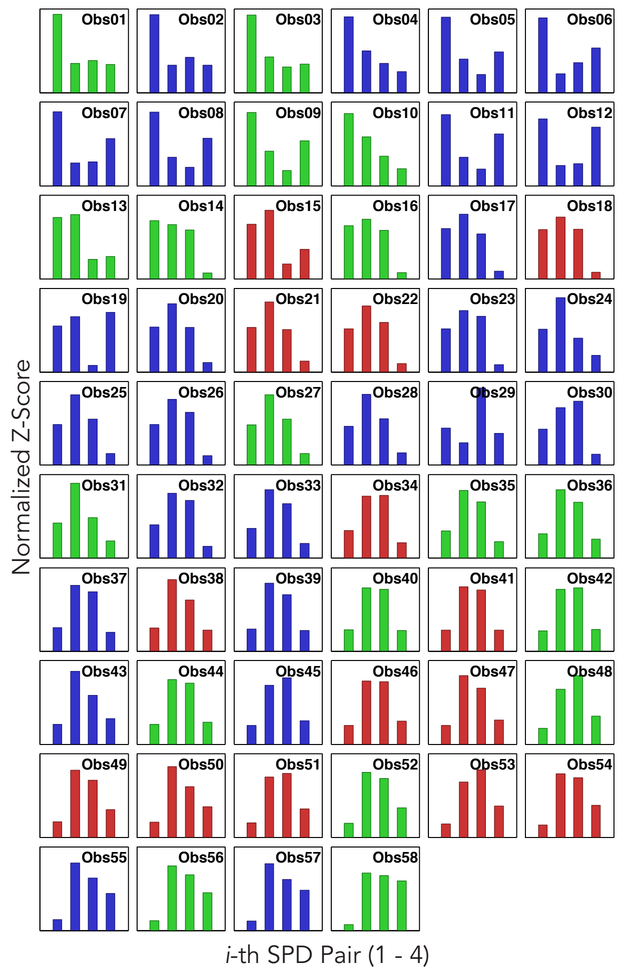

Figure 4. *Normalized z-scores obtained from 58 color-normal human observers. Observers are sorted based on the normalized z-score for pair 1. The bars are color-coded based on observer's age group: 20 year-old for blue, 30 year-old for green, 40 and more year-old for red.*

The normalized z-scores from human observers are plotted in Figure 4. Each plot corresponds to each observer's results. For each plot, the four bars represent the normalized z-scores for the four SPD pairs.

At a first glance, the observer variability is quite large, especially for young observers (age group 20 and 30). We can see at least two different groups. The majority of people forms mountain-shape, having larger values for pair 2 and 3. The second group forms valley-shape, having larger values for pair 1 and 4. Another thing to notice is that most of the older observers have valley-shape and they never form mountain-shape. The other possible factors that might affect results such as region, gender, total experiment time, and observer experience were also investigated and no dependence was found.

Assessment of Different Observer Functions

The simulation results for CIE 1931 and 1964 observers are shown in Figure 5. For CIE 1931 observer, pair 2 appears most different and pair 1 appears nearly matched, while pair 1 appears most different for CIE 1964 observer. The differences between the two observers come from the spectral differences of each pair. For instance, pair 1 has relatively large spectral difference in short-wavelength region (not shown in this paper) and CIE 1964 observer magnifies this difference whereas CIE 1931 observer compresses it.

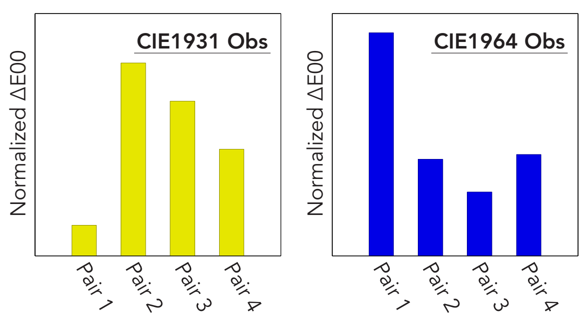

Figure 5. *Normalized ΔE_{00} for CIE 1931, 1964 observers.*

In a similar manner, the normalized ΔE_{00} are computed for Sarkar's observers and shown in Figure 6. Cat:1, cat:2, ..., cat:9 indicate Sarkar's i-th categorical observer. Note that the category 1 observer is same as CIE 1964 observer. Most of Sarkar's observers (category 2, 4, 5, 6) are similar to CIE 1964 observer since they were derived from 10°CMFs. Sarkar reported that category 8 and 9 are dominated by elderly people [18], which correlates with the fact that people in higher-age group always form mountain-shape in this experiment.

Regarding CIEPO06, it takes age and field size as input. Note that the age factor controls the lens pigment optical density and the field size mainly controls the macular pigment optical density. The effect of the age and field size changes on physiological factors are demonstrated in Figure 7. Different age and field size combinations were investigated to see how they change the prediction. The simulation results using CIEPO06 are shown in Figure 8. Increasing age factor in CIEPO06 causes the bars change from mountain-shape to valley-shape, which predicts the human observers' results. It is also found that decreasing field size has similar effect to prediction as increasing age factor. This is understandable since increasing lens pigment optical density (by increasing age factor) and increasing macular pigment optical

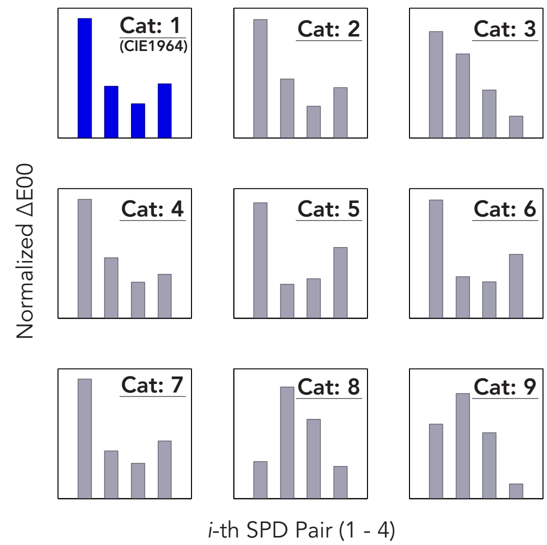

Figure 6. *Normalized ΔE_{00} for Sarkar's observers.*

density (by decreasing field size) both act as a yellowing filter in our eyes.

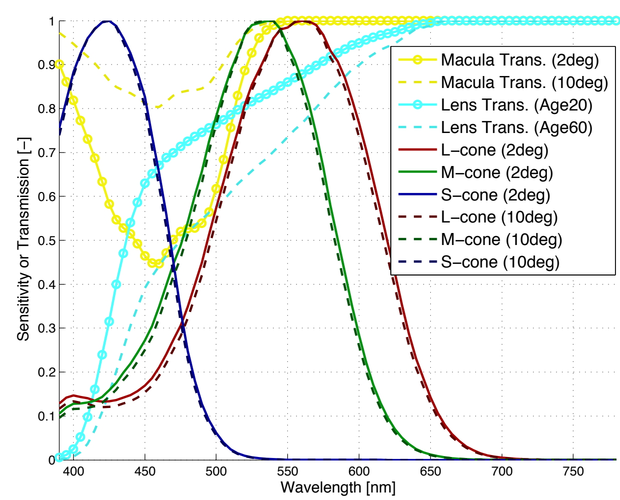

Figure 7. *Physiological factors in CIEPO06. Red, green, and blue lines indicate L, M, and S cone photopigment sensitivities. Yellow lines indicate macular transmission and cyan lines indicate lens transmission.*

In the human observers' results, the mountain-shape can be explained by CIEPO06 with smaller field size and/or higher age, and the valley-shape can be explained by CIEPO06 with larger field size and lower age. In other words, people in the mountain-shape group might have more yellow pigments than people in the valley-shape group.

However, if CIEPO06 is used to predict each observer's results using the observer's age and field size of 10°as input, the prediction is not satisfactory. For example, the results of people in age group 20 and 30 vary from mountain-shape to valley-shape. On the other hand, the predictions for CIEPO06 using age 20 to 30 and field size of 10°(row1-4 at column 1 in Figure 8) do not vary that much. This is probably because CIEPO06 models the performance of the average observer for a given age and field size, and would not be suitable to predict individual observer's results. Also, it is mentioned in CIE TC 1-36 report that since observers, even within the same age bracket, may differ, a fundamental observer must be a theoretical construct based on averages. Any "real observer" will be different from the "TC 1-36 Modified Colorimetric Observer" [16]. Another possible explanation would be that CIEPO06 does not incorporate peak-shifts in L and M cone photopigments due to genetic polymorphism [21, 22], which causes individual differences in CMFs.

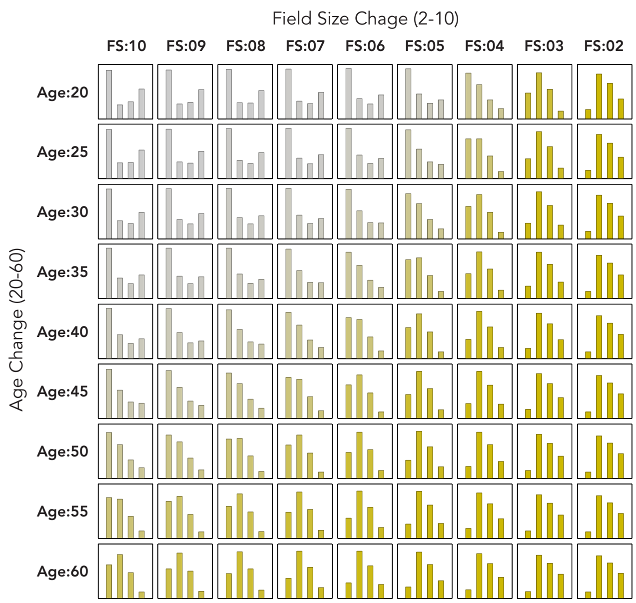

Figure 8. *Normalized ΔE_{00} for CIEPO06 with varying age and field size. The bar charts are color-coded based on the covariance with the prediction (age:20, FS:10). The bar becomes yellower as the covariance decreases.*

Relationship between Quattron Results and Estimated Individual Cone Fundamentals

We conducted another experiment where human observers performed five color matching and individual cone fundamentals were estimated from the results. The idea is similar to the work by Fairchild [23] and Viénot [24]. It is investigated if the results in this paper would be predictable by the estimated individual cone fundamentals.

To estimate the individual cone fundamentals, first, we extended CIEPO06, incorporating the peak-shifts in L and M cone photopigments. Next, we generated a number of cone fundamentals from the extended CIEPO06 with varying the four parameters (age, field size, peak-shift in L-cone, and peak-shift in M-cone). Then, we simply performed predictions using all the generated cone fundamentals. A set of parameters that predicts the five color matching results best is chosen as this person's physiological parameters. Finally, the individual cone fundamentals can be obtained using the estimated physiological parameters as input.

The predictions of CIE 1964 observer, CIEPO06, and the individualized cone fundamentals for four observers are shown in Figure 9. The observer numbers correspond to the ones in Figure 4. The maximum and minimum covariances are 1 and -1, respectively. Huge prediction improvement can be seen for observer 36. For other observers, the predictions of the individualized cone fundamentals are comparable to those of CIEPO06.

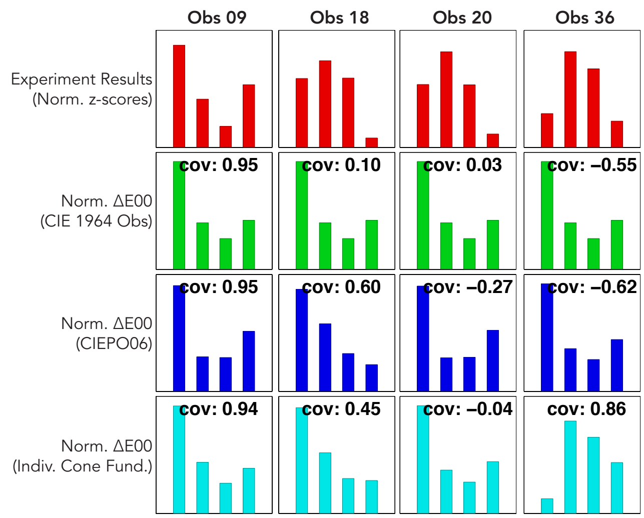

Figure 9. *The experimental results, the predictions of CIE 1964 observer, CIEPO06, and the individualized cone fundamentals for four observers. The covariance (max: 1, min: -1) with experimental result is shown for each prediction.*

Conclusion

A paired comparison experiment involving color difference judgments was designed and conducted for 58 human observers with normal color vision.

We found at least two different groups in our results: (1) those with high z-scores for pair 2 and 3, which forms mountain-shape, and (2) those with high z-scores for pair 1 and 4, which form valley-shape. The first group can be explained by CIEPO06 with more yellow pigmentation and the second group can be explained by CIEPO06 with less yellow pigmentation. All the people with higher age are in the first group. The variability in human observers' results is much larger than the one predicted from CIEPO06. The individualized cone fundamentals give better, at least comparable prediction to those of CIE 1964 observer and CIEPO06.

Acknowledgments

The research was supported by Technicolor. The modified Quattron display used in this experiment was loaned by SHARP.

References

[1] R.L. Alfvin and M.D. Fairchild. Observer variability in metameric color matches using color reproduction media. *Color Research & Application*, 22(3):174–188, 1997.

[2] M.D. Fairchild and D.R. Wyble. Mean observer metamerism and the selection of display primaries. In *Final Program and Proceedings-IS&T/SID Color Imaging Conference*, pages 151–156, 2007.

[3] R. Ramanath. Minimizing observer metamerism in display systems. *Color Research & Application*, 34(5):391–398, 2009.

[4] JR Aspland and P. Shanbhag. Samples suitable for metamerism-based color vision testing devices: The glenn colorule revisited. *AATCC review*, 6(4):32, 2006.

[5] P.K. Kaiser and H. Hemmendinger. The color rule: A device for color-vision testing. *Color Research & Application*, 5(2):65–71, 1980.
[6] H. Kalmus. Metameric 'color rule' matches of normal, colour deficient, cataractic, and aphakic observers. *Annals of Human Genetics*, 36(1):109–118, 1972.
[7] W.R. Biersdorf. The davidson and hemmendinger color rule as a color vision screening test. *Archives of Ophthalmology*, 95(1):134, 1977.
[8] H. Hemmendinger and C. Bottiger. Metamerism and its influence on attainable color tolerances. In *Proceedings of the 3rd AIC Congress Color 77*, pages 425–428, 1977.
[9] F.W. Billmeyer Jr and M. Saltzman. Observer metamerism. *Color Research & Application*, 5(2):72–72, 1980.
[10] M.A. Nardi. Observer metamerism in college-age observers. *Color Research & Application*, 5(2):73–73, 1980.
[11] S. Coren. A rapid method to assess crystalline lens pigment density in vivo. *Acta Ophthalmologica*, 65(5):575–578, 1987.
[12] E. Allen. An index of metamerism for observer differences. In *Proceedings of the 1st AIC Congress Color 69*, pages 771–783, 1970.
[13] CIE. Special metamerism index: Change in observer. *CIE Publication No.80*, 1989.
[14] A.D. North and M.D. Fairchild. Measuring color-matching functions. part ii. new data for assessing observer metamerism. *Color Research & Application*, 18(3):163–170, 1993.
[15] D.C. Rich and J. Jalijali. Effects of observer metamerism in the determination of human color-matching functions. *Color Research & Application*, 20(1):29–35, 1995.
[16] CIE. Fundamental chromaticity diagram with physiological axes - part 1. *CIE Publication No.170*, 2006.
[17] A. Sarkar, L. Blondé, P. L. Callet, F. Autrusseau, P. Morvan, and J. Stauder. Toward reducing observer metamerism issue in industrial applications: colorimetric observer categories and observer classification. In *Color Imaging Conference*, 2010.
[18] A. Sarkar. *Identification and Assignment of Colorimetric Observer Categories and Their Applications in Color and Vision Sciences*. PhD thesis, Université de Nantes, 2011.
[19] P. Morvan, A. Sarkar, J. Stauder, L. Blondé, and J. Kervec. A handy calibrator for color vision of a human observer. In *Multimedia and Expo (ICME), 2011 IEEE International Conference*, pages 1–4. IEEE, 2011.
[20] P.G. Engeldrum. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek Press, 2000.
[21] J. Neitz and G. H. Jacobs. Polymorphism of the long-wavelength cone in normal human colour vision. *Nature*, 1986.
[22] A. Stockman, L.T. Sharpe, et al. Tritanopic color matches and the middle-and long-wavelength-sensitive cone spectral sensitivities. *Vision research*, 40(13):1739–1750, 2000.
[23] M.D. Fairchild. A novel method for the determination of color matching functions. *Color Research & Application*, 14(3):122–130, 1989.
[24] F. Viénot, L. Serreault, and P.P. Fernandez. Convergence of experimental multiple rayleigh matches to peak l-and m-photopigment sensitivity estimates. *Visual neuroscience*, 23(3/4):419, 2006.

Author Biography

Yuta Asano is a Ph.D. student in Color Science program at Rochester Institute of Technology in Rochester, NY. He obtained his B.Eng. in Architecture (2009) and M.Eng. in Architectural Engineering (2011), both from Kyoto University in Kyoto, Japan.

Mark Fairchild is Associate Dean of Research and Graduate Education in College of Science, and Director of the Program of Color Science and Munsell Color Science Laboratory at Rochester Institute of Technology.

Laurent Blondé is a Principal Scientist at Technicolor, Rennes, France. In the Research & Innovation group he supports image science projects since 1988 on various technical and scientific topics including optics, CGI, digital video processing, TV broadcast, digital cinema, color management, and visual perception.