Evaluating the 1931 CIE Color Matching Functions

*Mark Shaw**
Munsell Color Science Laboratory
Rochester Institute of Technology
Rochester, NY 14623

Dr. Mark Fairchild
Munsell Color Science Laboratory
Rochester Institute of Technology
Rochester, NY 14623

*Current Affiliation – Applied Science Fiction, 8920 Business Park Drive, Austin, TX 78759

Abstract

The use of colorimetry within industry has grown extensively in the last few decades. Central to many of today's instruments is the work of the CIE system, established in 1931. Many have questioned the validity of the assumptions made by Wright [1] and Guild [2], some suggesting that the 1931 color matching functions are not the best representation of the human visual system's cone responses.

A computational analysis was performed to evaluate the CIE 1931 color matching functions against other responsivity functions using metameric data. The underlying principle was that an optimal set of responsivity functions would yield minimal color difference error between pairs of visually matched metamers. The difference of average color differences found in the six chosen sets of responsivity functions was small. The CIE 1931 2° color matching functions, on average, yielded the largest color difference, 4.56 ΔE*ab. The best performance come from the CIE 1964 10° color matching functions, which yielded an average color difference of 4.02 ΔE*ab.

An optimization was then performed to derive a new set of color matching functions visually matched using metameric pairs of spectral data. If one is to take all pairs, and perform an optimization that globally minimizes the average color difference, then one can hope to obtain an optimal set of responsivity functions. The optimum solution was to use a weighted combination of each of the different sets of responsivity functions. The optimized set, the 'Shaw and Fairchild' responsivity functions, was able to reduce the average color difference to 3.92 ΔE*ab.

The final part of the work was to build a computer-based simulation of the color differences between the different sets of responsivity functions. This simulation allows a user to load a spectral radiance, or spectral reflectance data file and display the tristimulus match predicted by each of the seven sets of responsivity functions.

Introduction

The use of colorimetry within industry has grown extensively in the last two decades. Some contributory factors have been the advancement of technology, hardware costs constantly decreasing, and the 'drive' for quality. Even though the technology is still in its infancy, many take for granted that instrumental tolerances will suffice and that human interaction can be minimized.

Central to many of today's instruments is the work of the CIE system (Commission Internationale de l'Éclairage) established in 1931. The CIE system allows the specification of color matches for a standard observer using color matching functions. These color matching functions for normal human observers are the fundamental basis of colorimetry. Studies performed by Wright[1] and Guild[2] are central to the work of the CIE, providing the foundation for the derivation of the CIE 1931 standard observer. Since the original work of the CIE, the standard has withstood an onslaught of technical pressures and remained a useful international standard for many years[3].

The 1931 Color Matching Functions were constructed from the relative color matching data of Wright[1] and Guild[2], with the assumption that they must be a linear combination of three functions including the 1924 CIE V(λ) luminous efficiency function. Stockman and Sharpe[4], Stockman, MacLeod and Johnson[5], and others, have questioned the validity of the assumptions made by Wright[1] and Guild[2]. Stockman and Sharpe[4] suggest that the 1931 color matching functions are not the best representation of the human visual system's cone responses. The implications of such questions are not known. Is the concern purely one of theoretical rigor, bearing no significance in the application of the current technology, or is the concern valid, one that may affect the whole of colorimetry as it stands to date?

If one is to assume that the CIE 1931 color matching functions are sufficient when they are not, then much of the work being done to build device models that minimize colorimetric error is attempting to attain an impossible goal. If the CIE 1931 color matching functions are not the ideal solution, would a new set of functions lead to better colorimetry? Or is the error within the bounds of statistical insignificance? This question is of great importance to the color science community. Applications that assume that the CIE 1931 color matching functions are sufficient, if they are not, will operate in error.

Existing works clearly document some of the issues that cause concern when using the CIE color matching functions and demonstrate their effects on the color community. Even so, little documented work has been done to compare the benefits of using the modified functions and sensitivities. The aim of this work was to evaluate the accuracy of the CIE 1931 color matching functions against more recent estimations and modifications, and determine whether the new cone fundamentals and color matching function derivations are truly better, or merely within the bounds of statistical insignificance. Six different sets of color matching functions and cone fundamentals were computationally analyzed using existing spectro-radiometric measurements of visually matched metameric pairs. The analysis indicated which set best approximated the observers' visual perception.

The work of the CIE

The CIE 1931 standard observer is based on two independent experiments performed by Wright[1] and Guild[2] that measured chromaticity coordinates for a total of 17 observers. Wright[1] used monochromatic primaries, whereas Guild[2] used broadband primaries. Since the primaries from one experiment can be specified in terms of tristimulus values from the other experiment, a linear transformation (3x3 matrix) is possible to change the results from the one set of primaries into the other. Thus a transformation was derived to convert both Wright and Guild's data into a set of common RGB primaries[3].

In order to eliminate the negative values in the color matching functions, the CIE derived a transformation to an imaginary set of primaries: XYZ. The goal of the transformation was to

eliminate the negative portions of the color matching functions and to designate one of the color matching functions to equal the CIE 1924 photopic luminous efficiency function, V(λ). Forcing one of the color matching functions to equal the V(λ) function served the purpose of incorporating the CIE system of photometry into the CIE system of colorimetry.

The validity of such decisions are still in much debate to date. Wright[6] suggests that no reference can be found to additivity experiments in the papers of Ives[7] and Guild[2] on the transformation of color mixture data, in spite of the fact that the assumption is fundamental to the legitimacy of such transformations. He writes that Guild, in his 1926 Survey, states: "Newton's law of mixtures follows directly from the geometry of the colour triangle. The law is that if any two colours are mixed, the colour produced lies on the line joining the two constituents on the colour chart and dividing it inversely as the quantities in the mixture." Guild was either relying on Newton's experiments to justify the principle, or claiming that the geometry of the colour triangle proved the law[6]. Wright adds that the additivity principle has been studied since 1931, and found to hold reasonably well for 2° color matching.

In 1951, Judd proposed a revision to the 1931 CIE standard observer[8], concluding that the CIE had given weight to early measurements that led to low average values at short wavelengths. Vos[9] has further refined Judd's modification by (i) making use of more precise computational procedures than those available to Judd in 1951, (ii) extending the data further into the far red, (iii) reducing by 0.2 log units the values of Judd's modified V(λ) function for wavelengths below 410nm, making them follow more closely to the values obtained by Stiles[10], (iv) slightly smoothing the CIE color matching functions from 380 to 400nm, and (v) truncating the color matching functions at 380nm, because the CIE 1931 data below that wavelength were extrapolated[9].

Wright believed that the integrity of the CIE 1924 luminous efficiency function V(λ) being incorporated into the color matching functions has not proved to be error free[6]. Stockman and Sharpe[4] reiterated this concern, "The 1931 CIE functions were constructed from the relative color matching data of Wright and Guild with the assumption that the Color Matching Functions must

be a linear combination of the V(λ) function. Not only is the validity of the V(λ) curve questionable, even after the corrections of Judd and Vos have been applied, but so too is the assumption that V(λ) must be a linear combination of the Color Matching Functions."

Alternative sets of Responsivity Functions

To save confusion between the differences of color matching functions and cone sensitivity functions, in this work both are called responsivity functions. Each set of responsivity functions claims to have its own benefits over the standard CIE set. Some are merely linear transformations of that derived by the CIE into cone space to better predict the cone responses, whilst others include modifications derived by the authors. A total of six sets were chosen to perform the computational analysis, including the work by the CIE in 1931; CIE in 1964; Stiles and Burch[10]; Demarco, Smith and Pokorny[11]; Stockman, MacLeod and Johnson[5]; and Vos and Walraven[12]. A tabulated set of the chosen responsivity functions can be found in the appendix. The most up to date sets can be found on the internet at the Color and Vision Research Laboratory home page: http://www-cvrl.ucsd.edu/

Experimental Data

This work utilized experimental data from previous visual experiments by Alfvin[13], Shaw and Montag[14], and Shaw and Montag[15]. The experiments shall not be discussed in great detail, as a more in depth summary can be found in the work of Shaw[16].

Alfvin and Fairchild[13] designed a visual experiment to permit observers to make critical color matches between prints or transparencies and a CRT display. With the use of a simple optical setup, consisting of an equilateral glass prism mounted on an optical bench, observers were able to simultaneously view both the soft and hard-copy matching stimuli. Both the CRT display and combination light-booth/light box were aligned with the optical prism and shielded from the observer. 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 darkened room. A neutral translucent diffusion material placed in

front of the CRT display eliminated the appearance of any visual texture in the soft copy stimulus, rendering the soft-copy stimulus identical to the hard-copy stimuli in terms of spatial characteristics. Observers were seated at a distance of approximately one meter from the matching stimuli. The 5x5 cm matching field subtending a visual angle of 2.9°. The observers were asked to adjust the color appearance of the soft-copy stimulus to create an exact match for each of the fourteen different hardcopy stimuli[13].

Shaw and Montag[14] designed a visual experiment to allow observers to perform a color match between a gray card of Munsell N5 and an ACS VCS 10 additive mixing device. The ACS VCS 10 consisted of seven colored discs, all rotating at high speed to simulate an additive integral visual response. An observer using the controls could then adjust the proportions of each colored disc to simulate the gray card. The seven discs in the ACS VCS 10 were White, Red, Green, Blue, Yellow, Purple and Black. Independent control was allowed of any three primaries at any one time by the user control panel. The goal was to generate a metameric match between the Munsell N5 paper and the three primaries. Of the five colored primary discs, two sets of three Primaries were chosen – Red, Green, Blue (RGB) and Blue, Yellow, Purple (BYP).

Observers participated in the Shaw and Montag[14] color matching experiment to assess inter- and intra-observer variability. Each observer performed the color matching experiment 10 times in succession for each primary triplet.

Using an identical experimental setup to that of Shaw and Montag[14], Shaw and Montag[15] repeated the visual experiment using different stimuli, and different primary sets. The stimulus was a neutral gray of $L^*=50$, created with a Fujix Pictrography 3000 color printer. The Fujix printer is a hybrid photographic/thermal-transfer continuous-tone digital printer. The primary sets chosen were Red, Green, Blue (RGB) and Green, Yellow, Purple (GYP).

Computational Analysis

One of the key underlying assumptions of this work is that an optimal set of responsivity functions will predict that the integrated response from a metameric pair are equal. This is

illustrated in Figure 1, yielding a minimum color difference in a color space like CIELAB. One can therefore evaluate the performance of a set of responsivity functions using metameric data, by calculating the color difference over all the metameric pairs in that color space. A problem arises when one wishes to compare the performance of different sets of responsivity functions. It is not sufficient to assume that one can use the CIELAB color space as the comparative space by just transforming the tristimulus values calculated by each set of color matching functions into CIELAB co-ordinates using the standard equations. One must find a common color space in which each set can be compared.

FIGURE 1

One must therefore consider how to compare the different sets of responsivity functions. One approach would be to derive a new color space in which all of the responsivity functions can be compared, but this introduces a new quality metric that may bear no significance to existing standards. An alternative approach is to assume an existing color space as standard, and transform all other sets of responsivity functions into the closest linear approximation.

Thus the problem of disparate responsivity functions was overcome by assuming the common color space to be CIELAB, using the CIE 1931 standard observer. In order to evaluate the other sets of functions accurately, a linear transformation was calculated to transform each set into an approximate CIE representation. A 3x3 transformation matrix was calculated for each set of color matching functions using least squares, shown in Equation 1. The transformed responsivity functions were then used to calculate a set of 'pseudo' tristimulus values for each set of functions.

$$\begin{bmatrix} x_{400} & y_{400} & z_{400} \\ \vdots & & \vdots \\ x_{700} & y_{700} & z_{700} \end{bmatrix}_{\substack{'pseudo' \\ xyz\ functions}} = \begin{bmatrix} r_{400} & g_{400} & b_{400} \\ \vdots & & \vdots \\ r_{700} & g_{700} & b_{700} \end{bmatrix}_{\substack{responsivity \\ functions}} \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}_{\substack{3x3 \\ transform}} \quad (1)$$

$$\mathbf{XYZ} = \mathbf{RGB} \cdot \mathbf{A} \qquad \text{where } \text{pinv}(\mathbf{W}) = \left(\mathbf{W}^{T} \cdot \mathbf{W}\right)^{-1} \mathbf{W}^{T}$$

$$\mathbf{A} = \mathit{pinv}(\mathbf{RGB}) \cdot \mathbf{XYZ}$$

By transforming each set into a CIE approximation, it enables one to use the standard quality metrics, including CIE ΔE_{ab}, and CIE ΔE_{94}. One must remember, though, that the CIELAB color space is not optimized for the other color responsivity functions, but can be considered a consistent color space for comparison. Any systematic shifts introduced by the cube-root transformation, or Von-Kries type chromatic adaptation normalization will be consistent for all responsivity functions.

The transformation of the Stiles set of responsivity functions is demonstrated in Figure 2. The diagram shows the original functions, the 3x3 transformation matrix, and the transformed Stiles and Burch functions overlaid on the CIE 1931 responsivity functions.

Tristimulus integration was used to calculate the tristimulus values for each of the metamer pairs from the spectral data. The 4nm wavelength increment spectral data was linearly resampled to a 5nm wavelength increment, and the wavelength range cropped to 400-700nm. The wavelength increment was chosen to correspond to that of the various sets of responsivity functions. Reference white tristimulus values for the measured radiance data were unknown. Therefore, since the gray patch in each data set had an L* of approximately 50, the reference white can be approximated using the tristimulus values of the reflectance spectra scaled by 5. CIELAB coordinates were calculated according to standard CIE methods, using each of the transformed sets of responsivity functions. CIE ΔL^{*}, Δa^{*}, Δb^{*}, ΔC^{*}, ΔH^{*}, ΔE_{ab} and ΔE_{94} values were calculated for each metameric pair.

Figure 2

Observer Variance

When considering metameric data from multiple observers, one must also consider the implications of observer variability. Inter-observer variations occur through differences in observer macular pigment, lens absorptions, and other pre-receptoral features. For the purpose of the work it was assumed that inter-observer variability was inherent within the data. The observer variance is considered an offset parameter that is present within the color difference results, and the best set of responsivity functions will minimize the color difference as well as possible. It is therefore important that one understands that the expectation of the work was not that an optimal set of responsivity functions would yield $0\Delta E^*_{ab}$ over all samples, but that it would yield the lowest mean color difference. Readers interested in understanding further the effects of inter- and intra-observer variability are advised to look to the work of North[17] and Alfvin[13].

Statistical Analysis

In order to evaluate the performance of each set of responsivity functions, two tailed t-tests were used. The t-test was used to compare the mean color difference vector (ΔL^*, Δa^*, or Δb^*) against a mean of zero. An ideal set of responsivity functions would yield a mean ΔL^*, Δa^*, Δb^* of zero, plus an offset for observer variance. A sample covariance matrix, S_{Lab}, defined by the sample covariances and variances of the ΔL^*, Δa^*, and Δb^* values, as shown in Equation 2, was calculated for each set of responsivity functions.

$$S_{Lab} = \begin{bmatrix} s^2_{\Delta L} & s_{\Delta L} s_{\Delta a} & s_{\Delta L} s_{\Delta b} \\ s_{\Delta a} s_{\Delta L} & s^2_{\Delta a} & s_{\Delta a} s_{\Delta b} \\ s_{\Delta b} s_{\Delta L} & s_{\Delta b} s_{\Delta a} & s^2_{\Delta b} \end{bmatrix} \qquad (2)$$

Assuming a multivariate normal distribution, the inverse of the sample covariance matrix S_{Lab} can be used to construct a 95% confidence region for the sample distribution of the CIE ΔL^*, Δa^*, and Δb^* multivariate data set[11]. An example of a Δa^*-Δb^* bivariate ellipse bound by a 95% confidence region for the sample distribution calculated by Equation 2 is shown in Figure 3, relative to the CIE 2° Standard Observer.

Figure 3

In addition to the Δa^*-Δb^* relationship shown in Figure 3, one should also consider the ΔL^*-Δa^*, and ΔL^*-Δb^* planes because the experimental color matches involved adjustments to each of the three independent variables defining the CIELAB color space. If any one of the three confidence regions centering at the sample means of the Δa^*-Δb^*, ΔL^*-Δa^* or ΔL^*-Δb^* planes do not contain the theoretical mean match for a given standard observer, the mean color matches are considered to be statistically significantly different from zero[11].

Results and Discussion

The computations discussed in the last paragraph were performed on each of the data sets individually, as well as all three data sets combined. As well as the final six sets of responsivity functions chosen, the Shaw and Fairchild set of responsivity functions derived in the work have been included in the results for ease of comparison. The numerical results and confidence ellipse plots for each data set can be found in the full report, available for download from the Munsell Color Science Laboratory home page[16].

Alfvin Data Set

As discussed earlier, the Alfvin 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 chosen by Alfvin as fixed matching stimuli. The seven colors were, red, green, blue, gray, cyan, magenta, and yellow. In order to evaluate how the region of color space also varies with the different sets of responsivity functions, the computation was performed on a color center basis also.

Combined Alfvin Data Set

The total Alfvin data set, containing all seven color centers, is comprised of 268 metamer pairs. The combined evaluation utilized all metamer pairs to determine the performance of each set of responsivity functions. A summary of the results can be found in Table I.

Table I

The results show that little difference can be found between the different sets of responsivity functions for the Alfvin Data. The CIE 1931 functions yield an average color difference of 4.39 ΔE_{ab}, 2.72 ΔE_{94}, with the best set of responsivity functions being that of Demarco, Smith and Pokorny[11], yielding an average color difference of 4.33 ΔE_{ab}, 2.67 ΔE_{94}.

The Shaw and Fairchild responsivity functions derived by optimization yield an interesting result, the average color difference is the second highest. This indicates that the contribution of the Alfvin data set to the optimization may have been less significant than the other data sets, with the optimization minimum being found even though an increase in the Alfvin data set occurred.

The statistical tests performed on the data show some interesting features. A two tailed t-test of the means was applied on each of the ΔL^*, Δa^*, Δb^* values with the null hypothesis that the means are equal to zero. The results shown in Table 2 indicate that the CIE 1931 functions failed the test in both the Δa^* and Δb^* planes, showing that there are systematic deviations from a mean of zero, implying that the functions are not optimal. The results of the Shaw and Fairchild responsivity functions fail to reject the null hypothesis in all three dimensions, indicating that the mean cannot be shown to deviate from zero.

Table 2

The Shaw and Fairchild confidence ellipse plots are shown in Figure 4, showing the 95% confidence ellipse and mean confidence ellipse for each of the CIELAB planes. It is clearly evident that the mean ellipse is far smaller than the variation of the data, in all cases. The shapes of the ellipses between different sets of responsivity functions change somewhat, but the change is usually only slight and of little significance.

Figure 4

Segmented Alfvin Data Set

The segmented analysis was done both by color center and by media type, yielding 14 separate analyses. Table III summarizes the average color difference results (ΔE_{ab}) for each separate analysis. It can be seen that some color centers yielded higher average color difference results than others, the blue transparency and yellow print being good examples.

Table III

Overall, the majority of mean color differences tended to vary around 3-4 ΔE_{ab}. This indicates that observer variance exists, irrespective of color center. The overall performance of each set of responsivity functions (Mean of means) is shown at the bottom of Table 3. One can see that the CIE 1931 color difference is lower than the Shaw and Fairchild color difference for this data set, again confirming the suspicion that the optimization minimization was not necessarily the optimal minimum for the Alfvin data set.

Shaw and Montag 98 and Shaw and Montag 99 Data Sets

The complete set of results and confidence ellipse plots of the Shaw and Montag 98[14] data, and the Shaw and Montag 99[15] data can be found in the full report[16]. Both sets of results exhibit interesting trends and have been summarized in Tables IV and V.

Table IV

Table V

The average color differences for both data sets is lowest when calculated using the CIE 1964 10° functions and the Shaw and Fairchild functions. This is interesting to note, since the general trend in the Alfvin data set was an increased color difference for both the CIE 1964 and Shaw and Fairchild functions. Also, the performance improvement is quite substantial, in both cases. Although the CIE 1964 functions provide the lowest average color difference, the statistical results show systematic deviations from a mean of zero in at least two of the three dimensions. This indicates that, although the average color difference is low, it is offset from a mean of zero. This is not surprising because the confidence ellipse plots of some of the responsivity functions,

shown in Figure 5, show that two distinct clouds exist. One possible reason can be linked to the two primary sets (of the additive mixture device) used to match the reference colors. The Shaw and Fairchild set also exhibit systematic deviations in two of the three dimensions, a similar explanation can be made.

Figure 5

All Three Data Sets Combined

It can be seen from Table VI that the Shaw and Fairchild responsivity functions yield the lowest color difference, averaged over all of the samples in the combined data set. But, it is clear that the difference between each of the sets is only slight, ranging from 4.56 ΔE^*_{ab} at worst, to 3.92 ΔE^*_{ab} at best.

Table 6

When looking at the statistical results, it was clear that no one set of responsivity functions performed optimally in all three dimensions. There was always at least one dimension in which it was possible to reject the null hypothesis. Although the statistics tend toward the decision that the means are significantly different from zero, one should also consider that in all cases the ΔL^*, Δa^*, Δb^* means were not far off zero.

The Shaw and Fairchild confidence ellipse plot for the combined data sets are shown in Figure 6. The mean confidence ellipse is very small in comparison with the outer ellipse, which is a result of the number of samples in the data set.

Figure 6

Thornton Metameric Pair

The recent works of Thornton[16-21] have focused on determining an observer's color matching functions using several sets of primary lights for each observer. Using one of Thornton's data

points – a pair of spectral power distributions that stress the deficiency of the standard observer (shown in Figure 7), the performance of each set of responsivity functions was calculated.

Table 7

Table VII shows the results of the computational analysis, it can be seen that the CIE 1964, and Shaw and Fairchild responsivity functions performed worst. Indicating that the Demarco, *et al.* and Vos and Walraven responsivity functions performed the best. It is important to note, though, that the Thornton data set only consisted of one metameric pair that was considered equivalent by eight observers. A more rigorous data set is needed to make any conclusive decisions.

Figure 7

Derivation of Optimized Responsivity Functions

The results discussed above have shown that the optimized weighting function performs very well on the Shaw and Montag 98 and Shaw and Montag 99 data sets, but not so well on the Alfvin data set. The result was surprising since the Alfvin data set comprised over half the samples in the combined data set on which the optimization was performed. It is also evident that the Shaw and Fairchild responsivity functions exhibit similarities to the CIE 1964 10° color matching functions, producing similar results in all three data sets. One could conceivably improve the fit of the Shaw and Fairchild responsivity functions to the Alfvin data by weighting the optimization.

Deriving an Optimized set of Responsivity Functions

The proceeding discussion documented the performance of the six sets of responsivity functions. Of the six sets, five were previously defined, and one was derived by the authors to satisfy the experimental objectives set forth in this work, the Shaw and Fairchild color matching functions. This section documents the approach taken in deriving the new set of color matching functions.

It is possible to consider that an optimal set of responsivity functions can be derived from visual color matching data by a modification of an existing set of responsivity functions using the visual data to weight the adjustments. Using the three combined metameric data sets, an optimization was performed to minimize the average color difference over all observations.

Thirteen different approaches were taken to find an optimal set of responsivity functions that best described the metameric data. A complete discussion of all the different techniques tested can be found in the full report[16].

Of the techniques tried it was immediately clear that some were unsuitable, those using linear regression, minimizing tristimulus error, performed worst. They could potentially be improved by weighting the radiance spectra with Neugebauer's CQF weighting[24, 25], but one would not expect an exceptional improvement. An ideal optimization would minimize color difference in a perceptual color space, such as CIELAB. The conversion of tristimulus values into CIELAB coordinates includes the cube-root function that is not accounted for in the tristimulus minimization. Other reasons for the failure of some techniques can be attributed to the dependency of the data itself. Ideally, when using a linear regression technique one assumes that each variable is independent. When using spectral data, an optimization of all wavelengths would require independence of all wavelengths, which is clearly not the case in the color matches. In Alfvin's experiment the color matches were created with a CRT, thus only giving three dimensions in which that data can vary. The same is true for the Shaw and Montag 98, and Shaw and Montag 99 experiments, with matches being made with only three primaries.

Of all the approaches tried, the optimization using existing sets of responsivity functions performed the best. The optimization function used the Levenberg-Marquardt technique, allowing constraints to be applied to different parameters in the optimization. The cost function and constraint applied are is shown in Equation 2.

$$\bar{x}_{optimum} = \sum_{j=1}^{6} w_j C_j \qquad \longrightarrow \qquad w_j = \text{Arg}\min \Delta E_{\text{rotated}}$$

where

$\bar{x}_{optimum}$ = Optimized weighting function (3)

w_j = jth Weight

C_j = jth Transformed Responsivity Function

constraint

$$w_1 = 1 - w_2 - w_3 - w_4 - w_5 - w_6$$

Where $\bar{x}_i$ is the i^{th} weighting function under evaluation, ($\bar{x}$, $\bar{y}$, or $\bar{z}$), $w_{i,j}$ is the j^{th} weight for the i^{th} weighting function that scales C_j, the j^{th} responsivity function. The optimization was constrained such that the weights $w_{i,1\text{-}6}$ sum to one, and the solution being found when a set of weights ($w_{i,j}$) were found that minimized the average color difference between the reference and sample CIELAB values.

This approach did not yield the lowest color difference of all the techniques, but it did yield the most realistic set of weighting functions. The final set of functions can be seen in Figure 8, resulting in an average color difference of 3.921ΔE^*_{ab} over all samples.

Figure 8

Computer Simulation

A computer simulation was written to visualize the differences between the different sets of responsivity functions. If one is to think of a visual experiment in which an observer is seated in front of a viewing booth and CRT, (see Figure 9). A spectroradiometer is positioned by the observers left shoulder, measuring the spectral radiance of samples positioned in the booth. The observer places a sample in the viewing booth, under the controlled viewing conditions, and measures the spectral radiance of the sample. The sample spectra is then passed to the computer and the simulation program displays the predicted CRT metameric match for that spectral

radiance, using a spectral CRT model and each set of responsivity functions. The observer can then look at the prediction of each set of responsivity functions and can study the relative differences between predictions. The influences of cross-media color matching are not considered to be of major importance here since the observer is comparing the different response functions displayed simultaneously on the screen.

Figure 9

In order to model the CRT, enabling one to display the results of different sets of responsivity functions, it was important to use a spectral model. Measuring the CRT primaries using a tristimulus colorimeter is insufficient since a characterization of the primaries must be done relative to each set of responsivity functions. Figure 10 outlines the process involved in characterizing the display for each set of weighting functions.

The spectral radiance of each primary, flare characteristics, and channel crosstalk properties were measured using a PhotoResearch PR650 spectroradiometer incorporating a half-height triangular bandpass of 4nm, recording the spectral radiance at 4nm wavelength intervals across the visible spectrum between 380 and 780nm.

Figure 10

Using each set of responsivity functions transformed into CIE representation, the tristimulus values were calculated. Then, for each set of tristimulus values, the CRT model parameters were calculated. First subtracting the flare, then building a flare free XYZ to RGB transform matrix, and finally determining the GOG model parameters. (Berns[26], CIE[27], and Berns *et al*[28]).

In order to display a spectral measurement on the CRT, the spectral radiance data was first used to calculate tristimulus values for each set of responsivity functions. Each set of tristimulus values had the flare subtracted by its respective flare measurement and the different sets of flare-free tristimulus values were then transformed into RGB scalars using the respective transform. The GOG model then converted each set of RGB scalars into RGB digital code values to be displayed

by the simulation. Figure 11 is an example of the same spectral radiance data modeled by each of the sets of responsivity functions, as displayed to the user.

Figure 11

Conclusions

A computational analysis was performed to evaluate the 1931 color matching functions against other responsivity functions using metameric data. The underlying principle being was that an optimal set of responsivity functions will yield minimal tristimulus error between a pair of visually matched metamers.

A common color space was used in order to compare the performance of the different responsivity functions, the color space used was CIELAB, based upon the CIE 1931 2° standard observer functions. The five other sets of responsivity functions were transformed using a linear 3x3 matrix into near-CIE approximations. The transformed functions were used to calculate 'pseudo' tristimulus values for each set, from which CIELAB coordinates were calculated using standard procedures. Color differences were calculated between each pair of CIELAB coordinates for each metameric match using the standard ΔE_{ab}, and ΔE_{94} color difference formulae.

The differences between the average color differences found in the six sets of responsivity functions were small. The CIE 1931 2° color matching functions, on average, provided the largest color difference, 4.56 ΔE_{ab}. The best performance came from the CIE 1964 10° color matching functions, yielding an average color difference of 4.02 ΔE_{ab}.

An optimization was then performed on the CIE 1931 color matching functions, using the concept that color differences between metamers can be used to improve predictions of color matching functions. If one is to take all pairs, and perform an optimization that globally minimizes the average color difference, then one can hope to obtain an optimal set of responsivity functions.

A total of thirteen optimization techniques were tested, but only two were found that were capable of both maintaining the integrity of the color matching functions, and reducing the average color difference. The optimal solution, the 'Shaw and Fairchild' responsivity functions, were able to reduce the average color difference to 3.92 ΔE^*_{ab}, using a weighted combination of each of the different sets of responsivity functions.

The final part of the work was to build a computer-based simulation of the color differences between the different sets of responsivity functions. The simulation allows a user to load a spectral radiance, or reflectance, data file and display the tristimulus match predicted by each of the seven sets of responsivity functions.

This work provides insight into many areas of color science. One can conclude from the work that the work of the CIE has stood its ground for the last sixty years, and is still, without doubt, a standard that performs well when compared with more recent research. It is clear that the standard is not perfect, having its areas of weakness.

It was found that the magnitude of observer variability was nearly eight times that of the variability found between the responsivity functions. One can conclude that one should be more concerned with the problems introduced by observer metamerism than the accuracy of the CIE 1931 functions themselves.

References

1. Wright W.D, (1928-29) A re-determination of the trichromatic coefficients of the spectral colours, Trans. Opt. Soc. 30, pp. 141-161

2. Guild J, (1931) The Colorimetric Properties of the Spectrum, Phil. Trans. Roy. Soc. (London), A 230, p. 149-187

3. Fairchild M.D, (1998) Color Appearance Models, Addison Wesley, Reading, Massachusetts, USA

4. Stockman A, Sharpe L.T, (1998) Human cone spectral sensitivities: a progress report, Vision Research, 38, 19, pp. 3193-3206

5. Stockman A, Macleod D.I.A and Johnson N.E, (1993) Spectral sensitivities of the human cones, Optical Society of America, 10, No 12, pp. 2491-2516

6. Wright W.D, (1981) 50 Years of the 1931 CIE Standard Observer for Colorimetry, AIC Colour 81, Paper A3

7. Ives H.E, (1923) The transformation of color-mixture equations from one system to another, J. Franklin Inst., vol 180, pp. 673-701 (1915) and vol 195, pp. 23-44 (1923)

8. Judd D.B, (1951) Colorimetry and Artificial daylight, Proc. 12th session CIE, Stockholm, I, Tech. Committee No 7, p. 11

9. Vos J.J, (1978) Colorimetric and Photometric Properties of a 2° Fundamental Observer, Col. Res. and Appl., 3, No. 3, pp. 125-128

10. Stiles W.S. and Burch J.M, (1955) Interim report on the Commission Internationale de l'Eclairage, Zurich, on the National Physical Laboratory's Investigation of Colour Matching, Optica Acta, 2, pp. 168

11. DeMarco, P., Pokorny, J., & Smith, V. C. (1992). Full spectrum cone sensitivity functions for X-chromosome linked anomalous trichromats. Journal of the Optical Society of America A, 9, 1465-1476.

12. Vos J.J and Walraven P.L, (1971) 'On the derivation of the foveal receptor primaries,' Vision Research, 11, pp. 799-818

13. Alfvin R, (1995) A computational analysis of observer metamerism in cross-medi color matching, MSc Thesis, Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, NY 14623 USA

14. Shaw M.Q and Montag E.D, (1998) Analyzing Observer Metamerism, Color Measurement Laboratory Project, Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, NY 14623 USA. Available from Author

15. Shaw M.Q and Montag E.D, (1999) Analyzing Observer Metamerism, Color Measurement Laboratory Project, Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, NY 14623 USA. Available from Author

16. Shaw M.Q, (1999) Evaluating the CIE 1931 Color Matching Functions, MSc Thesis, Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, NY 14623 USA. http://www.cis.rit.edu/mscl/ (see Publications - Student Research)

17. North A, (1991) Investigation of observer variability using a new method for determining color matching functions, MSc Thesis, Munsell Color Science Laboratory, Rochester Institute of Technology

18. Thornton W.A, (1992a) Toward a more accurate and extensible colorimetry. Part 1 – Introduction. The visual Colorimeter-Spectroradiometer. Experimental results, CR&A, **17**, No.2, pp. 79-121

19. Thornton W.A, (1992b) Toward a more accurate and extensible colorimetry. Part II – Discussion, CR&A, **17**, No.3, pp. 162-186

20. Thornton W.A, (1992c) Toward a more accurate and extensible colorimetry. Part III - Discussion, CR&A, **17**, No.4, pp. 240-262

21. Thornton W.A, (1997) Toward a more accurate and extensible colorimetry. Part IV – Visual experiments with bright fields and both 10° and 1.3° field sizes, CR&A, **22**, No.3, pp. 189-198

22. Thornton W.A, (1998b) Toward a more accurate and extensible colorimetry. Part VI – Improved weighting functions. Preliminary results, CR&A, **23**, No.4, pp. 226-233

23. Thornton W.A, Fairman H.S, (1998a) Toward a more accurate and extensible colorimetry. Part V – Testing visually matching pairs of lights for possible rod participation on the Angular-Stiles model, CR&A, **23**, No.2, pp. 92-103

24. Neugebauer H.E.J., (1956) Quality factor for filters whose spectral transmittances are different from color mixture curves and its application to color photography, *Jour. Opt. Soc. Amer.* **46,** No. 821, pp. 821-824

25. Engeldrum P.G, (1993) Color scanner colorimetric design requirements. *SPIE*. **1909**, pp. 75-83

26. Berns R.S, (1996) Methods for characterizing CRT displays, Displays, **16**, No. 4, pp. 173-182

27. CIE, (1996) The relationship between digital and colorimetric data for computer-controlled CRT displays, Central Bureau of the CIE, Vienna, CIE Publication No. 122

28. Berns R.S, Motta R.J and Gorzynski M.E, (1993) CRT Colorimetry. Part I: Theory and Practice, Col. Res. Appl., **18**, No. 5, pp. 299-325

Original Sets of CMFs / Cone Fundamentals

	CIE Lab 1931			CIE Lab 1964			Stiles			Demarco			Stockman			Vos		
Wavelength	X	Y	Z	X	Y	Z	r	g	b	l	m	s	l	m	s	l	m	s
400	1.43E-02	4.00E-04	6.79E-02	1.91E-02	2.00E-03	8.60E-02	9.63E-03	-2.17E-03	6.24E-02	2.66E-03	2.82E-03	1.08E-01	3.26E-03	2.97E-03	6.23E-02	1.52E-03	1.42E-03	1.08E-01
405	2.32E-02	6.00E-04	1.10E-01	4.34E-02	4.50E-03	1.97E-01	1.90E-02	-4.43E-03	1.32E-01	4.38E-03	4.75E-03	1.79E-01	6.69E-03	6.38E-03	1.32E-01	2.49E-03	2.41E-03	1.79E-01
410	4.35E-02	1.20E-03	2.07E-01	8.47E-02	8.80E-03	3.89E-01	3.08E-02	-7.20E-03	2.28E-01	6.89E-03	7.67E-03	2.85E-01	1.15E-02	1.14E-02	2.27E-01	3.89E-03	3.95E-03	2.85E-01
415	7.76E-02	2.20E-03	3.71E-01	1.41E-01	1.45E-02	6.57E-01	4.25E-02	-1.26E-02	3.59E-01	1.08E-02	1.24E-02	4.53E-01	1.59E-02	1.67E-02	3.59E-01	6.07E-03	6.52E-03	4.53E-01
420	1.34E-01	4.00E-03	6.46E-01	2.05E-01	2.14E-02	9.73E-01	5.17E-02	-1.67E-02	5.24E-01	1.58E-02	1.89E-02	6.59E-01	2.23E-02	2.58E-02	5.24E-01	8.88E-03	1.03E-02	6.59E-01
425	2.15E-01	7.30E-03	1.04E+00	2.65E-01	2.95E-02	1.28E+00	5.28E-02	-2.12E-02	6.86E-01	2.00E-02	2.54E-02	8.13E-01	2.64E-02	3.39E-02	6.85E-01	1.14E-02	1.48E-02	8.13E-01
430	2.84E-01	1.16E-02	1.39E+00	3.15E-01	3.87E-02	1.55E+00	4.43E-02	-1.99E-02	7.96E-01	2.33E-02	3.17E-02	9.08E-01	3.05E-02	4.34E-02	7.96E-01	1.38E-02	1.99E-02	9.08E-01
435	3.29E-01	1.68E-02	1.62E+00	3.58E-01	4.96E-02	1.80E+00	3.22E-02	-1.61E-02	8.95E-01	2.68E-02	3.95E-02	9.77E-01	3.52E-02	5.43E-02	8.94E-01	1.65E-02	2.67E-02	9.77E-01
440	3.48E-01	2.30E-02	1.75E+00	3.84E-01	6.21E-02	1.97E+00	1.48E-02	-7.35E-03	9.64E-01	3.01E-02	4.77E-02	1.00E+00	4.12E-02	6.75E-02	9.64E-01	1.96E-02	3.47E-02	1.00E+00
445	3.48E-01	2.98E-02	1.78E+00	3.87E-01	7.47E-02	2.03E+00	-2.34E-03	1.37E-03	9.98E-01	3.24E-02	5.55E-02	9.70E-01	4.56E-02	7.79E-02	9.98E-01	2.22E-02	4.28E-02	9.70E-01
450	3.36E-01	3.80E-02	1.77E+00	3.71E-01	8.95E-02	1.99E+00	-2.91E-02	1.96E-02	9.19E-01	3.43E-02	6.35E-02	9.10E-01	5.00E-02	8.84E-02	9.19E-01	2.47E-02	5.17E-02	9.10E-01
455	3.19E-01	4.80E-02	1.74E+00	3.43E-01	1.06E-01	1.90E+00	-6.07E-02	4.35E-02	8.25E-01	3.68E-02	7.31E-02	8.50E-01	5.69E-02	1.03E-01	8.26E-01	2.78E-02	6.20E-02	8.50E-01
460	2.91E-01	6.00E-02	1.67E+00	3.02E-01	1.28E-01	1.75E+00	-9.62E-02	7.10E-02	7.86E-01	4.12E-02	8.60E-02	7.99E-01	6.81E-02	1.25E-01	7.87E-01	3.28E-02	7.57E-02	7.99E-01
465	2.51E-01	7.39E-02	1.53E+00	2.54E-01	1.53E-01	1.55E+00	-1.38E-01	1.10E-01	6.67E-01	5.02E-02	1.07E-01	7.75E-01	8.33E-02	1.52E-01	6.69E-01	4.15E-02	9.55E-02	7.64E-01
470	1.95E-01	9.10E-02	1.29E+00	1.96E-01	1.85E-01	1.32E+00	-1.75E-01	1.51E-01	6.11E-01	6.27E-02	1.30E-01	6.89E-01	1.03E-01	1.85E-01	6.13E-01	5.54E-02	1.21E-01	6.89E-01
475	1.42E-01	1.13E-01	1.04E+00	1.32E-01	2.20E-01	1.03E+00	-2.13E-01	1.98E-01	4.88E-01	7.98E-02	1.57E-01	5.82E-01	1.25E-01	2.19E-01	4.92E-01	7.38E-02	1.50E-01	5.83E-01
480	9.56E-02	1.39E-01	8.13E-01	8.05E-02	2.54E-01	7.72E-01	-2.38E-01	2.40E-01	3.62E-01	1.02E-01	1.89E-01	4.68E-01	1.46E-01	2.48E-01	3.66E-01	9.70E-02	1.83E-01	4.68E-01
485	5.80E-02	1.69E-01	6.16E-01	4.11E-02	2.98E-01	5.70E-01	-2.57E-01	2.80E-01	2.66E-01	1.28E-01	2.24E-01	3.62E-01	1.67E-01	2.78E-01	2.71E-01	1.25E-01	2.20E-01	3.63E-01
490	3.20E-02	2.08E-01	4.65E-01	1.62E-02	3.39E-01	4.15E-01	-2.77E-01	3.34E-01	1.96E-01	1.62E-01	2.67E-01	2.76E-01	2.00E-01	3.22E-01	2.01E-01	1.59E-01	2.64E-01	2.76E-01
495	1.47E-02	2.59E-01	3.53E-01	5.10E-03	3.95E-01	3.02E-01	-2.91E-01	4.05E-01	1.47E-01	2.06E-01	3.24E-01	2.12E-01	2.48E-01	3.86E-01	1.54E-01	2.04E-01	3.21E-01	2.11E-01
500	4.90E-03	3.23E-01	2.72E-01	3.80E-03	4.61E-01	2.19E-01	-2.95E-01	4.91E-01	1.07E-01	2.63E-01	3.96E-01	1.64E-01	3.10E-01	4.63E-01	1.16E-01	2.61E-01	3.95E-01	1.64E-01
505	2.40E-03	4.07E-01	2.12E-01	1.54E-02	5.31E-01	1.59E-01	-2.97E-01	5.97E-01	7.67E-02	3.37E-01	4.91E-01	1.28E-01	3.88E-01	5.60E-01	8.66E-02	3.34E-01	4.87E-01	1.27E-01
510	9.30E-03	5.03E-01	1.58E-01	3.75E-02	6.07E-01	1.12E-01	-2.68E-01	7.02E-01	5.02E-02	4.23E-01	5.95E-01	9.56E-02	4.72E-01	6.58E-01	6.18E-02	4.22E-01	5.94E-01	9.55E-02
515	2.91E-02	6.08E-01	1.12E-01	7.14E-02	6.86E-01	8.22E-02	-2.17E-01	8.09E-01	2.88E-02	5.20E-01	7.06E-01	6.76E-02	5.62E-01	7.58E-01	4.21E-02	5.19E-01	7.06E-01	6.75E-02
520	6.33E-02	7.10E-01	7.82E-02	1.18E-01	7.62E-01	6.07E-02	-1.48E-01	9.11E-01	1.33E-02	6.17E-01	8.08E-01	4.74E-02	6.53E-01	8.54E-01	2.83E-02	6.16E-01	8.08E-01	4.74E-02
525	1.10E-01	7.93E-01	5.73E-02	1.73E-01	8.23E-01	4.31E-02	-3.52E-02	9.85E-01	2.12E-03	7.00E-01	8.84E-01	3.47E-02	7.32E-01	9.26E-01	1.84E-02	7.01E-01	8.87E-01	3.48E-02
530	1.66E-01	8.62E-01	4.22E-02	2.37E-01	8.75E-01	3.05E-02	1.06E-01	1.03E+00	-4.16E-03	7.73E-01	9.41E-01	2.56E-02	7.99E-01	9.75E-01	1.20E-02	7.73E-01	9.42E-01	2.56E-02
535	2.26E-01	9.15E-01	2.98E-02	3.04E-01	9.24E-01	2.06E-02	2.60E-01	1.05E+00	-8.30E-03	8.34E-01	9.78E-01	1.82E-02	8.47E-01	9.97E-01	7.84E-03	8.34E-01	9.79E-01	1.82E-02
540	2.90E-01	9.54E-01	2.03E-02	3.77E-01	9.62E-01	1.37E-02	4.20E-01	1.05E+00	-1.22E-02	8.83E-01	9.97E-01	1.24E-02	8.79E-01	9.97E-01	5.18E-03	8.83E-01	9.98E-01	1.24E-02
545	3.60E-01	9.80E-01	1.34E-02	4.52E-01	9.82E-01	7.90E-03	5.93E-01	1.05E+00	-1.40E-02	9.23E-01	1.00E+00	8.26E-03	9.15E-01	1.00E+00	3.45E-03	9.23E-01	1.00E+00	8.23E-03
550	4.33E-01	9.95E-01	8.70E-03	5.30E-01	9.92E-01	4.00E-03	7.90E-01	1.04E+00	-1.47E-02	9.54E-01	9.87E-01	5.45E-03	9.48E-01	9.92E-01	2.31E-03	9.54E-01	9.88E-01	5.43E-03
555	5.12E-01	1.00E+00	5.70E-03	6.16E-01	9.99E-01	1.10E-03	1.01E+00	9.98E-01	-1.49E-02	9.77E-01	9.61E-01	3.65E-03	9.66E-01	9.61E-01	1.56E-03	9.78E-01	9.62E-01	3.62E-03
560	5.95E-01	9.95E-01	3.90E-03	7.05E-01	9.97E-01	0.00E+00	1.23E+00	9.38E-01	-1.46E-02	9.93E-01	9.22E-01	2.53E-03	9.68E-01	9.09E-01	1.06E-03	9.93E-01	9.24E-01	2.50E-03
565	6.78E-01	9.79E-01	2.70E-03	7.94E-01	9.82E-01	0.00E+00	1.47E+00	8.80E-01	-1.38E-02	1.00E+00	8.70E-01	1.84E-03	9.77E-01	8.61E-01	7.27E-04	1.00E+00	8.71E-01	1.81E-03
570	7.62E-01	9.52E-01	2.10E-03	8.79E-01	9.56E-01	0.00E+00	1.75E+00	8.28E-01	-1.27E-02	9.97E-01	8.06E-01	1.44E-03	9.97E-01	8.18E-01	5.02E-04	9.97E-01	8.07E-01	1.41E-03
575	8.43E-01	9.15E-01	1.80E-03	9.51E-01	9.15E-01	0.00E+00	2.02E+00	7.47E-01	-1.14E-02	9.86E-01	7.32E-01	1.26E-03	9.95E-01	7.48E-01	3.48E-04	9.86E-01	7.33E-01	1.22E-03
580	9.16E-01	8.70E-01	1.70E-03	1.01E+00	8.69E-01	0.00E+00	2.27E+00	6.49E-01	-9.93E-03	9.65E-01	6.51E-01	1.16E-03	9.75E-01	6.62E-01	2.43E-04	9.65E-01	6.52E-01	1.12E-03
585	9.79E-01	8.16E-01	1.40E-03	1.07E+00	8.26E-01	0.00E+00	2.49E+00	5.63E-01	-8.41E-03	9.34E-01	5.64E-01	1.00E-03	9.57E-01	5.86E-01	1.71E-04	9.34E-01	5.65E-01	9.56E-04
590	1.03E+00	7.57E-01	1.10E-03	1.12E+00	7.77E-01	0.00E+00	2.67E+00	4.77E-01	-7.02E-03	8.94E-01	4.77E-01	8.12E-04	9.32E-01	5.09E-01	1.21E-04	8.94E-01	4.78E-01	7.64E-04
595	1.06E+00	6.95E-01	1.00E-03	1.13E+00	7.20E-01	0.00E+00	2.81E+00	3.85E-01	-5.74E-03	8.48E-01	3.93E-01	7.41E-04	8.92E-01	4.26E-01	8.58E-05	8.48E-01	3.94E-01	6.91E-04
600	1.06E+00	6.31E-01	8.00E-04	1.12E+00	6.58E-01	0.00E+00	2.87E+00	3.01E-01	-4.27E-03	7.95E-01	3.18E-01	6.10E-04	8.43E-01	3.48E-01	6.14E-05	7.95E-01	3.18E-01	5.60E-04
605	1.05E+00	5.67E-01	6.00E-04	1.09E+00	5.94E-01	0.00E+00	2.85E+00	2.29E-01	-2.91E-03	7.35E-01	2.50E-01	4.79E-04	7.84E-01	2.80E-01	4.41E-05	7.35E-01	2.50E-01	4.30E-04
610	1.00E+00	5.03E-01	3.00E-04	1.03E+00	5.28E-01	0.00E+00	2.76E+00	1.66E-01	-2.27E-03	6.70E-01	1.93E-01	3.12E-04	7.17E-01	2.19E-01	3.19E-05	6.70E-01	1.94E-01	2.65E-04
615	9.38E-01	4.41E-01	2.00E-04	9.51E-01	4.62E-01	0.00E+00	2.60E+00	1.14E-01	-2.00E-03	6.02E-01	1.47E-01	2.40E-04	6.44E-01	1.66E-01	2.32E-05	6.03E-01	1.47E-01	1.97E-04
620	8.54E-01	3.81E-01	2.00E-04	8.56E-01	3.98E-01	0.00E+00	2.37E+00	7.47E-02	-1.51E-03	5.30E-01	1.10E-01	1.98E-04	5.66E-01	1.24E-01	1.69E-05	5.30E-01	1.10E-01	1.58E-04
625	7.51E-01	3.21E-01	1.00E-04	7.55E-01	3.40E-01	0.00E+00	2.11E+00	4.65E-02	-9.38E-04	4.54E-01	8.08E-02	1.32E-04	4.87E-01	9.18E-02	1.24E-05	4.53E-01	8.08E-02	9.70E-05
630	6.42E-01	2.65E-01	0.00E+00	6.48E-01	2.84E-01	0.00E+00	1.81E+00	2.63E-02	-5.53E-04	3.80E-01	5.83E-02	9.03E-05	4.10E-01	6.62E-02	9.17E-06	3.80E-01	5.84E-02	6.04E-05
635	5.42E-01	2.17E-01	0.00E+00	5.35E-01	2.28E-01	0.00E+00	1.52E+00	1.27E-02	-3.17E-04	3.15E-01	4.18E-02	6.78E-05	3.37E-01	4.68E-02	6.80E-06	3.15E-01	4.18E-02	4.27E-05
640	4.48E-01	1.75E-01	0.00E+00	4.32E-01	1.80E-01	0.00E+00	1.25E+00	4.50E-03	-1.43E-04	2.56E-01	2.96E-02	5.25E-05	2.73E-01	3.29E-02	5.06E-06	2.56E-01	2.96E-02	3.17E-05
645	3.61E-01	1.38E-01	0.00E+00	3.44E-01	1.40E-01	0.00E+00	1.01E+00	9.66E-05	-4.08E-06	2.04E-01	2.07E-02	3.82E-05	2.17E-01	2.31E-02	3.79E-06	2.04E-01	2.07E-02	2.23E-05
650	2.84E-01	1.07E-01	0.00E+00	2.68E-01	1.08E-01	0.00E+00	7.86E-01	-1.96E-03	1.11E-04	1.59E-01	1.44E-02	2.51E-05	1.67E-01	1.62E-02	2.85E-06	1.59E-01	1.44E-02	1.50E-05
655	2.19E-01	8.16E-02	0.00E+00	2.04E-01	8.12E-02	0.00E+00	5.97E-01	-2.63E-03	1.92E-04	1.22E-01	1.01E-02	1.92E-05	1.26E-01	1.12E-02	2.15E-06	1.22E-01	1.01E-02	1.05E-05
660	1.65E-01	6.10E-02	0.00E+00	1.53E-01	6.03E-02	0.00E+00	4.43E-01	-2.63E-03	2.27E-04	9.14E-02	6.99E-03	1.44E-05	9.32E-02	7.69E-03	1.63E-06	9.14E-02	7.13E-03	7.36E-06
665	1.21E-01	4.46E-02	0.00E+00	1.12E-01	4.41E-02	0.00E+00	3.24E-01	-2.30E-03	2.15E-04	6.70E-02	4.85E-03	1.05E-05	6.79E-02	5.27E-03	1.24E-06	6.66E-02	4.88E-03	5.04E-06
670	8.74E-02	3.20E-02	0.00E+00	8.13E-02	3.18E-02	0.00E+00	2.35E-01	-1.87E-03	1.64E-04	4.82E-02	3.33E-03	7.56E-06	4.90E-02	3.62E-03	9.50E-07	4.82E-02	3.36E-03	3.46E-06
675	6.36E-02	2.32E-02	0.00E+00	5.79E-02	2.26E-02	0.00E+00	1.69E-01	-1.44E-03	9.72E-05	3.50E-02	2.33E-03	5.48E-06	3.52E-02	2.51E-03	7.30E-07	3.54E-02	2.37E-03	2.45E-06
680	4.68E-02	1.70E-02	0.00E+00	4.09E-02	1.59E-02	0.00E+00	1.21E-01	-1.08E-03	5.10E-05	2.57E-02	1.64E-03	4.02E-06	2.52E-02	1.76E-03	5.60E-07	2.57E-02	1.67E-03	1.73E-06
685	3.29E-02	1.19E-02	0.00E+00	2.86E-02	1.11E-02	0.00E+00	8.58E-02	-7.90E-04	3.53E-05	1.80E-02	1.11E-03	2.82E-06	1.78E-02	1.22E-03	4.30E-07	1.79E-02	1.15E-03	1.19E-06
690	2.27E-02	8.20E-03	0.00E+00	1.99E-02	7.70E-03	0.00E+00	6.03E-02	-5.68E-04	3.12E-05	1.24E-02	7.50E-04	1.94E-06	1.25E-02	8.47E-04	3.40E-07	1.24E-02	7.88E-04	8.10E-07
695	1.58E-02	5.70E-03	0.00E+00	1.38E-02	5.40E-03	0.00E+00	4.15E-02	-3.93E-04	2.45E-05	8.66E-03	5.17E-04	1.36E-06	8.62E-03	5.82E-04	2.60E-07	8.72E-03	5.49E-04	5.70E-07
700	1.14E-02	4.10E-03	0.00E+00	9.60E-03	3.70E-03	0.00E+00	2.81E-02	-2.62E-04	1.65E-05	6.21E-03	3.68E-04	9.70E-07	5.84E-03	3.98E-04	2.00E-07	6.20E-03	3.89E-04	4.00E-07

Table I : Average results of computational analysis for all color centers and observers in the Alfvin data set. It is clear to see that there is little difference between the different sets of responsivity functions.

Color Difference	CIE 1931	CIE 1964	Stiles	Demarco	Stockman	Vos	Shaw
ΔE^*_{ab}	4.39	4.88	4.37	4.33	4.38	4.33	4.56
ΔE^*_{94}	2.72	3.01	2.72	2.67	2.70	2.67	2.81

Table II : Results of statistical analysis of combined Alfvin data set using the CIE 1931 responsivity functions (a), and the Shaw and Fairchild responsivity functions (b).

CIE 1931 Color Matching Functions

	Mean	*Std Dev*	*Maximum*	*Minimum*
DE :	4.394	3.468	22.300	0.223
DE 94:	2.720	2.393	13.470	0.000
DL :	0.065	2.876	13.410	-13.100
Da :	0.473	2.876	15.440	-8.384
Db :	0.688	3.731	18.030	-21.000

	Critical Value	*Test Statistic*	*P Value*	*Decision*
DL :	1.969	0.371	0.356	Fail to Reject
Da :	1.969	2.653	0.004	Reject Null
Db :	1.969	3.019	0.001	Reject Null

(a)

Shaw and Fairchild Color Matching Functions

	Mean	*Std Dev*	*Maximum*	*Minimum*
DE :	4.563	3.602	24.360	0.404
DE 94:	2.814	2.475	14.100	0.000
DL :	0.147	2.941	13.650	-13.170
Da :	0.248	2.941	13.600	-9.139
Db :	-0.132	3.819	17.130	-23.810

	Critical Value	*Test Statistic*	*P Value*	*Decision*
DL :	1.969	0.816	0.208	Fail to Reject
Da :	1.969	1.249	0.107	Fail to Reject
Db :	1.969	-0.565	0.286	Fail to Reject

(b)

Table III : Mean color difference (ΔE^*_{ab}) results of computational analysis using the segmented Alfvin data set. Mean of means is the mean color difference of the averaged results for each color center, giving an indication to the overall performance of each set of responsivity functions.

Color Center		CIE 1931	CIE 1964	Stiles	Demarco	Stockman	Vos	Shaw
Gray	Print	3.69	3.65	3.28	3.38	3.28	3.38	3.31
	Trans.	3.08	4.20	3.24	3.03	3.25	3.03	3.49
Cyan	Print	3.62	4.13	3.54	3.34	3.54	3.34	3.44
	Trans.	3.32	5.95	3.88	3.33	3.88	3.32	4.65
Magenta	Print	4.06	4.43	3.92	4.04	3.91	4.05	4.54
	Trans.	4.42	4.42	4.32	4.43	4.32	4.43	4.31
Yellow	Print	5.78	7.18	5.69	5.75	5.69	5.76	6.34
	Trans.	4.61	4.77	4.44	4.61	4.44	4.61	4.41
Red	Print	3.63	3.54	3.52	3.57	3.52	3.57	3.51
	Trans.	4.61	4.87	4.75	4.60	4.76	4.60	4.78
Green	Print	4.43	4.39	4.36	4.34	4.36	4.34	4.38
	Trans.	4.43	5.13	4.60	4.20	4.60	4.20	4.62
Blue	Print	4.92	4.76	4.63	4.65	4.63	4.65	4.89
	Trans.	6.14	6.08	6.09	6.37	6.09	6.38	6.17
Mean of Means		4.34	4.82	4.30	4.26	4.31	4.26	4.49

Table IV : Average results of computational analysis for the SM98 data set.

Color Difference	CIE 1931	CIE 1964	Stiles	Demarco	Stockman	Vos	Shaw
ΔE^*_{ab}	3.80	2.03	2.90	3.39	2.90	3.38	2.08
ΔE^*_{94}	3.49	1.85	2.57	3.11	2.57	3.11	1.86

Table V : Average results of computational analysis for the SM99 data set.

Color Difference	CIE 1931	CIE 1964	Stiles	Demarco	Stockman	Vos	Shaw
ΔE^*_{ab}	4.53	3.07	3.86	4.36	3.86	4.36	3.42
ΔE^*_{94}	4.16	2.79	3.50	4.01	3.50	4.01	3.12

Table VI : Average results for the computational analysis of all three data sets combined

Color Difference	CIE 1931	CIE 1964	Stiles	Demarco	Stockman	Vos	Shaw
ΔE^*_{ab}	4.56	4.02	4.14	4.36	4.14	4.37	3.92
ΔE^*_{94}	3.41	2.82	3.01	3.24	3.00	3.24	2.78

Table VII : Results for the computational analysis using the Thornton metamer pair

Color Difference	CIE 1931	CIE 1964	Stiles	Demarco	Stockman	Vos	Shaw
ΔE^*_{ab}	5.35	9.75	6.75	4.84	6.74	4.83	7.96
ΔE^*_{94}	3.50	6.23	4.32	3.03	4.31	3.02	4.94

Figure Legends

FIG 1 : Ideally a true metameric pair would result in identical tristimulus values when using an optimal set of color matching functions / cone responsivities. It is on this basis that the computational analysis is performed, assuming that the set of responsivity functions that best describes human vision will yield the lowest average color difference when computing visual metamers.

FIG 2 : The rotation transformation used to rotate the Stiles and Burch responsivity functions to a close approximation of the CIE 1931 responsivity functions. It can be seen that the functions are not identical, a characteristic that is expected.

FIG 3 : The 95% (outer ellipse) and mean (inner ellipse) confidence region for the sample distribution in the $\Delta a^* \Delta b^*$ plane of the sample data.

FIG 4 : Confidence ellipse plots of the Shaw and Fairchild responsivity functions, using the whole Alfvin data set. 95% Confidence region - Outer ellipse, Mean – Inner ellipse.

FIG 5 : Confidence ellipse plots of the Shaw and Montag 98 data set, using the CIE 1931 standard observer in the $\Delta L \Delta a$ and $\Delta L \Delta b$ planes. The two 'clouds' of data points are thought to represent the color matches attained by the different primary sets used.

FIG 6 : 95% and mean confidence ellipses of Shaw and Fairchild responsivity functions, using all three combined data sets

FIG 7 : Thornton's metameric pair

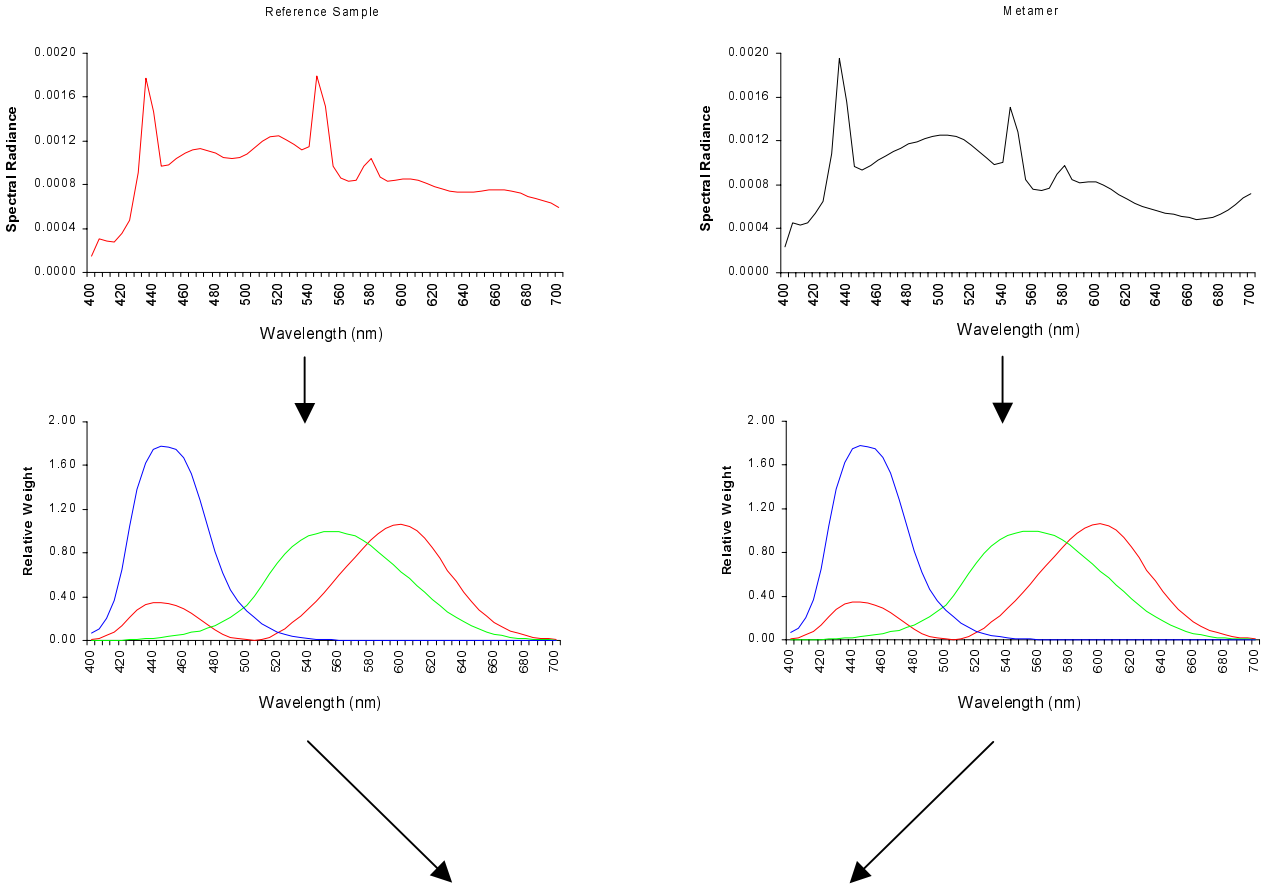
FIG 8 : Final selection of optimized color weighting functions

FIG 9 : Conceptual visual experiment for observer comparison of various sets of responsivity functions

FIG 10 : Characterization outline of CRT display using multiple sets of responsivity functions

FIG 11 : Simulation of each of the 7 responsivity functions using the spectral characterization of the CRT. The subtle differences are noticeable between the different sets of responsivity functions.

Figure 1



$$XYZ_{reference} = XYZ_{metamer}$$

Figure 2

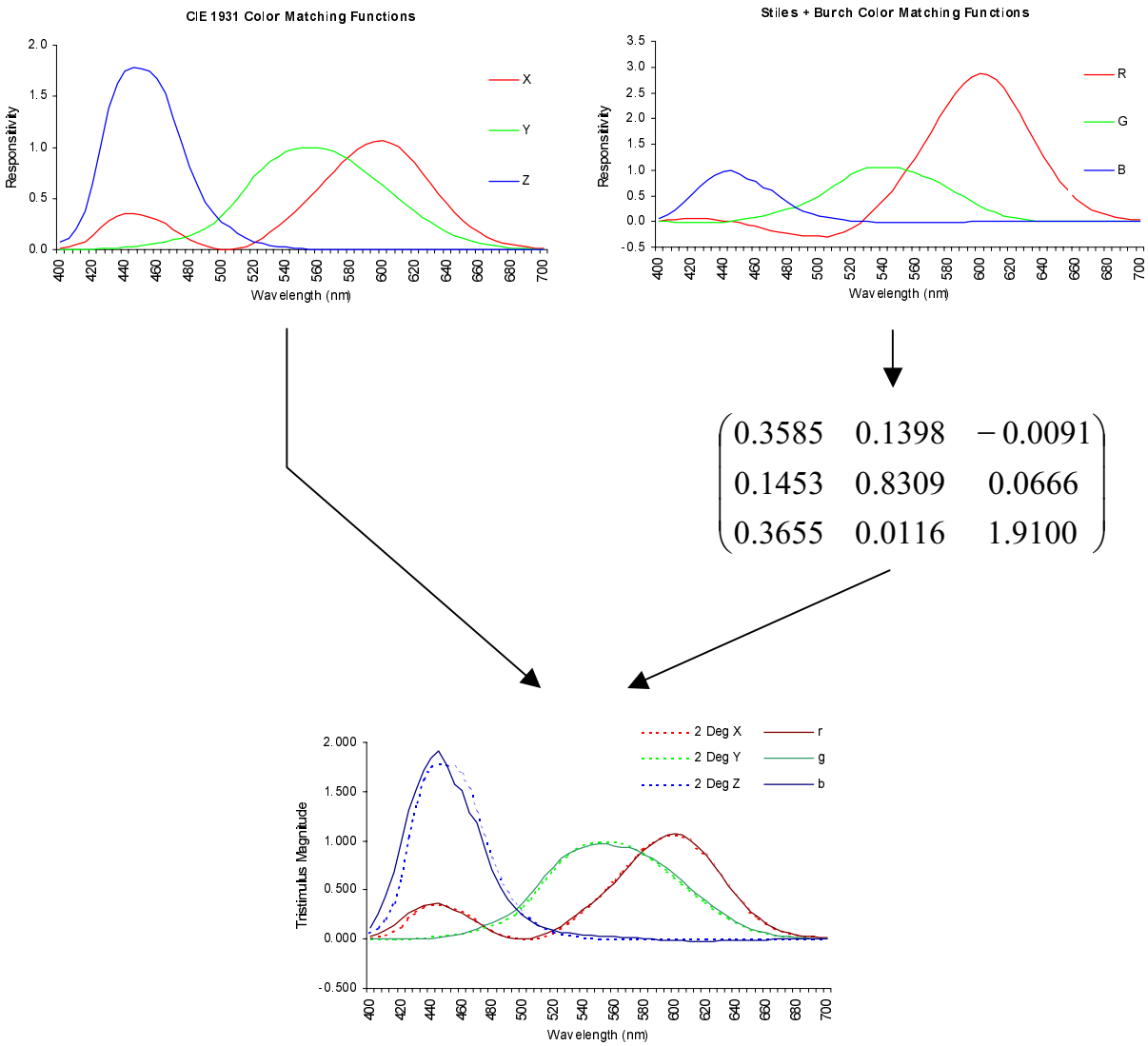

CIE 1931 Color Matching Functions
Responsitivity
Wavelength (nm)
X
Y
Z
Stiles + Burch Color Matching Functions
Responsitivity
Wavelength (nm)
R
G
B
$\begin{pmatrix} 0.3585 & 0.1398 & -0.0091 \\ 0.1453 & 0.8309 & 0.0666 \\ 0.3655 & 0.0116 & 1.9100 \end{pmatrix}$
2 Deg X
2 Deg Y
2 Deg Z
r
g
b
Tristimulus Magnitude
Wavelength (nm)

Figure 3

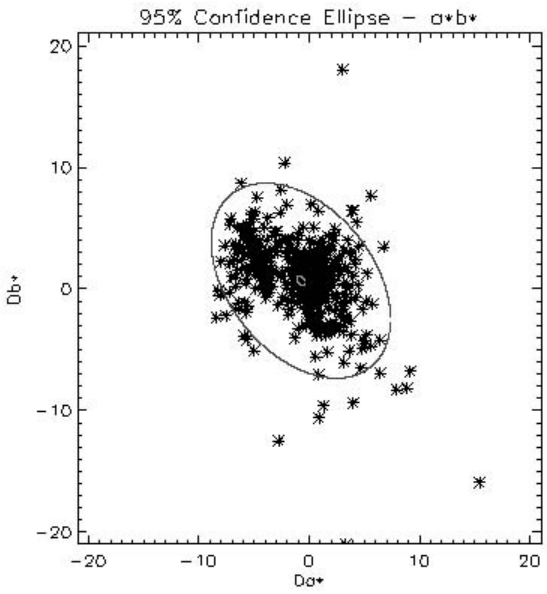


Figure 4

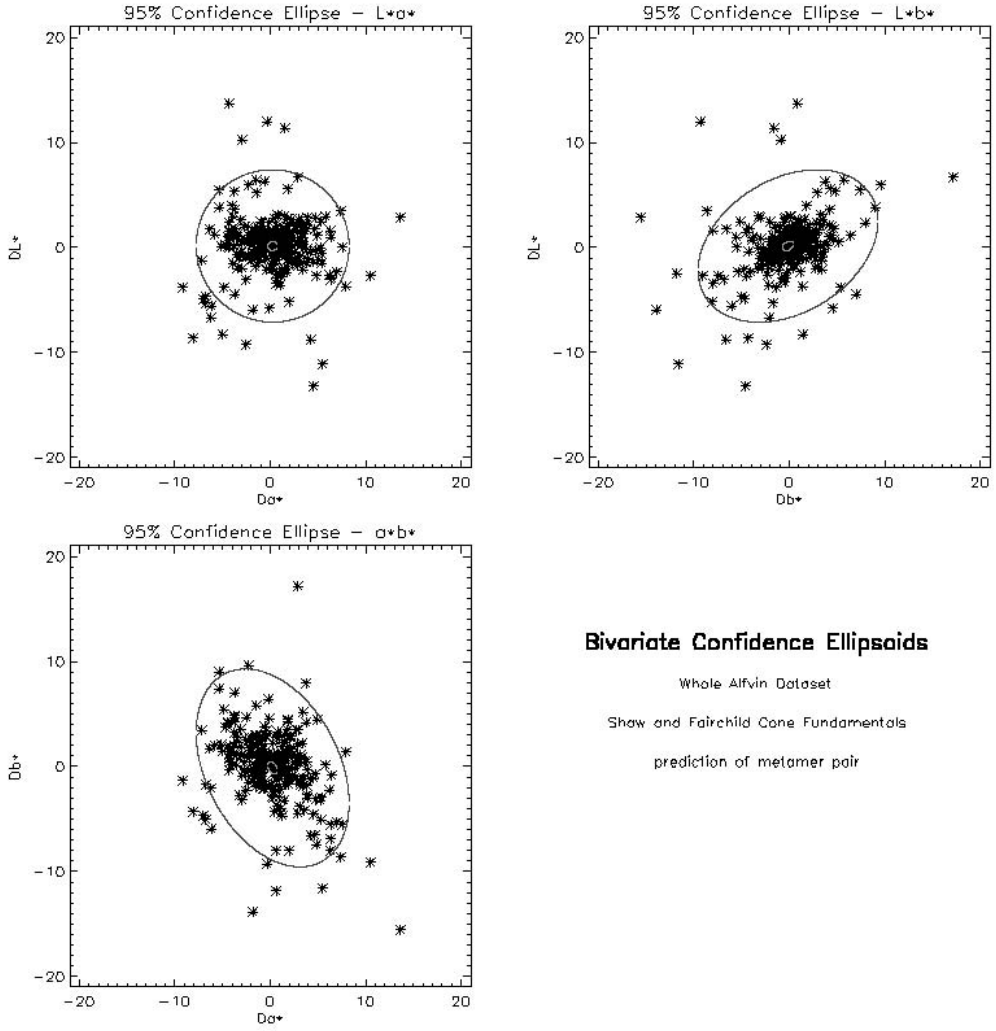

95% Confidence Ellipse – L*a*
95% Confidence Ellipse – L*b*
95% Confidence Ellipse – a*b*
DL*
Da*
Db*
Bivariate Confidence Ellipsoids
Whole Alfvin Dataset
Shaw and Fairchild Cone Fundamentals
prediction of metamer pair

Figure 5

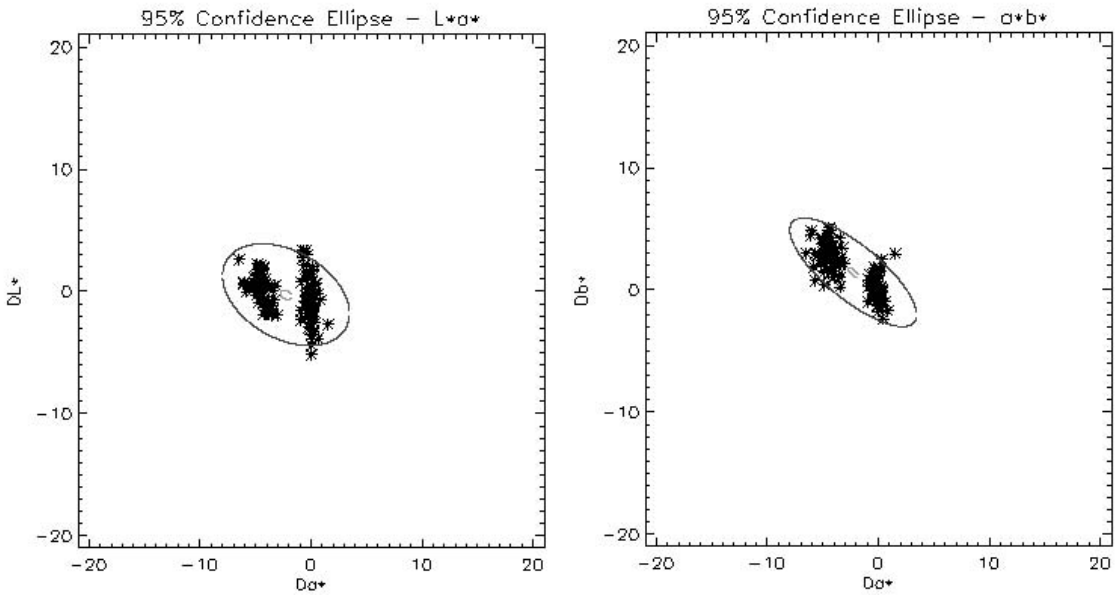

95% Confidence Ellipse – L*a*
95% Confidence Ellipse – a*b*
DL*
Da*
Db*
20
10
0
-10
-20

Figure 6

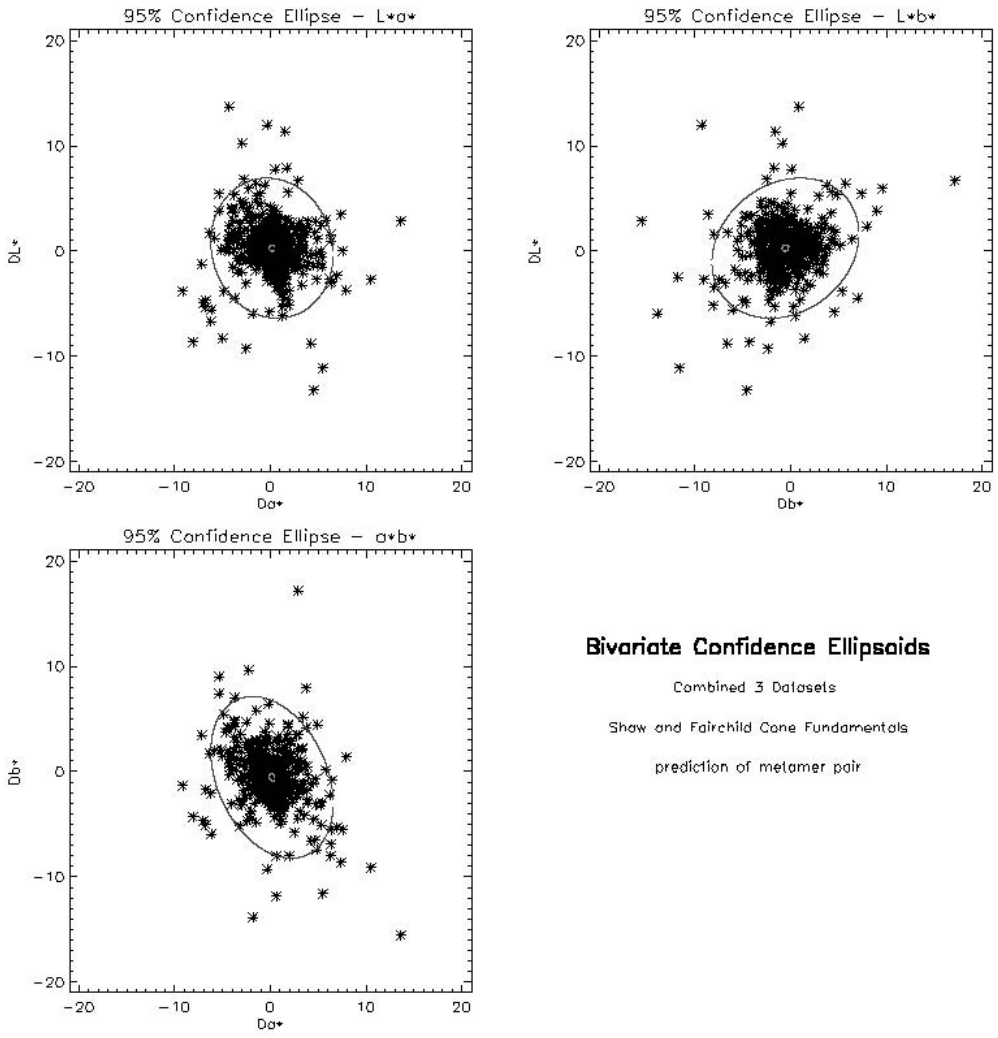

95% Confidence Ellipse – L*a*
95% Confidence Ellipse – L*b*
95% Confidence Ellipse – a*b*
DL*
Da*
Db*
Bivariate Confidence Ellipsoids
Combined 3 Datasets
Shaw and Fairchild Cone Fundamentals
prediction of metamer pair

Figure 7

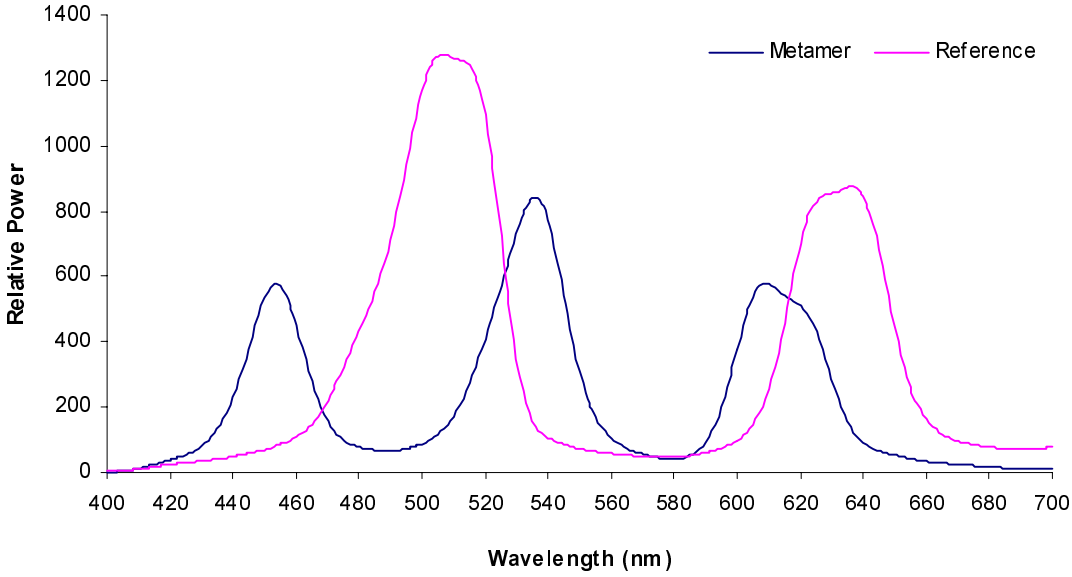


Figure 8

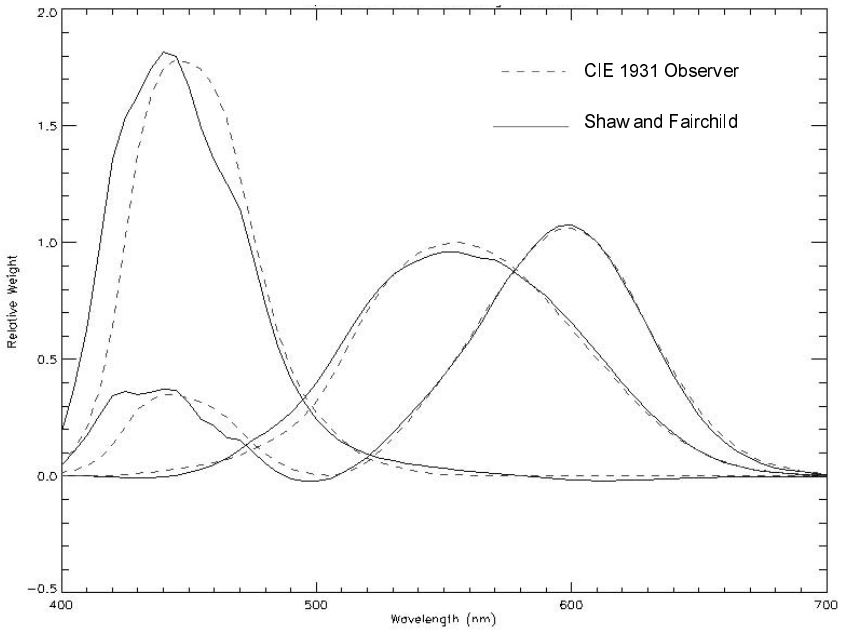


Figure 9

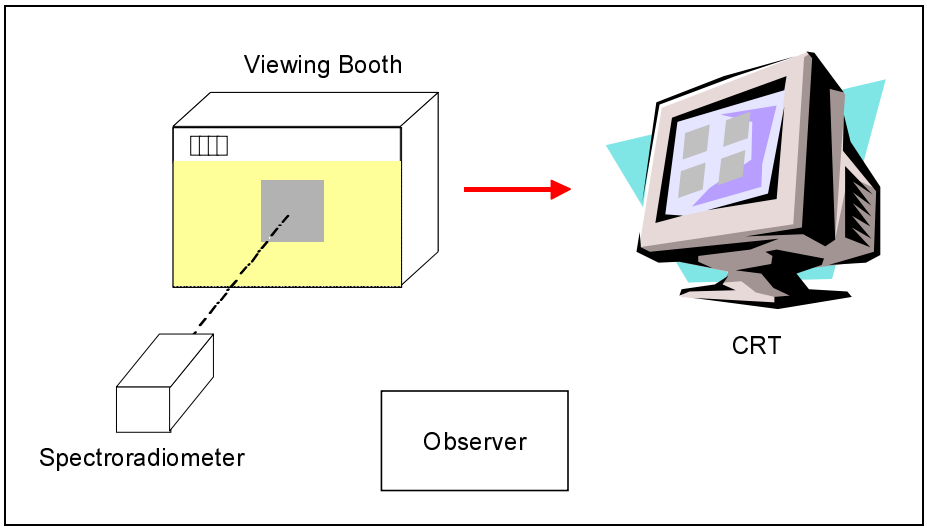


Figure 10

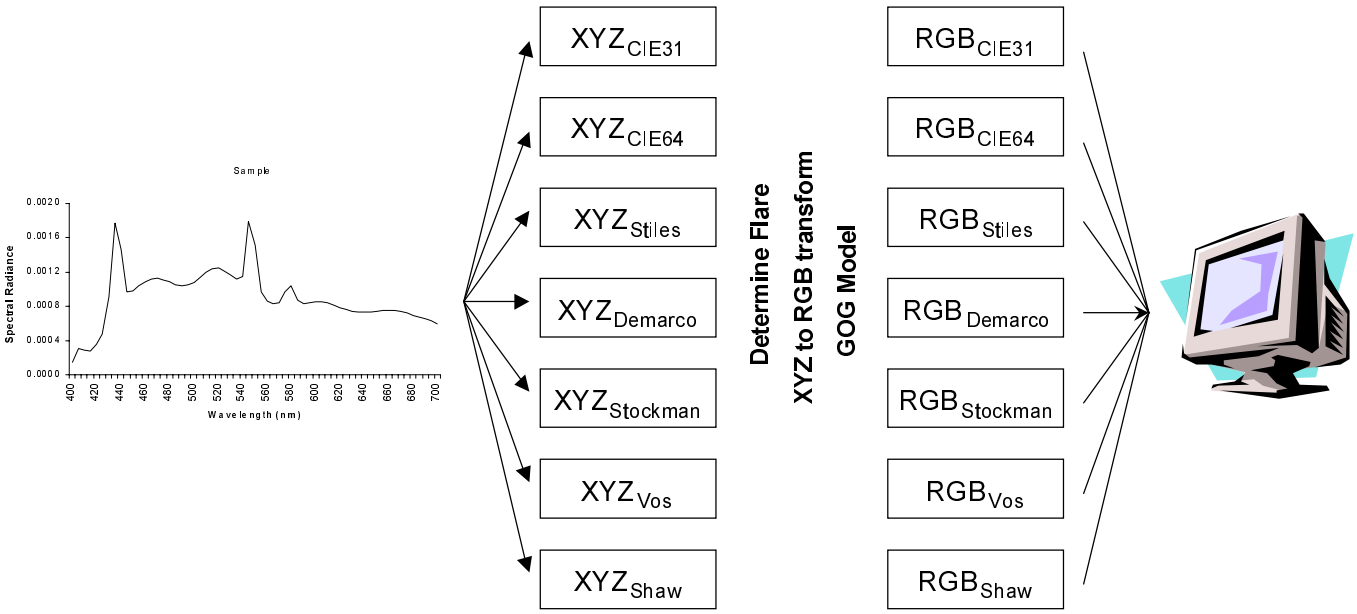
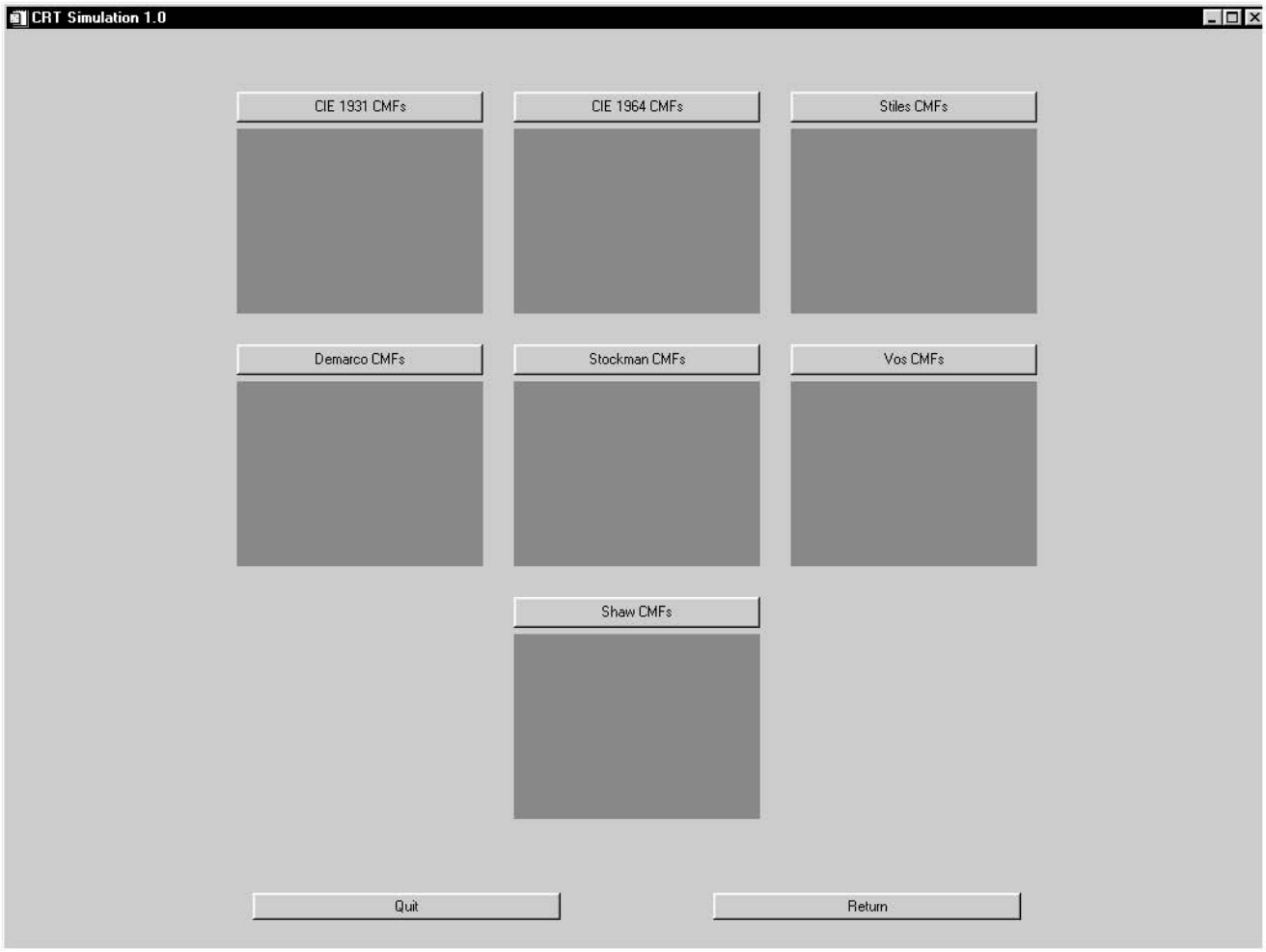


Figure 11


CRT Simulation 1.0
CIE 1931 CMFs
CIE 1964 CMFs
Stiles CMFs
Demarco CMFs
Stockman CMFs
Vos CMFs
Shaw CMFs
Quit
Return