Toward Reducing Observer Metamerism in Industrial Applications: Colorimetric Observer Categories and Observer Classification

*Abhijit Sarkar***, Laurent Blondé*, Patrick Le Callet**, Florent Autrusseau**, Patrick Morvan*, Jürgen Stauder**
**Technicolor Research, Rennes (France); **IRCCyN-IVC, Polytech Nantes, Nantes (France)*

Abstract

*The variability among color-normal observers poses a challenge to modern display colorimetry because of their peaky primaries. But such devices also hold the key to a future solution to this issue. In this paper, we present a method for deriving seven distinct colorimetric observer categories, and also a method for classifying individual observers as belonging to one of these seven categories. Five representative L, M and S cone fundamentals (a total of 125 combinations) were derived through a cluster analysis on the combined set of 47-observer data from 1959 Stiles-Burch study, and 61 color matching functions derived from the CIE 2006 model corresponding to 20-80 age parameter range. From these, a reduced set of seven representative observers were derived through an iterative algorithm, using several predefined criteria on perceptual color differences (delta E*00) with respect to actual color matching functions of the 47 Stiles-Burch observers, computed for the 240 Colorchecker samples viewed under D65 illumination. Next, an observer classification method was implemented using two displays, one with broad-band primaries and the other with narrow-band primaries. In paired presentations on the two displays, eight color-matches corresponding to the CIE 10° standard observer and the seven observer categories were shown in random sequences. Thirty observers evaluated all eight versions of fifteen test colors. For majority of the observers, only one or two categories consistently produced either acceptable or satisfactory matches for all colors. The CIE 10° standard observer was never selected as the most preferred category for any observer, and for six observers, it was rejected as an unacceptable match for more than 50% of the test colors. The results show that it is possible to effectively classify real, color-normal observers into a small number of categories, which in certain application contexts, can produce perceptibly better color matches for many observers compared to the matches predicted by the CIE 10° standard observer.*

Introduction

Conventional color reproduction relies on colorimetric data for a single "standard observer", representing an average colorimetric observer with normal color vision. The 1931 CIE 2° standard observer and 1964 CIE 10° standard observer are widely used in the industry (the latter is more suitable for large-field color stimuli encountered in most practical industrial applications). The use of a "standard observer" in colorimetric computations is essentially based on the assumption that the whole population of color normal observers can be reasonably represented by a single observer model, given by a set of three Color Matching Functions (CMFs). In 1989, CIE recognized the variability among individual observers by introducing the concept of standard deviate observer [1], but the model significantly under-predicted inter-observer variability [2], and was never adopted by the industry. Thus, applied colorimetry in its current form does not have any provision for incorporating observer variability into the computations.

The relevance of the observer variability issue is quite dependent on the application context. In cross-media color reproduction, the effect of observer variability has not been found to be significant [3]. This is presumably true for most of the industrial applications of the past several decades, be it printing, photography, painting or textile. Thus, taking into account individual observer variability in applied colorimetry did not warrant a serious consideration in the past.

However, this limitation has become non-trivial with the advent and wide-spread adoption of modern wide-gamut consumer displays. Colors on two displays with very different spectral characteristics are highly metameric in nature, often resulting in colors that are a satisfactory match for one observer, and an unacceptable match for another. This phenomenon, commonly termed as observer metamerism, is particularly significant when at least one of the two displays has narrow-band primaries. Many modern Liquid Crystal Displays (LCDs) are fitted with Light Emitting Diode (LED) backlight (or sometimes, laser primaries) in order to achieve more vivid, more saturated and brighter colors. These displays are particularly susceptible to observer variability [4][5], since their peaky primaries can cause noticeable shift in the chromaticities of perceived colors with relatively minor change in the visual characteristics of the observer. As a part of the current work on observer variability, a preliminary set of a color matching experiment [6] was performed on two displays, one being a reference studio CRT display with broad-band primaries, and the other being a Wide-Gamut, narrow-band LCD with LED backlight. Each of the ten observers made nine color matches on the two displays under dark and white surround conditions. CIE 10° standard observer predicted mean, maximum and the 90^{th} percentile color difference of individual observer matches were 1.4, 3.3 and 2.6 respectively. An average color match prediction error of 1.4 ΔE^*_{00} over all colors and all observers was reasonable, confirming that the CIE 10° standard observer is a reasonably good representation of an average observer. However, the maximum and the 90^{th} percentile ΔE^*_{00} values between individual observer matches predicted by the CIE 10° standard observer were considered significant, particularly for a carefully controlled experimental setup involving uniform color stimuli. For color

critical applications involving expert observers, such a difference will even be unacceptable under similar viewing conditions.

Not surprisingly, similar observer metamerism issue has been observed when narrow band RGB-LEDs were matched with broadband lights [7]. Note that none of the traditional industrial color applications mentioned before involved a color system with spectral characteristics similar to modern displays or the LEDs.

In the following sections, we discuss the limitations of an age-dependent observer in predicting observer metamerism, and present a novel statistical approach for deriving seven representative colorimetric observer categories. An experimental method for classifying individual observers as belonging to one of these seven categories is also presented. In this experiment, thirteen observers evaluated and ranked eight versions of fifteen test colors, corresponding to the seven observer categories and the CIE 10° standard observer. Based on these results, each observer is assigned to one of the seven observer categories.

CIE 2006 model and limitations of an age-dependent observer in an applied context

In 2006, CIE's (Commission Internationale de l'Éclairage) technical committee TC 1-36 published a report [8] (described hereafter as CIE06) on the choice of a set of Color Matching Functions and estimates of cone fundamentals for the color-normal observer. The CIE06 model is largely based on the work of Stockman and Sharpe [9]. Starting from 1959 Stiles-Burch (described hereafter as S&B) 10° CMFs [10], it defines 2° and 10° reference observers and provides a convenient framework for calculating average cone fundamentals for any field size between 1° and 10° and for an age between 20 and 80. Corresponding CMFs can be obtained through a linear transformation of the cone fundamentals.

However, CIE06 age parameter does not necessarily correspond to real observer ages. In other words, predicted model functions that best match the real observer data may not always be obtained using real observer ages. This may happen because of unsystematic observer variability, and/or because of the exclusion of one or more age-independent physiological factors from the CIE06 model [11]. CIE committee TC 1-36 also recognized this restriction by pointing out that CIE06 fundamental observer is a theoretical construct based on averages [8] (section 1.3).

In a recent work [12] by current authors, CIE06 age parameters that resulted in the best predictions of the cone fundamentals of individual S&B observers were determined. This was done by computing the correlation coefficients between the normalized cone fundamentals for each S&B observer and those corresponding to all possible CIE06 age parameter values between 20 and 80 (a total of 61). For each S&B observer, the corresponding CIE age was the one yielding the highest correlation coefficient for a given cone fundamental. This process was repeated for all three cone fundamentals and for all 47 S&B observers. No direct correspondence between the real and predicted ages was observed that could be represented through a mathematical model. It was also shown that for three different age-groups of S&B observers (with 6, 10 and 6 observers respectively), using real observer group ages in the CIE06 model led to larger error in intra-group average observer prediction than what resulted from using the CIE 10° standard observer. If two observers of similar ages, but having different peak wavelengths of long-wavelength sensitive cone photo-pigment absorption spectra are asked to evaluate colors on a modern wide-gamut display with narrow-band primaries (and to a lesser extent, in case of a display with broad-band primaries), they are likely to disagree on several color perceptions, particularly in the cyan and blue regions of the color space [11]. However, the CIE06 model will be unable to predict any variation since the peak-wavelength shift factor is not taken into account in the model. This discrepancy is likely to propagate to color matches across displays. While the CIE model is not meant to predict individual CMFs, intra-group averages should have been better predicted by the model. More observer data from various age-groups may be necessary for confirming this observation.

Nevertheless, the introduction of CIE06 model is perhaps one of the most fundamental contributions in the field of color science since the establishment of CIE 10° standard observer in 1964. With respect to this work, its significance lies in the fact that it provides an effective model for observer variability over all ages.

A Cluster Analysis of a combined set of S&B 47-observer data and CIE06 model predictions

A hypothesis of this work is that the CIE06 model predictions and the experimentally obtained visual color matching data from the 1959 S&B study, when combined together, incorporate most of the variability that can be found among the color normal population. The combined data set included 61 CIE06 cone fundamentals corresponding to 20-80 age parameter range, and the cone fundamentals corresponding to 47 Stiles-Burch observers, a total of 108 cone fundamentals. A theoretical analysis was performed to find a minimal set of average cone fundamentals that cover all possible variations in this combined dataset.

In terms of statistics, this is a problem of classification (i.e. grouping) within a complex data set. One of the methods appropriate for solving this problem is *cluster analysis* [13]. The purpose of the analysis is to arrange the functions into relatively homogeneous groups based on multivariate observations. In the current analysis, the total number of variables is 35 (normalized values at 35 wavelengths) and total number of observations is 108. A cluster analysis starts with undifferentiated groups and attempts to create clusters of objects (i.e. the CMFs) based on the similarities observed among a set of variables (i.e. CMF values at each wavelength). Variables must be selected that maximally discriminate among objects. Increasing dataset size results in increased cluster reliability. One of the cluster analysis methods commonly employed is the Partitioning method, also known as the K-means method. It begins by partitioning the actual data (rather than similarity measures) into a specific number of clusters. Then, objects are assigned and reassigned in an iterative method to simultaneously minimize intra-cluster variability and maximize

inter-cluster variability. This method was chosen as it is more likely to lead to a robust solution compared to other methods.

In the two-phase computational implementation in Matlab®, the first phase used batch updates, in which each iteration consisted of reassigning objects to their nearest cluster centroid, all at once, followed by recalculation of cluster centroids. The second phase used online updates, in which objects were individually reassigned if doing so would reduce the sum of distances, and cluster centroids were recomputed after each reassignment. Each cluster in the partition was defined by its member objects and by its centroid, or center. Suitable wavelength ranges (i.e. the number of variables) were chosen for L, M and S to avoid the influence of variations where functions had low amplitudes. Initial cluster centroid locations were selected by dividing 20-80 age range in equal parts and using corresponding CIE06 functions. Squared Euclidean distance measure (in cone fundamental space) was used in this analysis. The clustering was repeated 20 times (with different initial cluster centroid positions described above). Model functions were obtained by taking the mean of cluster members. The analysis was performed on LMS cone fundamentals, and the model LMS functions were then converted into CIE 10° standard observer equivalent CMFs through a 3x3 transformation. An approximate 3x3 LMS-to-XYZ transformation matrix (Eq. 1) was computed from the available 1964 10° $\bar{x}$ $\bar{y}$ $\bar{z}$ standard observer functions and $\bar{l}$ $\bar{m}$ $\bar{s}$ cone fundamentals of 47 Stiles-Burch observers.

$$\begin{bmatrix} \bar{x}_{10}(\lambda) \\ \bar{y}_{10}(\lambda) \\ \bar{z}_{10}(\lambda) \end{bmatrix} = \begin{bmatrix} 1.905378 & -1.321620 & 0.419512 \\ 0.698648 & 0.333043 & -0.013601 \\ -0.024300 & 0.040453 & 2.073582 \end{bmatrix} \begin{bmatrix} \bar{l}_{10}(\lambda) \\ \bar{m}_{10}(\lambda) \\ \bar{s}_{10}(\lambda) \end{bmatrix} \quad (1)$$

Table 1. Comparison of average and maximum color differences (ΔE^*_{00}) with respect to real observer (averaged over all 47 observers) for various average CMF sets

CMFs Under Comparison	Average ΔE^*_{00} for 240 patches	Maximum ΔE^*_{00} for 240 patches
CIE 10° standard observer	0.9	2.1
3 Model functions (total 27)	0.7	1.5
4 Model functions (total 64)	0.6	1.5
5 Model functions (total 125)	0.5	1.1
6 Model functions (total 216)	0.4	0.7

Derived model sets of CMFs were then used to predict 47 Stiles-Burch observer data. CIELAB coordinates were computed for all 240 color patches of the ColorChecker DC™ reference color chart with a CIE illuminant D65, by using i) real Stiles-Burch observer CMF data, ii) CIE 1964 10° standard observer functions and iii) all possible combinations of each of the model sets of CMFs derived from the above cluster analysis. Then, for each observer, color differences (ΔE^*_{00}) were computed between the CIELAB values obtained from real observer CMFs [case (i)] and those obtained from the predicted CMFs [case (ii) and (iii)]. Thus for each of the 47 Stiles-Burch observers, average color difference ΔE^*_{00} is computed out of the 240 patches. Lower the average color difference, the better is the model prediction. The analysis was repeated for 3, 4, 5 and 6 model sets of CMFs. All combinations of the CMFs (3 to 6) are compared to CIE 1964 10° observer (giving respectively $3^3 = 27$ to $6^3 = 216$ total possibilities). Note that for the model CMFs, the combination yielding best result was considered for individual observers (thus, each of the 47 observers had a corresponding best combination). Then the average and the maximum ΔE^*_{00} were computed, as shown in Table 1. Based on the accuracy of prediction, five model sets of CMFs were found to be the minimal to meet the goal of achieving close to one unit of maximum color difference (ΔE^*_{00}) for the 240 color patches of the ColorChecker DC™ reference color chart and the CIE illuminant D65, averaged over all 47 Stiles-Burch observers. With these 5 model sets of x-, y- and z- CMFs (or L-, M- and S- cone fundamentals), there can be 5x5x5, or 125 possible classes of observers.

Deriving reduced sets of model CMFs: seven observer categories

Out of the above-mentioned 125 possible observer categories (i.e. combinations of each of five x-, y- and z- CMFs), several categories can meet the goal of achieving any predefined ΔE^*_{00} criterion for a given observer. Thus, for the said constraint, fewer than 125 categories will suffice for achieving satisfactory result for all the 47 observers. Thus in this 2nd step, an iterative algorithm was implemented to pick the minimal number of observer categories such that at least one out of these categories satisfies the ΔE^*_{00} criterion for any S&B observer. The derivation of such reduced sets is dependent on the color data set and the color difference criterion. As before, the 240 color patches of the ColorChecker samples with the CIE illuminant D65 were used, since the samples cover a wide range of colors.

Note that while Euclidean distances in the LMS space were used in deriving the model CMFs, we use ΔE^*_{00} color difference equations for deriving the reduced sets of model CMFs. While we are well aware that this equation corresponds to CIE 10° standard observer and does not fully hold for other observer models, we hypothesize that the ΔE^*_{00} metric can be used as a reasonable baseline for the purpose of comparing the performance of various observer models. The error introduced in doing so cannot be more than that in case of using ΔE^*_{00} on the visual data of individual observers, which is done routinely. The use of ΔE^*_{00} was motivated by the need to use a perceptual metric while deriving the reduced set. Euclidean distance in the cone fundamental space does not satisfy that need.

Several criteria were established for selecting the reduced sets of CMFs. The same ΔE^*_{00} values computed in the previous step were used [case (i) and (iii)], but they were not averaged over all observers. Instead, for each observer the 90th percentile of the ΔE^*_{00} values for all the 240 color patches were considered. Thus, for each of the 47 observers there were 125 such percentile ΔE^*_{00}

values (hereafter referred to as DE in this section), corresponding to 125 possible observer CMF combinations. We must take into consideration that for some observers with atypical color vision characteristics, a given DE criterion may be hard to achieve with any of the 125 CMFs, while for some others, even a stricter criterion can be satisfied. Thus, an observer-dependent DE threshold was computed using the 10th or the 5th percentile of the 125 DE values, whichever was below 1.2. This meant the worst 5% or 10% DE values would not be considered while deciding which observer categories could be assigned to a given S&B observer. For six observers, the DE threshold computed this way was more than 2.0. However, these thresholds were still less than the DE values computed similarly with the CIE 10° standard observer, indicating that these specific S&B observers were far away from the average of the population.

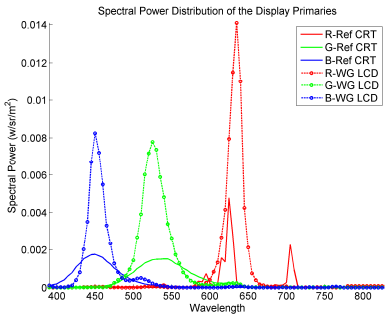

Figure 1. *Spectral Power Distribution of the CRT and the LCD used in the experiments*

The suitability of a given CMF combination for any S&B observer was determined by a "CMF Performance Index" (PI), based on the average percent deviation from the DE threshold (a positive PI indicated average DE was lower than the threshold). A CMF combination for the reduced sets was selected based on the highest number of observers with positive PI as well as the largest value of the PI.

Table 2. The reduced set of seven observer categories, their constituent average CMFs, and the total number of S&B observers assigned to the various categories

Iteration	Combination	X	Y	Z	Total Obs	%Obs Covered
1	2	1	1	2	17	36.2
2	58	3	2	3	14	66
3	6	1	2	1	8	83
4	33	2	2	3	4	91.5
5	81	4	2	1	2	95.7
6	63	3	3	3	1	97.9
7	76	4	1	1	1	100

Table 2 shows which of the 125 combinations, and their constituent x-, y-, z- functions were picked for the reduced sets of 7 observer classes. 4 x-CMFs, 3 y-CMFs and 3 z-CMFs constitute the reduced sets. Total number of Stiles-Burch observers assigned to each set, as well as cumulative percent of observers covered are listed. For example, combination 2 is made up of 1st x-CMF, 1st y-CMF and 2nd z-CMF, satisfying the aforementioned DE threshold for 17 observers, which is 36.2% of Stiles-Burch observer pool. Combination 58 met the DE threshold for another 14 observers, so combinations 2 and 58 together satisfied 66% of the S&B observers, so on and so forth. As shown, these combinations were selected in an iterative process, excluding the observers satisfied by the prior combinations in the subsequent iterations.

An experimental method for classifying color-normal observers

An experimental method for observer classification was implemented using two displays. The first was a 32" Sony BVM Cathode Ray Tube (CRT) display widely used as a studio reference display, and the second was an HP Dreamcolor (LP2480zx) Wide-Gamut Liquid Crystal Display (LCD) with LED backlight. For both displays, the luminance of the full white was set close to 97 cd/m^2. Spectral power distributions of the two displays are shown in fig 1. These displays were chosen because of the significant difference in their spectral characteristics, which meant a color match made on the two displays would be highly metameric in nature. The same experimental setup as in the color matching experiments was used, which is described in detail elsewhere [6].

The two displays were characterized using CIE 10° standard observer and each of the seven observer categories. Thus, corresponding to each of the eight sets of CMFs, a display forward and reverse model were determined.

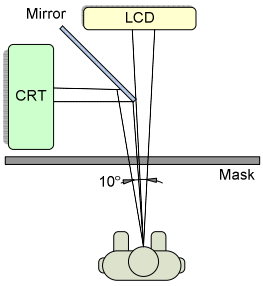

Figure 2. *Experimental setup*

In order to be able to identify the right category for a given observer, it is important that for each test color at least some of the seven versions of matches shown on the two displays are distinguishable from one another, and one (or possibly more) of these matches appear perceptibly better compared to the rest. This selection is limited by the spectral characteristics of the display primaries, since the displayed metameric colors are greatly affected by these characteristics. With this restriction in mind, there can be several possible ways to select the test colors. In this work, an algorithm was implemented to rank various colors based on the variance of tristimulus values corresponding to various observer categories. As before, the 240 ColorChecker patches were used. First, using display characterization data for the CRT and the LCD, seven pairs of XYZ tristimulus values were computed for each color. Thus for each of the 240 colors, there were seven sets of XYZ values predicted for the CRT, and seven corresponding sets of XYZ values predicted for the LCD. Root-mean-square (rms) distance of the two pairs of XYZ values were computed, which indicated how close the colors were in terms of respective tristimulus values for a given CMF-set. The variance (square of standard deviation) of these seven rms distances was used as a metric to determine if a color is suitable for observer classification. High variance indicated more variability in color matches among the seven versions of the test color. Note that even though XYZ values for various observer categories belong to different color

gamuts, the scales of the XYZ coordinate system are still the same (dependent on the wavelengths of monochromatic primaries in original Stiles-Burch experimental setup). This allowed us to compare these distances.

Once all the colors were ranked based on the variance metrics, fifteen colors were selected after a pilot test was performed with three observers to determine the suitability of the colors for the observer classification experiment. This visual test was necessary since the tristimulus space, in which the variance computation was performed, is not perceptual. Typically colors with relatively low chroma and low lightness turned out to be better candidates as test colors. Some of these 15 colors had similar hues, but different lightness levels.

An additional analysis was performed to identify the wavelength regions of x-, y- and z- functions with highest variability among the seven observer classes. Fig 3 shows the x-, y- and z- functions of the seven observer categories and the CIE 10° standard observer (black dots). Wavelength ranges where x-, y- and z- CMFs have highest variability are shown as vertical shaded lines. The vertical black lines correspond to the wavelengths where variances among the CMFs are the largest. Wavelengths around 580 nm, 520 nm and 430 nm have high variability in case of x-, y- and z- CMFs respectively. However, the variation in the x- CMFs of observer classes is not significant around 625-630 nm, where the red display primaries have sharp peaks. This further illustrates why the suitability of a color for observer classification is dependent on the spectral characteristics of the display primaries – one set of reference colors suitable for observer classification on one display may not be appropriate for another display. Note that these wavelength regions are specific to the observer categories and thus are independent of specific display primaries.

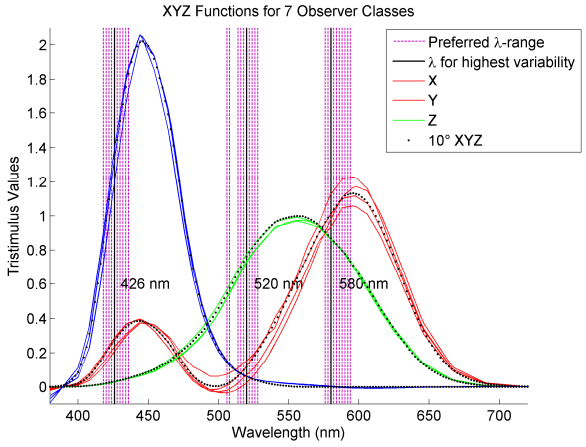

Figure 3. *Wavelength regions of x-, y- and z- functions with high variability among the seven observer categories*

Thirty observers took part in the observer classification experiment, including the ten observers who participated in the preliminary color matching experiments [6]. Both naïve and experienced observers participated. Ten observers were females. Many observers belonged to the age group 35-45. In separate trials, each observer was presented fifteen test colors. Each trial consisted of eight color-matches corresponding to the CIE 10° standard observer and the seven observer categories, which were shown on the CRT and LCD as uniform colors. The observers were able to conveniently browse through the eight versions using two buttons (forward and reverse) of a user control. The observer had no knowledge of the categories or the order in which they appeared. At the beginning of each trial, a random sequence was generated for the eight categories.

The observers were asked to assign the eight categories into one of three groups, namely, *unacceptable*, *acceptable* and *satisfactory*. This was accomplished in several steps, by: i) going through the eight versions to have an idea of the range of the color matches, ii) determining which of the eight color-matches have easily noticeable differences and thus are unacceptable matches; these were assigned to the *unacceptable* group and removed from the current trial, iii) determining which of the rest of the color-matches have perceptible differences, but are still acceptable matches; these were marked as *acceptable* and removed from the current trial, if needed, and iv) determining which color-matches have no perceptible difference; these versions were allocated to the *satisfactory* group. A software tool was developed that allowed the test administrator to assign or reassign any category to any of the above three groups. The tool also allowed removing or adding any category during the trial, a feature that was used in conjunction with random ordering for the verification of observer choices, when there was a sign of ambiguity or hesitation. The observers were free to assign any number of categories, none if needed, to any of the groups. For examples, in some cases no category was deemed as satisfactory.

The full session for each observer took between 45 minutes and one hour to finish.

Results and discussions

At the end of the test, a scoring table was formed for each observer by summing the total number of *satisfactory*, *acceptable*, and *unacceptable* scores for each of the eight categories, considering all 15 test colors. Table 3 shows two examples of such table for observers #1 and #8. Note that the category 1 is the CIE 10° standard observer. In determining the suitability of a category for any given observer, a high negative weight was assigned to the *unacceptable* counts, a small positive weight was assigned to the *acceptable* counts and a high positive weight was assigned to the *satisfactory* counts. Accordingly, an empirical performance score for each category was computed as per Eq. 2, and included in Table 3. Here, S, A and U represent fractional count (i.e. total counts divided by 15) of *satisfactory*, *acceptable*, and *unacceptable* groups respectively, R_i represents absolute scores of each category and R'_i represents relative scores, such that a score of 100 is assigned to the highest ranking category.

Through such scoring, the highest preference was placed on a category that was at least acceptable (i.e. acceptable or

satisfactory) for most of the test colors, followed by the higher number of *satisfactory* counts. For example, for observer #1, category 3 was preferred over category 5 since it was selected nine times as *satisfactory*, as opposed to seven times for category 5. For observer 8 on the other hand, category 2 received lower ranking than category 4 since the former was rejected once as *unacceptable*, even though they were judged *satisfactory* for the same number of times.

$$\begin{aligned} R_i &= 80S + 20A - 100U \\ U &= 1 - S - A \\ \Rightarrow R_i &= 180S + 120A - 100 \\ R_i' &= \frac{100R_i}{\sum_i R_i} \end{aligned} \qquad (2)$$

Thus, the objective of this analysis was to select a category that is more likely to result in an acceptable color match, even if it is not always the best possible match. This is graphically represented in fig 4, where each bubble corresponds to a category, and the area of a bubble is proportional to its relative score R_i'. The shaded bubbles are the assigned categories. Categories with non-positive scores, resulting from multiple *unacceptable* counts, are not plotted. Thus, the number of bubbles corresponding to a given observer and their relative sizes are indicative of the level of certainty with which we can assign a category to that observer. For example, there is higher uncertainty in category selection for observer #8 and little in case of observer #29. The observers belonging to the same categories are placed together for better visual interpretation of the results.

Table 3. Results for Observer 1 (top) and Observer 8 (bottom), showing for each category the total number of test colors belonging to various groups and the relative scores R'$_i$ for each category (category 1: CIE 10° standard observer)

Ranking \ CMF	1	2	3	4	5	6	7	8
Satisfactory	6	0	9	0	7	0	0	5
Acceptable	6	0	6	6	8	4	1	7
Unacceptable	3	15	0	9	0	11	14	3
Score	36	-179	100	-93	86	-121	-164	29

Ranking \ CMF	1	2	3	4	5	6	7	8
Satisfactory	9	12	2	12	10	0	11	0
Acceptable	5	1	12	2	4	4	3	9
Unacceptable	0	1	0	0	0	10	0	5
Score	82	88	40	100	88	-92	94	-32

For several observers, two categories received similar scores, while for observers #17, #22 and #29, even the best category was rejected for one or more test colors (not shown). These are expected since actual CMFs of an observer are not likely to exactly match with one of the categories, a difference that is manifested differently for various test colors, more so because these test colors are significantly influenced by the spectral characteristics of the display primaries. In such cases of ambiguity, it could be assumed that the chromaticities corresponding to various categories lied within the observer's tolerance, and so any of these categories, or their weighted mean could be used for classifying this observer. On the other hand, for observer #18, no category was deemed satisfactory for most colors, indicating the most suitable category for this observer is probably not included in the reduced set. It must be emphasized that this experimental setup is only meant to classify a given observer as belonging to one of the representative categories, and not to obtain his/her actual CMFs, which is impossible to achieve with such setup.

From fig 4, it is clear that the observer categories follow a definite pattern. For example, categories 5, 3 and 1 are closer to each other, while categories 2, 4 and 7 are closer to each other. Categories 6 and 8 are distinctly different from the others. With a very few exceptions, observers belonging to categories 3 and 5 rejected categories 2, 6, 7 and 8, observers belonging to categories 2 and 7 rejected categories 3, 5, 6 and 8, so on and so forth.

Also interesting is the fact that the CIE 10° standard observer (category 1) did not get the highest score for a single observer, although it was the 2nd best category for four observers. For observers #16, #17, #25 and #29, the standard observer color-matches were rejected for all 15 test colors (not shown). For all four, the categories could be determined with high certainty, indicating that the CIE 10° standard observer model is definitely inappropriate for these observers.

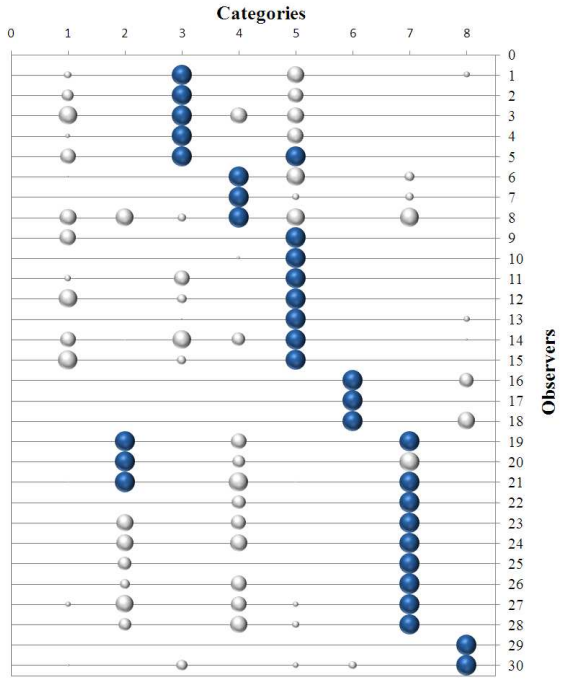

Figure 4. *Observer categories as determined through the observer classification experiment (category 1: CIE 10° standard observer)*

When considered alone, the CIE standard observer would probably produce an overall acceptable result for many of these 30 observers. But in comparison, other observer models produced better results relatively more often and thus were preferred over the CIE 10° standard observer. It is possible that given a choice, many observers would prefer a category different from the CIE 10°

standard observer. A possible explanation for the low preference for the CIE standard observer across the board lies in its derivation through the averaging over all CMFs, which results in a synthetic model that does not quite correspond to any real observer. Observers who are sufficiently different from the average unduly skew the results of the mean.

The two most popular categories are 7 and 5, representing 30% and 27% of observers respectively. Category 5 is somewhat close to the CIE standard observer as per our observer classification experiment. Category 7 is quite close to category 2, which, as per our previous analysis, was the dominant category for the Stiles-Burch observers.

Our results raise two fundamental questions: 1) should the "standard observer" be an average of the whole population, or should it be based on a statistical representation that better represents the majority of the population?, and 2) does a single "standard observer" continue to satisfy our need today, or is it time to have a provision for multiple observer models in applied colorimetry, and if so, how? In our ongoing work, we are attempting to address the second question. It is clear that multiple observer models may not be necessary, or even desirable, for industrial applications where observer metamerism is not a major issue, unlike modern wide-gamut displays and LED applications.

With respect to the first question, it is important to recognize that the best possible representation of the population of color-normal observers is critical, as the choice fundamentally affects our field. As far as an average match for all observers over the whole color space is concerned, the CIE 10° standard observer will probably still be reasonably good [6], but is it really the best possible representation of the color-normal population?

Conclusions and future work

The results from the first phase of observer classification experiment presented in this paper definitively confirm the existence of observer metamerism issue in modern displays with narrow-band primaries. But more importantly, they also show that such display systems can be exploited to better predict the variability in individual observers. The new method for observer classification described in this paper can help effectively address the issue of observer metamerism in industrial applications, and can also be a vital tool in fundamental color research. There is however no unique way to derive the observer categories. It is possible that there is some redundancy and/or insufficiency in our seven reduced sets of observer categories with respect to a large pool of color-normal observers. An inference can only be drawn when a sufficient number of observers are tested. However, our initial results confirm a key hypothesis of this work, that real, color-normal observers can be classified into a small number of categories by means of a practical experimental setup suitable for industrial applications. This is the first step in achieving our final goal of developing an *observer-dependent color imaging* method, where color workflow in a color reproduction device can potentially be tuned to one of several observer classes.

The immediate next steps in this ongoing work would be to further validate the observer categories through a wide-scale observer classification experiment, if needed, update the observer categories based on a large set of observers, and further refine the experimental method. It will also be important to establish a method for reversible transformations between the tristimulus space based on the CIE 10° standard observer and those obtained by using various observer categories. Finally, the effect and advantage of observer classification on complex images need to be investigated in viewing conditions that are typical in industrial applications.

References

[1] CIE, "Special Metamerism index: Change in observer", CIE Publ. No. 80, Central Bureau of the CIE, Vienna (1989).

[2] R.L. Alfvin and M.D. Fairchild, "Observer Variability in Metameric Color Matches Using Color Reproduction Media", Color Res. & Appl., 22(3): pg. 174-188 (1997).

[3] B. Oicherman, M.R. Luo, B. Rigg, and A. R. Robertson, "Effect of observer metamerism on colour matching of display and surface colours", Color Res. & Appl., 33 (5), pp. 346-359 (2008).

[4] M.D. Fairchild and D.R. Wyble, "Mean Observer Metamerism and the Selection of Display Primaries", Final Program and Proceedings-IS&T/SID Color Imaging Conference, pp. 151-156, Albuquerque, NM, USA (2007).

[5] R. Ramanath, "Minimizing Observer Metamerism in Display System", Color Res. & Appl., 34(5), pp. 391-398 (2009).

[6] A. Sarkar, L. Blondé, P. Le Callet, F. Autrusseau, P. Morvan, J. Stauder, "A color matching experiment using two displays: design considerations and pilot test results", Final Program and Proceedings, CGIV 2010 conference, Joensuu, Finland (2010).

[7] P. Csuti, and J. Schanda, "Colour matching experiments with RGB-LEDs", Color Res. & Appl., 33(2), pp 108-112 (2008).

[8] CIE Technical Report CIE 170-1:2006, "Fundamental Chromaticity Diagram with Physiological Axes - Part 1", 170-1:2006, (2006).

[9] A. Stockman and L.T. Sharpe, "The spectral sensitivities of the middle- and long- wavelength-sensitive cones derived from measurements in observers of known genotype", *Vision Research*, Vol. **40**, pp. 1711-1737 (2000).

[10] W.S. Stiles and J.M. Burch, "N.P.L. colour-matching investigation: final report", *Optica Acta*, Vol. **6**, pp. 1-26 (1959).

[11] A. Sarkar, L. Blondé, P. Le Callet, F. Autrusseau, J. Stauder, P. Morvan, "Study of Observer Variability on Modern Display Colorimetry: An Analysis of CIE 2006 Model", Final Program and Proceedings, the Congress of the International Colour Association (AIC) 2009, Sydney, Australia (2009).

[12] A. Sarkar, L. Blondé, P. Le Callet, F. Autrusseau, J. Stauder, P. Morvan, "Study of Observer Variability on Modern Display Colorimetry: Comparison of CIE 2006 Model and 10° Standard Observer", Final Program and Proceedings, the 11th Congress of the International Colour Association (AIC) 2009, Sydney, Australia (2009).

[13] R.A. Johnson and D.W. Wichern, Handbook of Applied Multivariate Statistical Analysis, 6th Ed., Pearson Education International, NJ, USA (2007)