Final Report for R1-54: Variability In Color-Matching Functions

Reporter: Abhijit Sarkar (IN)

Established: 2011

Terms of Reference
To document available data that describe the variation in color matching functions, together with an analysis of their variability.

Introduction
This reportership was established primarily in response to the new findings presented in this reporter’s doctoral thesis work [1], conducted at Technicolor Research & Innovation, and supervised by the University of Nantes, France. In that work, a method was proposed to classify color-normal human observers into a small number of categories. These categories were described by a set of color-matching functions derived through a theoretical analysis of existing large-field (10°) color-matching data, which were obtained experimentally as well as through mathematical modeling.

The goal of this reportership is to summarize these new findings and justify through scientific evidence why there is a need to look at the aspects of observer variability and the colorimetric observer from a fresh new perspective. For a more comprehensive literature review on observer variability, this reporter’s doctoral thesis [1] (Chapter 3) and Oicherman’s doctoral thesis [2] should be consulted.

The reader should note that the term “observer metamerism” has been avoided in this report, even though it is often used interchangeably with observer variability. Observer metamerism implicitly assumes the existence of two stimuli. We simply cannot define or describe observer metamerism without the context of color stimuli. Observer variability on the other hand is a more generic term, implying differences in the color vision characteristics (i.e. color-matching functions) among individual observers. We can think of observer variability as the cause, and observer metamerism as the effect.

Key sources of observer variability data
While many studies in the realm of color vision and sciences have produced valuable data on observer variability, following two are arguably the most comprehensive sources of such data, the first being experimentally determined functions for large-field color-matching, and the second being theoretical color-matching functions, resulting from mathematical modeling of physiological aspects of the human visual system.

Stiles and Burch’s color-matching experiment (1959)
More than fifty years ago from the time of writing this report, Stiles and Burch [3] undertook the most comprehensive, and arguably the most authoritative large-field color-matching experiment till date. This experiment, conducted at the National Physics Laboratory,
Variations of Rayleigh matches have also been examined experimentally on the Color & Vision Research Laboratory website [6].

CIE TC 1-36 Physiologically-based observer (2006)

In 2006, CIE’s technical committee TC 1-36 published a report [7] (described hereafter as CIEPO06, an abbreviation of CIE physiological observers 2006) on the choice of a set of color-matching functions (CMFs) and estimates of cone fundamentals for the color-normal observer. The CIEPO06 model is largely based on the work of Stockman and Sharpe [8]. Starting from 10° CMFs of 47 Stiles-Burch observers [3], the model defines 2° and 10° fundamental 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.

A detailed theoretical analysis [9] of CIEPO06 has shown that using real observer ages in the model can lead to large errors in intra-age group average Stiles-Burch observer CMF predictions. In its approach to construct a fundamental observer, the technical committee CIE TC 1-36 did not take into account individual variability [7] [10]. Thus, it is obvious that CIEPO06 model cannot and should not be used to predict individual color-matching functions. The above study [9] additionally concluded that in its current form, CIEPO06’s usefulness to even predict average colorimetric observer for a given age group might be limited. This conclusion needs further verification. Nevertheless, CIEPO06 provides an excellent framework for modeling overall observer variability over a wide age range.

Studies on individual cone fundamentals

A few studies [11] have dealt with individual variations of color-matching functions, analyzing the data collected by Stiles and Burch using 10° fields, examining the differences between the CMFs of the CIE 1931 standard colorimetric observer, the Judd’s revision of this set and the set of 2° CMFs collected by Stiles and Burch [12], comparing inter-individual and intra-individual variability of experimental CMFs [13]. Wyszecki and Stiles [5] (page 348) produced a global statistical analysis of the dispersion of the data collected by Stiles and Burch using 10° fields. In the last ten years, a few sets of matching results have been generated at low or moderate luminance levels to investigate intra- and inter- observer variability [14] [15], and test additivity and transformability of color matches [16] [17]. One study of nine observers’ color-matching functions concluded that a main cause of the individual differences was the difference of individual spectral lens density [18]. Individual variations of Rayleigh matches have also been examined experimentally [19] [20] [21] [22] or
theoretically [23]. Although all these studies have attempted to relate the variation of color matches to underlying physiological factors, they were not designed to model individual effects of these factors in a practical manner that could be implemented in applied colorimetry.

Quantifying observer variability: CIE standard deviate observer (1989)

Starting from the early eighties, several researchers attempted to quantify the extent of metamerism using the color-matching data from 20 observers, selected out of the 49 observers of the Stiles and Burch’s experiment [3]. The observers were selected based on their reliability and experience in trichromatic matching, not based on their actual results [5] (page 346).

Allen [24] was the first to propose the concept of a standard deviate observer and a general index of metamerism. The idea was to derive a standard deviate observer who has color-matching functions differing from the standard observer by amounts equal to standard deviations among the 20 sets of CMFs. It was a statistical construct involving analyses of variances and covariances of 20 sets of CMFs.

Mainly based on the works of Nayatani et al. [25] and Takahama et al. [26], the CIE published in 1989 a technical report titled Special Metamerism Index: Change in Observer [27] (henceforth referred to as the CIE standard deviate observer). The index was based on the computed color difference between the standard deviate observer and any of the standard colorimetric observers under a specified standard illuminant. Till date, it is the only official model that attempts to quantify observer variability.

However, the CIE standard deviate observer model did not perform well when evaluated with independent experimental data. Many researchers reported [15] [28] [29] [30] that the model under-estimated the variations in color-matching data of real observers. The suggested explanations for this failure were exclusion of some of the Stiles-Burch observers from the analysis which led to the development of the CIE standard deviate observer, and improper mathematical treatment of the original color-matching data [15]. Looking from the point of view of practical industrial applications, in particular hard-copy vs. soft-copy color-matching, some researchers [30] [2] have questioned the purpose and usefulness of an index of observer metamerism, and of a standard deviate observer. They suggested that individual variability in these conditions is governed by mechanisms of chromatic discrimination, and could be modeled by advanced color difference formulae with suitably adjusted parametric coefficients.

A Review of the thesis work on observer classification

The reporter’s doctoral thesis work [1] sought to offer a practical solution to the issue of observer variability in large-field color-matching. A key hypothesis of this work was that color-normal human observers could be classified into a small number of categories based on their color vision. Following description summarizes the key milestones achieved in the research:
1. Derivation of eight representative colorimetric observer categories through statistical analysis

An assumption in this work was that the CIEPO06 model predictions [7] and the experimentally obtained visual color-matching data from the 1959 Stiles-Burch study [3], when combined together, incorporated most of the variability that could be found among the color normal population. The combined dataset used in this study thus included 61 CIEPO06 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 two-step method was developed for deriving a minimal set of representative colorimetric observer models (or categories). In the first step, five distinct and representative long-wave sensitive (LWS), medium-wave sensitive (MWS) and short-wave sensitive (SWS) cone fundamentals (a total of 125 combinations) were derived through a cluster analysis on the combined dataset. Squared Euclidean distance measure (in cone fundamental space) was used in this analysis.

In the second step, a reduced set of eight representative observer models (or categories) was derived from the 125 combinations through an iterative algorithm. This derivation was based on several predefined criteria on perceptual color differences with respect to actual color-matching functions of the 47 Stiles-Burch observers and spectral power distributions of a large set of color stimuli.

A key aspect of the method used in deriving the observer categories was that both spectral and colorimetric attributes of the color-matching functions were considered to minimize model redundancy and ensure uniqueness of the selected categories.

However, it should be pointed out that combined dataset is still not perfect. The dataset does not take into account genetic influence on the peak LWS and/or MWS cone fundamentals [7]. So it is possible that observer variability predicted by the dataset is an under-estimation of what exists in the global population of color-normal observers. Although Stiles-Burch observer group did not have enough representation of elderly population (average observer age was 32), it is reasonable to assume that incorporating CIEPO06 predictions addressed that weakness.

2. Development of an observer classification method and implementation using two displays

An experimental method, employing two displays with widely different spectral characteristics, was developed in order to assign colorimetric observer categories to individual observers. The displays were characterized using the CIE 10° standard colorimetric observer and each of the colorimetric observer categories. Pairs of matching colors as predicted by various observer categories were shown on the two displays, and the observer was asked to choose the best matching pair through a multi-step experimental protocol. The chosen matching pair corresponded to a specific observer category. This process was repeated for several base colors. Finally, an empirical ranking system was used to determine the most appropriate observer category that resulted in superior color matches for most base colors for the given observer.

3. Development and testing of Observer Calibrator prototype

A portable, LED-based instrument prototype for observer classification was conceived during
the work. This prototype replicated the observer classification experimental setup based on two displays described in the previous section. The prototype has an LED driver that controlled the LEDs. The LED driver is interfaced to a computer. A software application residing in the computer can send appropriate signals to the LED driver in order to generate specific colors on both sides of the bipartite field. A user control device connected to the computer through Universal Serial Bus (USB) allows the observer to browse through various versions of a color match corresponding to individual categories.

Two collaborative experiments were performed in Germany and Hungary, involving a total of 49 observers (coincidentally the same number of observers as in 1959 Stiles-Burch study [3]). Results (Fig. 1) revealed dominant categories among the observer groups, which were different in the two experiments. Blue circles in the figure show the assigned categories for individual observers.

![Fig. 1. Results from observer classification experiment in Darmstadt, Germany (left) and Veszprem, Hungary (right) using the observer calibrator (category 1: CIE 10° standard colorimetric observer)](image)

A correlation analysis was also performed on observer classification data from the experiment in Germany, and suprathreshold color difference judgments obtained from an independent experiment involving the same set of observers. Some interesting correlation was observed between the observer categories and perceived color differences [1] (Chapter 6).
Discussion

1. Revisiting the concepts of standard and “deviate” colorimetric observers

According to the results of the experiments described in the previous section, the CIE 10° standard colorimetric observer was the chosen category for only around 8% of all observers tested - a result that was in congruence with the previous findings during the thesis work. In comparison, the three most popular categories were selected for 25%, 20% and 18% observers respectively. This raises the question if the current standard colorimetric observer has room for improvement. A possible explanation for the low preference for the CIE standard colorimetric observer across the board lies in its derivation through the averaging over all Stiles-Burch observer CMFs, which resulted in a mathematical model of an “average observer” that did not correspond to any of the real observers who participated in the observer classification experiments. In other words, Stiles-Burch observers who were sufficiently different from the average unduly skewed the results of the mean. Another important factor to keep in mind is that the standard colorimetric observer was purely a mathematical construct. Its derivation process neither involved additional verification and validation through psychophysical experiments, nor was there a competing observer model that could be used in a comparative visual study. Understandably, adopting such an extensive multi-step process would have been a difficult proposition back in 1964. Until now, such weaknesses were not of huge significance since most color system primaries have traditionally had relatively broad spectral characteristics that mitigated observer variability. That is no longer the case with the advent and widespread popularity of devices with narrow-band primaries, e.g. wide-gamut displays and Light-Emitting Diode (LED) or Laser-based sources.

Considering no single observer category satisfied even a quarter of the observer panel raises another important question: whether a single observer model can or should indeed be used for the whole population of color normal human observers, even when the application context demands better accuracy. The concept of a single average observer has been so fundamental to colorimetry that the representation of any observer who cannot reasonably be represented by an average has been conceived as a “deviate observer”. Granted, the terms “standard” and “deviate” were likely used by the scientific community in purely mathematical context. Nevertheless, non-experts often interpret the phrase “deviate observer” with a negative connotation. The predominant perception is that a human observer should have the same or similar color vision as represented by the “standard”; otherwise he or she has an abnormal color vision. In a way, the terminology used traditionally in color science community gives way to this undesirable misinterpretation. It is important to acknowledge and convey that it is perfectly natural for individual color normal observers to be different from each other. Moreover, an acknowledgment of this fact will encourage people to be less dismissive of others’ disagreement on color perception judgment. Accepting that a problem exists is the first step toward finding a resolution.

As has been shown in the thesis, observers belonging to a category closer to the average for a given group of observers do not necessarily represent the dominant categories with respect to a larger population. Over the past several decades enough progress has been made in the area of human color vision as well as digital color imaging technologies. This warrants a revisit to the definition and usage of observer models in colorimetry. There is really no unique way of defining a single “standard observer” or a “deviate observer”, and
no such attempt is probably necessary.

2. A more logical way of treating observer models in colorimetry

Defining the colorimetric observer models in colorimetry could be similar to what the CIE did to define the “standard illuminants”, by using terms like CIE standard illuminant A, B, D50, D65 etc. The observer models could be named based on their frequency of occurrence in a large population of color normal observers irrespective of their gender, race and genetics. A general agreement could be reached on using, for example, “standard colorimetric observer model A” under normal circumstances. One advantage of this method is no single model is claimed as “standard” or “deviate”, just like the CIE illuminants. The other advantage is that for a restricted population, a color or lighting professional can choose to use a more appropriate model. For example, two of the eight categories proposed in the thesis were considered to be predominant in the elderly population and rather infrequent in the general population of younger observers.

Of course such colorimetric observer categories cannot represent the color vision of individual observers precisely, but a carefully established set of representative observer models will give a more accurate individual color matching predictions than what is possible today with a single large-field standard colorimetric observer. It is important for the field of colorimetry to offer such improved accuracy to applications that need it. For applications that don’t need higher accuracy, current standard colorimetric observer or its improved version can be used, which could be one of the representative observer models.

3. Practical challenges in having multiple observer models in color science

One of the main challenges in having multiple observer models in colorimetry is in the implementation of a color imaging workflow. How do we accommodate multiple observer-dependent color representations without introducing unmanageable amount of complexity? To fully appreciate the complexity, we need to consider the fact that all color difference equations and perceptual color spaces are ultimately based on a single standard colorimetric observer. They may not hold true if we choose to work with a different observer model.

Another practical challenge is to address the concern that would invariably come up - do we really need such complexity? The fact is for many traditional color applications, issues arising due to observer variability are not really significant. Because of that reason, it might not be very easy to garner community-wide support for instituting multiple-observer models in colorimetry. Many would also be concerned about the “backward compatibility”, how we can continue to support systems and applications that may still need to use the current standard colorimetric observer.

In no way less challenging would be the task to define and agree upon a finite set of observer models, or categories, for use in colorimetry.

4. Need for a solution is real and urgent

The inability of applied colorimetry to account for observer variability is arguably its most serious drawback. It is becoming more and more evident that this shortcoming is non-trivial in several color-critical applications, for example in modern display colorimetry and various LED or Laser-based applications. There is a general agreement in the display industry that traditional color reproduction approach is leading to frequent disagreement on color-matches as judged by different individuals across various displays at least one of which has
narrow-band primaries. This was recently highlighted in a report submitted to the European Broadcasting Union’s (EBU) Strategic Program “FTV-Display”. Many Division 1 members have received this two-part report, prepared by Mr. Lars Haglund of Sveriges Television AB (SVT), the public service broadcaster in Sweden, highlighting the practical color reproduction issues faced by the display engineers in the recent years while employing the standard colorimetric observers. These reports are highly relevant for CIE, and should be reviewed alongside R1-54.

5. Thoughts on a possible approach to address observer variability

A potential practical solution is to implement an observer-dependent color imaging workflow at the device level. An example of such a workflow has been proposed in the thesis [1] (Chapter 7). Conceptually this is similar to the device-dependent color imaging, a well-established color management concept. Implementation of this workflow at the device level necessitates that colors corresponding to the generic standard colorimetric observer be converted into category-specific colors. This would require that the transformation of category-specific colors to generic observer-independent colors be standardized, along with the observer categories themselves.

This approach has the advantage that the choice of specific observer categories would be left to the device manufacturers, without affecting the generic color representation in an end-to-end color workflow. Applications where observer variability is deemed insignificant would thus be unaffected by the introduction of observer categories. It must be emphasized that the proposed approach is not the only solution, and its merits need to be validated by further research.

It is imperative that the observer categories be established not only based on theoretical analysis of classical color-matching data, but also be validated by psychophysical experiments involving a large number of observers. In this regard, the observer calibrator prototype developed in the thesis work is an ideal scientific instrument. It has nearly linear colorimetric characteristics, and thus can reproduce colors more accurately than what is possible in a typical display. Its bipartite field allows full control on the viewing condition, allowing quasi-symmetric color-matching. It can be used in unsupervised, computerized test that can be completed in less than an hour. Most importantly, it is portable, so it can be used in multiple laboratories around the world to test a vast number of observers. However, the prototype needs to be improved before it can be deployed for such purpose.

The Munsell Color Science Laboratory (MCSL) at RIT and Technicolor Research & Innovation are starting a collaborative research project in September 2012 to take the observer classification work forward. Some of the project goals are to independently verify the findings of the thesis [1], possibly improve the observer category set by minimizing redundancies, and improve the observer calibrator design. The project will be executed under the joint supervision of Prof. Mark Fairchild of RIT and Dr. Laurent Blondé of Technicolor.

Prof. János Schanda and Prof. Ronnier Luo expressed interest in testing the observer categories in their research laboratories. Possibly some other research teams would also come forward in near future to collaborate on this topic.
Conclusion and Recommendation

The main purpose of this report was to summarize the recent findings on observer variability. Hopefully the foregoing discussion will encourage the reader to revisit the fundamental aspects of the colorimetric observer with a fresh new perspective.

It seems reasonable to recommend that R1-54 be closed and a Technical Committee be formed under Division 1 to investigate various aspects of standard colorimetric observers. Two key aspects need to be investigated: i) if the current standard colorimetric observer for large-field color-matching can and should be improved, and ii) if multiple colorimetric observer models (i.e. observer categories) can and should be instituted, and if so, how they should be used. It is better to leave the formulation of the Terms of Reference to the Division 1 officers. The Terms of Reference could possibly encompass the small-field (2°) standard colorimetric observer, but it was not within the scope of the findings reported here. A rough guideline for future work that has been presented at the end of the thesis could act as the starting point for a discussion at the CIE.

If the CIE Division 1 decides to form a Technical Committee, it should coordinate and facilitate future research activities in various laboratories on this topic, and ensure that a concerted effort be made toward a common goal.

The recommendations in this report may seem to contradict those of R1-43, which suggested that more resources not be allocated to the study of variability of quasi-symmetric color-matching. However, the context of those recommendations was cross-media color-matching. Current report looks at the issue of observer variability from a more fundamental standpoint, and advocates for a new approach in applied colorimetry. A comprehensive solution must start at quasi-symmetric color-matching without involving color appearance issues. Perceptual color difference equations and perceptual color spaces must be built upon a robust colorimetry capable of adequately handling observer variability.

Acknowledgment

Following individuals were involved in reviewing the thesis work in various capacities, and provided extensive feedback and corrections. Since a major part of this report is based on what came out of that work, I would like to acknowledge their indirect contributions to R1-54:

Dr. Laurent Blondé, Dr. Florent Autrusseau, Prof. Patrick Le Callet, Prof. Françoise Viénot, Prof. Mark Fairchild.

I must also acknowledge my ex-colleagues at Technicolor R&I, Mr. Patrick Morvan and Dr. Jürgen Stauder, for their significant contributions to the research project.
References


[27] CIE, Special metamerism index: Change in observer, CIE Publ 80, Central Bureau of the CIE, Vienna (1989).

