

# Method for Optimizing the Human Observer Functions

Christopher J. Edge; Eastman Kodak GCG Labs; Oakdale, MN USA

## Abstract

During the last 20 years, there has been discussion within the color research community regarding the accuracy of the  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  human observer functions and the variations among observers. Some studies have indicated disagreement between numeric vs. visual metameric matches when comparing white light sources or comparing hard copy vs. soft proofs. This paper describes a method for systematically improving the human observer functions in response to new data from proposed experiments that can be reproduced at multiple locations.

## Introduction

The CIE observer functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$ , which are used to calculate CIEXYZ, are the basis for all color measurements that require the matching of colors. By combining CIEXYZ with non-linear color appearance models (CAMs) such as CIELAB and CIECAM96s, complex color images as well as simple color patches can be reproduced with great success, at least between stimuli that do not contain significantly different spectral properties.

The raw color matching function data upon which the observer functions are based demonstrate significant differences between the individual observers, [1] whether due to true differences or other factors such as skill in color adjustment. At least some studies, such as those by Thornton, Shaw, Fairchild, and others have indicated differences between visual matches of metamer pairs and numerical matches calculated using the current human observer functions.[2-8]

One approach proposed by Fairchild and others to resolve the apparent discrepancy is to model the anticipated variations between observers due to such factors as age and field of view.

The method proposed in this paper attempts to optimize the current standard observer functions based on the average results from populations of observers where pairs of stimuli with dissimilar spectra were deemed to match. The resulting optimized functions are only moderately different from the current standard functions, yet can greatly improve the apparent discrepancies in these data sets.

Recognizing that there are many complex factors involved, including questions regarding instrument accuracy and whether 1 nm resolution or better is required, we recommend that new metameric matching pair data be acquired using newly available synthetic spectral light sources. This can permit replication of results at multiple locations to eliminate questions of data validity. This can also lead to improved estimates of the average color matching functions and statistical variation about the average for a large population of observers.

## Background

Plots of the human observer functions are well known. However, plots of the original CMF data upon which the observer functions are based are less commonly published. The following is an example of the individual CMFs of twenty observers at 10°

which are listed in the form of tables in the Appendix of Wyszecki:[9]

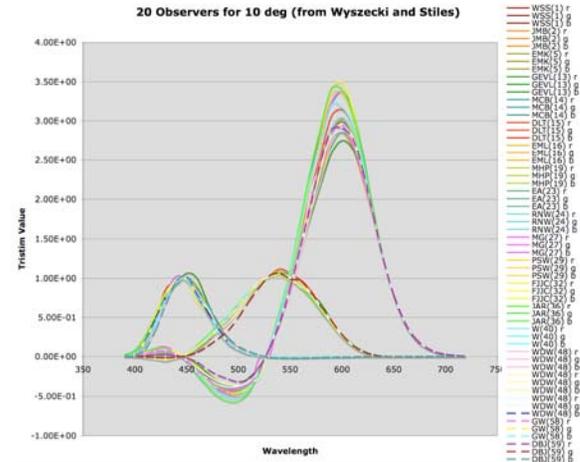


Figure 1. Raw CMF data of 20 observers from the work of Stiles and Burch in 1959

The human observer functions (both 2° and 10°) are linear transforms of the average CMFs from data sets that are similar to the above with regards to variation among observer data. Regardless of the source of the variation, one can easily see that there exists a +/- sigma uncertainty in estimating the average CMFs for the human population. In particular, the existing 2° standard observer functions are based on the work of Guild and Wright (performed in 1928 – 1931) involving only 17 individuals. Repeat plots of one individual such as figure 1 in North and Fairchild [10] would indicate that at least some of the variation may be regarded as random as opposed to systematic observer differences.

In the early 1990's, Thornton created a apparatus for performing metameric matches between a neutral white reference and spectral doublets created from thirty single-pass-band interference filters. [2-6] His results, which were summarized in part IV of his series of articles, indicated that visually matching pairs of white metamers could differ significantly in their chromaticities [4] which was apparently confirmed in the study performed by North and Fairchild [11]. However, whereas Thornton concluded that a revision to the CIE observer functions was required, North and Fairchild concluded that the variation was able to be modeled using the deviant observer function, which allows adjustment for such factors as age and field of view.

In 1995, Alfvén and Fairchild (RIT) performed similar experiments comparing fixed colors (prints and transparencies in a viewer) to adjustable colors on a CRT, followed by Shaw and Montag in 1998 and 1999 comparing red, green, blue, or blue, yellow, purple colors on a spinning disk to neutral gray reflective

samples in a viewer. This work was summarized and analyzed by Shaw in order to assess and possibly improve the human observer functions [7-8], due to the non-zero calculated  $\Delta E$ 's between the visually matching metamer pairs.

In light of the above experiments and the degree of variation between the multiple observers shown in figure (1), it is reasonable to consider the possibility of modifying or improving the human observer functions. Newly available synthetic spectral light sources should permit much easier replication of the worst-case examples of visual and numerical disagreement documented in the work performed by Thornton, North, Fairchild, et. al. Such tests should confirm the magnitude of systematic error, if any, in the observer functions as well as the sigma of the population due to observer metamerism, i.e. the difference between observers.

The method described below was developed based on historical data in anticipation of future experimental data that should be confirmed independently at multiple locations. The method can be then be used to improve the observer functions in the event that the data confirms the need to do so as well as to model the characteristics of individual observers.

## Modeling the Human Observer Functions

The CMF experiments of Guild and Wright (1931) and Stiles and Burch (1959) resulted in CMFs that were converted via linear transformations to the human observer functions that we know today. Later, Hunt cites the work of Estevez et. al. to define the Hunt-Pointer-Estevez matrix for converting the  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  observer functions to approximations of the red, green, blue cone responses of the retina known as  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  for long, medium, and short wavelength sensitivities (also referred to as  $\rho$ ,  $\gamma$ ,  $\beta$  to avoid confusion with the "L" of luminance):[12]

$$M_{XYZ \rightarrow LMS} = \begin{pmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0 & 0 & 1.00 \end{pmatrix} \quad (1)$$

It can be shown that the spectral cone responses  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  will uniquely determine whether colors match for a fixed choice of conversion between LMS and XYZ. This is helpful for optimizing the observer functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  because the cone responses  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  are simpler to model and each have only one maximum value, unlike the observer functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$ .

With this in mind, we consider whether the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  functions can be accurately modeled using a small set of parameters. If this is the case, one should be able to update the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  functions in a progressive manner using CMF data from 1931, 1955, as well as data from recently measured metamer pairs. One should be able to combine data from recent experiments without contradicting data from the original CMF experiments.

Studying the plots for  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  derived from  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$ , it seems clear that to first order they can be considered asymmetrical Gaussians, which would be typical behavior of a quantum transition which is Doppler-broadened by rotational and vibrational energy levels. The following very primitive model was used by the author of this paper to parameterize the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  functions:

$$f_{lms}(\lambda, \alpha_i, \lambda_i, \Delta\lambda_{i1}, \Delta\lambda_{i2}, \gamma_{i1}, \gamma_{i2}, \delta_{i1}, \delta_{i2}, \Delta\gamma_{i1}, \Delta\gamma_{i2}) \quad (2)$$

$$= \alpha_i (\delta_{i1} + (1 - \delta_{i1}) e^{-(\lambda - \lambda_i) / 2\Delta\lambda_{i1} (\gamma_{i1} + \Delta\gamma_{i1} |\lambda - \lambda_i|)}) \text{ for } \lambda < \lambda_i$$

$$= \alpha_i (\delta_{i2} + (1 - \delta_{i2}) e^{-(\lambda_i - \lambda) / 2\Delta\lambda_{i2} (\gamma_{i2} + \Delta\gamma_{i2} |\lambda_i - \lambda|)}) \text{ for } \lambda > \lambda_i$$

where  $i=0,1,2$  for  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$ . The parameter  $\lambda_i$  defines the wavelength of maximum sensitivity for  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$ . The parameter  $\Delta\lambda_{i1}$  defines the width of the quasi-Gaussian on the side where  $\lambda < \lambda_i$ ,  $\Delta\lambda_{i2}$  defines the width of the quasi-Gaussian for  $\lambda > \lambda_i$ . The exponents  $\gamma_{i1}$  and  $\gamma_{i2}$  (which are nominally of value 2 for a Gaussian distribution) in a similar fashion define the steepness of the curve shape for a given Gaussian-like width for  $\lambda < \lambda_i$  and  $\lambda > \lambda_i$ . The scaling parameter  $\alpha_i$  defines the relative height of the sensitivity for  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$ . In all the following calculations, we assume that for any set of parameters defining the shape of each  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  cone response, the amplitudes will be uniquely defined by normalizing the integrals  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  such that each are equal to 1.0, thereby ensuring an equal values of LMS when integrated with an equal energy response.

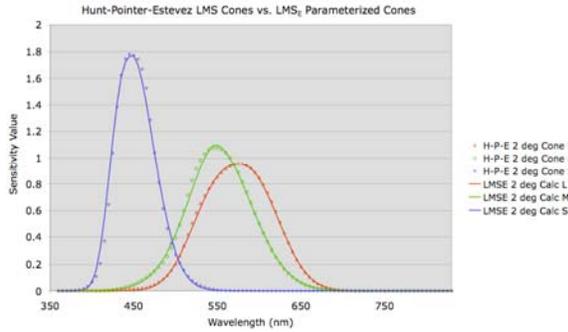
The values of  $\delta_{i1}$  and  $\delta_{i2}$  (which are nominally 0, and which are assumed to reduce to 0 outside the visible spectrum) allow control over the minimum value of the  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  functions – this is important since very small values of LMS  $\rightarrow$  XYZ can have a big impact on CIELAB due to the non-linear functions that define it. Finally, the correction parameters  $\Delta\gamma_{i1}$  and  $\Delta\gamma_{i2}$  (which are also nominally 0) allow a gradual increase or decrease in the power law of the exponent to optimize the correlation between the parameterized  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  and  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  vs. the existing standard functions. Like the other parameters, they correspond to  $\lambda < \lambda_i$  and  $\lambda > \lambda_i$  respectively.

Thus, a parameterized  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  can be created from a parameterized  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  via the matrix

$$M_{LMS \rightarrow XYZ} = M_{XYZ \rightarrow LMS}^{-1} \quad (3)$$

from equation (1) above. We will use  $\overrightarrow{xyz_E}(\lambda)$  to denote the modified human observer functions  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  and  $\overrightarrow{XYZ_E}$  to denote the integrated tristimulus values XYZ calculated from the color stimulus and  $\overrightarrow{xyz_E}(\lambda)$ .

Example plots for parameterized  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ ,  $\bar{s}(\lambda)$  are as follows:

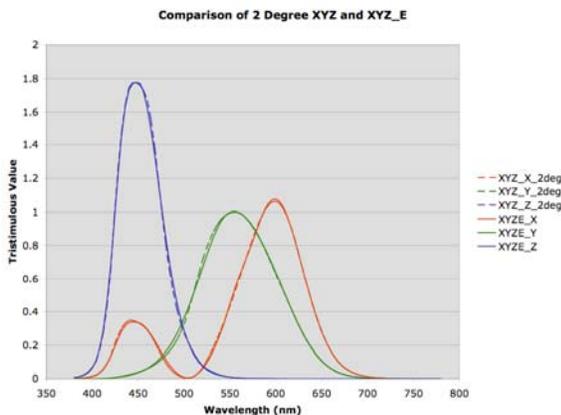


**Figure 2.** Comparison of LMS functions calculated from XYZ using the H-P-E matrix with parameterized cone responses

This simple parameterization of  $\bar{l}(\lambda), \bar{m}(\lambda), \bar{s}(\lambda)$  above gives surprisingly good results. A least squares fit was performed in order to optimize the parameters above using the spectral CMF data of Guild and Wright. The cost function to be minimized was a combination of  $\Delta E$  difference between CIE XYZ and  $\overline{XYZ}_E$  for monochromatic light for D50 illumination on a white reflector and for 101 monochromatic light source spectra of power 1/3 of the corresponding white reflector ranging from 380 nm to 730 nm and the difference between the spectral response of CIE XYZ and  $\overline{XYZ}_E$  with a weighting factor of 100 in order to correspond roughly to the range of CIELAB.

LSF error minimization using 4 nm increments gave an average error of 2.1  $\Delta E$  and a maximum error of 5.76  $\Delta E$  between CIE XYZ for the 2 degree observer and the  $\overline{XYZ}_E$  for D50 white and for the extreme case of the monochromatic light stimuli (noting that typical values of chroma were 100 – 250).

Test data comprising of 262 reflective spectral measurements of Matchprint™ Digital Halftone samples, including all permutations of 0%, 40%, 70%, and 100% tints for CMYK sampled at 10 nm increments were also used to compare CIE XYZ and  $\overline{XYZ}_E$ . Average and max errors of 0.2  $\Delta E$  and 0.6  $\Delta E$  respectively were calculated for the 262 test Matchprint™ colors. The following plot shows the comparison to the CIE  $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  functions:



**Figure 3.** Comparison of CIE XYZ with parameterized XYZ generated from cone responses

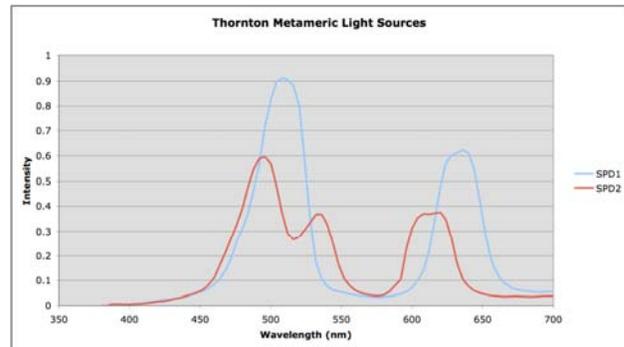
In a similar fashion, the parameterization was performed on the CIE 10° observer functions, resulting in an average and max error between the model and the standard observer of 1.63  $\Delta E$  and 5.1  $\Delta E$  for D50 white and for the monochromatic colors of the visible spectrum.

The fact that the model can simulate either the 2° or 10° observer to an average of 2  $\Delta E$  for the extreme case of monochromatic colors is a very good indicator that the model should be satisfactory for optimizing the existing CIE observers based on all available CMF data.

Note that when we apply the same quality criteria for comparing the consistency of the 2° and 10° standards to one another, we find that the two observers disagree by an average and maximum error of 11  $\Delta E$  and 76  $\Delta E$  for saturated colors. In this paper, we have not yet confirmed to what degree the different observers imply disagreement of metameric matches vs. differences in relating LMS to XYZ.

## Optimizing the Human Observer Functions

By digitization of plots from the published works for Thornton, matching white SPDs have been obtained. The plots digitized were figures 1 – 5 from his final paper in the series published in CRA [6]. Two of these data extractions were compared to the original SPD files that Thornton shared with Mark Shaw and which were used by Shaw in his calculations, with good correlation. The following example plot shows one pair of Thornton’s metameric light sources:



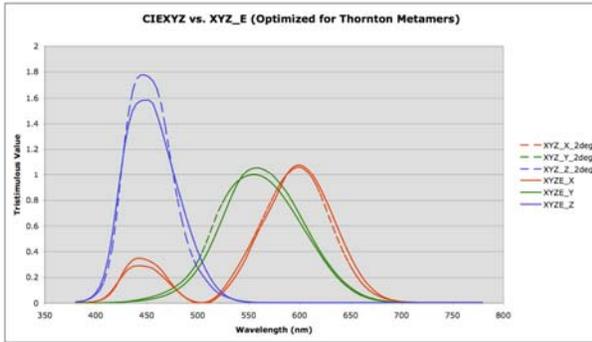
**Figure 4.** Example matching SPDs from Thornton experiment

As indicated by Thornton, the 2° observer calculates significant  $\Delta E$  differences between the 5 matching white light sources. The calculated values of  $L^*a^*b^*$  result in an average and maximum  $\Delta E$  error of 10 and 29 respectively.

Optimizing the parameters to minimize this large  $\Delta E$  error can be performed via LSF. Since the original CMF data was obtained using saturated colors and had significant variation as shown in figure (1) above, and since Thornton’s set of 5 metameric colors were all matching whites as confirmed by 8 observers (white balance being very sensitive to the eye), a weighting factor of 5 was chosen to give preference to minimizing the  $\Delta E$  error for the matching whites vs. the  $\Delta E$  error calculated from the CMF data of the 2° observer. This weighting factor was achieved by dividing each of the two summations in the cost function by  $N_l$  and  $N_s$  (number of wavelengths, number of

metameric pairs) in order to define the average sum squared error for each of the two summations, then multiplying the latter summation by 25.0 (i.e.  $5^2$  since a weighting factor of 5 in  $\Delta E$  implies a weighting factor of 25 in the square of  $\Delta E$ ).

The resulting minimization of the cost function gave the following plots of parameterized  $x(\lambda), y(\lambda), z(\lambda)$  vs. the 2° CIE standard observer:



**Figure 5.** Comparison of CIE XYZ functions to  $XYZ_E$  optimized with Thornton data

The improvement to the calculated  $\Delta E$ 's between the 5 matching metameric whites is significant: average  $\Delta E = 1.75$ , maximum  $\Delta E = 2.55$ .

## Preliminary Conclusions

Our preliminary conclusion for the above analysis is that even if future experiments confirm the significant  $\Delta E$  discrepancy indicated by the Thornton experiment, the existing CIE 2° observer can be slightly modified while maintaining reasonable agreement with the original data upon which it was based. If the Thornton experiment is confirmed not to be accurate, and if large observer differences are confirmed, the method described above should be helpful for characterizing the observer differences.

The author gratefully acknowledges Mark Fairchild of RIT, who connected the author to his former student Mark Shaw of HP, who in turn searched through boxes of archived research in order to find the data that made this analysis possible.

## References

- [1] Gunter Wyszecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2<sup>nd</sup> Edition, (John Wiley&Sons, NY, 1982), Figure 3(5.5.6), page 345.
- [2] W.A. Thornton, "Toward a more accurate and extensible colorimetry. Part I. Introduction. The visual colorimeter-spectroradiometer. Experimental Results," *Color Research and Application*, 17, 79-122 (1992).
- [3] W.A. Thornton, "Toward a more accurate and extensible colorimetry. Part II. Discussion," *Color Research and Application*, 17, 162-186 (1992).
- [4] W.A. Thornton, "Toward a more accurate and extensible colorimetry. Part III. Discussion (continued)," *Color Research and Application*, 17, 240-262 (1992).
- [5] W.A. Thornton, "Toward a more accurate and extensible colorimetry. Part IV. Visual Experiments With Bright Fields and Both 10° and 13° Field Sizes," *Color Research and Application*, 22, 189-198 (1997).
- [6] W.A. Thornton, "Toward a more accurate and extensible colorimetry. Part VI. Improved Weighting Functions. Preliminary Results," *Color Research and Application*, 23, 226-233 (1998).
- [7] Mark S. Shaw, *Evaluating the 1931 CIE Color Matching Functions*, RIT Master's Thesis, June 1999.
- [8] Mark S. Shaw and Mark D. Fairchild, "Evaluating the 1931 CIE Color Matching Functions", *Color Research and Application* 27, 316-329 (2002).
- [9] Gunter Wyszecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2<sup>nd</sup> Edition, (John Wiley&Sons, NY, 1982), pp. 817-821.
- [10] Amy North and Mark D. Fairchild, "Measuring Color Matching Functions. Part II. New Data for Assessing Observer Metamerism", *Color Research and Application*, 18, 163-170 (1993).
- [11] *Ibid.*, p.166
- [12] Robert Hunt, *The Reproduction of Colour*, (Fountain Press, England, fifth edition, 1995), p. 707.

## Author Biography

*Christopher Edge received his undergraduate degree and Ph.D. in physics from the University of Virginia in 1978 and 1988 respectively. He began his career with Kodak GCG (then 3M graphic prep systems) 1986 by helping to develop the optics and RIP for one of the first digital halftone proofing systems. During the 1990's he led the development of color technology for the highly successful Rainbow Desktop Proofer™, which became the basis for the Color Fidelity Module™ and Color Locking Software™ technologies. Beginning in 2000, he led the development of the core technologies for Kodak Matchprint Virtual™, which has inspired his recent efforts into the development of improved human observer functions.*