Identification and Assignment of Colorimetric Observer Categories and Their Applications in Color and Vision Sciences

Abhijit Sarkar

UNIVERSITÉ DE NANTES

Ecole polytechnique de l'Université de Nantes

———

ÉCOLE

Année 2011

N° attribué par la bibliothèque

Identification and Assignment of Colorimetric Observer Categories and Their Applications in Color and Vision Sciences

———

THÈSE DE DOCTORAT

Discipline : AUTOMATIQUE ET INFORMATIQUE APPLIQUEE

Spécialité : TRAITEMENT DU SIGNAL ET DES IMAGES COULEUR

Présentée
et soutenue publiquement par

Abhijit Sarkar

Le 26 octobre 2011, devant le jury ci-dessous

Président	Mme. Sabine Süsstrunk	Professeur, School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Suisse
Rapporteurs	M. Mark Fairchild	Professeur, Centre for Imaging Science, Rochester Institute of Technology, Etats-Unis
	Mme. Françoise Viénot	Professeur MNHN (Muséum National d'Histoire Naturelle), Centre de Recherche sur la Conservation des Collections, Paris
	M. Laurent Bédat	Maître de conférences, IETR, INSA de Rennes
Examinateurs	M. Patrick Le Callet	Professeur, IRCCyN, Polytech Nantes, Université de Nantes
	M. Florent Autrusseau	Ingénieur de Recherche, IRCCyN, Université de Nantes
Invité	M. Laurent Blondé	Principal Scientist, Technicolor R&I, Rennes

Directeur de thèse : M. Patrick Le Callet

ED : …….

(Uniquement pour STIM et SPIGA)

Acknowledgments

Moving to France for a 3-year industrial PhD, and adjusting to a new culture with no prior knowledge of the language have been great adventures for me. The person who made a lot of things a lot easier was my supervisor at Technicolor R&I, Dr. Laurent Blondé. It is mainly because of him I survived the agonizing experiences with the exceedingly difficult French administrative system. He and his wonderful family helped me integrate with the French culture and lifestyle, a process that has been an amazing experience for me. On the professional front, Laurent played a critical part in bringing this thesis to the state where it stands today.

I must also thank my colleagues, in particular Mr. Patrick Morvan and Dr. Jürgen Stauder, who extended a tremendous amount of support in every possible way throughout these years. I feel lucky to have had colleagues like them. I am also thankful to Technicolor management for allowing me all the freedom a corporate researcher can dream of.

The guidance and supervision of my academic supervisors Prof. Patrick Le Callet and Dr. Florent Autrusseau at IRCCyN ensured my work remained focused and had adequate technical depth. I am also thankful for all their support at the personal level, particularly during my initial days in France.

I would like to thank the members of my jury, Dr. Laurent Bédat, Prof. Mark Fairchild, Prof. Sabine Süsstrunk and Prof. Françoise Viénot, for accepting to be part of the committee. In particular, I am indebted to my reviewers, Prof. Fairchild and Prof. Viénot, for their valuable and comprehensive feedback on the thesis manuscript. Prof. Fairchild was my advisor at the Munsell Color Science Laboratory, and continues to be a source of inspiration. Prof. Viénot extended an incredible amount of help and advice during the final phase of my PhD, and I am highly appreciative of all the time she generously devoted to our numerous discussions.

This research would not have been possible without the active participation and cooperation of all my 79 observers at Technicolor R&I, Rennes, at the Technische Universität Darmstadt, Germany and at the University of Veszprém, Hungary. I express my heartfelt gratitude to all of them.

As I complete my eight years of graduate studies, I recall the unwavering support of my family for all these years − all the encouragement and inspiration that motivated me to chase my wild dreams.

Finally, I would like to acknowledge Dr. Stefan Winkler for compiling a list of many wonderful quotations on his website, http://stefan.winkler.net/, some of which I have used in this thesis.

Abstract

The most fundamental aspect of applied colorimetry is the trichromacy of our visual system. Trichromacy leads to observer metamerism, in which two stimuli with very different spectral power distribution can produce a color match for a given observer, but will result in a mismatch for another observer with different color vision characteristics. This variability among observers with normal color vision poses a challenge to various modern industrial applications, including wide-gamut displays with narrow-band primaries, and Light-Emitting Diode (LED) or Laser based applications. Thus, the main objective of this thesis is to offer a practical solution to this problem for color-critical industrial applications.

This work starts by conducting a comprehensive theoretical analysis on various aspects of the physiologically-based observer model (CIEPO06) proposed by the Technical Committee TC 1-36 of the Commission Internationale de l'Éclairage (CIE). In the context of color perception on modern narrow-band displays, the performances of the CIEPO06 model and of the CIE 10° standard colorimetric observer in predicting average Stiles and Burch (1959) observer data were evaluated. Some weaknesses of both observer models were identified, and an improvement of the CIEPO06 model was proposed based on a nonlinear optimization.

In the next stage, several color-matching experiments were performed on two displays with very different spectral characteristics, one was a Cathode Ray Tube (CRT) display, and the other was a Liquid Crystal Display (LCD). The results confirmed the effect of observer metamerism in display color matches.

Working toward a solution, a statistical analysis was performed on existing experimental and physiological datasets of color-matching functions. A set of eight colorimetric observer categories was proposed for use in color science and vision. Subsequently, an experimental observer classification method using two displays was developed. Through visual experiments it was proved that human observers with normal color vision can be classified into a small number of categories based on their color vision. This was followed by the development of a compact, inexpensive proof-of-concept prototype, described as the *Observer Calibrator* in this thesis. Using this prototype, two collaborative observer classification experiments involving a total of 49 observers were performed with researchers in Germany and Hungary. A correlation analysis was 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. The consistency between observer categories and color difference data gave an indirect validation of the observer classification method.

Finally, an implementation of colorimetric observer categories in a practical color imaging workflow has been proposed. This workflow, described in this thesis as the *observer dependent color imaging* (ODCI), involves conversion of tristimulus values corresponding to CIE 10° standard colorimetric observer, into the tristimulus values corresponding to individual observer categories. Nonlinear transformations that result in accurate color transformations have been derived.

The observer classification method, together with the compact and economical prototype, is the enabling factor for the practical implementation of *observer dependent color imaging* workflow in industrial applications. It is also hoped that the contributions of this thesis will be valuable for scientific research in the domains of color and vision sciences.

Table of Contents

Twenty years from now you will be more disappointed by the things you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor. Catch the trade winds in your sails. Explore. Dream. Discover.

- Mark Twain

When we walk to the edge of all the light we have and take the step into the darkness of the unknown, we must believe that one of two things will happen: there will be something solid for us to stand on or we will be taught to fly. ~ Patrick Overton, Faith

1. Introduction

1.1 Motivation

When two color stimuli produce the same visual response, a visual match is obtained. Two stimuli with very different spectral power distribution can give rise to identical cone response, leading to a color match. However, such a match established by one observer can, and quite often does lead to a mismatch for a different observer, as the second observer has a different set of color-matching functions (CMFs) than the former. This phenomenon is commonly termed as observer metamerism.

Various studies in the past, both classical and applied, have provided significant amount of insight into the issue of observer variability in color-matching, and its ramifications in basic color science and applied color technology. While over the past couple of decades our knowledge of underlying physiological reasons for individual variability in human color vision has been enriched considerably, we are yet to come up with a practical solution accounting for variability in applied colorimetry. Being constrained to a single average observer model, colorimetry is unable to predict how individual color matches might differ from those of an average match. The consequence is non-trivial for certain color-critical industrial applications.

One example is the color adjustment process (called color grading in industrial parlance) in cinema and television post-production applications where the raw movie content at the post-shooting stage is modified to achieve the right color effect. The Colorist has to work with the Director of Photography (DP) to adjust the colors in the original content so as to achieve color coherence and homogeneity throughout various scenes, while maintaining the artistic expressions originally envisioned by the Film Director and the DP. However, if the Colorist and the DP have different color vision characteristics, they will perceive colors differently, and the colors that look similar to one will look perceptibly different to the other. While the art of colorists fills the gap, conventional colorimetry will fail to account for this difference in color vision.

The broad objective of this work is to propose a framework and a color imaging workflow that takes individual observer variability into account, and provides a practical solution for industrial applications.

1.2 Research hypothesis

A principal hypothesis of this work is that human observers with normal color vision can be classified into a small number of categories based on their color vision. These observer categories, characterized by specific color-matching functions, can be identified through an appropriate statistical analysis of a large set of individual observer data. Based on such categorization of the whole observer population, multiple colorimetric observer models can be established for use in applied colorimetry.

An associated hypothesis, without which a practical application of colorimetric observer categories will be impossible, is that such categories can be adequately identified. In other words, it is hypothesized that there exists a simple, practical means to experimentally determine which of several categories can be assigned to a certain color normal human observer. A keystone of this work is the premise that *the solution lies in the problem itself* - meaning that devices that are fraught with observer variability and metamerism issues, are our best bet in coming up with a solution to these problems. Examples of such devices include modern wide-gamut displays with narrow-band primaries.

With regard to the first hypothesis, it is expected that the spectral characteristics of the color-matching functions specific to a given observer category would not match exactly to individual observers who are assigned that category. However, with proper category identification, overall colorimetric results obtained by using the assigned category for any given observer can be expected to be more accurate than the results yielded by any other category or a standard colorimetric observer. Accordingly, usage of colorimetric observer categories in colorimetry can reduce the problem of observer metamerism.

Another point needs to be made with regard to the second hypothesis. While the application of colorimetric observer categories may not be meaningful in an application where several observers are simultaneously viewing colors on a device or medium, under certain conditions it might be useful. As an example, when all the observers concerned are in the higher age group, it would be more appropriate to use the categories that are more prevalent among higher age-group observers. Indeed, prevalence of certain categories among higher age-group observers is supported by the results obtained in this thesis. However, more direct benefit of the concept of observer categories seems to exist in applications where accurate color reproduction for individual observers is desired.

1.3 Organization of the thesis

This thesis is organized in eight chapters.

Chapter 2 reviews several fundamental concepts and principles of color science and color vision, focusing mainly on those aspects that are relevant for this thesis. It starts with a review of the anatomy

and physiology of the human visual system. Then, it deals with various aspects of the perception of color, followed by an introduction to the colorimetry and visual color-matching. The universally accepted colorimetric system proposed by the CIE is described next. The chapter concludes with enumerating various physiological sources of individual differences in color-matching.

Chapter 3 presents a literature review on the phenomenon of observer metamerism, and how individual observer variability can affect color-matching. Both classical color matching experiments involving monochromatic stimuli and applied color-matching experiments involving narrow-band and broad-band stimuli are reviewed.

A comprehensive theoretical analysis on the age-dependent physiological observer model recently proposed by the CIE Technical Committee 1-36 (henceforth CIEPO06) constitutes Chapter 4. The chapter starts by discussing the colorimetric observers. Next, various physiological factors on display color perception are discussed, followed by a comparative analysis on the performance of the CIEPO06 model and 1964 CIE 10° standard colorimetric observer in predicting the average observer data within a given age group. The analysis considers spectral sensitivity data as well as colorimetric data in the context of displays. A nonlinear optimization of the CIEPO06 model is performed, and the results are analyzed.

Chapter 5 explores the issue of observer variability in the context of display color-matching. A color-matching experiment is performed on two displays with very different spectral characteristics, one with narrow-band primaries, and the other with broad-band characteristics. Detailed description of the experimental design is presented, as well as the results and analysis.

The most important contribution of this work, namely the development of colorimetric observer categories, is presented in Chapter 6. An observer classification method using two displays (the same two described in Chapter 5) is proposed. Further, the *Observer Calibrator* prototype developed during the course of this work is described. Finally, results from collaborative experiments performed with two research laboratories in Germany and Hungary are presented. One of these experiments provides indirect validation of observer classification method.

Chapter 7 presents the concept of Observer-dependent color imaging workflow. The implementation aspects of the workflow are described. The advantages of such a workflow are discussed.

Finally Chapter 8 draws conclusions from this work.

I would rather live in a world where my life is surrounded by mystery than live in a world so small that my mind could comprehend it. ~ Harry Emerson Fosdick

2. A review of color vision and color science fundamentals

In this chapter, several fundamental concepts and principles of color science and color vision are reviewed, focusing mainly on those aspects that are important for a good comprehension of this thesis. It is not meant for experts in the topical area of this thesis, but for those interested readers less familiar with the field. The chapter starts with a review of the anatomy and physiology of the human visual system. Then, it deals with various aspects of the perception of color, followed by an introduction to the colorimetry and visual color-matching. The universally accepted colorimetric system proposed by the CIE is described next. Finally, various physiological sources of individual differences in color-matching are described.

2.1 The human visual system

The human visual system is an enormously complex and sophisticated biological organ. It is estimated that around 80-90% of all neurons in the human brain interact with visual signals [1]. Not surprisingly, it took us many centuries to develop a scientific understanding of the functioning of the visual system. Galen (AD 130 - 200), a Roman physician, surgeon, and philosopher considered to be the most accomplished of all medical researchers of antiquity, attempted to explain this functioning. He proposed that the light rays emanated from the eye, interacted with the object, and then returned to the eye, wherein the rays interacted with a "visual spirit" that flowed from the brain to the eye and back, carrying with it the replicas of perceived objects [2] (page 24). As amusingly unscientific as the proposition was, to Galen's credit, he at least correctly anticipated the involvement of brain in our visual functioning, as has been established by the modern day vision science.

2.1.1 *The eye: anatomy and physiological optics*

Hubel, co-winner of the Nobel Prize in Physiology/Medicine in 1981 for mapping the visual cortex, elegantly describes [3] the sophistication of eye as a sensory organ: "*The eye has often been compared to a camera. It would be more appropriate to compare it to a TV camera attached to an automatically tracking tripod—a machine that is self-focusing, adjusts automatically for light intensity, has a self-cleaning lens and feeds into a computer with parallel-processing capabilities so advanced that engineers are only just starting to consider similar strategies for the hardware they design.*"

The two eyes in a human, sitting in the hemispherical eye sockets, are able to undergo rotations through three pairs of *extraocular muscles*, which are controlled by the brain. Voluntary fixation mechanism allows high speed movements to steer the eye from one direction to the other, commonly known as saccades. An involuntary fixation mechanism allows the eye to fixate at a point. The visual fields from the two eyes overlap, allowing binocular vision and depth perception.

The optical system of the human eye is composed of the cornea, the aqueous humor, the lens, and the vitreous humor, as shown in Fig. 2-1.

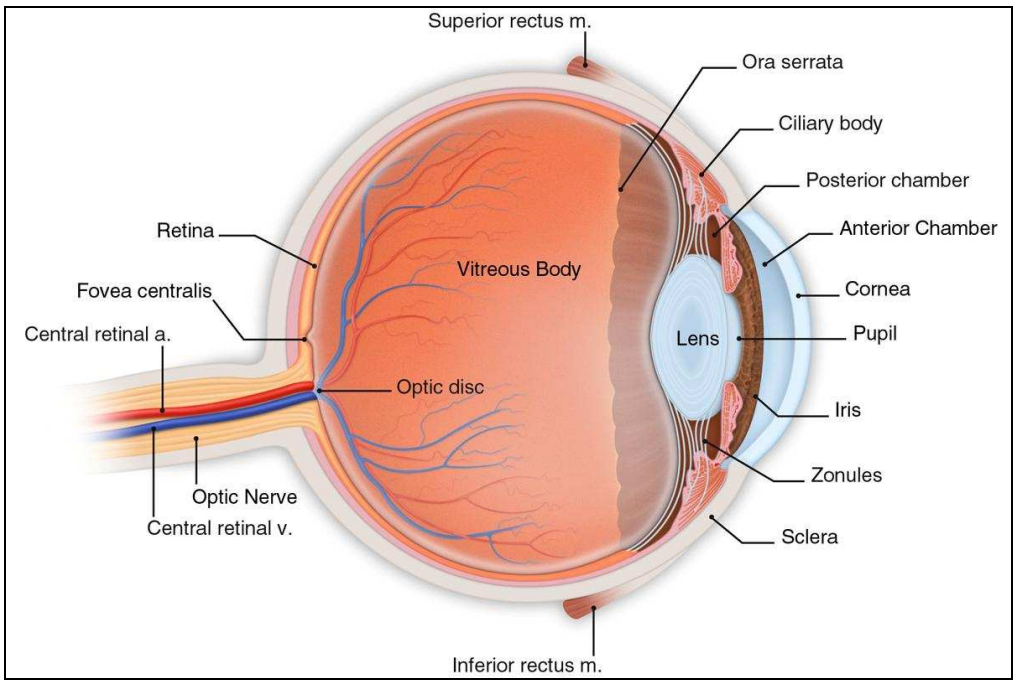


Fig. 2-1. An anatomical drawing of a human eye in cross-section (http://www.newsomeye.com/patient-education/anatomy-of-the-eye/)

First, light enters the transparent layer of *cornea*, behind which is the anterior chamber filled with a transparent liquid called *aqueous humor*. About two-third of the optical power of the eye (the ability to bend incoming light) needed for focusing takes place at the air-cornea transition. The *lens* has only a third of the total refractive power of the eye, due to optically similar characteristics (refractive indices) of the surrounding elements. However, its main responsibility is to make necessary adjustment in order to focus objects at various distances. The *lens* has an automatic, adjustable focusing ability through the *ciliary muscles*. This ability, commonly called *accommodation*, allows the eye to focus at objects at various distances from the eye. When the axial length of the eye does not fall within the range of *accommodation*, the eye is unable to focus on near objects. If eye's axial length is too long, the subject is unlikely to be able to focus on nearby objects, a condition called *myopia*. If eye's axial length is too short, the subject will be unable to focus on distant objects, a condition called *hyperopia*. With age, the lens can gradually lose its elasticity to be able to focus on

nearby objects, a condition known as *presbyopia*. All these conditions can be rectified by using corrective eyeglasses.

Beyond the *aqueous humor*, the light passes through the *pupil*, the eye's aperture. It is the circular opening in the opaque *iris*, a set of involuntary muscles controlling the amount of light entering the eye, and giving the eye its color. After the iris, the light passes through the *lens* and then through another transparent liquid called *vitreous humor*. Finally, after passing through *vitreous humor*, the light strikes the retina at the back of the eye's inner wall.

Since the cornea is not perfectly symmetric, the optical properties of the eye are not homogeneous in different directions. Thus, the light stimuli coming from different directions cannot all be focused with same accuracy, a condition called *astigmatism*. When this condition is significant enough to interfere with perception, corrective eyeglasses are needed. Like the directional inhomogeneity, the optical properties of the eye are not spectrally homogeneous either. Thus stimuli of different wavelengths do not get focused in the same way, an effect known as *chromatic aberration*. Chromatic aberration is not a unique characteristic of the eye, it happens in any lens in general.

There are many other sources of eye malfunctioning. These are beyond the scope of this chapter.

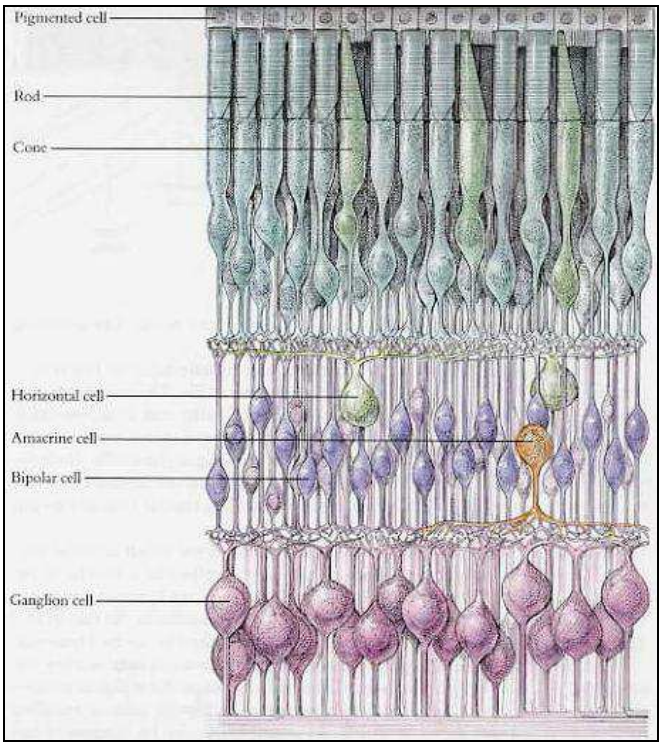


Fig. 2-2. A cross section of the retina, about midway between the fovea and far periphery, where rods are more numerous than cones. From top to bottom is about 0.25 mm. (illustration from [3])

2.1.2 *The retina*

The retina, whose cross-section is shown in Fig. 2-2, is part of the central nervous system that converts light (in the form of packets of energy, or photons) into neural signals. This conversion is carried out by two types of photoreceptor cells, *rods* and *cones*, residing at the back of the retina. The receptors' names reflect their shape. The *rods* are more numerous (120 million as opposed to 8 million cones), and are responsible for our vision under low-light level (scotopic condition), thus highly sensitive to light. *Cones* do not function under dim light, but are responsible for color vision and visual acuity under normal light level (photopic condition). Right at the center of the eye there is a small region of about 0.5 mm diameter called *fovea* (see Fig. 2-1). This region contains a high density of *cones* but virtually no *rods*. Conversely, as we move away from fovea, the density of *cones* decreases rapidly (see Fig. 2-3), although they are present throughout the retina. This contributes to the fact that we see fine details of objects that are at the center of the visual field, whereas objects seen through peripheral vision are relatively blurry. However, the brain structure also plays a role in it. The central area of the visual field gets greater representation than the periphery in the visual pathway from retina leading to the brain, and later in the visual cortex of the brain. Visual pathway and visual cortex are briefly discussed in the next subsection.

Because of the higher concentration of the rods in the periphery of the retina, and as they have a higher sensitivity than cones at low light levels, we see better with our peripheral vision in the dark than with our central or foveal vision.

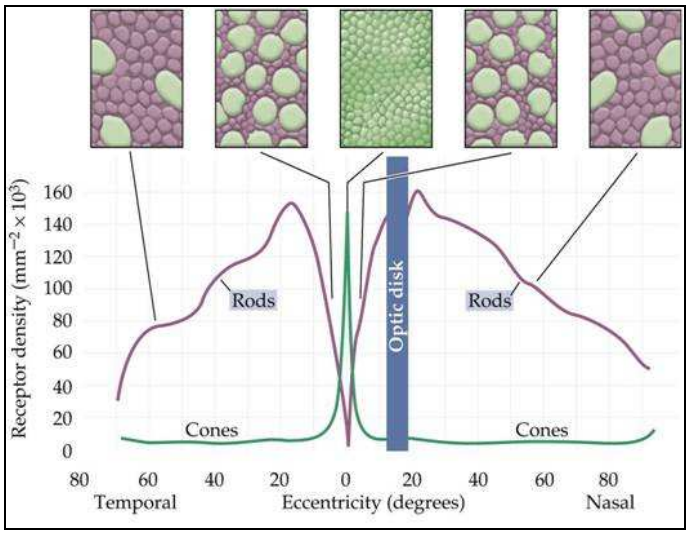


Fig. 2-3. Distribution of rods and cones on the human retina (http://www.rci.rutgers.edu/~uzwiak/NBSummer11/NBSummerLect4.html)

The middle layer of retina contains three types of nerve cells, called *bipolar cells*, *horizontal cells*, and *amacrine cells*, while the front layer contains the *retinal ganglion cells*. *Bipolar cells* receive inputs from the receptors, and pass onto the *retinal ganglion cells*. However, this transmission can also take place through an indirect path, through the involvement of other two types of cells. *Horizontal cells* connect receptors and *bipolar cells*, while *amacrine cells* link *bipolar cells* and *retinal ganglion cells*. Near the fovea, a single cone connects to a *bipolar cell*, which in turn connects to a single *retinal ganglion cell*. However, moving away from fovea toward the periphery, several receptors feed one *bipolar cell*, and many *bipolar cells* connect to a ganglion cells. This allows around 1 million *ganglion cells* in the retina to interface with nearly 128 million *rods* and *cones*.

Coming back to the photoreceptors, both rods and cones contain light-sensitive pigments. Rods have only one type of pigment (called rhodopsin), while the cones are of three types, with each type has a different pigment absorbing different wavelengths of light. The receptors respond to light through a process called *transduction*, in which a molecule of visual pigment absorbs a photon, and through a complex biochemical reaction results in change in electrical potential in the outer membrane of the photoreceptor. This leads to the release of a chemical transmitter that then affects the next nerve cell, or neuron. In this regard, it is relevant to describe another process called *pigment bleaching*, where a large amount of rhodopsin molecules is isomerized by too much light.

When many photons are absorbed within the same receptor, the response is not linear, but a logarithmic function of the number of photons absorbed [4]. This explains why our eye is relatively less sensitive to brightness change at high luminance level, compared to that at low luminance level.

On absorbing a photon, a pigment molecule cannot absorb additional photons. It can be restored to the prior unbleached state through the action of enzymes in the pigment epithelium behind the retina (see Fig. 2-2), containing a black pigment called *melanin* [3] (Chapter 3). This pigment layer also absorbs any photons that remain unabsorbed past the receptor layer. The retinal structure is such that the photoreceptors are located at the back of the retina, necessitating light to pass through other cell layers in the front and middle layers of retina before it can reach the receptors in the back. This oddity of the retinal structure originates from the organogenesis of the eye and brain. The retina which is part of the central nervous system sprouts from the embryonic brain, with the future photoreceptor cells in the front, which reach the ocular cavity of the eye and ultimately lands in the back of the eye. All the nerve cells, *horizontal*, *bipolar* and *amacrine cells*, in the front of the retina are transparent and do not interfere with the incoming light. Further, in the fovea where the visual acuity is the highest, these cell layers are displaced to the side to expose the cones [3], resulting in the fovea taking the shape of a shallow pit (Fig. 2-1).

The long, thin projections (or axons) of the ganglion cell bodies pass across the surface of retina, collect in a bundle at the *optic disc* (or optical nerve head) (Fig. 2-1) and leave the eye to form the *optic nerve*. The *optic disc* forms the *blind spot* located at 10 to 15 degree from the foveal direction on the nasal side (Fig. 2-3). It does not contain any receptor cells.

2.1.3 *Visual pathway and visual cortex*

The optic nerve, on coming out of the *optic disc*, forms what is known as *optic chiasm* (chiasm means crossing in Greek), shown in Fig. 2-4. This results in a cross-mapping of the visual field , left part of the visual field goes to the right half of the cortex, and vice versa. In each cerebral hemisphere, two pathways emerge from the *optic chiasm*. The smaller pathway ends in a visual center located outside the cerebral hemisphere called *superior colliculus*, and is thought to be responsible for eye movement. The other pathway goes through the lateral geniculate nucleus (LGN) to the occipital cortex, also known as primary visual cortex or V1, and situated on the occipital lobe (Fig. 2-5). Individual neurons (nerve cells) in the LGN can be activated by any change in brightness or color within the area of view (receptive field) of any one eye. Neurons in V1 transmit visual information to various distinct cortical regions located in the posterior temporal and parietal cortex. Almost half of the cortex is involved in visual function [2] (page 24).

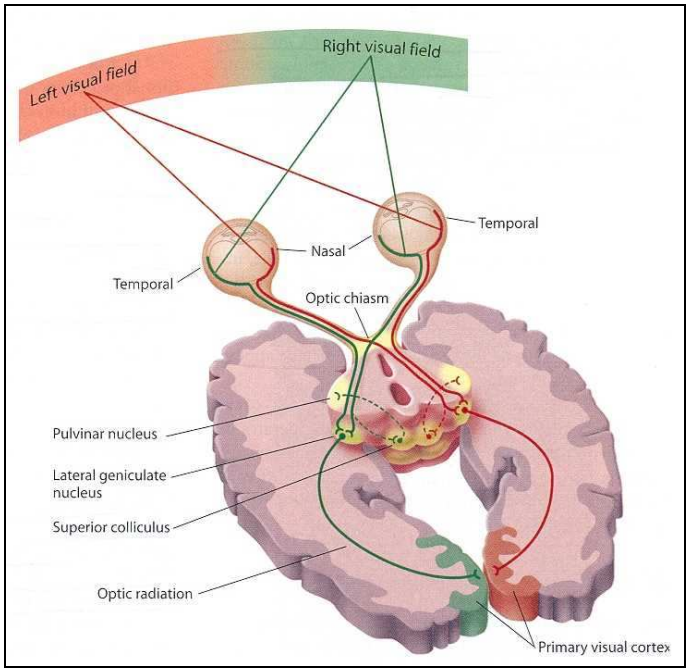


Fig. 2-4. Human visual pathway carrying sensation from the eye to the cerebral cortex (http://www.edoctoronline.com/medical-atlas.asp?c=4&id=21964)

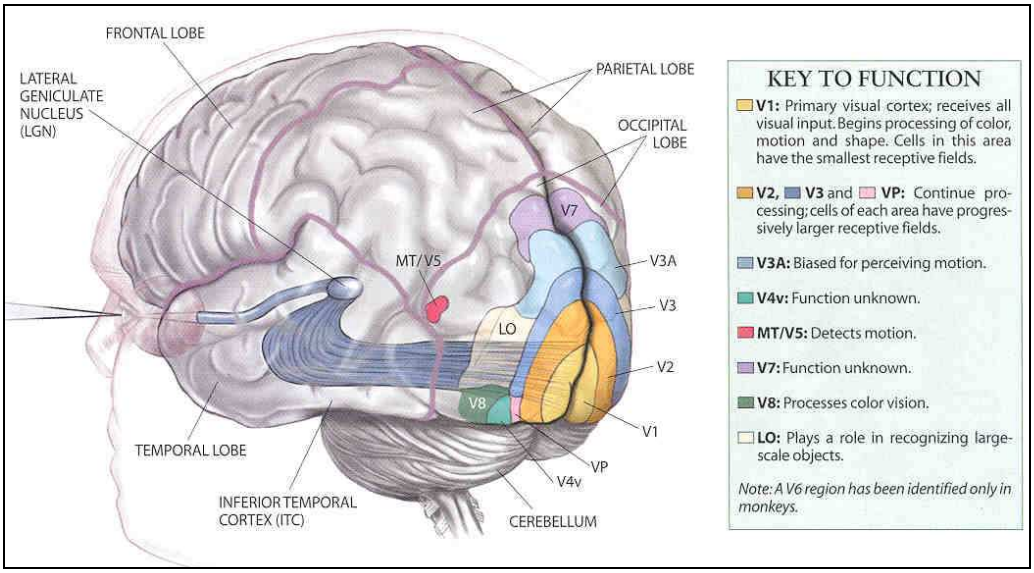


Fig. 2-5. Various regions of visual cortex responsible for vision (illustration from [5])

As the visual function of brain is not within the scope of this thesis work, this review does not delve any further into various anatomical, physiological or functional aspects of the visual system that extends beyond the eye and the retina. However, it is important to point out the higher order processes beyond the retina are equally important for visual functions. For a discussion on those aspects of visual function, and also for an in-depth discussion on the topics reviewed here, the reader is directed to more comprehensive references by Palmer [2], Wandell [6], Hubel [3] and chapters 2 [7] and 6 [8] of the book *The Science of Color*.

2.2 Perception of color

2.2.1 *Light as a physical quantity and its photometric counterpart*

Color is a result of complex interactions between physical light and our visual system. Different aspects of the visual system as it relates to color perception have been reviewed in the previous section. However, it is important to describe light as a physical quantity in order to better understand color perception.

Modern color science started its journey in the 17th century when the legendary English physicist Sir Isaac Newton conducted experiments with his glass prisms and incident sunlight, and concluded in his "New Theory of Colours" (1671): "*The Rays to speak properly are not coloured. In them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Colour…So Colours in the object are nothing but a Disposition to reflect this or that sort of Rays more copiously*

than the rest." [2] This explanation of the physical properties of light might seem obvious and rather simplistic given how much we know today, but this was the first fundamental insight of the role of light in interacting with objects to stimulate our color vision. Today, the dual nature of light as wave and particles is well established. The quantum nature of light is important to understand how a photon with a given energy has a probability to generate an electric signal in a cone, given the pigment absorptance. The wave nature of light is important for the understanding of color vision. In its most basic representation, a *photon* is a very small packet of vibrating electromagnetic energy characterized by its wavelength (the photon energy is $E = \frac{hc}{\lambda}$, where *h* is Plank's constant, *c* the speed of light in vacuum, λ the wavelength). Its unit is 1 nanometers (or nm in short), which is 10^{-9} meters. Sometimes wavenumber, which is reciprocal of wavelength ($\tilde{\nu} = 10^7/\lambda$, where wavenumber $\tilde{\nu}$ is in cm^{-1} and wavelength λ is in nm), is also used (typical unit cm^{-1}). Light from any source can be described in terms of the relative power emitted at each wavelength. Visible energy forms only a small part of electromagnetic spectrum, as shown in Fig. 2-6.

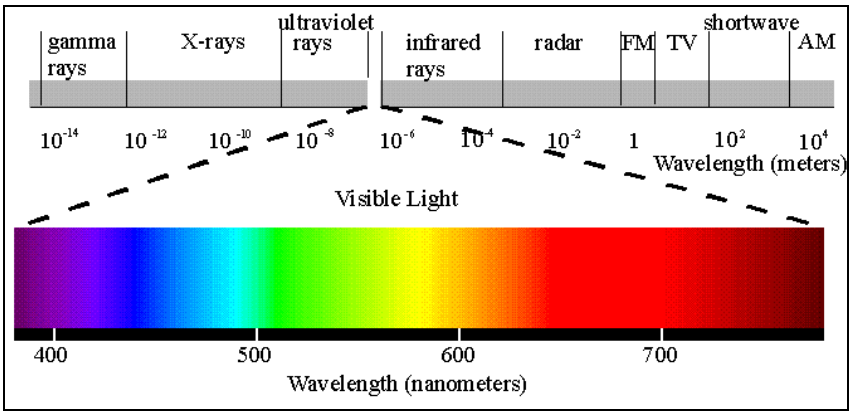


Fig. 2-6. Electromagnetic spectrum (http://www.yorku.ca/eye/spectru.htm)

A stimulus is an event that induces some response from our visual system. The light stimulus (whether visible or not) is such an event, and is quantified by radiometry. The most fundamental radiometric quantity is radiant energy, which is a measure of the total amount of light and expressed in joules. Radiant power in a particular location and propagating in a particular direction is called radiance, whose unit is watts per steradian per meter squared (watts = joules/second). Photometry on the other hand relates to the quantification of visible stimuli, taking into account the spectral sensitivity of the visual system. There are several references with a detailed discussion on radiometry and photometry [9] [7]. Here, a couple of photometric quantities need to be described since they have been used quite frequently in this thesis.

Luminance is the luminous flux (i.e. visible radiant power) in a beam emanating from a surface or falling on a surface in a given direction, expressed per unit of projected area of the surface as viewed from that direction, per unit solid angle [10] (page 63). The luminance values are obtained by using luminous efficiency functions, discussed later. *Illuminance* is the luminous flux incident per unit area [10] (page 63).

It is useful to express stimuli seen by the visual system in terms of a metric that takes into account the effect of eye's pupil. Retinal illuminance *Troland* is obtained by multiplying the luminance of a visual stimulus (in cd/m^2) by the area of the pupil in mm^2. Retinal illuminance can be photopic or scotopic, depending on which luminous efficiency function is used. When the luminance is below 0.001 cd/m^2 the condition is said to be scotopic, above 10 cd/m^2 the condition is considered as photopic, and in between the two, the condition is considered as mesopic.

2.2.2 *Color resulting from cones responses*

Each photoreceptor in our retina, rod or each of the three cones, contains a different kind of visual pigment. As explained in Section 2.1.2, the three types of cones are responsible for our trichromatic vision. The photoreceptors transduce arriving photons into the temporal and spatial patterns of electrical signals that eventually lead to color perception [11]. The pigments in the photoreceptors have different chemical compositions, and consequently vary in their relative ability to absorb light of different wavelengths. Thus, color is the consequence of unequal stimulation of the three types of cones. Having three types of cone receptors help us discriminate colored light from white light. The pigments in the three cone types have their peak absorptions at about 430, 530, and 560 nanometers, and are thus referred to as short-, medium- and long-wave sensitive cones respectively. The peak wavelengths thus fall in the violet, green and yellow-green parts of the spectrum respectively. The absorption curves of the cones, plotted in a logarithmic scale against the wavelengths and normalized to unity at the peak wavelength, are commonly referred to as the spectral sensitivity functions. As will be explained in Section 2.2.5, spectral sensitivity functions of the cones at the corneal plane are referred to as cone fundamentals.

2.2.3 *Color as a psychological phenomenon and its description*

Color is a psychological phenomenon that simply cannot be described without considering an observer. Color can be defined as a perception that depends on the response of the human visual system to light, or a physical stimulus resulting from the interaction of light with objects.

Thus color is essentially a subjective experience. Any color experienced by an observer with normal color vision can be expressed in terms of three dimensions. These dimensions form a three-

dimensional coordinate system defining a *color space*. A mode of representation close to the usual description of colors by observers uses *hue*, *saturation*, *lightness* as dimensions.

2.2.3.1 Hue

Hue is defined as the attribute of a visual perception based on which an area appears to be similar to one of the colors: red, yellow, green and blue, or to a combination of adjacent pairs of these colors considered in a close ring [10] (page 22). In the cylindrical color space, it corresponds to the angular direction around the central vertical axis, as shown in Fig. 2-7.

2.2.3.2 Lightness

Lightness is defined as the attribute by which a perceived color is judged as equivalent to one of the series of grays ranging from black to white [10] (page 22). Lightness, sometimes referred to as value, is the vertical axis in the color space (Fig. 2-7).

2.2.3.3 Saturation

Saturation can be defined as the chroma divided by lightness. Chroma is defined as the color attribute that indicates the degree of departure of the color from a gray of the same lightness [10] (page 22). In color space, saturation corresponds to the distance outward from the central axis to the point representing a given color (Fig. 2-7).

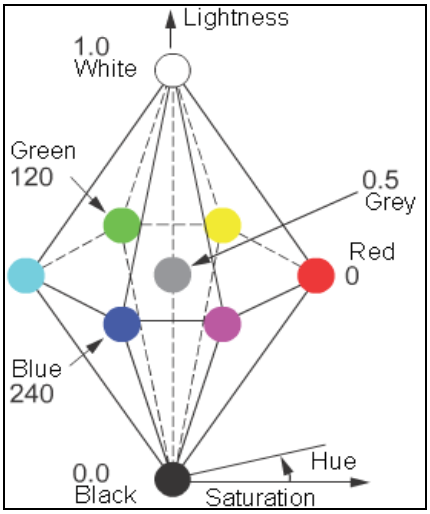


Fig. 2-7. Color space showing three dimensions, namely hue, saturation and lightness (http://www.ccs.neu.edu/course/cs4300/L5/L5.html)

The mean wavelength of the physical stimulus corresponds to hue. Here, the mean wavelength refers to the peak of the normal spectral distribution of the stimulus. Likewise, the area under the spectral function represents the lightness, and the variance is linked to the saturation of the stimulus. Here, the variance refers to the width of the normal distribution.

2.2.4 *Theories of color vision*

In the late 18th century, two major theories emerged that attempted to describe the complex process of color perception. According to Mollon [12], it was George Palmer [1740-1795] who first proposed in 1777 that there were "three physical kinds of light and three corresponding particles in the retina". This proposition took a more concrete shape when in 1802 Thomas Young suggested a link between the three primaries and sensory physiology. Following major contributions from Hermann von Helmhotz in 1852 and James Clerk Maxwell in 1855, the trichromatic theory was established. This theory, often called Young-Helmholtz trichromatic theory, says that there are three types of color receptors in the eye with overlapping functions, so any given wavelength can stimulate the three receptor systems to different degrees [2]. The trichromatic theory is able to explain why the color space is three-dimensional, how physically distinct combinations of wavelengths can lead to the same pattern of activation across the three receptor types. The latter is probably the most important fundamental property of the visual system: metamerism. The trichromatic theory also explains the basic forms of color blindness resulting from one receptor type missing, namely protanopia (long wavelength receptor missing), deuteranopia (medium wavelength receptor missing) and tritanopia (short wavelength receptor missing).

The trichromatic theory based on Young, Helmholtz and Maxwell's work was not universally accepted. It was observed that colors missing from the perception of color blind people always occurred in pairs, for example, red and green, or blue and yellow. Further, subjective experience of yellow seemed to suggest that it was more like a primary color, and not a mixture of red and green. The trichromatic theory was also unable to explain why a color does not appear to be simultaneously red and green, or simultaneously blue and yellow. Physiologist Ewald Hering proposed in 1878 three opponent mechanisms involving three receptor types, one of which responded oppositely to red and green colors, the other responded oppositely to blue and yellow colors, and the third responded oppositely to white and black. Hering thought each of these antagonistic pairs were associated with the dissimilation or assimilation of a "specific visual substance in the eye or visual system" [12].

Hering's opponent process theory could explain a lot of phenomenological facts mentioned before that the trichromatic theory could not. However, it was evident that both competing theories had some merits, and there was severe disagreement on adopting one over the other. However, reconciliation came through the proposition of dual process theory by Leo Hurvich and Dorothea Jameson in 1957

[13]. The theory stated that color processing happened in two stages, the first stage involving an embodiment of Young-Helmholtz's trichromatic theory, and the second stage employing a version of Hering's opponent process theory. Both stages of the dual process theory have been confirmed to occur in the retina.

2.2.5 *Cone spectral sensitivities and cone fundamentals*

The study of cone spectral sensitivities dwell in the realm of many allied fields, including psychophysics, biophysics, physiology, electrophysiology, anatomy, physics, and molecular genetics. Out of these, Psychophysics gives the most reliable spectral sensitivity data [11]. While psychophysical methods attempt to measure the sensitivity of the eye toward the entering light at the corneal level, other methods do the same measurement directly at the photoreceptor level. As we have seen in Section 2.1, light has to travel through the ocular media before reaching the photoreceptor. In the course of this travel, light gets absorbed by the lens and macular pigment at the fovea. This reduces the overall sensitivity of the eye with respect to the cones' absorption (see Section 2.5 for a review of various physiological factors influencing the cone spectral sensitivity). Thus, it is important to define cone spectral sensitivity in such a way that takes into account this light loss.

Cone fundamentals are defined as the spectral sensitivity functions of long-wave sensitive (LWS), medium-wave sensitive (MWS) and short-wave sensitive (SWS) cones, measured in the corneal plane [14]. According to the principle of univariance, a photoreceptor is essentially a sophisticated photon counter, the output of which varies according to the number of photons it absorbs, independent of their wavelengths [11]. Brindley proposed the quantal hypothesis [15], which states that a foveal color match is obtained when the quantal catch rate is equivalent for each of the three active photopigments, thus making such a match trichromatic and photopigment-limited. Any linear transformation of color-matching data obtained from a color-matching experiment (see Section 2.3) describes the color-matching properties of the eye. Thus, cone fundamentals can be obtained through a linear transformation of the color-matching functions, as shown in Fig. 2-8.

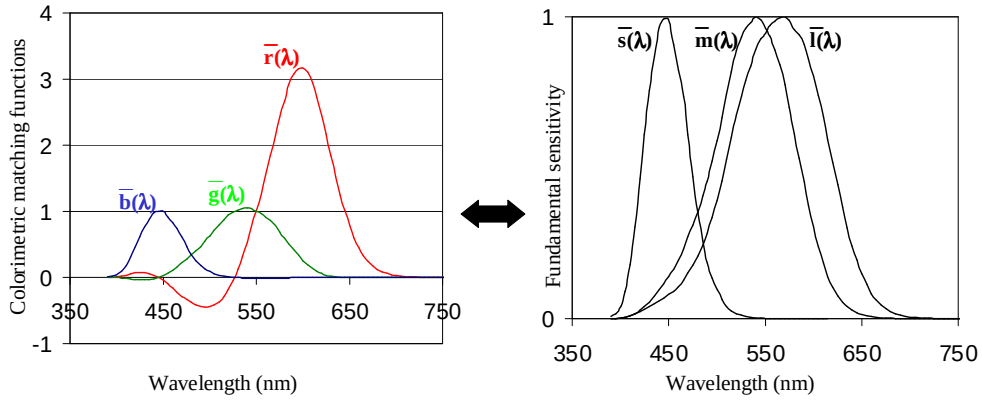


Fig. 2-8. A linear transformation exists between the color-matching functions and the cone fundamentals (Courtesy: Françoise Viénot)

Mean cone spectral sensitivity data are shown in Fig. 2-9. The data were collected from normal trichromats (people with normal color vision) as well as dichromats (people with color deficiency due to a missing receptor). This will be further elaborated after discussing some aspects of color deficiency in the next subsection. In Chapter 4, a detailed discussion on cone fundamentals and their derivation has been presented.

2.2.6 *Color deficiency*

Color deficiency can be congenital or acquired. Acquired color deficiency is outside the scope of this discussion. Congenital color deficiency represents a hereditary, permanent condition that is characterized by an abnormality of color matching and/or color discrimination ability. It is thought to be due to mutation, rearrangement and deletion of the opsin genes that determine the structure and function of the cone visual photopigments [16] (page 138). There are three major types of color deficiencies (dichromacy) resulting from a missing receptor type, namely protanopia (long wavelength receptor missing), deuteranopia (medium wavelength receptor missing) and tritanopia (short wavelength receptor missing). The first two are more common than the tritan defect, and mainly affect male population as they are X-chromosome linked defects. For example, in Europe, 8% of male and 0.4% of female population are affected by these deficiencies [16] (page 138). Apart from the most frequent dichromacy, another color deficiency is monochromacy, where two out of three cones are missing, or rod monochromatism (achromatopsia) due to the absence of all three cones. Both are extremely rare in the human population. Fig. 2-10 shows how the color spectrum is perceived by someone with normal color vision, and those with various color deficiencies.

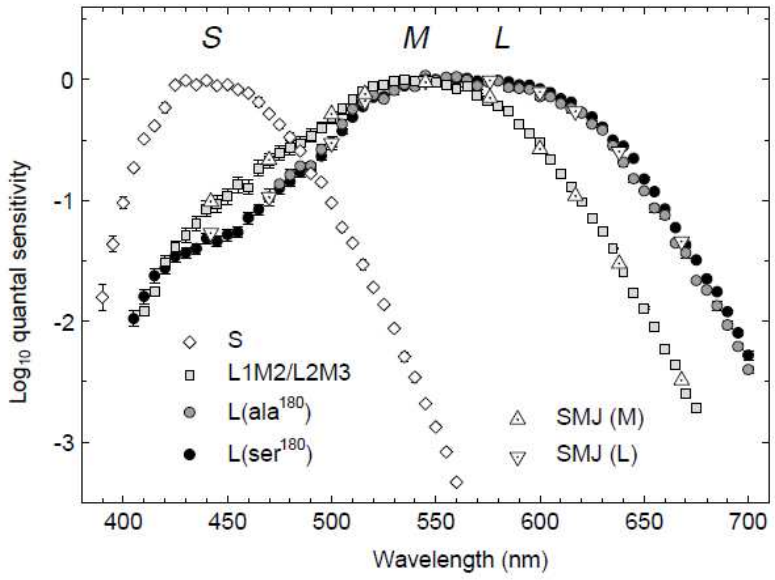


Fig. 2-9. Mean cone spectral sensitivity data obtained from different experiments. Mean long- (L-), medium- (M-) and short-wave (S-) sensitive functions were obtained from several deuteranopes, protanopes, and monochromats respectively, in addition to normal trichromats ([11], page 59)

Color defects have been studied since the 1800s [17]. The famous English chemist, John Dalton himself was a deuteranope and according to Mollon [12] (page 23), was the first to give an account of the phenomenon of dichromacy in 1794 . Even though his hypothesis of the presence of a blue-colored filter in the eye was later proved to be invalid, his name is forever associated with this topic through the term daltonism, which refers to color deficiency in many languages. König, a student of Helmholtz, hypothesized that the dichromatic forms of color defect represent reduced forms of normal trichromatic color vision. This hypothesis is critical in the field of color vision. Since the spectral sensitivities of the three cone types overlap extensively throughout the spectrum, measurement of a single cone type of a normal trichromat poses great challenge, and requires employment of special isolation procedure to measure a single cone type [18][11]. Assuming that two unaffected cone spectral sensitivities in a dichromat resemble those of a normal trichromat allows scientists to measure individual cone spectral sensitivities. Modern cone fundamentals are based on König hypothesis [16] (page 117) and are thus called König fundamentals [19].

Quantitative and qualitative anomalies in color perception can be measured using an instrument called an anomaloscope. This instrument, introduced in the early 20th century by W. A. Nagel (according to Mollon [12]), can be used for classification of color deficiency.

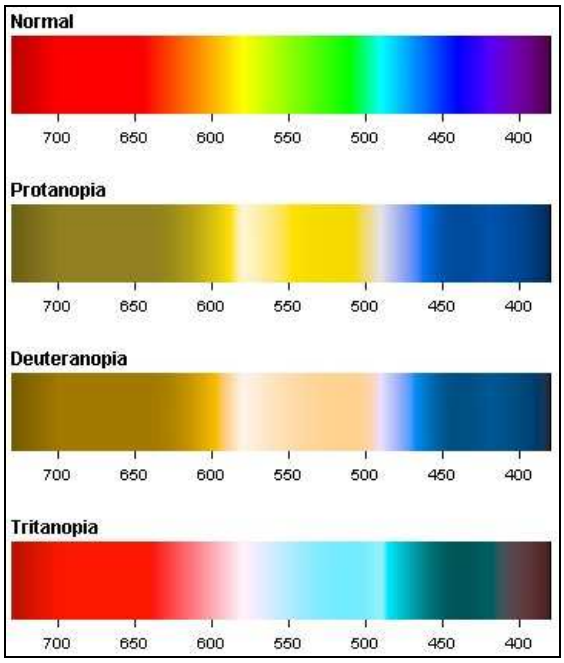


Fig. 2-10. The spectrum as perceived by individuals with normal color vision, protanopia, deuteranopia and tritanopia (http://www.internettg.org/mar99/accessibility_color_challenged.html, [20], [21])

With the recent advances in molecular genetics, it is now possible to select protanopes and deuteranopes with the appropriate M- or L-cone photopigment gene(s), for spectral sensitivity measurements [22]. Fig. 2-9 introduced in Section 2.2.5 summarizes results from various such

experiments. Long-, medium- and short-wave sensitive functions were obtained from a number of deuteranopes, protanopes, and monochromats respectively. In each case, cone spectral sensitivity data from normal trichromats were also collected by employing special cone isolation procedures [11]. The means of various experimental datasets are plotted in Fig. 2-9.

Some trichromats show less severe color vision deficits, but exhibit some similarities with protanopes and deuteranopes. These trichromats have all three cone types present, but exhibit an altered form of normal color vision. Such color vision is known as anomalous trichromacy, or more specifically protanomaly and deuteranomaly, signifying protan defects and deutan defects respectively. The defects can be simple or extreme, depending on the severity. A protanomalous trichromat is said to have a short-wave sensitive (SWS) cone photopigment and two medium-wave sensitive (MWS or MWS-like) cone photopigments, usually differing by a small shift in spectral peak [23]. On the other hand, the deuteranomalous trichromacy, the most common form of all congenital color deficiencies, is characterized by the presence of SWS cone photopigment and two LWS-type cone photopigments. Interestingly, while deuteranomalous trichromats do not have MWS-cone functionality (according to Neitz and Neitz [23], two-thirds of deuteranomalous men did not have MWS-cone functionality), they still have the genes responsible for the MWS cone photopigment. This is considered to be one of the most important unanswered questions with regard to the molecular genetics of color vision defects [23].

In the recent decades, we have come to know a great deal about the role of molecular genetics of the opsin genes in causing these color deficiencies. A comprehensive treatise on this topic is offered by Sharpe et al. [24] and Neitz and Neitz [23].

2.3 Colorimetry and visual color-matching

Colorimetry is the branch of color science that deals with numerical specification of the color of a physically defined visual stimulus [9] (page 117). It provides a system of color measurement and specification based upon the concept of equivalent-appearing stimuli. At the core of colorimetry is the concept of metamerism, whereby lights of dissimilar spectral characteristics appear identical to a given observer. Two stimuli that result in identical cone signals will match in color, irrespective of their spectral characteristics. Metamerism is discussed in detail in Chapter 3. In the current section, several fundamental laws constituting the principles of colorimetry will be reviewed. But before that, the notions of additive and subtractive mixing must be presented.

2.3.1 *Color mixture*

Color can be mixed in two different ways, additive mixing and subtractive mixing [25]. When two light stimuli are added together from different parts of the spectrum or of different spectral composition, it is called additive mixing. As a result, the radiant power of the output stimuli at any wavelength interval is equal to the sum of the powers of the constituent stimuli. Additive mixing occurs when displays, projectors or optical devices project beams of colored light on to the same area, and individual colors merge at the retinal receptor level to form a unified perception of color.

On the other hand, subtractive mixing occurs when dyes or pigments are mixed together, or when two or more color filters are placed in series. If a beam of white light is projected on to such pigments or filters, a part of the spectrum is absorbed (or subtracted) by each component dye or pigment which in turn determines the color of the reflected, diffused or transmitted light.

2.3.2 *Principles of colorimetry: Grassmann's laws*

A fundamental concept in colorimetry is *trichromatic generalization*, which follows from the trichromacy theory described in Section 2.2.4. *Trichromatic generalization* states that over a wide range of viewing conditions, several color stimuli can be matched completely by mixing three fixed primary stimuli whose powers have been appropriately adjusted. *Trichromatic generalization* leads to the following four linearity laws first proposed by Hermann Grassmann in his laws of additive color mixture [26]:

Symmetry: If $\boldsymbol{A} = \boldsymbol{B}$ then $\boldsymbol{B} = \boldsymbol{A}$

Transitivity: If $\boldsymbol{A} = \boldsymbol{B}$ and $\boldsymbol{B} = \boldsymbol{C}$ then $\boldsymbol{A} = \boldsymbol{C}$

Proportionality: If $\boldsymbol{A} = \boldsymbol{B}$ then $\boldsymbol{kA} = \boldsymbol{kB}$

Additivity: If $\boldsymbol{A} = \boldsymbol{B}$ and $\boldsymbol{C} = \boldsymbol{D}$ then $\boldsymbol{A} + \boldsymbol{C} = \boldsymbol{B} + \boldsymbol{D}$

If $\boldsymbol{A} = \boldsymbol{B}$ and $\boldsymbol{A} + \boldsymbol{C} = \boldsymbol{B} + \boldsymbol{D}$ then $\boldsymbol{C} = \boldsymbol{D}$

Where *A, B, C* and *D* are color stimuli.

These laws are some of the most fundamental principles in color science, and thus have been subjected to intense scrutiny for many decades. Under certain circumstances, these laws do not hold well [27] [28]. One instance of such failure has to do with the technique used in establishing a color match, and will be discussed in Section 2.3.3.2. Another instance is the condition under which the rod

photoreceptors in the retina actively contribute to the color perception. This will be described in Section 2.5.4.

2.3.3 *Color-matching experiments*

Color-matching experiments are of fundamental importance in colorimetry and color science, since from color-matching data we can obtain color-matching functions (CMFs) of individual observers, from which we can obtain the average CMFs that can be used in colorimetric computations. Colorimetric systems are described later in Section 2.4.

Different matching procedures can be followed while establishing a color match. These procedures differ in their objectives and implications. Some of the key procedures are described in the following subsections. A more detailed description can be found in Wyszecki and Stiles' color science book [9] (page 279).

2.3.3.1 Asymmetric and quasi-symmetric matching

Asymmetric matching refers to the situation where the two test stimuli being viewed are not the same in all respects, for example if they are not imaged on identical areas of the same retina, if their physical characteristics differ, if the conditioning stimuli are different compared to the test stimuli, or if the viewing of the two test stimuli are not independent [9] (page 281). Most color-matching viewing conditions are asymmetric. Determination of equivalence by strict substitution is the only *symmetric matching* condition. An asymmetric match could be *indirect*, where two test stimuli are judged independently based on their appearance quality like hue, chromaticness etc, or it could be *direct*, where the appearances of the two test stimuli are judged in the same observation. I*ndirect asymmetric match* is an extension of a *symmetric match*. Most color-matching experiments employ *direct asymmetric matching*, for example in a bipartite field. Note that in such a case, the stimuli are imaged in closely adjacent, but different areas in the retina.

In many cases however, one can assume that two stimuli matched asymmetrically, would also match in a symmetric matching procedure by strict substitution. Such a matching procedure is termed as *quasi-symmetric*. A carefully designed color-matching experiment with a bipartite field is likely to fall in this category.

2.3.3.2 Maxwell and maximum saturation techniques

There are two main methods typically used in color-matching experiments, the Maxwell method and the maximum saturation method. Fig. 2-11 explains the two methods. Both use a bipartite field, either horizontal or vertical, and a mixture of three different monochromatic primary stimuli ***R***, ***G*** and ***B*** in

the blue, green and red regions of the spectrum. In the Maxwell method, a mixture of monochromatic test stimulus L with variable wavelength λ is mixed with any two primaries (***R*** and ***G*** in the figure) to match the fixed reference white stimulus ***W***. Here, *R(λ)*, *G(λ)* and *L(λ)* represent the amount of primaries ***R*** and ***G***, and the test stimulus ***L*** needed to arrive at the match. On the other hand, in the maximum saturation method, one primary (***B*** in the figure) is desaturated with the test stimulus ***L*** in order to match a mixture of the remaining two fixed primary stimuli (***R*** and ***G*** in the figure). The amounts *R(λ)*, *G(λ)* and *B(λ)* of the three primary stimuli are called the tristimulus values of the test stimulus ***L***. The $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$ color-matching functions can then be obtained by Eq. (2-1) in case of maximum saturation method and by Eq. (2-2) in case of Maxwell method. In the latter equation, R_W, G_W and B_W are the radiant powers of the primary stimuli ***R***, ***G*** and ***B*** providing the fixed reference white stimulus ***W***.

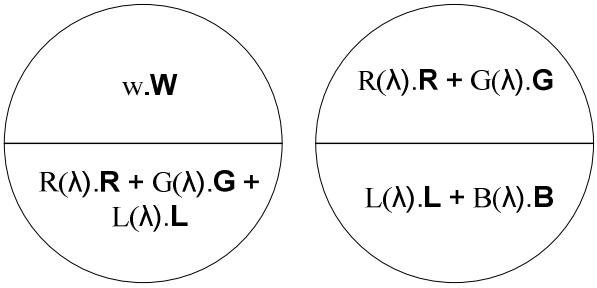


Fig. 2-11. Color-matching by the Maxwell method (left) and the maximum saturation method (right) ([9], page 384)

$$\bar{r}(\lambda) = \frac{R(\lambda)}{L(\lambda)}, \quad \bar{g}(\lambda) = \frac{G(\lambda)}{L(\lambda)}, \quad \bar{b}(\lambda) = \frac{B(\lambda)}{L(\lambda)} \tag{2-1}$$

$$\bar{r}(\lambda) = \frac{R_W - R(\lambda)}{L(\lambda)}, \quad \bar{g}(\lambda) = \frac{G_W - G(\lambda)}{L(\lambda)}, \quad \bar{b}(\lambda) = \frac{B_W - B(\lambda)}{L(\lambda)} \tag{2-2}$$

Out of the two methods, the maximum saturation method is more common. Both the CIE 2° and CIE 10° standard colorimetric observers are based on data collected from color-matching experiments employing the maximum saturation method. According to Grassmann's laws of additivity and proportionality [26], both methods should result in the same color-matching functions for a given observer. However, this is not always the case. It was first shown by Blottiau [29] that adding an equal amount of red desaturating stimulus to both halves of a matching bipartite field resulted in a mismatch particularly in the blue part of the spectrum. Blottiau's color-matching experiment involved

a modified version of Donaldson instrument ([9], page 478) employing the Maxwell method. Trezona later [30] [31] replicated Blottiau's experiment using his Wright colorimeter ([9], page 476) and confirmed the failure of additivity in the blue tristimulus values. However, the deviations were thought to be not significant in comparison to just discriminable color differences, leading to the conclusion that the failure was due to poor discrimination in the blue region.

Nonetheless, subsequent studies by Crawford [32], and followed by Lozano and Palmer [33] showed that failure of additivity in large field color-matching was indeed real. The spectrum loci obtained by the Maxwell method and the maximum saturation method deviated from each other in the blue-green region of the spectrum (Fig. 2-12). For a field size smaller than 10°, say 1° or 2°, the effect was somewhat reduced, but did not disappear. Thus rod intrusion alone could not explain this effect.

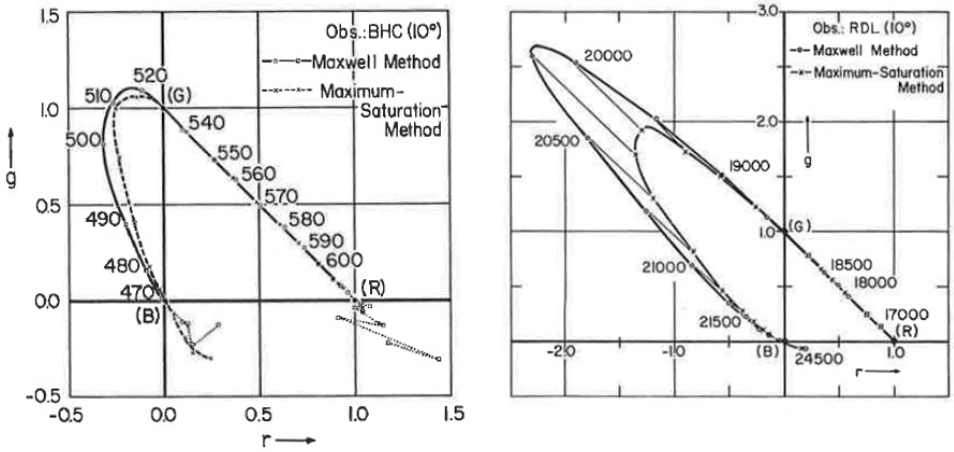


Fig. 2-12. Spectrum loci derived from color matches made in a 10° bipartite field by one individual observer using the Maxwell method and the maximum saturation method. Left figure is from Crawford's study [32] and the right figure is from Lozano and Palmer's study [33] (Reproduced from [9], page 385)

If the validity of the additivity and proportionality laws is in question, then Maxwell method should be preferred since in this case the matches are made in a field of constant luminance and chromaticity, ruling out the possibility of nonlinearity introduced by a change of these attributes. However, matches for the white reference stimulus in the Maxwell method have been reported [32] [33] as having higher uncertainty compared to a match derived through the maximum saturation method. This uncertainty is further amplified when the corresponding point on the spectrum locus is derived from the white match [9] (page 386).

Several attempts have been made to explain the discrepancies in the color-matching data obtained by the Maxwell vs. maximum saturation method [34]. More recently, the CIE Technical Committee TC 1-56 [28] has taken an in-depth look into this aspect, as part of its investigation into the problem of failure of Grassmann additivity. The committee made several observations. It noted that at low

luminance (~3 cd/m²) the Grassmann's additivity might not hold, potentially due to Mesopic color mechanisms. But even at high luminance (~300 cd/m²), a discrepancy of the spectrum locus derived from the Maxwell vs. maximum saturation color matches was observed. However, the issue of discrepancies between the results from the Maxwell vs. maximum saturation method remains unresolved.

2.3.4 *Chromaticity diagram*

The data from a color-matching experiment can be expressed in terms of vectors in a three-dimensional space, representing the tristimulus values (say, C_x, C_y, C_z). The three primaries form the axes in the three-dimensional space. The tristimulus values can be converted into quantities whose sum always equals unity, as shown in Eq. (2-3).

$$c_x = \frac{C_x}{C_x + C_y + C_z}$$
$$c_y = \frac{C_y}{C_x + C_y + C_z} \qquad (2\text{-}3)$$
$$c_z = \frac{C_z}{C_x + C_y + C_z}$$

The quantities (c_x, c_y, c_z) are called chromaticity coordinates. Since their sum is always unity, any two sufficiently describes a color in a two-dimensional space. The two-dimensional representation of color-matching data is called a chromaticity diagram. Fig. 2-13 is an example. The horseshoe shape is called the spectrum locus, representing the chromaticities of monochromatic stimuli at various wavelengths. The line joining the two ends of the spectrum locus is called the purple line, representing the locus of the chromaticities of additive mixtures of deep blue and deep red stimuli.

The curved line at the center is the locus of chromaticity coordinates of a blackbody radiator at various color temperatures. A blackbody radiator is any surface that emits radiant energy identical in all respects with that from a small aperture in a constant temperature energy absorbing enclosure. The correlated color temperature is defined as the temperature of an ideal blackbody radiator whose chromaticity most nearly resembles that of the light source.

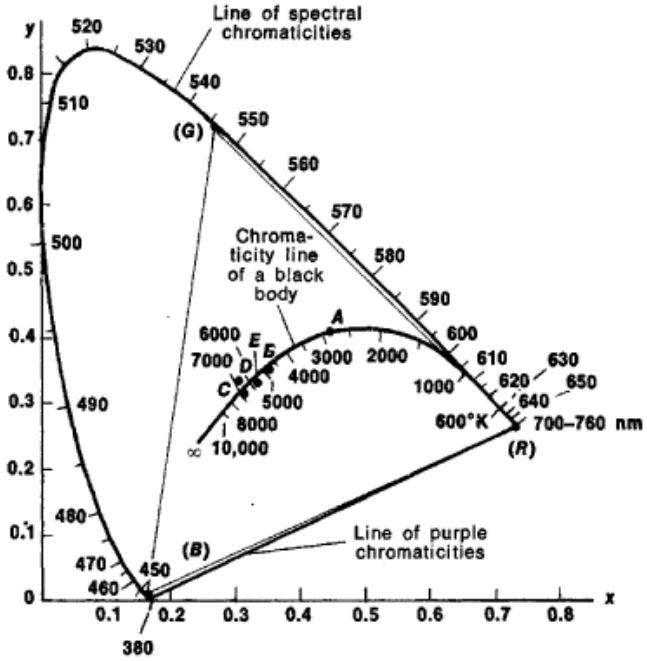


Fig. 2-13. An example of a chromaticity diagram
(http://encyclopedia2.thefreedictionary.com/Color+Measurement)

2.3.5 *Physiologically based Chromaticity diagram*

A physiologically based chromaticity diagram can be constructed in which the cone spectral sensitivities, i.e. the cone fundamentals form the rectangular axes. An advantage of such a chromaticity diagram is that it represents relative cone excitation. One such chromaticity diagram was proposed by MacLeod and Boynton [35], where the projective plane is an equiluminant chromaticity plane with coordinates (l_{MB}, s_{MB}) [see Eq. 2-4]. A basic assumption in forming the MacLeod-Boynton chromaticity diagram is that short-wavelength sensitive cone fundamental $\bar{s}(\lambda)$ does not contribute to luminance. In this diagram, as a consequence of this assumption, the abscissa $l_{MB} = L/(L+M)$ represents the equal and opposite change in LWS and MWS cone excitations, i.e. an increase in the LWS luminance is counterbalanced by an equal decrease in MWS luminance, but the sum is unity. The ordinate $s_{MB} = S/(L+M)$ denotes the level of short-wave sensitive (SWS) cone excitation at a constant retinal illuminance.

$$l_{MB} = \frac{L}{L+M}; s_{MB} = \frac{S}{L+M} \tag{2-4}$$

Here, L, M, and S represent tristimulus values obtained by integrating the respective cone fundamentals by the relative spectral power of the stimulus. When L, M, and S are obtained by integrating long-wave, medium-wave and short-wave sensitive cone fundamentals respectively with a monochromatic stimulus of unity length, coordinates (l_{MB}, s_{MB}) will be a function of wavelength, defining the spectrum locus.

Boynton and Kambe proposed a new unit for the cone excitation space called cone trolands [36]. Cone troland is obtained by multiplying cone excitations, expressed in terms of chromaticity coordinates (l_{MB}, s_{MB}), by retinal illuminance expressed in troland. Thus, the amount of L-cone trolands and M-cone trolands indicates the respective contribution of LWS and MWS cone excitations to the retinal illuminance. Since SWS cones do not contribute to luminance in this represntation, the scale for S-cone troland must be appropriately defined. In the representation proposed by Boynton and Kambe [36], one troland of the equal energy spectrum amounts to one S-cone troland.

Fig. 2-14 is the Macleod-Boynton chromaticity diagram obtained by using Smith and Pokorny 2° cone fundamentals [37], where the ordinate s_{MB} has been arbitrarily set at unity at its peak. The spectrum locus for the Smith and Pokorny 2° observer is shown in the diagram, along with the chromaticities of monochromatic stimuli at various wavelengths. To understand the meaning of the straight lines, we need to first describe *copunctal points*.

Based on König's hypothesis introduced in Section 2.2.6, a dichromat (an observer missing one of the three cones) needs only two primary colors to make any color match. Normalized dichromatic data, when plotted on the chromaticity diagram, result in straight lines called *confusion lines*. *Confusion lines* converge at a point in the chromaticity space called *copunctal points*. Recall that protanopes lack long-wave sensitive (L-) cones, deuteranopes lack medium-wave sensitive (M-) cones, and tritanopes lack short-wave sensitive (S-) cones. The copunctal points for each of these categories of dichromats represent the cone spectral sensitivities of the missing fundamentals. Thus protan, deutan and tritan copunctal points define the three cone-based physiological primaries L, M and S respectively [38].

In Fig. 2-14, coordinates (1, 0) and (0, 0) represent protan and deutan copunctal points. Protan and deutan confusion lines are represented by dashed arrows. Tritan confusion lines are represented by a set of parallel, vertical lines (not shown). The point EES represents the chromaticities of hypothetical *equal energy spectrum*, with unity spectral power at all wavelengths.

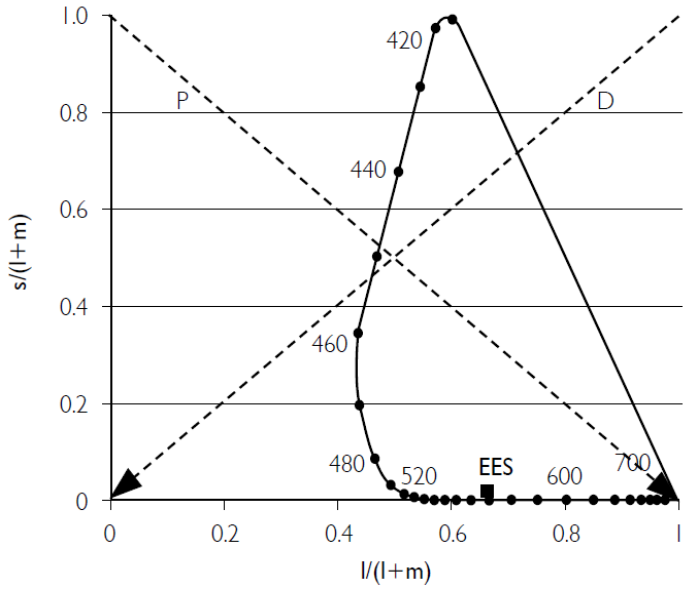


Fig. 2-14. Macleod Boynton chromaticity diagram (Reproduced from [16], page 119)

Derivation of Macleod-Boynton chromaticity coordinates for Stockman-Sharpe 10° cone fundamentals [22] will be described in Chapter 4.

2.4 CIE colorimetric system

At the heart of colorimetry is the concept of an ideal trichromatic observer, whose color-matching properties are expressed by three independent functions of wavelength. These are the color-matching functions (CMFs) of the ideal observer. Color matches made by the ideal observer always follow Grassmann's laws (see Section 2.3.2). The ideal observer is an average of a group of normal trichromats, and so this observer's CMFs are likely to differ from those of individual observers. The extent of the difference depends on individual observers, for some observers it can be negligible, for some others it can be rather significant. Thus, the ideal observer is essentially a mathematical construct.

For the color science community, it is important to universally agree upon an ideal observer, established by an internationally recognized scientific body. With this goal, Commission Internationale de l'Éclairage (CIE) was set up in 1913 [9] (page 131). The CIE specifications of standard observers are discussed next.

2.4.1 *CIE standard colorimetric observers*

In 1931, the CIE defined a standard observer for colorimetry. In doing so, it decided to combine the photometric and colorimetric properties of the standard colorimetric observer into one set of function. Accordingly, 2° bipartite color-matching data from Wright's [39] and Guild's [40] studies were used, along with the luminous efficiency function defined by CIE in 1924 [41] based on the works of Coblentz and Emerson [42] and Gibson and Tyndall [43]. Wright measured CMFs of ten observers using monochromatic primary lights of 650 nm, 530 nm and 460 nm wavelengths [39]. Guild on the other hand measured CMFs of seven observers using broadband lights as primaries. The units of the primary stimuli were based on equal-energy white, a specific white stimulus of 4800 K color temperature (designated as NPL white), whose three chromaticity coordinates were equal to each other [40]. Additionally, for the ease of computations, the tristimulus values were converted so that all values were positive. The *CIE 1931 Standard Colorimetric Observer*, shown in Fig. 2-15, is the main observer model on which much of colorimetry is based. Even though it is recommended only for small fields of 1° - 4° field-of-view, this restriction is not always followed very strictly in the industry.

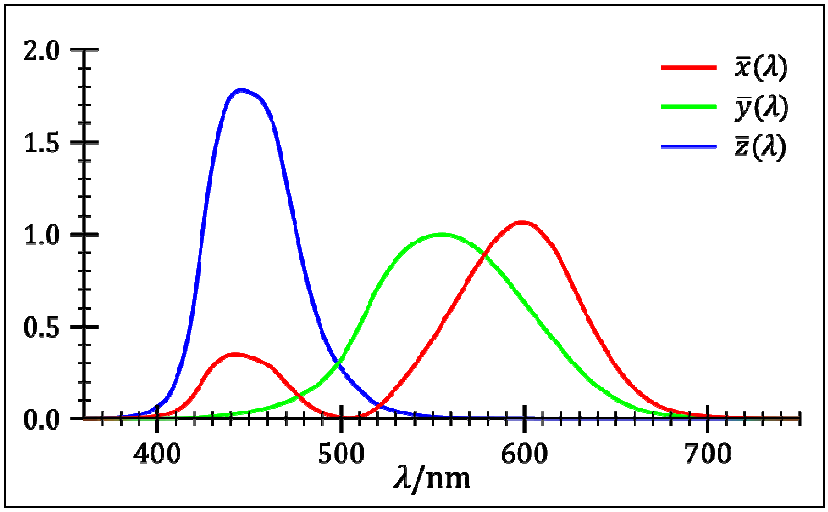


Fig. 2-15. Color-matching functions of the CIE 1931 2° standard colorimetric observer

Fig. 2-13 presented earlier is actually the CIE 1931 (x,y) chromaticity diagram. The points A, E, C etc on the blackbody locus represent chromaticity coordinates corresponding to various CIE standard illuminant (A, E, C etc) and CIE 1931 standard colorimetric observer.

In 1964, the CIE recommended an alternative set of standard CMFs $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ as a supplement to the 1931 standard observer for applications involving large-field visual color-matching. These functions were based on the 10° color-matching experiments of Stiles and Burch [44] and Speranskaya [45] and are referred to as *CIE 1964 Supplementary Standard Colorimetric Observer*. Stiles and Burch used a trichromatic colorimeter (described later in Chapter 3) with monochromatic

primary stimuli at wavelengths 645.2 nm, 526.3 nm and 444.4 nm, and measured CMFs of 49 observers. A different set of primary stimuli was also employed for certain part of the spectrum, but the final results were all transformed to the primaries mentioned above. To reduce rod intrusion (see Section 2.5.4), the luminance of the matching field was kept high. A minor mathematical correction was also applied to reduce rod intrusion.

On the other hand, Speranskaya [45] used 27 observers for her experiment with broadband primaries. The central 2° of the field was masked off to avoid the maxwell spot (a central nonuniformity in the field-of-view due to the contributions of macular pigment, which is highly concentrated in the fovea). The luminance of the visual field was 30-40 times lower than that of Stiles and Burch study, thus the results were significantly affected by the rod intrusion, particularly in the longer wavelengths of the $\bar{b}(\lambda)$ function. Although Speranskaya used 640 nm, 545 nm and 465 nm as primary wavelengths, the data were later transformed to the same system of primaries as used by Stiles and Burch.

Judd [46] [47] averaged the two sets of data after correcting for rod intrusion in Speranskaya's data. A ratio of 3:1 weighting was assigned to the two sets, with more weight assigned to the Stiles and Burch data. The ratio was also changed toward the end of the spectrum since Stiles and Burch dataset had greater spectral range. Smoothing and extrapolation were also used to arrive at the final all-positive average CMF. Thus, the CIE 1964 standard colorimetric observer does not come directly from the original Stiles and Burch data, but after significant amount of mathematical processing.

Note that the precision of large-field color-matching is generally more than that of small-field color matching. For example, for 10° field-of-view, color matching is expected to be two or three times more precise than the 2° field-of-view [9] (page 132). The precision of color matching is indicated by reduced intra-observer variability.

2.4.2 *CIE XYZ tristimulus values*

The most commonly used mathematical way of describing color is through the CIE tristimulus values, X, Y and Z, using the CIE 1931 colorimetric system. To compute these values, contributions of relative spectral power of a CIE standard light source [S(λ)], the spectral reflectance of the viewed object [R(λ)], and the CIE 1931 standard colorimetric observer [$\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$] are multiplied at each wavelength, product weighted by the difference between two subsequent wavelengths, and then summed over all wavelengths (λ). The computation is shown in Eq. (2-5), where *k* is a normalization factor. Here, *Q(λ)* represents spectral power distribution of the light reflected from the object.

$$
\begin{aligned}
Q(\lambda) &= S(\lambda)R(\lambda) \\
X &= k\sum_{\lambda} Q(\lambda)\bar{x}(\lambda)\Delta\lambda \\
Y &= k\sum_{\lambda} Q(\lambda)\bar{y}(\lambda)\Delta\lambda \\
Z &= k\sum_{\lambda} Q(\lambda)\bar{z}(\lambda)\Delta\lambda
\end{aligned}
\qquad (2\text{-}5)
$$

If we replace [$\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$] in Eq. (2-5) by [$\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$], we obtain X_{10}, Y_{10}, and Z_{10}, the tristimulus values for CIE 1964 colorimetric system.

The first line in Eq. (2-5) is used for object-colors stimuli, when the spectral reflectance of the object is known, from which we can compute the spectral power distribution of the light reflected from the object. For self-luminous stimuli however, spectral power distribution of the stimulus [$Q(\lambda)$] is known, so the first line is skipped.

The factor k is generally defined as in Eq. (2-6). The expression in the denominator is computed as explained before. This assigns the tristimulus value Y of white stimulus an arbitrary value of 100. In case of object-color stimulus, this white has a spectral reflectance of unity at all wavelength [$R(\lambda) = 1$], and is called a perfect reflecting diffuser. In case of CIE 1931 colorimetric system, the Y tristimulus value represents the luminance factor in terms of cd/m^2.

$$
k = \frac{100}{\sum_{\lambda} S(\lambda)\bar{y}(\lambda)\Delta\lambda} \qquad (2\text{-}6)
$$

2.4.3 *CIELAB color space*

CIELAB is one of the most common color space used in color applications. The CIELAB coordinates can be obtained from CIEXYZ values of an object-color stimulus (either CIE 1931 or CIE 1964 colorimetric system) by using Eq. (2-7).

$$
\begin{aligned}
L^* &= 116\, f\left(\frac{Y}{Y_n}\right) - 16 \\
a^* &= 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\
b^* &= 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right]
\end{aligned}
\qquad (2\text{-}7)
$$

Where X, Y and Z are the CIE XYZ tristimulus values, X_n, Y_n and Z_n are the tristimulus values of the reference white, and the function $f(\omega)$, where ω is (X/X_n), (Y/Y_n) or (Z/Z_n), is given by Eq. (2-8).

$$f(\omega) = \begin{cases} \omega^{1/3} & \omega > (6/29)^3 \\ 7.787(\omega) + 16/116 & \omega \le (6/29)^3 \end{cases} \tag{2-8}$$

The perceptual correlates of chroma and hue are given by Eq. (2-9).

$$\begin{aligned} C^*_{ab} &= \sqrt{a^{*2} + b^{*2}} \\ h^*_{ab} &= \tan^{-1}\left(\frac{b^*}{a^*}\right) \end{aligned} \tag{2-9}$$

However, the CIELAB color space is not quite perceptually uniform, particularly in the blue region of the color space. As a result, the Euclidean distance in CIELAB space between two colors does not always correspond to the perceived color difference. This non-uniformity issue was addressed by the CIE by establishing an advanced color difference equation, as described below.

2.4.4 *CIEDE2000 advanced color difference formula*

In 2000, CIE proposed an advanced color difference formula [48] (henceforth CIEDE2000). The aim was to improve the correlation between computed and perceived color differences in industrial applications compared to what was provided by the erstwhile color difference formula of 1994 [49] (henceforth CIEDE94). Like CIEDE94, CIEDE2000 incorporates specific corrections for non-uniformity of CIELAB space, namely the weighting functions S_L, S_C, S_H for lightness, chroma, and hue respectively. Three parametric factors k_L, k_C, k_H account for the influence of illuminating and viewing conditions in color-difference evaluation. For these formulas, parametric factors are set as 1.0 for a given set of reference conditions. The CIEDE2000 color difference formula is given by Eq. (2-10).

$$\Delta E_{00} = \left[\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right) \times \left(\frac{\Delta H'}{k_H S_H}\right) \right]^{0.5} \tag{2-10}$$

CIEDE2000 includes a rotation term [the last term in Eq. (2-10)] that accounts for the interaction between chroma and hue differences in the blue region. It also alters the (a^*) axis of CIELAB, which mainly affects colors with low chroma (neutral colors). The primes in the color difference terms ($\Delta L'$), ($\Delta C'$), and ($\Delta H'$) denote corrections for neutral colors in lightness, chroma, and hue differences respectively.

CIEDE2000 (also indicated by symbol ΔE_{00}) has been used in this thesis as the color difference metric whenever appropriate. Note however that this metric is only valid for 2° or 10° standard colorimetric observer.

2.5 Sources of individual differences in color-matching

In Section 2.4.1, the standard colorimetric observers were introduced. However, the color-matching properties of individual observers differ from those of a standard or an average observer. The extent of this variation depends on the individual observers. This section presents a brief review of some of the most important physiological factors responsible for the individual deviations in color-matching. Further discussions on pre-retinal filters, photopigment optical density and photopigment absorption spectra can be found in Chapter 4. A more complete discussion on different sources of individual variability can be found in [9] (page 347).

2.5.1 *Pre-receptoral filters in the eye*

Two major sources of variations in color-matching are due to lens and other ocular media optical density and macular pigment optical density.

2.5.1.1 Lens optical density

Lens absorption constitutes almost all of total ocular media absorption. The lens optical density varies significantly from one individual to the other, and also increases substantially with age [50]. Van Norren and Vos [51] showed that individual ocular media absorption of Crawford's [52] 50 observers aged between 17 and 30 years varied from the average results by about 25% at the short wavelengths. Adult lens transmission is thought to have two components, one being age-dependent, and the other being age-independent. After age 30, lens transmission reduces at all wavelengths because of an increase in the internal scattering of light. Additionally, there is an increase in the pigment density that causes strong absorption to take place at short wavelengths, as well as an increase in lens thickness [7]. Mathematical model of lens optical density will be presented in Chapter 4.

2.5.1.2 macular pigment optical density

The macular region in the human retina (see Section 2.1.2) contains a photo-insensitive pigment that selectively filters light arriving at the base of the photoreceptors, absorbing most strongly from 400 to 550 nm with a peak near 458 nm. The macular pigment optical density has a high degree of individual variability, with most of the variability occurring between 400 and 525 nm and peak optical density varying from 0 to over 1.2 log units [7]. The optical density of the macular pigment is highest at the center of the fovea, and decreases exponentially with retinal eccentricity [16]. As a result, macular

pigment can cause the color-matching data to vary depending on the retinal position used in establishing the color-match [53] [54] [55]. As an example, in case of color-matching in a 30° bipartite field, the observer match was found to be a balance between a peripheral and a central match, with a slight bias toward peripheral assessment [56].

High density of macular pigment in the central 1°-2° foveal region gives rise to a well-documented perceptual effect called the *Maxwell spot* [9] (page 133). For certain color stimuli in large-field viewing (say 10° or larger), a color inhomogeneity may appear at the area of fixation in the form of an ill-defined ellipse with major axis horizontal and spanning 1° or 2° [16] (page 116). It has been found that the exclusion of the affected central region does not have a significant impact on color matching [56]. In color-matching experiments, the observers are often instructed to ignore the Maxwell spot.

As we conclude the discussion on the effect pre-receptoral filters on color-matching, it is worthwhile to mention the seminal works of Moreland [57] [58], who showed that specific pairs of wavelengths allow one to obtain matches that are robust to the variation of lens and macular pigment. These wavelengths were obtained through optimization of blue and green primaries, and were intended to be used in tritanomaloscopy. Following a thorough study on tritan matches, Moreland developed a new type of anomaloscope employing Moreland equations. Moreland anomaloscope is used quite frequently in the field of color vision.

2.5.2 *Photopigment optical density*

Once light (in the form of photons) reaches the photoreceptor layer at the back of the retina (see Section 2.1.2), it must be absorbed in the photopigments in order to enable the visual perception. The concentration of the pigments and the length of the photopigment-filled outer segments of photoreceptors affect the absorption spectrum. These two factors determine the effective optical density of the photopigment. Due to the longer photoreceptors in the center of the fovea, the effective optical density of photopigments is the highest at the foveal center and decreased exponentially with retinal eccentricity independent of age and cone type. As a result, the effective optical densities of photopigments decrease as the field sizes increase [16]. While some psychophysical studies suggested [59] [60] that LWS-cone photopigment optical density was higher than that of MWS-cones, some other researchers [61] concluded that the LWS- and MWS-cone photopigment optical density did not differ across the population.

The effective optical density of photopigments also varies among individuals [62]. Differences in optical density of the LWS and the MWS photopigments can account for variability in chromaticity coordinates normalized in the manner described by W. D. Wright [63]. This method, generally referred to as WDW normalization [9] (page 134) in the literature, discounts individual variability due

to prereceptoral filtering (due to ocular media or macular pigment optical densities), leaving the variability that is only due to the photoreceptor system, However, this is true only for monochromatic stimuli.

High axial photopigment optical density results in the flattening of the spectral absorptance of photopigments. This happens since the absorption is wavelength dependent, with absorption at the peak being maximum. According to Rodieck [4], maximum absorption efficiency is around 2/3, out of three photons that reach the molecules of rhodopsine, two trigger an isomerization. When axial photopigment optical density increases, absorptions at the longer and shorter wavelengths increase while at the peak wavelength it is still at its maximum, resulting in a flatter absorption function. This in turn broadens the underlying spectral sensitivities of the photoreceptors. The dependence of the photopigment spectrum on optical density is known as *self-screening* [7]. Once a pigment molecule is bleached, it can no longer absorb photon. Light level can substantially affect the concentration of unbleached pigment molecules. Thus, as the light level increases, for example when the eye goes from the dark-adapted state to the light-adapted state, the spectral absorptance functions of cones narrow. This change in the shape of the absorptance spectrum is reflected in the color matching functions.

2.5.3 *Variability in the photopigment peak wavelength (λ_{max}) due to genetic polymorphism*

In recent years, significant progress has been made in understanding molecular biology responsible for human color vision, including the identification of the genes that encode the LWS- and the MWS-cone photopigments. These photopigments can show polymorphism in the amino-acid sequences of their opsin genes [64] [65] [24]. Such polymorphisms can affect the λ_{max} of the photopigment spectra. The most common polymorphism is a single amino-acid substitution (Alanine for Serine or vice versa) at position 180 of the LWS-photopigment opsin genes, resulting in a peak wavelength shift of up to 4 nm [66] [67]. As an example, Sharpe et al. [68] estimated the difference in photopigment λ_{max} from the mean L(ser^{180}) and L(ala^{180}) spectral sensitivities as around 2.7 nm. The data are plotted in Fig. 2-9 as gray and black circles.

Other than serine-alanine polymorphism, the largest shifts in λ_{max} are produced by substituting alanine for threonine at position 285 (up to 14 nm) and phenylalanine for tyrosine at position 277 (up to 7 nm) [24] (page 9). An LWS photopigment has a tyrosine in position 277 and a Threonine in position 285. On the other hand, an MWS pigment has a phenylalanine in position 277 and an Alanine in position 285. These two substitutions contribute the most to the large shift between the LWS and the MWS photopigments. Moreover, the substitution of an alanine for a serine in position 180 in LWS and/or MWS pigments may exaggerate the difference between their peak wavelengths, as a serine makes the peak wavelength shift a little toward longer wavelengths in either pigment [69].

2.5.4 *Rod participation*

A color mixture is treated as a linear system in colorimetry, following several key properties including additivity and proportionality (Grassmann's laws). This is particularly true for small (e.g. 2°) foveal fields. According to Brindley's quantal hypothesis [15], a foveal color match is obtained when the quantal catch rate is equivalent for each of the three active photopigments, thus making such a match trichromatic and photopigment-limited. However, for larger matching field or in case of parafoveal viewing (where the field is imaged outside the fovea in the eye), a fourth photoreceptor, the rod, becomes active under certain viewing conditions. As a result, color-matching may not always follow Grassmann's laws [26] (see Section 2.3.2), even though they remain trichromatic [70]. Nevertheless, for a pair of matching stimuli in a bipartite field, if the spectral radiant power in both half-fields were reduced or increased by the same amount independent of the wavelength, the match would still be valid [9] (page 356), provided the match is photopic.

In large-field color-matching, a hypothetical match where the rod receptors are somehow suppressed from responding to the stimuli is called rod-suppressed match. CIE 1964 supplementary standard colorimetric observer, described in Section 2.4.1, attempts to define the matching properties of a rod-suppressed retina by mathematically correcting for rod intrusion [9] (page 357).

Several experimental studies have attempted to account for rod intrusion by balancing for rod responses in the matching fields in a large bipartite field, so that color matches remain stable at various stimulus levels. The color-matching process thus becomes tetrachromatic [71] [72] [73]. Wyszecki and Stiles' reference contains a detailed discussion on tetrachromatic color matching [9] (page 366).

2.6 Conclusions

In this chapter, various fundamental concepts and understanding of color science and color vision were reviewed. Starting with a discussion on the key components of the human visual system, we reviewed the basic theories and established knowledge on the perception of color as we know today, followed by a discussion on the principles of colorimetry and visual color matching. Then various aspects of the CIE colorimetric system were reviewed. Finally, some of the most common sources of individual differences in color matching were enumerated. The goal of this chapter was not a complete and comprehensive discussion on all these topics, but to provide a concise review of them, with appropriate references for further reading.

Accordingly, this chapter sets the foundation for more in-depth discussions on the issue of observer variability in color science and color applications, which will be presented in the subsequent chapters.

The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" ("I found it!"), but rather "Hmmm... that's funny..." ~ Isaac Asimov

3. Observer Metamerism and Individual Observer Variability in Color-Matching: A Review

3.1 Introduction

When two color stimuli produce the same visual response, a visual match is obtained. Two stimuli with very different spectral power distribution can give rise to identical cone response, leading to a color match. However, such a match established by one observer can, and quite often does lead to a mismatch for a different observer, as the second observer has a different set of color-matching functions (CMFs) than the former. This phenomenon is commonly termed as observer metamerism.

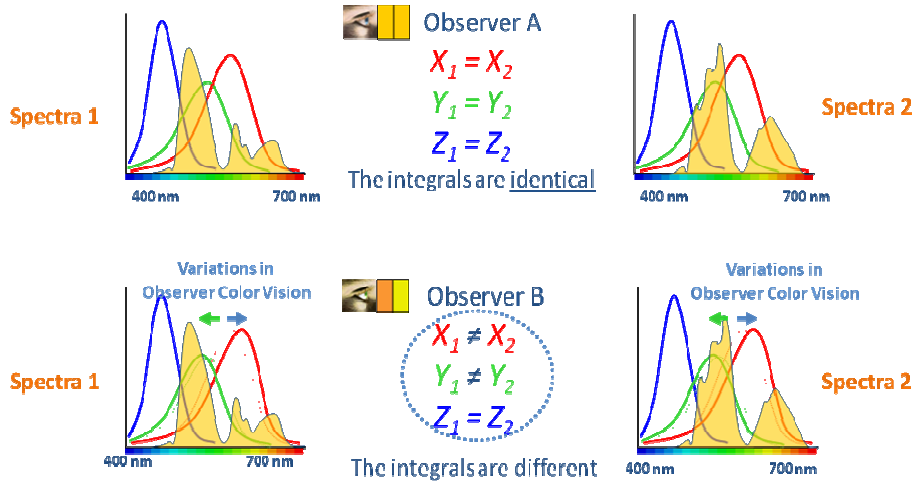


Fig. 3-16. Illuminant (top) and observer (bottom) metamerism (Courtesy: Laurent Blondé)

The origin of the metamerism lies in the trichromacy of the visual system. A metameric color match between two stimuli, either objects or illuminations, is conditional. If the stimuli do not match upon a change in illumination, the pair is said to exhibit illuminant metamerism (Fig. 3-16 top), which results from a change in the spectral power distribution of the illuminant. If changing observer causes a mismatch, the pair is said to exhibit observer metamerism (Fig. 3-16 bottom), since the mismatch is caused by a change in the CMFs of the observer. If the spectral characteristics of the primary

colorants of two color reproduction devices are not the same, any color match made on these devices is metameric in nature, and thus may not hold when one observer is replaced by another.

While in literature some researchers have used the terms observer metamerism and observer variability interchangeably, there is a subtle difference between them. 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 (in this context, the CMFs) among individual observers. We can think of observer variability as the cause, and observer metamerism as the effect.

In the next section, various studies aimed at understanding, quantifying and modeling observer metamerism are reviewed. In the two sections that follow, several color-matching experiments leading to a better understanding of the individual observer variability are reviewed. Of these, Section 3.3 deals with classical color-matching experiments and Section 3.4 deals with applied color-matching experiments. Here, classical color-matching experiments refer to those that involve monochromatic stimuli generated by a monochromator or a similar instrument, while applied color-matching experiments refer to those that involve displays or similar devices employing either broadband or narrow-band primaries.

3.2 Quantifying observer metamerism

The practical consequence of individual variability in CMFs is observer metamerism. It poses a significant challenge in many industrial applications, since its effect is that a satisfactory color reproduction across various devices and media is often not guaranteed for all consumers and clients. Therefore, from practical applications' point of view, it is of interest to somehow model and quantify observer metamerism. Following subsections summarize various attempts toward achieving this goal.

3.2.1 *Color Rule as a metric of observer metamerism*

The D&H Color Rule is a device in which two series of paint patches with different spectral characteristics slide against each other. It was originally produced by Davidson & Hemmendinger, subsequently by the Munsell Color Co., but is no longer available [74]. In this Color Rule, the observer can view two patches, one from each series, side by side in a rectangular window. The patches were selected in such a way that for a given combination of illuminant and an observer, two patches from the two series would make a metameric match. Either with change in illuminant or with change in observer the matching pair changed. The observer's task was thus to slide the rules to find the right matching pair under a given illumination.

In one of the early attempts to quantify variations in observer color vision, Kaiser and Hemmendinger [75] analyzed D&H Color Rule data from various past studies and argued that normal trichromats made responses with the Color Rule that were dependent on age. They found that the yellowing of the lens in the human eye resulted in similar change in responses on the Color Rule as by an illuminant change of 50–75 reciprocal mega Kelvins (MK^{-1}). Thus, Kaiser and Hemmendinger found a strong correlation between age and lens density in the human eye.

3.2.2 *Spectral characteristics as metrics of metamerism*

There have also been attempts to quantify the extent of observer metamerism possible for a given set of two stimuli. A widely popular theory is that two metameric stimuli with identical tristimulus values for a standard colorimetric observer require at least three crossovers of the stimulus functions at different wavelengths in the spectral domain [76][77][78][79]. Berns and Kuehni [80] argued that these crossover locations depend exclusively on the spectral properties of the metameric stimuli, and that: “any relationship between crossover wavelengths and properties of the visual system such as maximal responsivities appears coincidental”. However, there is a disagreement in the scientific community over this assertion, with a counter-argument that a crossover near the peak sensitivity and a crossover far away from the peak sensitivity are unlikely to have similar implications on observer metamerism [81].

In a related mathematical approach by Kuehni and Ramanath [82], observer CMFs were interpreted as dimension reduction functions. The magnitude of squared difference between stimulus functions was considered to be an approximate measure of the degree of metamerism. Consequently, the maximal three-crossover metameric pair was defined as a neutral gray with uniform function value of 0.5 and a metamer with three sharp transitions between 0 and 1. The wavelengths at which such transition occurred were called transition wavelengths. Differences between observer CMFs were predicted with the help of transition wavelengths for such metameric pair, and it was shown that the transition wavelengths could be used effectively in comparing and distinguishing CMFs of individual observers.

However, an essential requirement of the above methods is that the observer CMFs must be known. These methods aim to identify either the observer CMFs or the stimuli that lead to high observer metamerism. Their purpose is not to offer a solution to the problems encountered in practical applications, but to provide a better understanding of the phenomenon of observer metamerism.

3.2.3 *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 [44]. The observers were selected based on their reliability and experience in trichromatic matching, not based on their actual results [9] (page 346).

Allen [83] 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.

In a different statistical approach, Nayatani et al. [84] performed a singular-value decomposition analysis on the 20-observer data and derived four deviation functions characterizing the variations of color-matching functions of color normal observers. The new standard deviate observer was tested on two sets of metameric spectral reflectance values of 12 and 68 metamers. Only the first deviation function was used to evaluate the degree of observer metamerism. A subsequent study by Takahama et al. [85] expanded the method by using the first deviation to evaluate the index of observer metamerism. All four deviations were used to construct the confidence ellipsoids defining the range of mismatches expected for a given pair of metamers, viewed by actual observers with normal color vision but different from the reference. In an independent study, Ohta [86] performed a nonlinear optimization of the 20-observer data to formulate a standard deviate observer model. The model was close to the one obtained by Nayatani, and was assessed to well represent the original 20 observers.

Mainly based on the works of Nayatani et al. [84] and Takahama et al. [85], the CIE published in 1989 a technical report titled *Special Metamerism Index: Change in Observer* [87] (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 metamerism.

However, the CIE standard deviate observer model did not perform well when evaluated with independent experimental data. As we will see shortly, many researchers reported [88] [74] [89] [90] 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 [88] and improper mathematical treatment of the original colour matching data [89]. In this regard, it is interesting to note that studies like those of Katori and Fuwa [91] and Nayatani [92] that reported much smaller observer variability compared to other studies in the US and Europe were all conducted in Japan, prompting some authors [74] to speculate genetic or ethnic influence in the apparent contradiction of experimental results. However, there is no substantial evidence to support this speculation as yet. Looking from the point of

view of practical industrial applications, in particular hard-copy vs. soft-copy color-matching, some researchers [90] [27] have questioned the purpose and usefulness of an index of observer metamerism, and 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.

3.3 Observer variability in classical color-matching experiments

Numerous researchers have conducted color-matching experiments with a variety of experimental setups and goals [9] (page 288). Many experiments were performed with a small field-of-view. For example, the 2° color-matching experiments by Guild [40] and Wright [39] [93] are some of the most authoritative experimental works in color science, leading to 1931 CIE 2° standard colorimetric observer functions [94] [95]. These studies and their subsequent evaluations have shown a great deal of observer variability [9] (page 343). In one of the first attempts to model the uncertainties involved in the color-matching data, Nimeroff et al. [96] proposed a statistical model they termed as Complete Standard Colorimetric Observer System. The model included the mean of the color-matching functions of various observers, as well as variance and covariance of these functions derived from the intra- and inter-observer variability. Their analysis showed the ratio of inter- and intra-observer variability was about 5.7. More recently, in a computational analysis of CIE 2° standard colorimetric observer and other CMFs, Shaw and Fairchild [97] found that the magnitude of observer variability was nearly eight times that of the variability found between various CMFs, and concluded that the problem of observer metamerism was more of a concern than the accuracy of the CIE 2° standard colorimetric observer itself.

Similarly, a preliminary experiment conducted by Stiles [98] showed that the CMFs of different observers varied by as much as two log units. The overall standard deviation of the collected data was found to be much larger than the standard deviation of one or two individual observers. Unlike Kaiser and Hemmendinger [75], Stiles found only a weak correlation between age and lens density in the human eye, and also stronger influence of macular pigment on observer variability compared to the aging of the eye lens.

The following subsections outline some of the most notable classical large-field, trichromatic color-matching experiments. Note that small-field (for example 2°) color-matching experiments, as well as various theoretical studies on color-matching data are not reviewed here. Also excluded from the scope of this discussion are the tetrachromatic color-matching experiments with four primary stimuli, aimed at investigating rod participation in large-field color-matching.

3.3.1 *Stiles and Burch's experiment (1959)*

More than fifty years ago from the time of writing this thesis, Stiles and Burch [44] conducted at the National Physics Laboratory at Teddington, England the most comprehensive, and arguably the most authoritative large-field color-matching experiment till date involving a total of 49 normal trichromats. These data, together with those obtained by Speranskaya [45] eventually led to the CIE 1964 supplementary standard observer for large-field viewing, which is referred to as CIE 10° standard colorimetric observer throughout this thesis (see Chapter 4 for further discussion). Stiles and Burch used three double monochromators with subtractive dispersion. Such configuration ensures that the spectral dispersion at the exit slit of the second monochromator is essentially zero, and the light leaving its exit slit is spectrally uniform. The monochromators were mounted vertically on top of each other, as shown in Fig. 3-17. A movable, narrow slit in the middle level was used to select the monochromatic primary stimulus to form one half of the test field provided in the photometer cube. Three fixed slits in the upper level selected the monochromatic primaries, which after recombination in the second level provided the comparison field in the photometer cube. The lower level employed a similar mechanism as the top level using the same primaries, with the effect of de-saturating the test stimulus.

The radiances of primary stimuli could be independently controlled by several neutral density filters placed next to the slits in the middle level. Beyond the photometric cube, all the light concentrated within a square area of two millimeter size and collected in the pupil of the observer. The observer saw the horizontally divided bipartite field by the method of Maxwellian view [9] (page 478). A 14° surround with the same spectral composition as the test stimulus was provided.

Color-matching functions of the observers were measured at wave-numbers from 14000 cm^{-1} to 25500 cm^{-1} at intervals of 250 cm^{-1}. The monochromatic red, green and blue primary stimuli were located at wavenumbers 15500, 19000 and 22500 cm^{-1}, respectively, which translate to wavelengths of 645.2 nm, 526.3 nm and 444.4 nm respectively. The retinal illuminance values of the test stimuli at these wavelengths were around 794, 1585 and 63 photopic trolands respectively. A detailed description of the experimental variables is given in [9] (page 338).

Stiles and Burch investigated intra-observer variability by repeating measurements for two observers four and five times respectively. Intra-observer variation was large in the blue region of the color space and relatively low in regions where corresponding tristimulus values were the largest. Variability between individual observers is illustrated in Fig. 3-18. Singularities at the wavelengths of primary stimuli indicate the locations of the primaries. For example, for all observers the short-wave sensitive color-matching function was set to a value of unity at 444.4 nm, and other two color-matching functions had the value zero, thus resulting in zero standard deviations. In analyzing the

variability in the individual observer data, Stiles and Burch considered possible contributions from various physiological factors. They noted that the variations could not be completely explained by the absorption due to the filter pigments in the eye. They also took into account rod participation in large-field color-matching [9] (page 354) and tried to account for it.

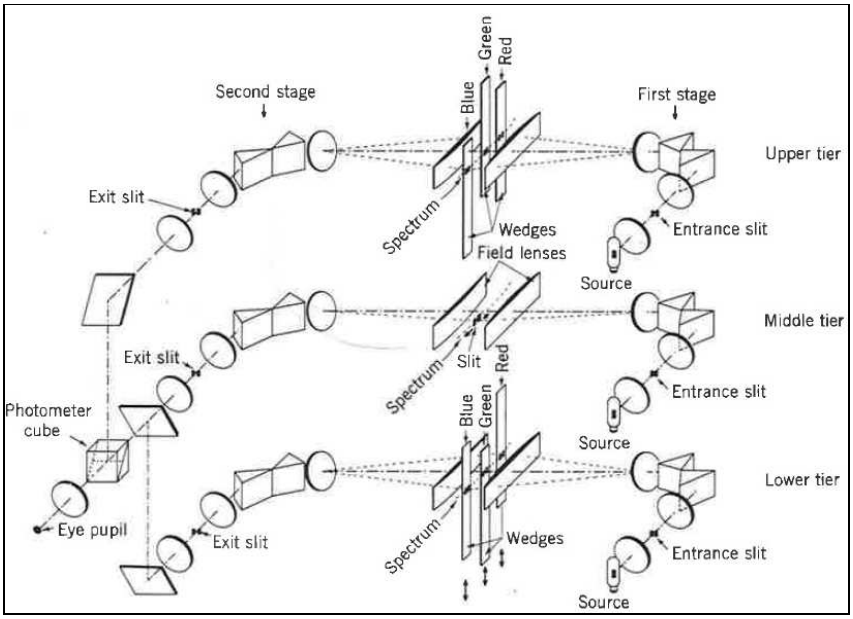


Fig. 3-17. Schematic diagram of Stiles trichromator (from [9], page 476)

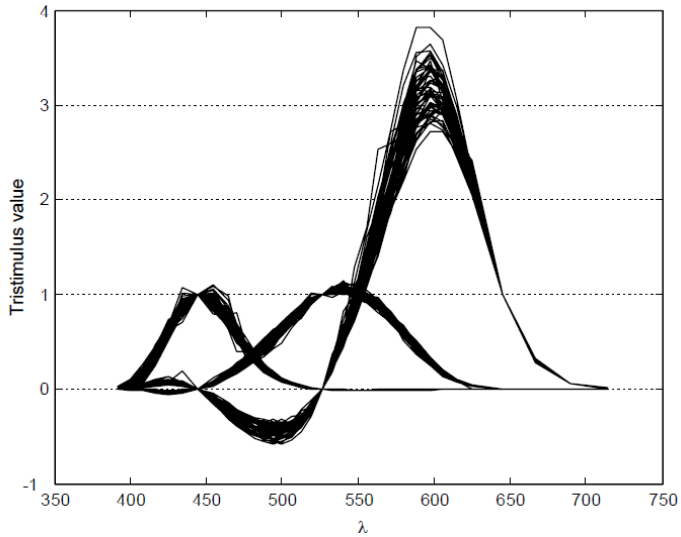


Fig. 3-18. 10° color-matching functions of 49 Stiles-Burch observers (from [44])

The data from the large-field experiment by Stiles and Burch [44] have been used extensively in this thesis. The results are discussed in Chapter 4 and 6. The Stiles and Burch dataset has recently been recompiled and made available electronically on the Color & Vision Research Laboratory website [99].

3.3.2 *Viénot's experiments (1977)*

Viénot [100] designed an optoelectronic instrument for the measurement of color-matching functions using the Maxwell method (see Section 2.3.3). An additive mixture of two color primaries were used on one side of the 10° bipartite field, while on the other side an additive mixture of a third primary and a monochromatic light was used. A 30° surround was used, along with an intermittent white stimulus alternating with the colored beam by means of a flicker device. The white stimulus was meant to break temporal adaptation. The primaries were obtained through several interference filters held before a high-pressure xenon arc lamp. The observers (two observers participated) were able to make the two halves of the bipartite field match by moving three photometric wedges to control the colors on both halves, and the luminance on the left half.

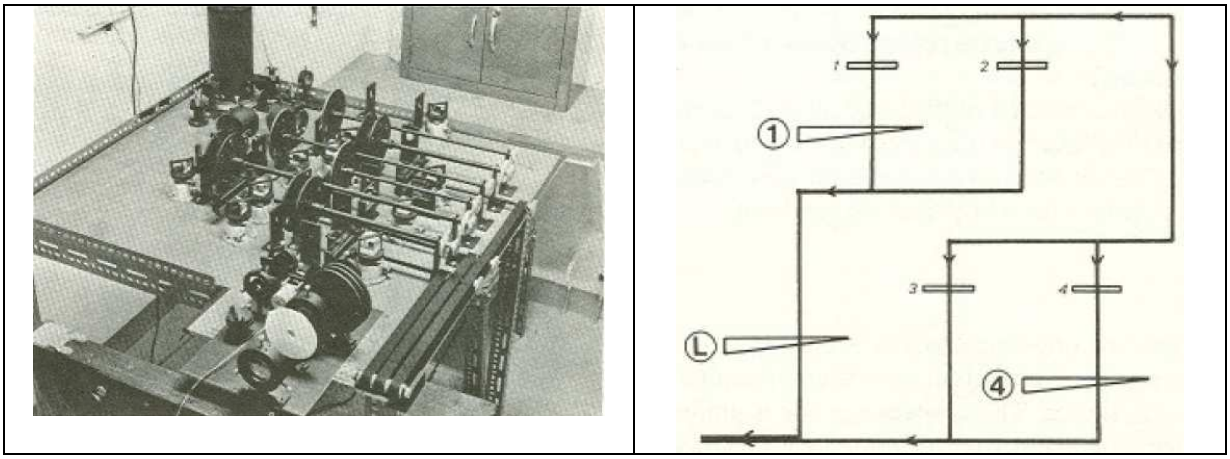


Fig. 3-19. Viénot's color-matching instrument (left) and its schematic diagram (right) (from [100])

The instrument, shown in Fig. 3-19, was subsequently used to measure color-matching functions of 10 observers [101]. The luminance levels of the test field varied widely from 150 trolands for 695 nm stimulus to 4250 trolands for 480 nm stimulus. The intra- and inter-observer variations in the data were compared with the results of Stiles and Burch [44], and their possible explanations were explored. Inter-observer variations were significantly more than those reported by Stiles and Burch [44] in the red and blue extremities of the color spectrum, while for other parts of the spectrum they were comparable. The high variability in the higher wavelengths (red extremity) was attributed to low luminance of blue flux and the relative insensitivity of the short-wave sensitive color-matching function. For the blue extremity of the spectrum, the effect of Maxwell spot (see Section 2.5.1), and differences in the experimental method used in the two studies, particularly different operating luminance levels, were thought to be the reasons behind higher inter-observer variability. Viénot further analyzed and concluded that (r, g) chromaticity diagram was not convenient for comparison of such inter-individual variability. With regard to intra-observer variations, the retinal heterogeneity

over the 10° field was proposed as a possible cause for such variability. According to Pokorny et al. [102], the color-matching is determined neither by the fovea nor by the perimeter of the retina, but by the intermediate area between them. Viénot [101] argued that variations of inert filter pigments in the eye and cone length in this intermediate area result in the balance between two half-fields being unstable, leading to intra-observer variations during color-matching.

3.3.3 *Katori and Fuwa's experiment (1979)*

With an instrument similar to that of Stiles and Burch, Katori and Fuwa [91] conducted a 10° color-matching experiment with 10 normal trichromats at the Electrotechnical Laboratory in Tokyo, Japan. The aim was to derive a 10° luminous efficiency function from the measured color-matching functions and heterochromatic brightness matches with flicker photometry. Some discrepancies were observed in the mean results as compared to the mean color-matching functions from the Stiles and Burch study, particularly in the short- and long-wave sensitive regions. This was attributed to the differences in the luminance level and to rod intrusion.

3.3.4 *Thornton's experiments (1992)*

Thornton [103] [104] [105] performed several 10° color-matching experiments with his visual colorimeter-spectroradiometer instrument, using disparate sets of spectral primaries. Maxwell method of color-matching was used (see Section 2.3.3), where the reference field was always white. Three primary sets were used: "prime-color" (PC) in the spectral region of 452-533-607 nm, "antiprime" (AP) in the spectral region of 497-579-653 nm, and "nonprime" (NP) in the spectral region of 477-558-638 nm. Thornton observed that the CIE 10° standard colorimetric observer performed relatively poorly in the presence of spectral content in the AP region whereas better performance was achieved when incoming light was composed of a matching combination of the PC primaries. He further observed that "a single computed chromaticity fails to represent a set of lights pronounced metameric by a normal human observer. Conversely, some members of a set pronounced metameric by the Standard Observer may mismatch grossly to a normal human observer." [104]

While some of the discrepancies in the results regarding perceived brightness or matching condition of two lights were attributed to the mathematical construct that is the CIE standard colorimetric observers, the discrepancies reported in Thornton's papers went well beyond observer metamerism. For example, he noted that "the large chromaticity errors among the 28 Maxwell-Method matching lights…are present even in color diagrams constructed from the observer's own maximum-saturation matching data from the same (PC) primary-set." He thus concluded "errors in computation of the tristimulus values (which should be identical for visually-matching lights) must thus be due to some basic shortcoming in the use of CMFs as weighting functions on the SPD of the incoming light." A

major inference from his work was that Grassmann's law of additivity [26], the very basis of much of colorimetry as we know, did not hold for transformation of certain primaries. In other words, when color-matching data obtained by using one set of primaries were used to predict color matches obtained by using a different set of primaries, discrepancies were observed between computed and measured tristimulus values. Thornton's findings led to an intense debate in the scientific community. In fact, an entire CIE Symposium [106] was devoted to Thornton's findings, and led to the formation of the CIE Technical Committee TC 1-56 [28] and several independent investigations [107] [108] [109]. While this aspect of Thornton's work, the apparent additivity failure of color-matching data, is outside the scope of this thesis, it has been dealt in detail in Oicherman's PhD thesis [27].

3.3.5 *Color-matching experiments to compare the Maxwell and maximum saturation method (1965-72)*

The two different methods of color-matching, namely the Maxwell method and the maximum saturation method, were discussed in Chapter 2. Several large-field classical color-matching experiments to probe the discrepancies in the data obtained using these two methods were conducted back in the 1970's, which are briefly mentioned here for the sake of completeness. Implications of the results from these experiments were already discussed in Chapter 2.

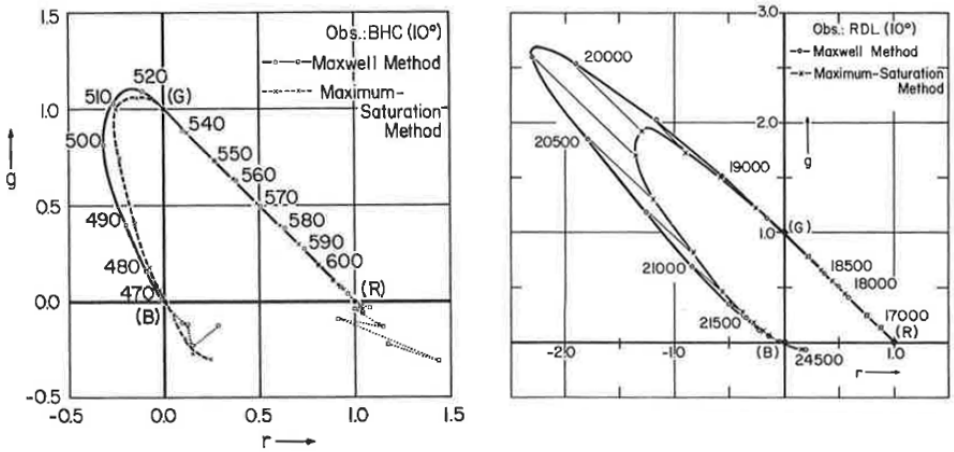


Fig. 3-20. Spectrum loci derived from color matches made in a 10° bipartite field by one individual observer using the Maxwell method and the maximum saturation method. Left figure is from Crawford's study [32] and the right figure is from Lozano and Palmer's study [33] (Reproduced from [9], page 385)

Crawford [32] conducted a color-matching experiment with six observers using the same monochromator used by Stiles and Burch. Results from his experiments, described earlier in Chapter 2, are reproduced here in Fig. 3-20 (left). Both Maxwell and maximum saturation methods were employed on large 10° field as well as 1° and 2° foveal field color-matching. Narrow-band primaries at 650, 530 and 460 nm were used in the experiments. The spectrum loci obtained by the Maxwell

method and the maximum saturation method deviated from each other in the blue-green region of the spectrum. The effect was smaller with 1° and 2° field, but nonetheless present, thus Crawford ruled out the possibility of rod intrusion playing a role in this discrepancy.

Lozano and Palmer [110] also conducted similar experiments using the Stiles and Burch's colorimeter. CMFs were measured for four observers using the maximum saturation method. The observers additionally matched 20 broadband stimuli having a wide range of chromaticities using the Maxwell method. Results similar to those of Crawford were reported, and are shown in Fig. 3-20 (right). Observed blue tristimulus values for some observers were often underestimated in computations. While the intra-observer variability was found to be around 3%, the discrepancies were around 20%. In a subsequent study [33], the CMFs of one observer were measured using the Maxwell method at high luminance level of 160 Td, and using the maximum saturation methods at both high (160 Td) and low (10 Td) luminance levels. Similar trends in the results were observed as before.

In several pilot tests, Wyszecki [9] (page 386) found similar results using a color-matching instrument at the National Research Council, Canada having the same design as Stiles' monochromator (Fig. 3-17). 2° and 9° visual fields, at an illuminance level of 1000 Td, were used. The data showed similar features as those of Crawford [32] and Lozano and Palmer [33], even though Crawford's finding that the effect of additivity failure was less pronounced for smaller field size could not be confirmed. The magnitude of the effect was found to be dependent on the wavelength.

3.4 Observer variability in applied color-matching experiments

While a huge amount of research has been conducted in the past to identify the sources and magnitudes of individual variations in color-matching, the evidence of a significant effect of these variations from the perspective of applied colorimetry was scarcely documented until early 1990's. In the following subsection, several key applied studies on observer variability are reviewed.

3.4.1 *Maxwell-type color-matching experiment using a CRT and a tungsten-halogen lamp*

North and Fairchild [111] conducted a Maxwell-type color-matching experiment using an instrument utilizing Cathode Ray Tube (CRT) display primaries on one half of a 2° bipartite field, and a tungsten-halogen lamp with interference filters on the other half, simulating daylight. The observers controlled the CRT primaries in the lower half to match the daylight reference in the top half. Filter wavelengths were mixed with the CRT primaries in order to determine the color-matching data at specific wavelengths. Color-matching data at seven wavelengths were obtained for 18 observers in the age range of 20 to 40, including one observer who performed 20 repetitions. The authors estimated

the color-matching functions of each observer through a mathematical model, starting from experimental data obtained at the seven wavelengths.

In the analysis of their data, the authors made several conclusions [88]. First, they found that the mean data for the 18 observers were consistent with the 1931 CIE 2° standard observer and Stile's 2° mean observer [98], and concluded that the CIE standard colorimetric observer was an appropriate representation of the average color normal human observer. Second, they concluded that inter-observer variability in their data, which was significantly more than the intra-observer variability for a single observer, was much larger than what was predicted by the CIE standard deviate observer [87]. Next, their method showed little difference in the color-matching data of two individuals over a 20-year period. Finally, no correlation was found between the observer age and any of the model coefficients for lens and macular optical density. One researcher [112] subsequently questioned the accuracy of North and Fairchild's method and soundness of their conclusions. Indeed, the inter-observer variability reported by North and Fairchild (see Fig. 6 of [88]) seem to be significantly more than what has been observed in subsequent studies, including in this thesis research. Further, color-matching functions of an individual can reasonably be expected to vary over a 20-year period, particularly in the short wavelengths. While the unexpected results could have originated from the approximations in the mathematical modeling, there is enough documented evidence that color-matching data obtained by the Maxwell method have more uncertainty than the data obtained by the maximum saturation method [9] (page 386). However, North and Fairchild's conclusion about the CIE standard deviate observer's [87] under-estimation of the inter-observer variability was corroborated by Rich and Jalijai [74], and in other subsequent studies reviewed below.

3.4.2 *Cross-media color-matching experiment using a CRT and color prints/transparencies*

In order to better quantify observer variability in color matches between CRT displays and printed materials, Alfvin and Fairchild [89] conducted a visual experiment on color matches between color prints or transparencies and a CRT display. The objective was to quantify the precision and accuracy of three sets of color-matching functions, and also the magnitude of inter- and intra-observer variability. An optical apparatus consisting of an equilateral glass prism was used to allow the observers to view simultaneously both the soft- and hard-copy stimuli in a vertically symmetric bipartite field. The equiluminant stimuli with an absolute luminance of 50 cd/m² were viewed as unrelated and self-luminous colors at a visual angle of 2.9°. The observers were asked to adjust the color appearance of the soft-copy stimulus by adjusting the color along CIELAB (L*, a*, b*) dimensions in order to create an exact color-match for each of the hard-copy stimuli. Spectral radiances of the stimuli were measured after each match. Results from the experiment showed that the variability of inter-observer color matches was approximately twice as large as the intra-observer variability in the color matches. The mean color difference from mean (MCDM) for inter-observer

variation was 2.7 CIELAB units. The results refuted a previous study by Pobboravsky [113] where the effect of observer metamerism on color-matching between hardcopies and soft-proofs was shown to be insignificant. Alfvin and Fairchild concluded that the existing CIE 2° and 10° standard colorimetric observers were a good representation of the population of normal trichromats, but the inter-observer variability was significantly larger than the prediction of the CIE observer metamerism index [87].

3.4.3 *Observer variability prediction using Davidson & Hemmendinger Color Rule*

Diaz et al. [114] studied how the metameric match changed when each physiological parameter responsible for variations in color vision was altered. The authors used Davidson & Hemmendinger (D&H) Color Rule [75] to predict the matches of a theoretical observer with normal color vision. Color matches were also predicted for deviate observers by first deriving the cone fundamentals of the theoretical normal observers, then by changing experimentally determined values of lens and macular pigment density, and finally by accounting for a shift in the long-wavelength sensitive photopigments. For determining the lens density, a Maxwellian view at 8 td was used in a 10° foveal field. For estimating the macular pigment density, heterochromatic flicker photometry matches of 466 nm test stimulus and 558 nm reference stimulus were measured on the periphery of a 10° foveal field and were compared with their values on a 2° foveal field, at photopic illuminance. Observations were made in Maxwellian view at 2.40 log td. Matches from eight observers obtained under monocular vision using D&H Color Rule were compared to the matches predicted by computing their personalized cone fundamentals from several independent psychophysical measurements. The authors performed a quantitative assessment of the effects of various sources of individual variation in color vision on a metameric color match. In conclusion, it was suggested that a match could be better predicted by using personalized corrections of various physiological parameters than by using a theoretical model.

3.4.4 *Cross-media color-matching experiment using paint samples and two displays*

Oicherman et al. [115] investigated the contribution of various sources of variability in color-matching by conducting a color-matching experiment of maximum-saturation type. They conducted an asymmetric color-matching experiment [107] where eleven observers were asked to match the colors displayed on a CRT and an LCD to the colors of two achromatic and eight chromatic paint samples placed inside a light booth one at a time. A 6° viewing angle was used, with the maximum luminance level set at 120 cd/m². As in the study of Alfvin and Fairchild [89], the colors were adjusted in CIELAB (L^*, C^*_{ab}, h^*_{ab}) for chromatic stimuli and (L^*, a^*, b^*) for achromatic stimuli. The results showed a discrepancy between the CIE 10° standard colorimetric observer and the mean of real observer data in the form of a blue shift. The authors hypothesized that this discrepancy was due to additivity failure caused by adaptation.

Large variability was observed between the Stiles and Burch (1958) color-matching dataset and the results obtained by the authors. They also reported a significant under-prediction of the observer variations of color-matching data by the CIE standard deviate observer [87], accounting for only 15% of inter-observer variability. The main difference between the inter- and intra-observer variability was found to be in the lightness dimension. Since differences in physiological factors have a major effect on color perception and relatively minor effect on lightness perception, the authors argued that the inter-observer variability in this case (i.e. cross-media color-matching) was not governed by observer metamerism. As per their argument, mechanisms operating asymmetric color-matching are potentially different from those of direct comparison of cone signals, and thus the degree of observer metamerism does not correspond to the degree of variability of matches between spatially separated stimuli. They suggested that an optimization of the CIEDE2000 (ΔE_{00}) parametric coefficients was more appropriate approach to model the observer variability in cross-media color reproduction.

3.4.5 *Color-matching experiment using broad-band stimuli and LEDs*

Csuti and Schanda [116] conducted a Maxwell-type color-matching experiment in a 2° x 3° bipartite field, where one half of the field was illuminated by filtered incandescent lamp, while the other half was illuminated by an additive mixture of RGB LEDs. The dominant wavelengths of LED primaries were 626 nm, 525 nm, and 476 nm. Colored filters were used to generate specific colors on the reference field with the incandescent lamp. Luminance of the reference field varied between 90 and 400 cd/m². Six observers performed color-matching by changing hue, brightness and saturation (the color space used was not mentioned in the paper) of the LED primaries. The authors reported large visual mismatches, particularly in the blue part of the chromaticity diagram, when the CIE 2° standard colorimetric observer was used in the computation. However, the chromaticity error in (u', v') coordinate system could be reduced by around 50% by using color-matching functions derived from CIE 2° physiological cone fundamentals [14]. In a subsequent step, the authors optimized the 2° physiological cone fundamentals to obtain further improvement in the results [117].

3.4.6 *A new generation of color-matching instruments*

Before concluding this discussion on applied color-matching experiments, it is pertinent to mention the new spectrally programmable light engines that are likely candidates for a new generation of colorimeters, which could be used in near future for conducting color-matching experiments. One such instrument is OneLight Spectra™ [118]. It is based on Texas Instruments' DLP™ Technology [119], which employs a microprocessor fitted with up to two million tiny, hinge-mounted microscopic mirrors with precision digital control. In case of OneLight Spectra™, these mirrors reflect light with a specific spectral power distribution into a liquid light pipe with 5 nm aperture. The instrument operates in the 380 nm - 720 nm range and provides software control of the intensity at each

wavelength independently. While the instrument has a spectral accuracy of 1 nm, the current spectral bandwidth is 14 nm. Thus the output stimulus is fairly narrow-band, but not monochromatic.

The advantage of such an instrument is the incredible flexibility in spectrum generation without requiring several expensive optical components like photometric wedges and interference filters, and their elaborate and cumbersome mounting. The disadvantages, at least for the time being, are the high price (we will require two instruments for generating a bipartite field), and a relative high signal bandwidth (cannot be used in classical experiments). Nevertheless, these instruments demonstrate high prospect for being adaptable for the purpose of color-matching experiments.

3.5 Conclusions

As this chapter demonstrates, various studies in the past, both classical and applied, have provided significant amount of insight into the issue of observer variability in large-field color-matching, and its ramifications in basic color science and applied color technology. While over the past couple of decades our knowledge of underlying physiological reasons for individual variability in human color vision has enriched considerably, we are yet to come up with a practical solution in applied colorimetry. Being constrained to a single average observer model, colorimetry is unable to predict how individual color matches might differ from those of an average match. The consequence is non-trivial for certain color-critical industrial applications.

Oicherman in his PhD thesis [27] (Chapter 2.7) aptly highlights the lack of progress with regard to offering an industrially viable solution to the problem of observer metamerism: "…*almost complete absence of studies on evaluation of observer metamerism in industrially-relevant conditions is very surprising. It seems that there is a marked discrepancy between the declared significance of observer metamerism in industry, and interest of researchers in carrying out studies on quantifying and characterising the phenomenon*."

Current thesis research attempts to bridge this gap.

The wonder of science is not in the answers it provides but in the questions it uncovers. For every miracle it finally explains, ten thousand more miracles come into being. ~ John Pielmeier, Agnes of God (1978)

4. Colorimetric Observers and Observer Variability

4.1 Introduction

The most fundamental aspect of applied colorimetry is the trichromacy of our visual system, which allows us to represent any color in terms of its tristimulus values. Computing tristimulus values for any object color requires the use of the spectral reflectance of the object color, the spectral power distribution of the scene illuminant, and the spectral characteristics of a colorimetric observer. For the color imaging community, it is of interest to investigate which is a better representation of real observer data, CMFs derived from CIE 2006 physiological observer model, or the CIE 10° standard colorimetric observer. This issue has been explored through a theoretical analysis performed in the context of display colorimetry.

4.1.1 *CIE 2° and 10° Standard Colorimetric Observers*

In 1931, the CIE (Commission Internationale de l'Éclairage) defined a standard observer for colorimetry, based on Wright's [39] and Guild's [40] 2° color matching data. However, the basic datasets were transformed to incorporate V(λ), the luminous efficiency function of the CIE standard photometric observer [41], into the standard colorimetric observer. Incorporating both photometric and colorimetric characteristics was motivated by a need to simplify hardware computations [10], but this has been a major source of criticism of CIE 1931 standard colorimetric observer, since the CIE standard photometric observer was based on an entirely different set of psychophysical task than color matching [16] (page 110). CIE 1931 standard colorimetric observer led to spectral estimation error caused by the underestimation of luminosity at short wavelengths with the CIE standard photometric observer. Revisions of the CIE standard photometric observer $V(\lambda)$ function below 460 nm were proposed by Judd [120] in 1951, and further revision below 410 nm was proposed by Vos [121] in 1978. The former was widely accepted in the vision science community, and the latter resulted in a CIE recommendation in 1988 in the form of a supplementary observer $V_M(\lambda)$ for photometry [122], but the color imaging industry continued to use the original CIE 1931 standard colorimetric observer derived from 2° color matching data, applicable to small fields.

In 1964, CIE recommended a large-field standard colorimetric observer based on the work of Stiles and Burch [44] and Speranskaya [45]. Stiles and Burch maintained high photopic luminance of the matching fields and incorporated mathematical corrections to exclude the effect of rod intrusion in long-wavelength color matches. The color-matching function $\bar{y}(\lambda)$ represents the relative spectral luminous efficiency function of the CIE 10° standard colorimetric observer, but the photometric standard still uses $\overline{y_2}(\lambda)$ from the CIE 2° standard colorimetric observer to define luminance, even for large-field stimuli.

For many practical industrial applications, the use of the 2° standard colorimetric observer is questionable, as the field-of-view is typically much larger than 2°. Indeed, many industrial engineers have chosen to use $\bar{y}_{10}(\lambda)$ in colorimetric applications. However, because of the absence of rod contribution, and more importantly, because of individual differences in the visual system, it has been observed that even the CIE 10° standard colorimetric observer does not always correspond to real observer matches for large fields.

4.1.2 *CIE 2006 Physiologically-Based Observer*

In 2006, CIE's technical committee TC 1-36 published a report [14] (described hereafter as CIEPO06, an abbreviation of CIE 2006 physiological observers) 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 [22]. Starting from 10° CMFs of 47 Stiles-Burch observers [44], 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.

4.1.3 *Individual cone fundamentals*

In its approach to construct a fundamental observer, technical committee CIE TC 1-36 has ignored individual variability [14] [123]. A few studies [124] 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 [125], comparing inter-individual and intra-individual variability of experimental CMFs [101]. Wyszecki and Stiles [9] (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 [111] [88] and test additivity and transformability of color matches [115] [108]. One study of nine observers' color-matching functions

concluded that a main cause of the individual difference was the difference of individual spectral lens density [126]. Individual variations of Rayleigh matches have also been examined experimentally [64] [127] [128] [129] or theoretically [130]. Although these studies have attempted to relate the variation of color matches to underlying physiological factors, they failed to model individual effects of these factors in a practical manner that could be implemented in industrial applications.

This chapter takes advantage of the framework developed in CIEPO06 to examine through theoretical analysis the effect of age on the CMFs of individual observers and on individual color matches as viewed on displays.

4.1.4 *General colorimetric transforms*

Each set of CIEPO06 cone fundamentals can be converted to CMFs through a linear transformation. At the time of this work, the final 3x3 transformation matrix for such conversion was not yet made available by CIE TC 1-36. Two approaches could yield a proper linear transformation. An approximate 3x3 LMS-to-XYZ transformation matrix was computed from the available CIE 1964 10° $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ standard colorimetric observer functions and the average $\bar{l}^{N}_{SB_{10}}(\lambda)$, $\bar{m}^{N}_{SB_{10}}(\lambda)$, $\bar{s}^{N}_{SB_{10}}(\lambda)$ cone fundamentals of 47 Stiles-Burch observers each normalized to unity. The transformation matrix is given below:

$$\begin{bmatrix} \bar{x}_{10}(\lambda) \\ \bar{y}_{10}(\lambda) \\ \bar{z}_{10}(\lambda) \end{bmatrix} = \begin{bmatrix} 1.905378 & -1.321620 & 0.419512 \\ 0.698648 & 0.333043 & -0.013360 \\ -0.024300 & 0.040453 & 2.073582 \end{bmatrix} \begin{bmatrix} \bar{l}^{N}_{SB_{10}}(\lambda) \\ \bar{m}^{N}_{SB_{10}}(\lambda) \\ \bar{s}^{N}_{SB_{10}}(\lambda) \end{bmatrix} \tag{4-11}$$

The above matrix was used at all times for converting any normalized L, M, S cone fundamentals from Stiles-Burch dataset into CIE XYZ like CMFs similar to 10° $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ functions. It is reasonably close to the matrix published earlier by other researchers [116]. Note that in [116], a negative sign was accidentally omitted in the 1st row, 2nd column of the transformation matrix (Eq. 4-11).

Another approximate 3x3 LMS-to-XYZ transformation matrix was computed from the CIE 10° standard colorimetric observer functions and the CIEPO06 model cone fundamentals $\bar{l}_{CIE06_{10}}(\lambda)$, $\bar{m}_{CIE06_{10}}(\lambda)$, $\bar{s}_{CIE06_{10}}(\lambda)$ applicable for an age of 32 and 10° field size without any normalization of the cone fundamentals, as shown in Eq. 4-12.

$$\begin{bmatrix} \bar{x}_{10}(\lambda) \\ \bar{y}_{10}(\lambda) \\ \bar{z}_{10}(\lambda) \end{bmatrix} = \begin{bmatrix} 0.006873 & -0.005386 & 0.005550 \\ 0.002520 & 0.001358 & -0.000181 \\ -0.000089 & 0.000167 & 0.027432 \end{bmatrix} \begin{bmatrix} \bar{l}_{CIE06_{10}}(\lambda) \\ \bar{m}_{CIE06_{10}}(\lambda) \\ \bar{s}_{CIE06_{10}}(\lambda) \end{bmatrix} \quad (4\text{-}12)$$

This transformation matrix was used in the analysis of the effect of various physiological factors on CIEPO06 cone fundamentals, where normalization is not desirable. If normalized CIEPO06 cone fundamentals are used, the resulting transformation matrix is very close to that of Eq. 4-11.

4.1.5 *The CIEPO06 model*

The CIEPO06 model is a convenient and effective mathematical tool for understanding how various physiological factors affect the cone fundamentals, and thus the CMFs. A brief review of the model will be helpful in better understanding the analysis that follows.

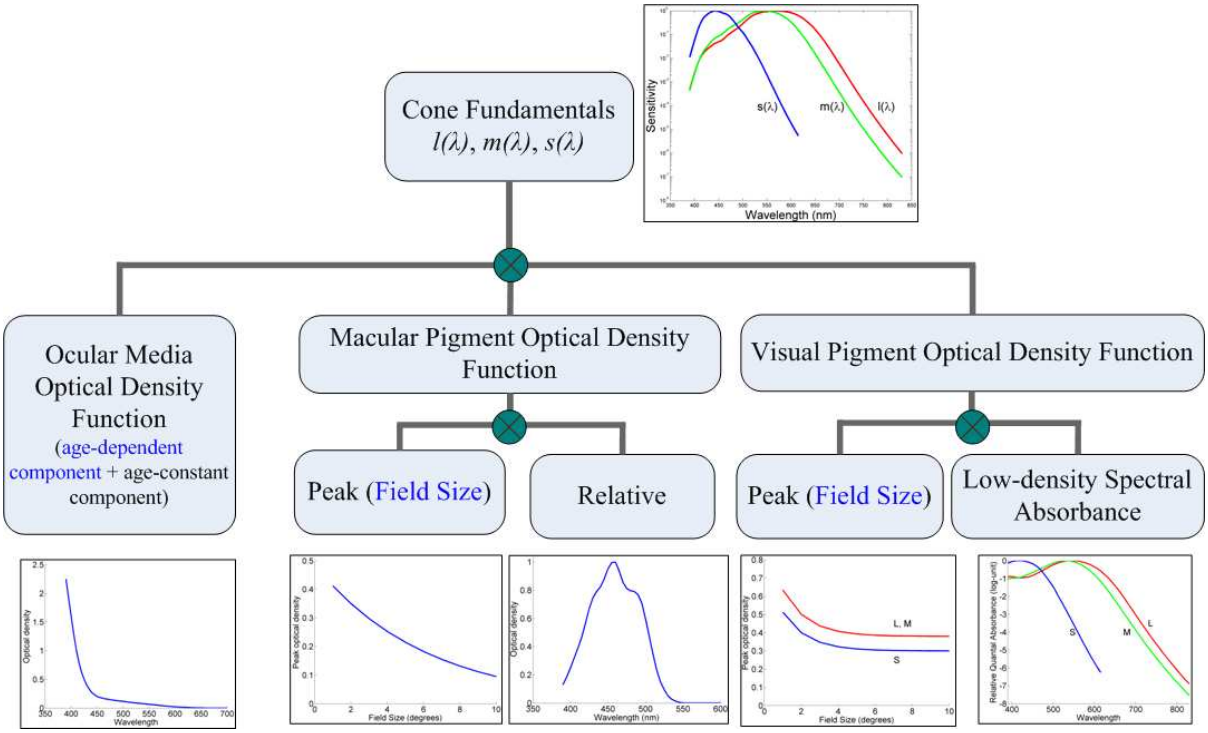


Fig. 4-21. A block diagram of the CIEPO06 framework

CIEPO06 framework [14], shown in Fig. 4-21, involves two parameters, namely, the field-size, varying between 1° and 10°, and the observer age, varying between 20 and 80. Three physiological factors have been incorporated in the CIEPO06 model, in the form of spectral optical density functions for: a) lens and other ocular media absorption, b) macular pigment absorption, and c) visual pigments in the outer segments of photoreceptors. Out of these, the ocular media optical density function has an age-dependent and an age-independent component. The macular pigment optical

density function consists of a peak function and a relative function, where only the peak function varies with the field size. Similarly, the visual pigment optical density has two components, the peak as a function of the field size, and the low-density spectral absorbance that is independent of any parameters.

The CIEPO06 cone fundamentals can be written in a simplified form as in Eq. 4-13 [14]. $A_l(\lambda)$, $A_m(\lambda)$, $A_s(\lambda)$ are the low-optical density spectral absorbance for long-, medium- and short-wave sensitive cones respectively. $D_{vis,l}$, $D_{vis,m}$ *and* $D_{vis,s}$ are peak optical densities of the visual pigments for three cones. $D_{mac}(\lambda)$ *and* $D_{ocul}(\lambda)$ are the optical densities of the macular pigment and the ocular media (including the lens) respectively, with the optical density (or absorbance) being the $\log_{10}$ function of the inversed transmission of the media:

$$\begin{aligned}\bar{l}(\lambda) &= \left[1-10^{-D_{vis,l}.A_l(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)} \\ \bar{m}(\lambda) &= \left[1-10^{-D_{vis,m}.A_m(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)} \\ \bar{s}(\lambda) &= \left[1-10^{-D_{vis,s}.A_s(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)}\end{aligned} \tag{4-13}$$

While these three physiological factors are important contributors to observer variability, there is another important but more complex source of variability that has not been included in the CIEPO06 model. A number of studies have suggested that individual differences in the color vision are partly due to the variations in the peak wavelength (λ_{max}) of the cone photopigment [131]. These differences can be due to individual variability, but can also be due to a variation in genetic composition or polymorphism, for example, a single amino-acid substitution (Alanine for Serine) at position 180 of the long-wave sensitive (LWS) photopigment opsin genes [24].

The rest of the chapter is organized as follows. In the next section, a theoretical analysis investigating the relative importance of various physiological factors on display color perception is presented. In Section 4.3, the average Stiles-Burch observer data from three different age-groups are compared with the corresponding CIEPO06 model predictions and the CIE 10° standard colorimetric observer. The perceptual effect of the prediction errors in these two cases are then explored in the context of display colorimetry in Section 4.4. Next, Section 4.5 presents a constrained nonlinear optimization of the CIEPO06 model, performed in an attempt to improve the prediction errors for various age-groups. The chapter concludes by summarizing the results obtained from these theoretical analyses in Section 4.6.

4.2 Effect of various physiological factors on display color perception

Individual variation in color perception depends on the spectral characteristics of the stimuli. As Smith and Pokorny [132] have observed, "*With the generally broadband spectra of reflective*

materials, factors such as lens transmission or macular pigment density provide correlated changes in the spectral distribution of light arriving at the retina from different samples. Thus there may be a translation of color axes but little rotation…Specification based on narrow-band trichromatic primaries may be more or less subject to individual variation, depending on the relation between the spectra of the biological variables and the spectra of the colorimetric primaries". In view of this observation, two questions arise: i) how do various physiological factors described in the previous section affect the color perception on a given display? And, ii) how do these effects vary between a display with broadband primaries and another with narrow-band primaries? The purpose of the analysis described in this section was to investigate these two issues.

4.2.1 *Displays used in the analysis*

The effects of various factors were compared in terms of color perception on two displays with different spectral characteristics. The first was a Sony BVM32 Cathode Ray Tube (CRT) display widely used as a reference studio display (hereafter referred to as Ref-CRT). The second was a Hewlett-Packard DreamColor LP2480zx professional 30-bit Wide-Gamut Liquid Crystal Display (LCD) with LED backlight (hereafter referred to as WG-LCD).

The spectral power distributions of the primaries of the two displays are shown in Fig. 4-22. There is a significant difference in the spectral characteristics between the two displays. WG-LCD is representative of modern wide-gamut displays with peaky primaries, and Ref-CRT is representative of a typical CRT display, and of HDTV broadcasting standard references. 3x3 primary tristimulus matrices of the two displays were computed, which represented the linear relationship between the XYZ tristimulus values and the RGB channel values. Note that normally the digital counts first need to be corrected (linearized) for the display nonlinearity (gamma correction) before computing the primary tristimulus matrix. However, since this analysis is strictly theoretical, and since gamma correction does not affect rest of the computations, display nonlinearity has been ignored in this work. Thus, using the primary tristimulus matrix of a given display, any set of XYZ values could be converted into the corresponding set of RGB channel values and vice versa.

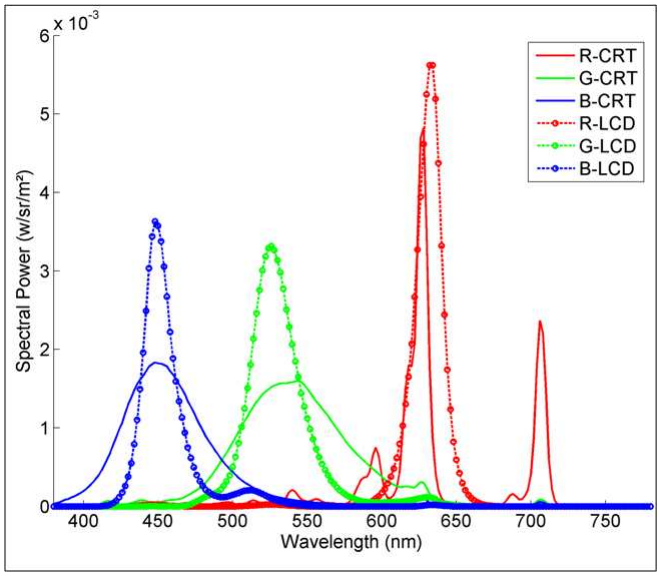


Fig. 4-22. Spectral Power Distributions of the two displays used in the analysis

4.2.2 *Method of analysis*

In this work, the relative importance of the four physiological factors described earlier on the cone fundamentals were explored within the framework of CIEPO06. Cone fundamentals for 10° field size and an age of 32 were computed by independently modifying the contribution of individual factors as follows:

i) mean optical density of ocular media varied by ±25%

ii) peak optical density of macular pigment varied by ±25%

iii) peak optical density term for low-density photopigment relative absorption spectra varied by ±25% (0.38 is nominal)

iv) peak wavelength shift of the cone photopigment optical density in the outer segment of the photoreceptor: a) LWS peak shift by -4 nm (toward shorter wavelength), b) medium wave-sensitive (MWS) peak shift by +4 nm (toward longer wavelength)

Such modifications of optical densities by the same percentage allow us to compare the effect of various factors. For cases (i) and (iv), the modifications are the same as those reported by Smith and Pokorny [132]. A high optical density in case (iii) signifies higher photoreceptor self-screening, resulting in the broadening of the photopigment relative absorption spectra [7] (page 65-66), while case (iv) signifies LWS and MWS polymorphism described earlier. For case (iv), the peak wavelength λ was first shifted in the wavenumber scale ($\tilde{\nu} = 10^7/\lambda$, where $\tilde{\nu}$ is in cm^{-1} and λ is in nm), the cone absorptance spectra were re-sampled, modified cone fundamentals were computed and converted

from the quanta to energy units, and finally were renormalized. Note that case (iv) considers LWS and MWS peak wavelength shifts independently.

For each planned variation of these four factors, a set of modified CIEPO06 cone fundamentals was computed, and compared to corresponding CIEPO06 cone fundamentals under normal conditions. The difference between the two sets of functions indicates the contribution of a given physiological factor. The difference was computed in terms of Euclidean distance in the cone fundamental space. Note also that CIEPO06 10° cone fundamentals have been used here, unlike Smith and Pokorny 2° cone fundamentals as in [132].

In order to simulate the effect of various physiological factors when viewing color stimuli on different displays, chromaticities of these stimuli for a given display and a given set of modified CIEPO06 cone fundamentals must be computed. In this analysis, seven test stimuli were selected from various parts of the common gamut of the CRT and the LCD. These stimuli were chosen such that they covered the whole common display gamut in the CIE 1976 (u', v') coordinate system (Fig. 4-23). The chromaticity of the seventh stimulus was close to that of display white. These coordinates were converted to XYZ colorimetric system through a straightforward transformation, as shown in Eq. 4-14 below. The chromaticity coordinates are listed in Table 4-1.

$$x = \frac{9u'}{6u' - 16v' + 12}; y = \frac{4v'}{6u' - 16v' + 12}; \qquad (4\text{-}14)$$
$$X = \frac{x}{y}Y; Z = \frac{z}{y}Y$$

Table 4-1. CIE 1964 xy and CIE 1976 (u', v') chromaticity coordinates for seven test stimuli and the display whites

Stimulus	x_{10}	y_{10}	Y (cd/m^2)	u'_{10}	v'_{10}
TS-1	0.35	0.48	25	0.1737	0.536
TS-2	0.45	0.39	25	0.2655	0.5177
TS-3	0.53	0.32	25	0.3668	0.4983
TS-4	0.24	0.27	25	0.1667	0.4219
TS-5	0.32	0.21	25	0.2623	0.3873
TS-6	0.18	0.14	25	0.1667	0.2917
TS-7	0.32	0.34	25	0.1988	0.4752
Full White-CRT	0.3093	0.3260	97.36	0.1966	0.4662
Full White-LCD	0.3070	0.3240	97.01	0.1957	0.4648

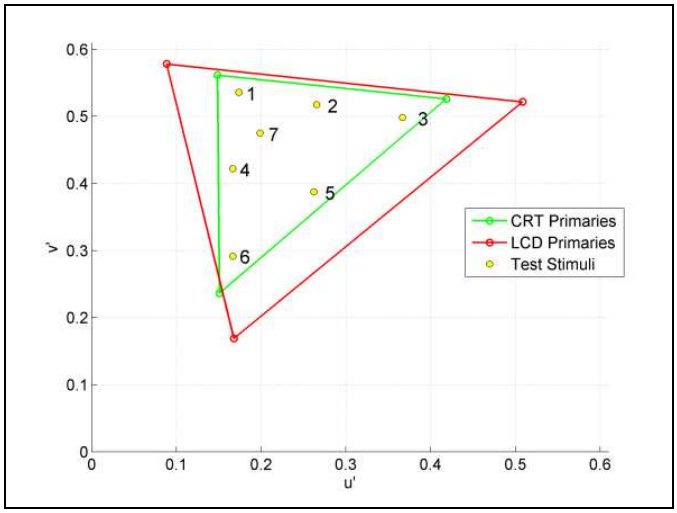


Fig. 4-23. Seven test stimuli in (u', v') chromaticity diagram

Smith and Pokorny [132] investigated the effects of different physiological factors on two sets of chromaticities at a nominal luminance of 8 cd/m², varying along the horizontal and vertical lines in the cone-troland chromaticity diagram. This luminance level is rather low for most industrial applications, thus a constant luminance of 25 cd/m² was used for seven distinct chromaticities described in the next section. Further increase in the luminance resulted in out-of-gamut colors for the displays in the Macleod-Boynton space.

From tristimulus values (X_{10}, Y_{10}, Z_{10}) of the test stimuli, the RGB channel values (R, G, B) required to produce these colors on the two displays were computed using the display primary tristimulus matrices, as shown in Eq. 4-15. The primary tristimulus matrix for a display is formed by the tristimulus values of peak primaries.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} X_{r,\max} & X_{g,\max} & X_{b,\max} \\ Y_{r,\max} & Y_{g,\max} & Y_{b,\max} \\ Z_{r,\max} & Z_{g,\max} & Z_{b,\max} \end{bmatrix}^{-1} * \begin{bmatrix} X_{10} \\ Y_{10} \\ Z_{10} \end{bmatrix} \tag{4-15}$$

The product of the RGB values for each channel and the spectral data of the corresponding display primaries [$P_{pri\text{-}R}(\lambda)$, $P_{pri\text{-}G}(\lambda)$, $P_{pri\text{-}B}(\lambda)$], when added for all three channels, gave the spectral power distribution of the test stimuli for a given display, as per Eq. 4-16. These spectral data were used to compute tristimulus values in the subsequent step, described next. In computing the spectral power distribution of the test stimuli, it is assumed that the displays have perfect additivity and proportionality, and also stable primaries.

$$P_{stim}(\lambda) = \begin{bmatrix} R & G & B \end{bmatrix} * \begin{bmatrix} P_{pri-R}(\lambda) \\ P_{pri-G}(\lambda) \\ P_{pri-B}(\lambda) \end{bmatrix} \tag{4-16}$$

4.2.3 *Derivation of cone troland coordinates from a given set of display channel values*

The derivation of cone troland coordinates from Smith-Pokorny 2° cone fundamentals has been described in detail elsewhere [38] [133] [35]. The method used in this study for deriving the cone troland coordinates corresponding to a given set of display *channel values* and the CIEPO06 10° cone fundamentals is described now.

As mentioned in Chapter 2, MacLeod and Boynton [35] proposed a chromaticity diagram (l_{MB}, s_{MB}) [see Eq. 2-4], where the projective plane is an equiluminant chromaticity plane. A basic assumption in forming the MacLeod-Boynton chromaticity diagram is that short-wavelength sensitive cone fundamental $\bar{s}(\lambda)$ does not contribute to luminance. In this diagram, the abscissa $l_{MB} = L/(L+M)$ represents the equal and opposite change in LWS and MWS cone excitations, i.e. an increase in the LWS luminance is counterbalanced by an equal decrease in MWS luminance, but the sum is unity. The ordinate $s_{MB} = S/(L+M)$ denotes the level of short-wave sensitive (SWS) cone excitation.

$$l_{MB} = \frac{L}{L+M}; s_{MB} = \frac{S}{L+M} \tag{4-17}$$

In order to scale the ordinate axis, the concept of cone trolands has been introduced. Since the troland is a unit used to express a quantity proportional to retinal illuminance, the amount of L-cone trolands and M-cone trolands indicates the respective contribution of LWS and MWS cone excitations to the retinal illuminance. Since it is assumed that the SWS cones do not contribute to luminance, S-cone troland must be appropriately defined. In the representation proposed by Boynton and Kambe [36], one troland of the equal energy spectrum amounts to one S-cone troland.

In case of CIEPO06 cone fundamentals which are the same as Stockman-Sharpe 10° cone fundamentals [22] each scaled to unity peak, the luminous efficiency function [99] is given by Eq. 4-18. However, as this analysis involves comparing normal and modified cone fundamentals, any normalization must be avoided since it can unduly shift the peak wavelength of modified cone fundamentals, making it difficult to infer whether such shift is due to a physiological factor or because of normalization.

When cone fundamentals are not normalized to unity peak, luminous efficiency function can be obtained by adding LWS and MWS cone fundamentals in 1.98:1 ratio (same ratio as in Eq. 4-18), thus LWS cone fundamentals were scaled by 1.98 to begin with (Eq. 4-19). No scaling was used for SWS cone fundamentals. Next, the product of scaled cone fundamentals and the test stimulus spectral power distribution [$P_{stim}(\lambda)$] obtained from Eq. 4-16 was computed for each wavelength and summed over the whole wavelength range, resulting in LMS tristimulus values in the cone fundamental space

(Eq. 4-20). The resulting tristimulus values were specific to a given display and a given set of modified cone fundamentals, computed from various normal and modified CIEPO06 10° cone fundamentals. Macleod-Boynton chromaticity coordinates (l_{MB}, s_{MB}) were then obtained from LMS tristimulus values as described before (Eq. 2-4).

$$V_{SS,10}(\lambda) = 0.692839\bar{l}(\lambda) + 0.349676\bar{m}(\lambda) \tag{4-18}$$

$$l_{SC}(\lambda) = \bar{l}(\lambda) * 1.98 \tag{4-19}$$

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} l_{SC}(\lambda) \\ m(\lambda) \\ s(\lambda) \end{bmatrix} * P_{stim}(\lambda) \tag{4-20}$$

Again, to comply with the definition of S-cone trolands, Macleod-Boynton s-coordinates (s_{MB}) were scaled such that s-coordinate of equal energy white would be equal to unity. In case of CIEPO06 10° cone fundamentals, the computed scale factor was 21.7209.

The luminance values [Y_{stim}] of the test stimuli were obtained by vectorially adding the peak primary luminance values [Y_{Rmax}, Y_{Gmax}, Y_{Bmax}] scaled by the respective channel values, as shown in Eq. 4-21.

$$Y_{stim} = \begin{bmatrix} R & G & B \end{bmatrix} * \begin{bmatrix} Y_{R\max} \\ Y_{G\max} \\ Y_{B\max} \end{bmatrix} \tag{4-21}$$

Using the above method, relative cone trolands were computed for the seven test stimuli and are plotted in Fig. 4-24.

Using an observer model different from the 10° standard colorimetric observer is likely to distort the uniformity of u'v'Y color space, the extent of which depends on the specific observer model used. However, in this analysis it is hypothesized that in a small region of three-dimensional space around a given color, the Euclidean distances for various observer CMFs can be compared. Because of this issue, use of more complex color space like CIELAB and color difference equations was avoided as they could possibly amplify uniformity distortions. u'v'Y was chosen over xyY because of better visual uniformity.

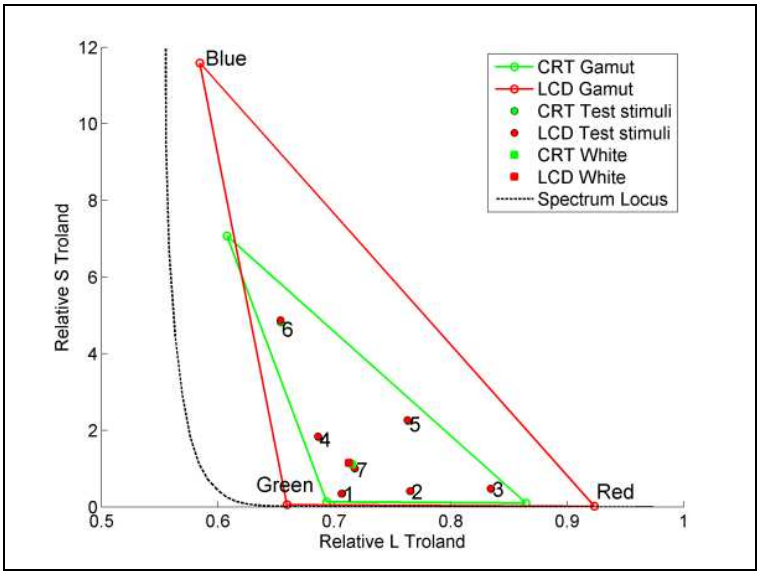


Fig. 4-24. Seven test stimuli in Boynton-Kambe relative cone troland coordinates based on CIEPO06 10° cone fundamentals

4.2.4 *Results*

In Fig. 4-25 the (*u*', *v*') chromaticity shifts of the seven test stimuli are shown, depicting the effects of modified cone fundamentals on chromaticities of the colors on the CRT (green symbols) and the LCD (red symbols). The squares represent a 25% increase in the optical density of the ocular media (Fig. 4-25a), of the macular pigment (Fig. 4-25b) and of the cone photopigment (Fig. 4-25c), and in Fig. 4-25d, a shift of the peak LWS cone wavelength by 4 nm toward shorter wavelengths (see Section 3B). The triangles represent a 25% decrease in the optical density of the ocular media (Fig. 4-25a), of the macular pigment (Fig. 4-25b) and of the cone photopigment (Fig. 4-25c), and in Fig. 4-25d, a shift of the peak MWS cone wavelength by 4 nm toward longer wavelengths. Fig. 4-26 shows the same chromaticity shifts in the cone troland chromaticity diagram. Table 4-2 represents the root-mean-square (RMS) of the (*u*', *v*') coordinate shifts of seven displayed stimuli due to each of the four factors (scaled by 1000). Mean and maximum RMS differences are shown for both displays.

Table 4-2. (u', v') RMS distance (x1000) from average cone fundamental

Source of Variability		RMS (x 1000) [CRT]		RMS (x 1000) [LCD]	
		Mean	**Max**	**Mean**	**Max**
Ocular Media Peak Optical Density	0.25%	10.25	16.45	9.40	14.53
	-0.25%	11.28	17.79	9.68	14.86
Macular Pigment Peak Optical Density	0.25%	2.93	4.69	3.25	5.04
	-0.25%	2.96	4.72	3.29	5.08
Photopigment Peak Optical Density	0.25%	13.51	26.59	13.85	25.60
	-0.25%	20.23	36.57	20.31	35.47
Photopigment Peak Wavelength Shift	L - 4 nm	8.42	19.97	10.00	22.50
	M + 4 nm	8.72	20.01	5.84	15.09

All four factors do not affect the target specification to the same extent. Out of all four factors, photopigment peak optical density affects the observer color perception the most, as evident from Table 4-2. In case of ocular media and macular pigment absorption, the change in color perception occurs along the same direction: toward yellow-green when the optical density is increased and toward blue when it is decreased (Figs. 4-25a and 4-25b). This is true even for the test stimulus close to the display white. These directions of change are in line with Wyszecki and Stiles' results [9] (pp 352). However, the effect of macular pigment absorption is significantly less than ocular media absorption, in fact it is the least significant physiological factor when compared to the others, as per Table 4-2. The change due to macular pigment absorption is marginally larger for LCD as compared to the CRT, and is the opposite in case of ocular media absorption. That ocular media optical density plays a dominant role in observer variability, even within the same age-group, has been reported by several vision researchers. Pokorny et al [50] observe that "*studies which include a large number of observers of similar age indicate that there is considerable variability in estimated lens density at any given age. For example, van Norren and Vos noted that the difference between the five highest and five lowest of Crawford's 12 observers was greater than one log unit at 400 nm. This variation may be even more pronounced in an older group of subjects*". Note that in terms of cone excitation, largest change due to modification of ocular media and macular pigment absorption occurs for the blue color (test stimulus 6).

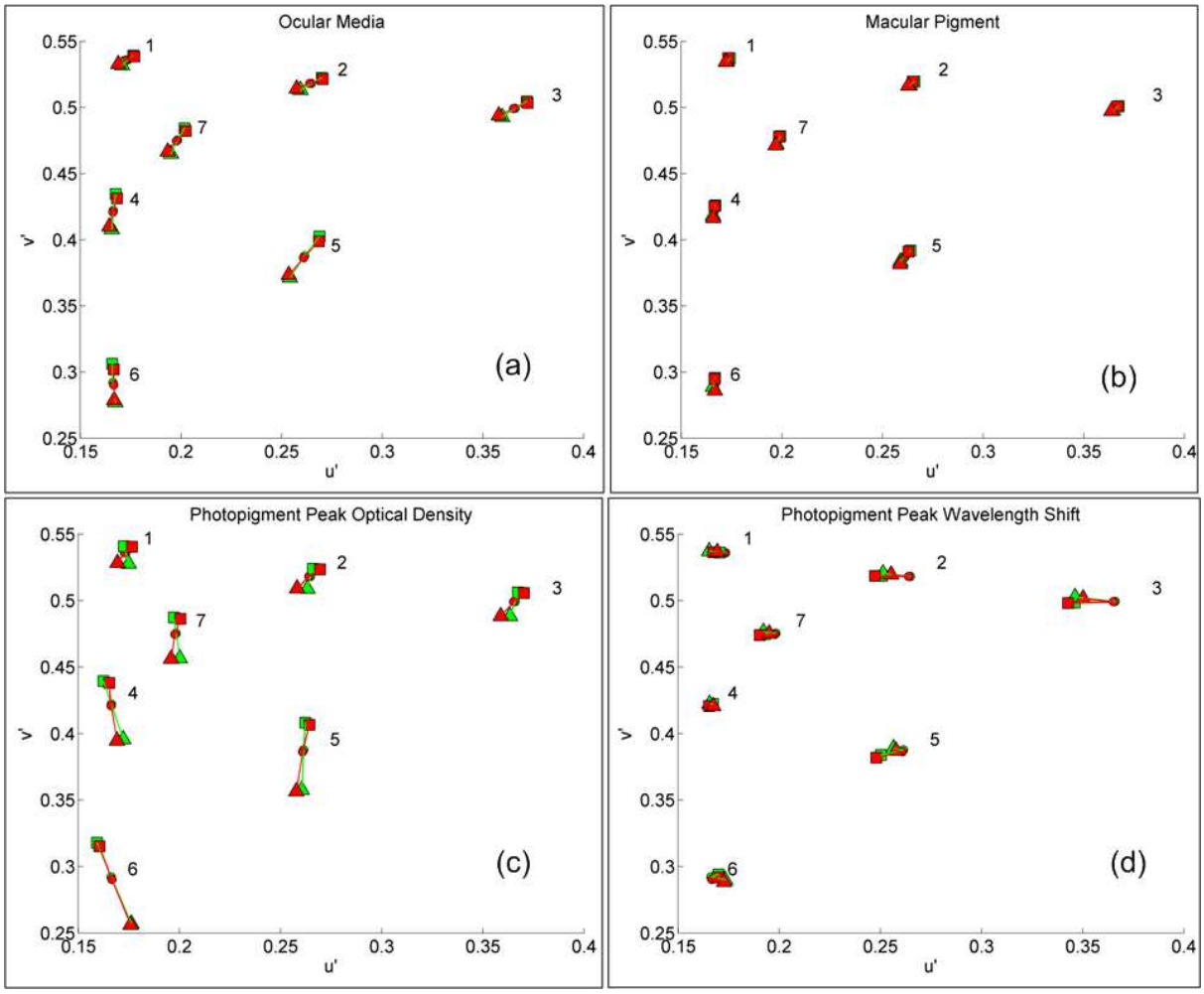


Fig. 4-25. Simulated chromaticity shift for seven test stimuli due to modified cone fundamentals in (u',v') chromaticity diagram. Increase (squares) and decrease (triangles) of the peak optical density by 25% are shown for ocular media (a), for macular pigment (b) and for photopigment peak optical density (c). Peak wavelength shift of LWS cone photopigment by 4 nm toward shorter wavelengths (squares) and of MWS cone photopigment by 4 nm toward longer wavelengths (triangles) are shown in (d). Green symbols correspond to the CRT and red symbols to LCD.

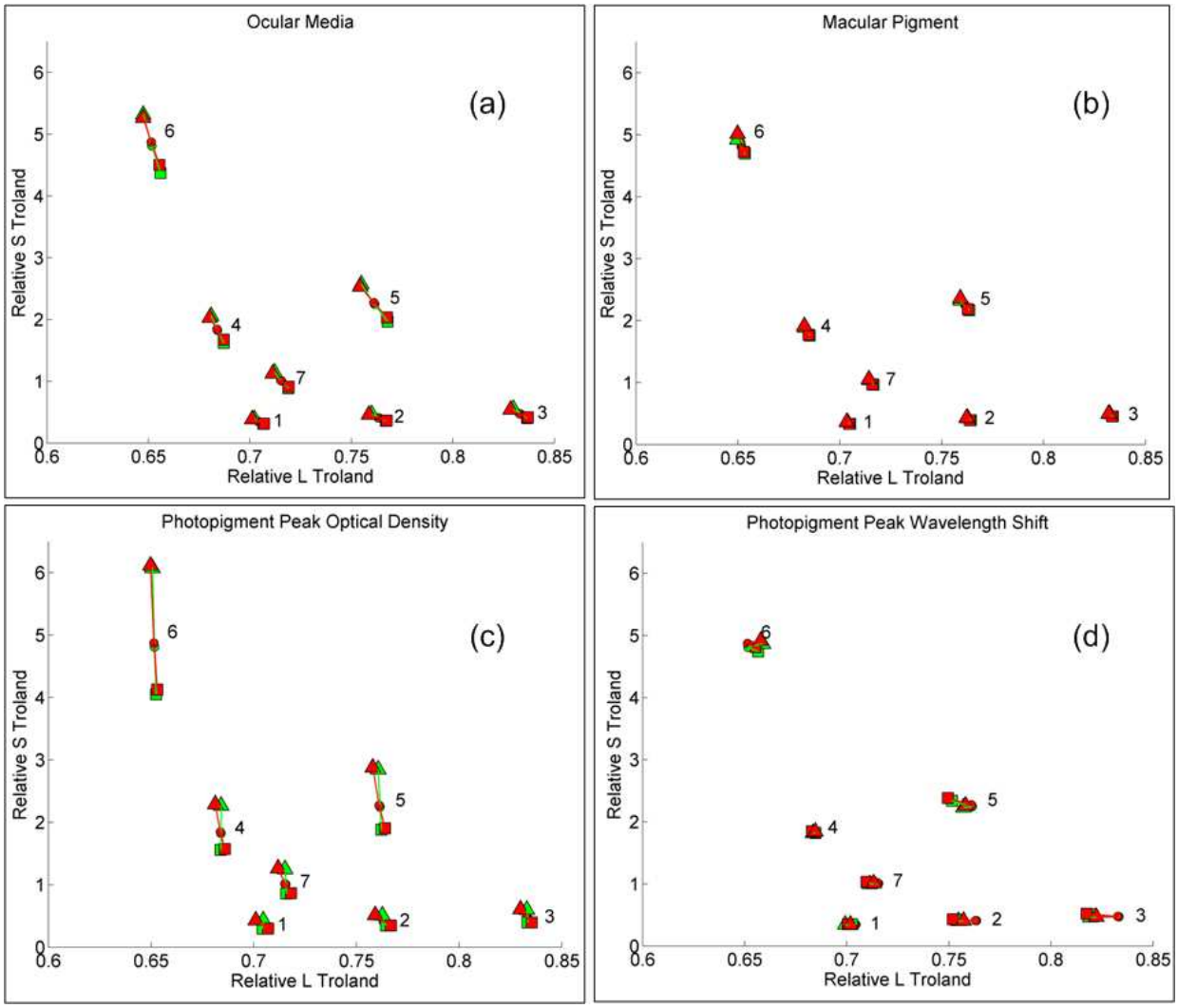


Fig. 4-26. Simulated chromaticity shift for seven test stimuli due to modified cone fundamentals in relative cone troland space. Increase (squares) and decrease (triangles) of the peak optical density by 25% are shown for ocular media (a), for macular pigment (b) and for photopigment peak optical density (c). Peak wavelength shift of LWS cone photopigment by 4 nm toward shorter wavelengths (squares) and of MWS cone photopigment by 4 nm toward longer wavelengths (triangles) are shown in (d). Green symbols correspond to the CRT and red symbols to LCD.

Finally, in case of CRT, the effect of photopigment peak wavelength shift is as large as that of ocular media absorption, particularly the LWS cone shift. In case of LCD, the LWS cone peak wavelength shift is by far the second most important factor in influencing display color perception, after photopigment peak optical density.

4.2.5 *Analysis of results*

Interestingly, the direction of change due to the modification of the photopigment peak optical density is different for the CRT and the LCD, both in terms of (u', v') chromaticity coordinates (Fig. 4-25c) and relative cone trolands (Fig. 4-26c). This difference is more apparent in green-red region of color

space and reduces as we go toward blue. We can assume that relative position of display primaries with respect to the cone fundamentals has an influence on such difference in directional effects between the two displays. However, other physiological factors do not show such trend. Another observation is that for the reddish-yellow (test stimulus #2), red (test stimulus #3) and magenta (test stimulus #5), the directions of change due to LWS and MWS peak wavelength shifts (Figs. 5d and 6d) are the same. An explanation of this observation is that the LWS and MWS peaks move toward each other. For other stimuli, the effect of peak wavelength shifts is not significant

Since the photopigment peak optical density has the strongest influence in display color perception compared to other factors, and since largest chromaticity shift due to this factor occurs in blue, we can assume that individual variations in the color vision of a large population of real observers will have a significant impact on the perception of blue.

This analysis also shows that the photopigment peak wavelength shift is an important physiological factor affecting display color perception, particularly in case of modern displays with narrow-band primaries (Table 4-2). The difficulty in modeling this factor imposes serious limitation on the age-dependent observers of CIEPO06. Observer variability within a given age-group due to such factors cannot be predicted, even though this variability can be more significant than the effects of some of the factors already included in the model.

This analysis has some inevitable constraints. It is difficult to predict the extent to which various physiological factors affect the color perception of an individual observer. It is also difficult to ascertain what amount of peak wavelength shift should physiologically correspond to a 25% change in peak optical densities. A peak wavelength shift of 4 nm was assumed since this is the largest shift observed due to the serine-alanine amino acid substitution at position 180 of the photopigment opsin genes, a common form of polymorphism [24]. It is hypothesized that the conditions analyzed here all represent extreme changes in four physiological factors, and thus are reasonable to compare. In spite of the above constraint, this analysis highlights the relative importance of various factors in affecting color perception on displays.

4.3 Intra-age group average observer prediction with CIEPO06 model and the CIE 10° standard colorimetric observer

As already suggested in the past, question arises whether it could be worthwhile to explore if the observed inter-subject differences in color matches could be predicted by adjustment of more of the CIEPO06 parameters. In this study, experimental data from the 1959 Stiles-Burch study [44] involving 47 observers were re-examined, since this is the most comprehensive visual dataset for color vision available to date.

4.3.1 *CIEPO06 Age Parameters for Real Observers*

The age parameter was introduced in the CIEPO06 model to take into account the difference in absorption in the ocular media, in particular the lens, between the aged and the young observers. At this time, the age dependencies of the absorption by the macular pigment as well as the densities of the visual pigments were considered of minor influence. The two-component age function of the CIEPO06 model originated from several experimental bases which were thought to be representative of large groups of observer [50]. Thus, the CIEPO06 age parameter does not necessarily correspond to the age of the real Stiles-Burch observers. 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 random observer variability, and/or because of the exclusion of one or more physiological factors from the CIEPO06 model. These factors could be age-independent, like the peak wavelength shift of the LWS or MWS cone photopigment as discussed earlier, or these could be age-dependent physiological factors not considered in CIEPO06. CIE committee TC 1-36 also recognized this restriction by pointing out that CIEPO06 fundamental observer was a theoretical construct [14]. In this analysis, the CIEPO06 age parameters that resulted in the best predictions of each individual Stiles-Burch observer cone fundamental data were determined. For each individual Stiles-Burch observer, three CIEPO06 age parameters were derived so as to fit as closely as possible the three cone fundamentals, respectively. Two different methods were used. In the first method, the correlation coefficients were computed between the normalized cone fundamentals for each Stiles-Burch observer, using Eq. 4-11 as explained in Section 4.1.4, and those corresponding to all possible CIEPO06 age parameter values between 20 and 80 (a total of 61). The corresponding CIEPO06 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 Stiles-Burch observers. In the second method, corresponding CIEPO06 age for each Stiles-Burch observer was predicted by minimizing the RMS errors between the normalized cone fundamentals for each Stiles-Burch observer, and those corresponding to all possible CIEPO06 age parameter values between 20 and 80.

4.3.2 *Comparison of CIEPO06 predicted and real ages of Stiles-Burch observers*

In Fig. 4-27, the CIEPO06 predicted ages obtained using the correlation coefficient (CORR) method have been plotted against the actual ages of 47 Stiles-Burch observers. The second method (RMSE) produced very similar results. No direct correspondence was found between the real and predicted ages.

The gain offered by the adjusted CIEPO06 age over the real age could be validated by examining the prediction of matches of equal-energy white. Fig. 4-28 shows (x, y) chromaticity of equal-energy white computed with CMFs derived from CIEPO06 cone fundamentals for each Stiles-Burch

observer. CIEPO06 cone fundamentals were obtained by using corresponding ages from both methods (CORR and RMSE) as well as by using actual observer ages. Matches obtained with real observer cone fundamentals are also plotted. While CIEPO06 with age correspondence (with either method) yields greater observer variability than CIEPO06 with actual observer ages, it fails to explain all the variability in the real observer data, particularly along the ordinate.

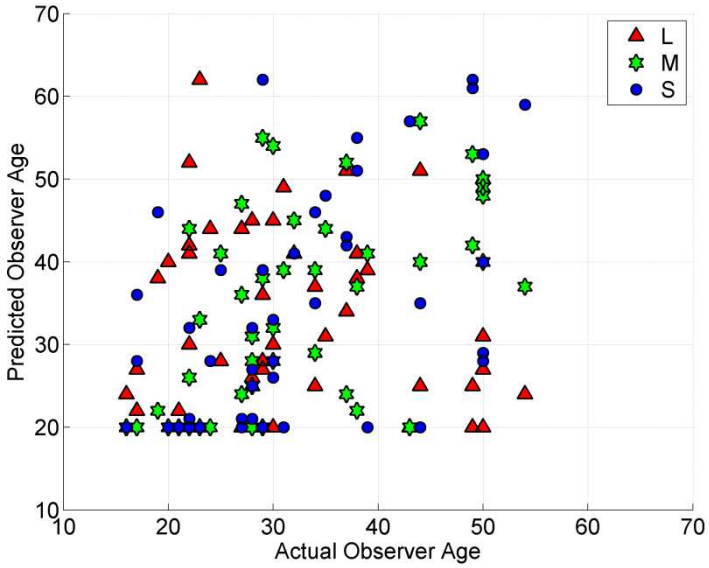


Fig. 4-27. Age correspondence between CIEPO06 model's best prediction and 47 Stiles-Burch observers

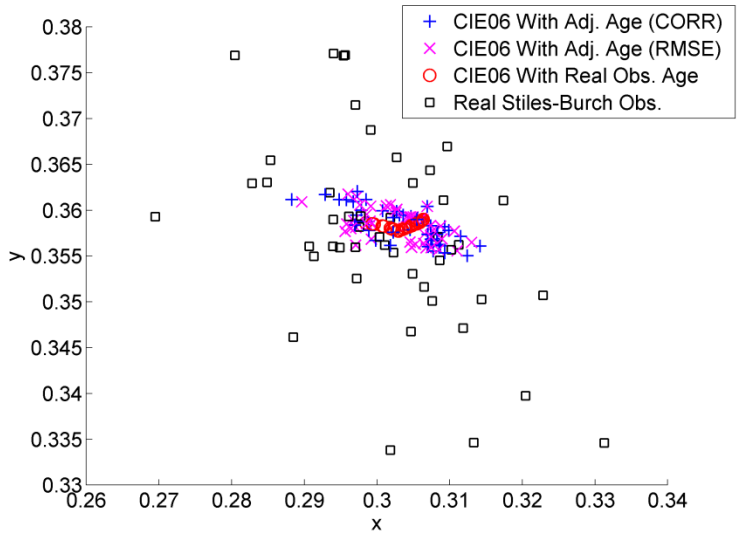


Fig. 4-28. Chromaticities of matches of equal-energy white, computed using cone fundamentals from the 47 Stiles-Burch observer data and CIEPO06 predictions, with two adjustment methods for age (CORR and RMSE) as well as with actual observer age

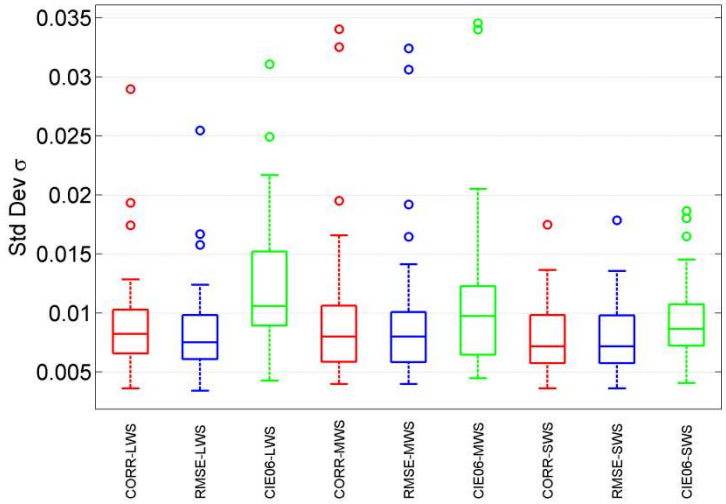


Fig. 4-29. Mean standard deviation of CIEPO06 cone fundamentals from the 47 Stiles-Burch observer data, with two adjustment methods for age (CORR and RMSE) as well as with actual observer age. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points that are not considered outliers, while outliers are plotted individually as small circles.

The mean standard deviations of the CIEPO06 cone fundamentals from the 47 Stiles-Burch observer data averaged over all observers are plotted in Fig. 4-29. The LWS, MWS and SWS cone fundamentals obtained by using corresponding ages from the two methods (CORR and RMSE) and by using actual observer ages are shown. Mean (central mark), as well as the 25th and 75th percentiles (dotted bars) of standard deviations are higher when real observer ages are used in the model. The error is higher for LWS and MWS cone fundamentals than for SWS cone fundamental. This further shows that by adjusting the age parameter, the CIEPO06 prediction of real Stiles-Burch observer data is improved.

4.3.3 *Grouping Stiles-Burch Observers with respect to age*

To conform to the age-dependent observer model of CIEPO06, three dominant age-groups among the Stiles-Burch observers were identified. The groups were formed in such a way that the age difference between observers within any group was not more than two years. This constraint allowed grouping of only 22 out of 47 observers. Six observers with ages between 22 and 23 formed Group-1, ten observers with ages between 27 and 29 consisted Group-2 and another six observers with ages between 49 and 50 were placed in Group-3. In the rest of the analysis, these three observer groups are used. For each group, CIEPO06 age correspondence for the average data was established using the correlation coefficient method for the average Stiles-Burch cone fundamentals for the group and CIEPO06 cone fundamentals for all possible ages. In the following sections, two sets of CIEPO06

CMFs for each observer group were obtained, CIEPO06 CMFs obtained by using adjusted age parameter values given by the correlation coefficient method, and CIEPO06 CMFs obtained by using actual average observer ages.

4.3.4 *Comparing CIEPO06 Model Prediction and 10° Standard Colorimetric Observer with Intra-Group Average*

Once three groups of observers were identified, the variability of CMFs was examined within each group. The examination put more emphasis on the regions of the spectrum where $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ peak. In Fig. 4-30, intra-group minimum, maximum and average CMF values are shown along with the 10° standard colorimetric observer CMFs, the CIEPO06 model predictions, with age correspondence and with real ages. Table 4-3 lists the results of a statistical comparison of the Stiles-Burch observer CMFs, 10° standard colorimetric observer and CIEPO06 model predictions with age correspondence and with real ages. Values corresponding to $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ functions, in the corresponding long-, medium- and short- wavelength ranges for each group are shown. The 3rd column in Table 4-3 shows the intra-group standard deviation of the Stiles-Burch data (note that standard deviation has the same units as the data), signifying intra-group observer variability. Following three columns list absolute difference of various functions from the intra-group mean, averaged over all wavelengths. The three functions considered here are i) 10° standard colorimetric observer, ii) CIEPO06 with real observer ages as input, and iii) adjusted CIEPO06 ages with age correspondence as input. The absolute differences of the functions were multiplied by three weighting functions (for LWS, MWS and SWS respectively) before averaging over all wavelengths. The weighting functions were computed by dividing the three intra-group average Stiles-Burch observer CMFs by their respective sum over all wavelengths. The role of the weighting functions was to assign more weights to the values around the peak than those in the lower end of the ordinate, while ensuring the weights were proportional to original observer data. Note that since the $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, $\bar{z}_{10}(\lambda)$ CMFs do not have the same ordinate scale, the rows should not be compared as such.

As shown in Fig. 4-30 and Table 4-3, in case of x-CMFs for Group-1 and -3, both original CIEPO06 model predictions with real ages and 10° standard colorimetric observer deviate from the intra-group average. CIEPO06 model with real observer ages generally performs similar to or worse than the 10° standard colorimetric observer $\bar{x}_{10}(\lambda)$ and $\bar{y}_{10}(\lambda)$ CMFs. For Group-1 and -3, the age correspondence method mostly improves CIEPO06 predictions, and is mostly better than the standard colorimetric observer. For Group-2, the prediction error is relatively low even without age

correspondence, indicating CIEPO06 model's age parameter works well for the age group of 27-29. This is not surprising since the average observer age in the Stiles-Burch study, on which CIEPO06 is based, was 32. For Group-3 concerning aged observers, CIEPO06 performs worse than the standard colorimetric observer for $\bar{x}_{10}(\lambda)$ and $\bar{y}_{10}(\lambda)$ CMFs. The errors in the original model prediction are comparable to the intra-group standard deviation, indicating that the prediction errors are statistically significant.

Table 4-3. Deviations of CMF data from intra-group average Stiles-Burch observer, 10° standard colorimetric observer and CIEPO06 model predictions with age correspondence and with real ages

CMF	Grp. No.	Mean Intra-group Stiles-Burch Std. Dev.	Mean Scaled Abs. Diff. From Mean Intra-group Stiles-Burch Data		
			CIE 10° Std. Col. Obs.	CIEPO06	
				Model with Real Ages	Model with Age Corres.
$\bar{x}(\lambda)$	1	10.11	5.68	6.53	2.51
	2	11.28	2.54	1.74	1.99
	3	9.12	9.93	10.58	6.06
$\bar{y}(\lambda)$	1	6.02	2.81	4.73	1.13
	2	6.68	2.28	2.42	2.43
	3	5.41	2.12	4.21	2.5
$\bar{z}(\lambda)$	1	22.7	19.25	8.22	7.55
	2	25.54	10.88	6.2	6.17
	3	21.43	11.71	5.21	3.99

As far as the $\bar{z}_{10}(\lambda)$ CMF is concerned, the CIEPO06 model produces markedly better results compared to the CIE 10° standard colorimetric observer, even without age correspondence. On an average, the reduction in mean absolute difference is more than 50%. $\bar{z}_{10}(\lambda)$ CMF also shows high standard deviation compared to $\bar{x}_{10}(\lambda)$ and $\bar{y}_{10}(\lambda)$, indicating that the high prediction error of the standard colorimetric observer is, at least partially, due to observers having short wavelength cone sensitivity significantly different from the average. As explained in Section 4.2.4, there is high variability in ocular media optical density among observers, which is more pronounced among higher age-group observers [50]. Presumably, this variability will manifest more significantly in the blue region of color space. It is logical to hypothesize that in the process of averaging over whole population of all ages, observers significantly different from the majority unduly affect the average. Within the constraints of

current analysis, CIEPO06 seems to offer an improvement over the 10° standard colorimetric observer in predicting intra-age group average z-functions.

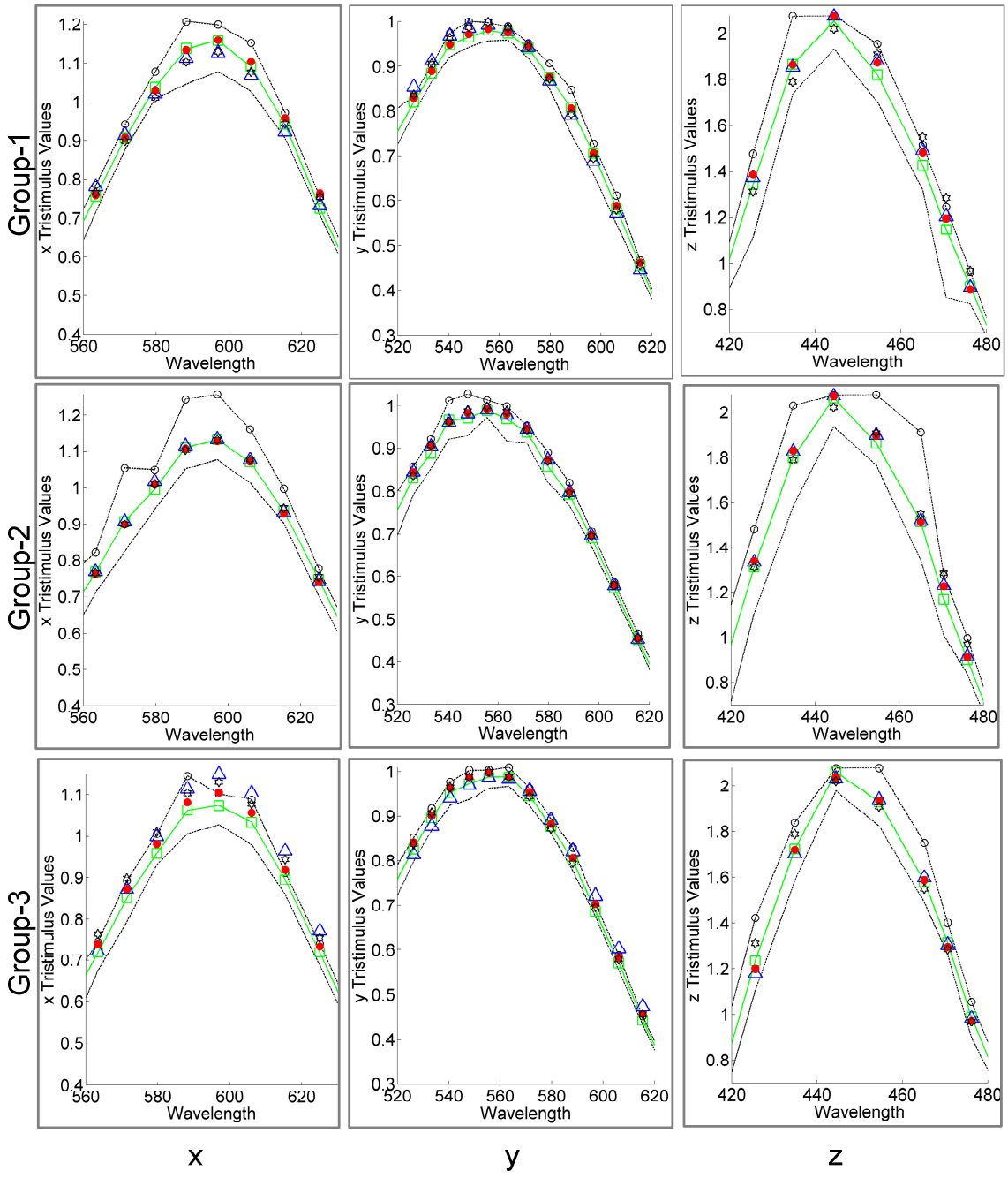


Fig. 4-30. CMFs for the Stiles-Burch intra-group average observer (green line with squares), CIEPO06 model predictions (blue triangles), CIEPO06 model predictions with age correspondence (red filled circles) and CIE 10° standard colorimetric observer (black star) for Group-1 (top row), -2 (middle row) and -3 (bottom row). Stiles-Burch Observers' intra-group minimum (black line) and maximum (black line with circles) are also shown. Each plot shows the CMFs around the peak only.

4.4 Display colorimetry: comparison of CIEPO06 CMFs and the CIE 10° standard colorimetric observer

Any statistical method used to compare the model predictions with real observer data is incomplete without an analysis of the perceptual effect of the prediction errors. Thus, an additional analysis was performed to simulate the effect of the deviations of CIEPO06 model predictions and the CIE 10° standard colorimetric observer from the average intra-group observer data on display color perception. The same method of computation of (u', v') tristimulus values for the seven test stimuli was followed as was used for analyzing the effect of various psychological factors, described earlier. The only difference in this case is in the last step. The spectral power distributions of the test stimuli, obtained from the *channel values* and the spectral data of the display primaries, were integrated with either the CIEPO06 CMFs with age correspondence, or the CIEPO06 CMFs with real ages, or the CIE 10° standard colorimetric observer to obtain the (u', v') specification. (u', v') RMS distances were computed between coordinates corresponding to Stiles-Burch intra-group average ($u'_{av,SB}$, $v'_{av,SB}$) and those corresponding to various model predictions (u'_{pred}, v'_{pred}), as shown in Eq. 4-22. In this equation, the distances are normalized by ($u'_{av,SB}$, $v'_{av,SB}$), the coordinates for Stiles-Burch intra-group average data. Such normalization allows us a comparison of relative magnitudes of various distances.

$$rms = 100 \cdot \sqrt{\left(\frac{u'_{pred} - u'_{av,SB}}{u'_{av,SB}} \right)^2 + \left(\frac{v'_{pred} - v'_{av,SB}}{v'_{av,SB}} \right)^2} \qquad (4\text{-}22)$$

Table 4-4 lists these normalized distances computed for the LCD. For the CRT, the RMS distance differences between chromaticities predicted by the CIE 10° standard colorimetric observer and CIEPO06 model were less apparent and are not shown. Note that all these distances are computational color differences between actual and model-predicted chromaticities, and simply help us compare model prediction errors in a perceptual space. The distances in different parts of the color space are not comparable since the (u', v') space is not perceptually uniform, but small distances corresponding to various CMFs can be compared. So the values in the Table 4-4 should be compared row-wise, and not column-wise.

The shaded entries in Table 4-4 represent the cases where the original CIEPO06 model with real ages predicted the intra-group averages better than those of the CIE 10° standard colorimetric observer (lighter shade), as well as cases where CIEPO06 model with age correspondence predicted the intra-group averages better than the original CIEPO06 model (darker shade). While for Group -1 and -3, original CIEPO06 model predictions are generally better than the CIE 10° standard colorimetric

observer, the model mostly performs worse in case of Group-2. Applying the age correspondence generally improves the model prediction in case of Group-1 and -2. For Group-3 however, age correspondence mostly degrades the original model prediction quite significantly. This shows reducing overall RMS error in the cone fundamental or tristimulus space does not necessarily result in improved prediction of color perception in a chromaticity space. Another possible explanation is that the observer variability in higher age-group observers is not well modeled in CIEPO06 (see Fig. 4-30), thus intra-group average prediction is adversely affected by the poor prediction of color matches for observers significantly different from the average.

Table 4-4. (u', v') normalized RMS distances (x100) of predicted chromaticity values from Stiles-Burch intra-group average CMFs, computed for seven test stimuli as viewed on the LCD. Predicted chromaticity values were obtained using CIE 10° standard colorimetric observer CMFs, CIEPO06 model CMFs with real ages and CIEPO06 model CMFs with age correspondence. Shaded values indicate improvement in the prediction of chromaticities corresponding to intra-group average CMFs, either by the CIEPO06 original model compared to the CIE 10° standard colorimetric observer (lighter shade), or by the CIEPO06 model with age correspondence compared to the original CIEPO06 model (darker shade).

Test Stim.	Group-1			Group-2			Group-3		
	CIE 10° Std Col. Obs	CIEPO06 with real ages	CIEPO06 With Age Corres.	CIE 10° Std Col. Obs	CIEPO06 with real ages	CIEPO06 With Age Corres.	CIE 10° Std Col. Obs	CIEPO06 with real ages	CIEPO06 With Age Corres.
TS-1	4.52	5.23	2.00	2.89	2.80	2.43	1.79	0.41	3.70
TS-2	2.36	1.85	1.90	1.47	1.84	1.74	1.77	1.79	1.83
TS-3	1.11	0.49	1.40	0.82	1.16	1.22	1.46	1.89	0.79
TS-4	4.19	4.34	0.61	2.68	2.72	1.84	0.81	0.63	4.62
TS-5	1.97	0.81	0.92	1.29	1.91	1.42	1.30	2.15	2.59
TS-6	3.54	3.13	1.15	2.27	2.64	1.23	1.80	1.47	5.42
TS-7	3.51	3.35	1.48	2.22	2.43	1.96	1.37	0.68	3.35

Now, how could we correlate the observations from Table 4-3 (see Section 4.3.4) and Table 4-4? Note that Table 4-3 lists scaled prediction errors around the peak regions of individual x-, y- and z- CMFs, while Table 4-4 lists normalized RMS distances in predicting several test stimuli reproduced on the LCD in two-dimensional (u', v') chromaticity space. Although it is not surprising that the observations are not always congruent with each other, two inferences can be drawn by taking into account results from both analyses.

Overall, the CIEPO06 model in its original form does not always offer an improvement over the 10° standard colorimetric observer in predicting intra-age group average observer data. Using values different from actual observer ages in CIEPO06 model can achieve better overall correlation between actual and model predicted CMFs in the tristimulus or cone fundamental space, but does not

necessarily result in improved prediction of individual color matches, particularly when the stimuli do not have a flat spectral characteristics. While the short wavelength CIEPO06 CMFs consistently perform better than the 10° standard colorimetric observer for all three age-groups, the model's prediction errors in medium and long wavelengths are significantly higher for Group-3. Why does the model not work well for higher age-group Stiles-Burch observers at longer wavelengths? This issue is further investigated in the next section.

4.5 Optimized CIEPO06 cone fundamentals for Stiles-Burch observer groups

As discussed in the preceding sections, CIEPO06 model does not satisfactorily predict the intra-group Stiles-Burch average observer color-matching functions $\bar{x}_{10}(\lambda)$ and $\bar{y}_{10}(\lambda)$ in the long- and medium- wavelength range, particularly for higher age-group observers. These observations thus raise the question: can we improve the model performance in the longer wavelengths? If so, how can we achieve that?

This prediction error can result from many potential sources. For example, it could be due to individual observer's LWS or MWS photopigment peak wavelength shift resulting from genetic polymorphism (as discussed in Section 4.2), or it could be due to poor modeling of cone absorptance spectra in longer wavelengths. As far as the prediction error at higher wavelengths is concerned, we can rule out the role of ocular media and macular pigment optical density factors, since their influences are insignificant beyond 550 nm. Note that ocular media optical density is the only physiological factor in CIEPO06 model that changes with age. To probe possible ways to improve the CIEPO06 model prediction at higher wavelengths, a constrained nonlinear optimization was performed under two different conditions.

4.5.1 *Method of optimization*

In the first case, only the peak wavelength shifts of the LWS or MWS photopigments were allowed to vary, keeping all other parameters constant. In the second case, a weighting function for the low-optical density absorption spectra was introduced, which was then optimized. In both cases, the original CIEPO06 functions at the short-wavelengths were not altered.

The equations for CIEPO06 cone fundamentals were introduced in Eq. 4-13. In the first optimization, only the peak wavelength of $A_l(\lambda)$ and $A_m(\lambda)$ functions were allowed to shift, keeping all other parameters constant. The optimized cone fundamentals can thus be represented by Eq. 4-23:

$$\bar{l}_{opt}(\lambda) = \left[1-10^{-D_{vis,l}\cdot A_{shifted,l}(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)}$$
$$\bar{m}_{opt}(\lambda) = \left[1-10^{-D_{vis,m}\cdot A_{shifted,m}(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)} \quad (4\text{-}23)$$

Here, $[D_{vis,l}A_{shifted,l}(\lambda)]$ and $[D_{vis,m}A_{shifted,m}(\lambda)]$ terms are mathematical representation of the peak wavelength shift due to polymorphism. In the actual implementation of the optimization method, the peak wavelength λ was first shifted in the wavenumber scale ($\nu = 10^7/\lambda$, where ν is in cm^{-1} and λ is in nm) independently for LWS and MWS photopigment, next the cone absorptance spectra were re-sampled, then modified cone fundamentals were computed and converted from the quanta to energy units, and finally were renormalized. In the objective function, the Root Mean Square Error (RMSE) over the whole wavelength range was computed between the modified CIEPO06 cone fundamentals and Stiles-Burch intra-group average cone fundamentals, and was minimized iteratively by changing the amount of peak λ shift. This shift was constrained between +250 and -250 cm^{-1}, with a starting value of 100 cm^{-1}. Thus, the optimization process left the contributions of macular pigment and ocular media unaltered; only the contributions of LWS and MWS cone absorption spectra were changed. The SWS cone fundamental was not modified. The optimization was terminated after 10000 iterations, or below an error of 10^{-6}, whichever was earlier.

In the second optimization, weighting functions $w_l(\lambda)$ and $w_m(\lambda)$ for the low-optical density spectral absorbance terms $A_l(\lambda)$ and $A_m(\lambda)$ respectively were introduced beyond 550 nm [Eq. 4-24]. As before, the SWS cone fundamental was unaltered.

$$\bar{l}_{opt}(\lambda) = \left[1-10^{-D_{vis,l}\cdot A_l(\lambda)\cdot w_l(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)}$$
$$\bar{m}_{opt}(\lambda) = \left[1-10^{-D_{vis,m}\cdot A_m(\lambda)\cdot w_m(\lambda)}\right]\cdot 10^{-D_{mac}(\lambda)}\cdot 10^{-D_{ocul}(\lambda)} \quad (4\text{-}24)$$

While some authors have already questioned the CIEPO06 SWS cone fundamental at short wavelengths [117], for the current work, there are two reasons for restricting optimization above 550 nm. Firstly, we are primarily interested in reducing prediction errors at higher wavelengths. Secondly, the ocular media and macular pigment optical densities have significant contributions to the cone fundamentals below 550 nm. Thus, even if we introduce a weighting function below 550 nm and obtain better results, it is difficult to isolate a single physiological factor as the source of prediction error.

As in the first optimization, the Root Mean Square Error (RMSE) between the modified CIEPO06 cone fundamentals and Stiles-Burch intra-group average cone fundamentals were minimized in the objective function.

4.5.2 *Results*

As a result of the first optimization, the peak wavelengths of $A_l(\lambda)$ and $A_m(\lambda)$ functions were shifted differently for different groups. For Group-1, LWS and MWS peak wavelength shifts were 3.6 nm and 1.3nm respectively, both toward shorter wavelengths. For Group-2, only the LWS function was shifted by 0.1 nm toward shorter wavelengths. For Group-3 on the other hand, the shifts were toward longer wavelengths, 4.1 nm and 0.3 nm for respectively LWS and MWS functions.

The second optimization resulted in different LWS and MWS weighting functions for the three groups. These functions are shown in Fig. 4-31. The optimized function is obtained by multiplying the original CIEPO06 model function by the respective weighting function. Thus a weighting of unity does not affect the original model function. As shown in Fig. 4-31, the LWS weighting functions have higher values than those of MWS cones. What is interesting is that for both LWS and MWS, the weighting functions for Group-1 and -3 are somewhat symmetrical around the unity weights. To remind the reader, these two groups consist of younger (22-23 years) and older (49-50 years) observers respectively, while Group-2 observers have average age in the middle (27-29 years). For higher age group observers, peak optical density is reduced by the optimization process, and is increased for the lower-age group.

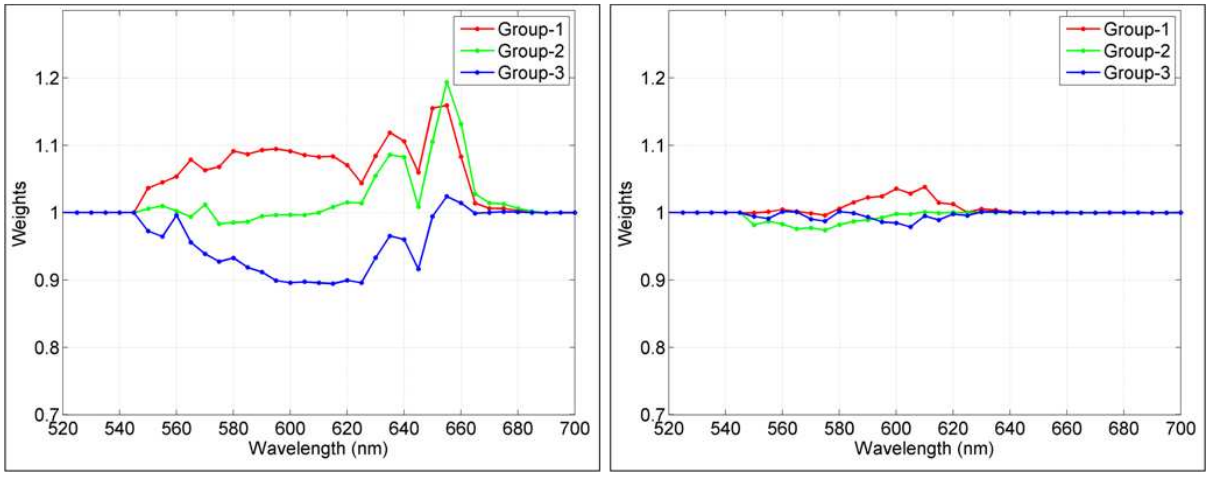


Fig. 4-31. Weighting functions for optimizing the LWS (left) and MWS (right) low density spectral absorbance. Optimization was performed above 550 nm.

Results of both optimization processes are incorporated in Table 4-5, introduced earlier in Section 3D (see Table 4-3). Both $\bar{x}(\lambda)$ and $\bar{y}(\lambda)$ intra-age group average color-matching functions of Stiles-Burch observers of Group-1 and -3 are better predicted by the optimized model.

Table 4-5. Comparison of deviations of CMF data from intra-group average Stiles-Burch observer, 10° standard colorimetric observer, CIEPO06 original model predictions and optimized CIEPO06 model with modified low density absorbance spectra

CMF	Group No.	Mean Intra-group Stiles-Burch Std. Dev.	Mean Scaled Abs. Diff. From Mean Intra-group Stiles-Burch Data		
			CIE 10° Std. Obs.	CIEPO06 Model	
				Original	Optimized (Low Density Abs. Spectra)
$\bar{x}(\lambda)$	1	10.11	5.68	6.53	2.01
	2	11.28	2.54	1.74	2.17
	3	9.12	9.93	10.58	2.01
$\bar{y}(\lambda)$	1	6.02	2.81	4.73	1.34
	2	6.68	2.28	2.42	1.4
	3	5.41	2.12	4.21	1.12

The improvement in model performance is also substantiated in Table 4-6. The shaded entries in Table 4-6 represent the cases where the original CIEPO06 model with real ages predicted the intra-group average data better than those of the CIE 10° standard colorimetric observer (lighter shade), as well as cases where the optimized CIEPO06 model predicted the intra-group averages better than the original CIEPO06 model as well as the CIE 10° standard colorimetric observer (darker shade). These values were computed in the same way as described in Section 4.4. Overall, the peak wavelength shift optimization did not lead to better prediction of average data. An effect of polymorphism on the average data is not apparent in any of the three groups. This supports Webster's conclusion [131] that no polymorphism effect among the Stiles-Burch observers could be confirmed. However, this depends on the observer group involved in the study. Viénot [128] showed that a shift in the wavelength of peak sensitivity of the cone photopigments could account for the variability in multiple Rayleigh matches from color normal observers.

In the analysis reported here, significant improvement was achieved by optimizing the low-density photopigment spectral absorbance functions for Group-1 and -3. On an average, for Group-1, the average RMS prediction error for the seven stimuli reduced by more than 70% as compared to the CIE 10° standard colorimetric observer, while for Group-3, the improvement was around 45%. Only exception is the blue test color (TS-6) for group-3, which in any case does not have significant spectral power in the wavelengths beyond 550 nm. In general in case of Group-2, the optimization did not improve the results.

Table 4-6. (*u'*, *v'*) normalized RMS distances (x100) from Stiles-Burch intra-group average chromaticities computed for seven test stimuli as viewed on the LCD. Results are shown for the CIE 10° standard colorimetric observer, original CIEPO06 model predictions and optimized model prediction through peak wavelength shift and weighted low-density photopigment spectral absorbance. Shaded values indicate improvement in the prediction of chromaticities corresponding to intra-group average CMFs, either by the CIEPO06 original model compared to the CIE 10° standard colorimetric observer (lighter shade), or by the optimized CIEPO06 model compared to both original CIEPO06 model and CIE 10° standard colorimetric observer (darker shade).

Test Stim.	Group-1				Group-2				Group-3			
	CIE 10° Std. Col. Obs.	CIEPO06	CIEPO06 Optim. (Peak λ Shift)	CIEPO06 Optim. (LD Abs Spectra)	CIE 10° Std. Col. Obs.	CIEPO06	CIEPO06 Optim. (Peak λ Shift)	CIEPO06 Optim. (LD Abs Spectra)	CIE 10° Std. Col. Obs.	CIEPO06	CIEPO06 Optim. (Peak λ Shift)	CIEPO06 Optim. (LD Abs Spectra)
TS-1	4.52	5.23	5.21	0.58	2.89	2.80	3.54	2.96	1.79	0.41	2.18	0.11
TS-2	2.36	1.85	3.88	0.25	1.47	1.84	2.23	1.12	1.77	1.79	2.23	0.26
TS-3	1.11	0.49	2.86	0.72	0.82	1.16	1.08	0.11	1.46	1.89	1.60	0.46
TS-4	4.19	4.34	3.99	0.86	2.68	2.72	3.06	2.84	0.81	0.63	0.97	0.73
TS-5	1.97	0.81	2.70	0.77	1.29	1.91	2.25	1.30	1.30	2.15	1.78	1.10
TS-6	3.54	3.13	2.47	0.86	2.27	2.64	2.36	2.43	1.80	1.47	2.16	2.36
TS-7	3.51	3.35	4.18	0.52	2.22	2.43	2.95	2.23	1.37	0.68	1.76	0.36

4.5.3 *Discussion*

Foregoing discussion leads to a hypothesis that a major source of the CIEPO06 model prediction errors at higher wavelengths is in the model's cone absorptance spectra, which has two components, photopigment low-density spectral absorbance function and the peak optical density of visual pigment. Fig. 4-31 indicates that cone absorptance spectra should have an age-dependent component, which would cause the cone absorptance spectra to reduce as the age is increased. This component should have different values in the long- and medium- wavelength range.

What could be the physiological explanation for such a component, which is missing from the model? As explained in section 5.8 of the CIE TC 1-36 report [14], there are some indications that the peak optical density of the visual pigment decreases gradually as a function of age. However, because of insufficient or contradictory data to support this hypothesis [134] [62] [61], such dependence has been ignored in the model. A logical argument would be that the age dependence of this factor has a significant effect on cone fundamentals and color matches, and that its exclusion from the CIEPO06 model leads to prediction errors of intra-age group average at higher wavelengths. This argument appears to contradict Webster and Macleod's [135] observation that none of the factors extracted through a factor analysis of the Stiles-Burch 10° data corresponded to differences in photopigment density, and only a weak role of density differences was suggested by the fits to the correlation

matrix. They concluded the peak wavelength shift of photopigment density was a more salient determinant of individual differences in the matches. A key difference between that study and current analysis is that Webster and Macleod were investigating individual variability without regard to age groups, while current analysis focused on intra-age group average prediction. For the latter, differences in photopigment optical density does seem to be an important factor.

It should be emphasized that the optimization method described in this section is purely mathematical. Deriving a physiologically-based correction function was beyond the scope of current study. However, this analysis isolates the likely source of a major flaw in the CIEPO06 model, correcting which can lead to a significant improvement in model performance, particularly for observers in higher age-groups compared to the Stiles-Burch observers' average age of 32.

4.6 Conclusions

In this chapter, a theoretical analysis on various aspects of the physiologically-based observer model proposed by CIE TC 1-36 (CIEPO06) was presented. In the context of color perception on modern narrow-band displays, we evaluated the performance of the CIEPO06 model in predicting the average data for three different age-groups of Stiles-Burch observers and compared the results with the CIE 10° standard colorimetric observer. Here, the goal was to determine if an age-dependent observer provides an advantage over a single average observer. Several conclusions can be drawn from the current study as listed below:

i) The photopigment peak optical density has the strongest influence in display color perception compared to other physiological factors. This finding assumes further significance in light of Smith et al.'s [63] [132] observation that a variation of ±0.2 unit of photopigment optical density from the mean could account for 99% of the individual variance in the Stiles-Burch pilot data [44]. Photopigment peak wavelength shift is another factor having significant contribution to observer variability, but is not within the scope of the CIEPO06 model. ii) Using real observer ages in the model leads to large errors in intra-age group average Stiles-Burch observer CMF prediction, making it difficult to use this model in practical applications. iii) CIE 10° standard colorimetric observer z-function has a large error with respect to intra-age group average z-functions of all three Stiles-Burch age-groups studied, namely 6, 10 and 6 observers in age range 22-23, 27-29 and 49-50 years respectively; in all three cases, CIEPO06 model provides significant improvement, iv) x- and y-CMFs derived from the CIEPO06 model for the observer age group of 49-50 years show high deviation from the intra-group average, the error being comparable to intra-group standard deviation. v) In terms of predicting average color perception for different age-groups on a display with narrow-band primaries, the CIEPO06 model in its original form does not always offer an improvement over the 10° standard colorimetric observer. This limitation is particularly apparent for higher age-group

observers in the red-green part of the color space. vi) A constrained nonlinear optimization of the CIEPO06 model shows that only peak wavelength shifts of the LWS and MWS photopigment density fails to improve intra-age group average prediction, while weighting functions for the photopigment density functions above 550 nm significantly improves this prediction both in the spectral domain and chromaticity space, for both age groups of 22-23 and 49-50 years. This weighting function is different for different age-groups and also different for LWS and MWS cone photopigment densities. It is proposed that the peak optical density of visual pigments be made an age-dependent function in the CIEPO06 model and be defined independently for LWS and MWS cone photopigments.

As a final note, the above conclusions are based on an analysis of the Stiles-Burch observer data. While this is the most comprehensive visual data available till date, it will be of interest to validate these conclusions using an independent visual dataset.

A couple of months in the laboratory can frequently save a couple of hours in the library. ~ Westheimer's Discovery

5. An investigation of Observer Variability in Display Color Matching

5.1 Introduction

In Chapter 3, a detailed review of various studies on observer variability in visual color matching was presented. The effect of inter-observer variability has often been found to be significant in scientific studies on color matching, both in the classical and in applied contexts. Observer variability and metamerism can also be a nontrivial issue in industrial applications involving critical color matching tasks. This is particularly true for those applications that involve various kinds of modern display devices. One example is the color adjustment process (called color grading) in post-production applications where the raw movie content at the post-shooting stage is modified to achieve the right color effect. The Colorist has to work with the Director of Photography (DP) to adjust the colors in the original content so as to achieve color coherence and homogeneity throughout various scenes, while maintaining the artistic expressions originally envisioned by the Film Director and the DP. However, if the Colorist and the DP have different color vision characteristics, they will perceive colors differently, and the colors that look similar to one will look perceptibly different to the other. Conventional colorimetry will fail to account for this difference.

Further, the film may have to be converted to a version suitable for television or DVD (a process known as digital mastering). This then becomes a cross-media color reproduction issue, where we are trying to reproduce the colors, as seen on a theatre screen, to equivalent colors on a specific reference display with a certain color gamut. Processes like color grading and digital mastering are color critical, requiring high-fidelity color reproduction, often involving displays. Presently in the post-production stage, two digital mastering tasks are undertaken – one is for the large-screen (film and/or digital), and the other is for the small-screen (i.e. television, DVD). Because of the wide disparity in the color gamuts of theatre projectors and television displays, significant/complete digital re-mastering is required for the small-screen version. Even though film studios have principally relied upon reference CRTs, a rapid market adoption of wide-gamut, high-definition displays and projectors and gradual discontinuation of manufacture of CRTs may soon require the studios to employ these modern displays for post-production operations.

Very recently, studios have also started offering remote color grading services, which means multiple devices being used by various professionals at multiple locations for color grading, a trend that is sure to make the issue of observer variability even more pertinent in the media and entertainment industry.

Thus, it is of interest to study the effect of observer variability in color matching across conventional and modern displays, and to acquire experimental data in such a context. The data so collected can subsequently be used to better model the observer variability, and to find solutions to associated problems.

5.2 Investigating observer variability: color matching experiments using two displays

Observer metamerism is not only an important consideration in cross-media color reproduction where the primary objective is to achieve faithful color reproduction, but it is also a critical color imaging issue when various devices of the same category are reproducing colors using primaries with widely varying spectral characteristics. As explained in the previous section, the primary focus of this work is on modern display systems, where observer metamerism happens to be more evident than traditional industrial applications like printing, paint, textile etc. Cathode Ray Tube (CRT) displays of the past decades used primaries that were relatively flat, which allowed minimal spectral differences during color reproduction, thereby reducing observer metamerism. However, many modern displays tend to use narrow-band primaries in order to achieve wider color gamuts and greater luminance contrast. This makes these displays more susceptible to observer metamerism. This was the motivation to investigate the effect of observer metamerism in modern display applications through visual experiments. In the next subsections, the experimental design aspects are discussed in detail.

5.2.1 *The setup*

Two displays were used in these experiments. 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. The spectral power distributions of the two displays are shown in Fig. 5-32. There is a significant difference in the spectral characteristics between the two displays, so, a color match made on the two displays is metameric in nature. This justifies the choice of these two displays for the observer variability study. The LCD is representative of modern wide-gamut displays with peaky primaries. The CRT has a 10-bit HD/SDI input and the LCD has an 8-bit DVI input. The two displays were controlled independently through a specially-designed hardware, integrated with the software developed for the color matching experiments.

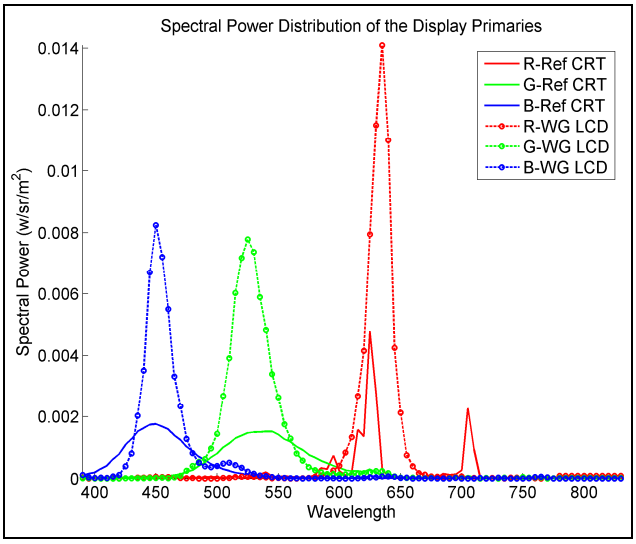


Fig. 5-32. Spectral Power Distribution of the CRT and the LCD used in the experiments

The displays were placed perpendicular to each other, as shown in Fig. 5-33. A front-surface reflection mirror was placed in front of the CRT at 45° to the observer's line-of-sight, which was perpendicular to the LCD screen to avoid the directionality issue of the LCD. The observer's visual field consisted of a 10° bipartite field, the right half of which was the LCD screen, and the left half was the CRT screen, seen through the mirror. A mask was placed between the observer and the displays to block the view of the displays and the mirror, allowing the observer to see only two solid self-luminous color patches on the two sides of the field when looking at the mask from its centered normal. The mirror also blocked lights from the CRT to fall on the LCD screen. The distance between the observer and the mask was 69 cm (2.3 ft), and that between the mask and the LCD screen was 68 cm (2.2 ft).

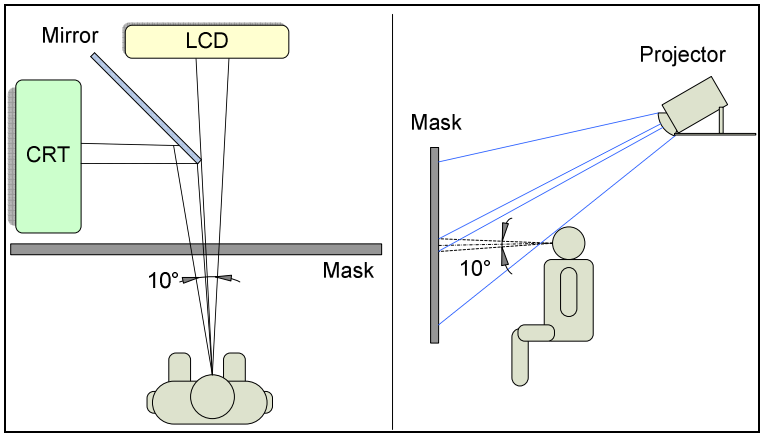


Fig. 5-33. Experimental setup

The width of the mirror formed a 0.02° vertical black field separation at the observer's eyes, unavoidable for mechanical reasons. Luminance discrimination is best when the two half fields are precisely juxtaposed. By introducing the field separation, red-green chromatic discrimination remains the same, but discrimination based on differential short-wavelength sensitive cone excitation improves [16] (page 136). The effect of the separation introduced by the mirror edge on the color matching was outside the scope of current study.

The displays were characterized before the experiment. The display lookup tables (LUTs) thus obtained were used to determine the initial RGB digital counts that would result in specific chromaticities on the displays for the CIE standard colorimetric observer. However, during the adjustment of the CRT color by the observer, a simple linear transform from XYZ to RGB was preferred over the display LUT, as this allowed the observer to have a better control over the adjustment in a linear scale. The mirror was included in the characterization of the CRT, to account for any spectral absorption or transmission by the mirror surface. At the beginning of each session, the luminance of the full-white of both displays was measured to ensure that they were close. While both displays were found to be quite stable in terms of full white luminance, radiometric data for both displays were collected after each color match (except for the pilot test, as explained later). Thus, the experimental results were independent of the stability of display characterization, or of the assumption of the validity of the display additivity and proportionality. For the measurement, a spectroradiometer was placed directly behind the observer at the eye level, and two displays were measured in succession. The spectroradiometer PhotoResearch PR-670 used in this work was factory-calibrated three months before the experiment with a NIST traceable light source. The luminance as well as radiometric uncertainty relative to NIST was ±2% and spectral wavelength uncertainty was less than ±2 nm.

5.2.2 *Observer task*

The observer was asked to adjust the color on the left half of the bipartite field (matching field - CRT) to match the color on the right half (test field - LCD). The observers were aware that they were matching colors on two displays. Since the CRT had a 10-bit channel resolution (i.e. 1024 levels of R, G and B channels), it was chosen as the matching field, and the LCD was used as the test field. Thus, the color matching task was a quasi-symmetric matching procedure.

However several experimental design issues were encountered.

5.2.2.1 Which parameters to adjust?

Several possibilities for adjustment of the colors were explored. Adjustment in chroma, hue and lightness was thought to be more intuitive and was preferred over the direct RGB channel adjustment

[111], or the adjustment of opponent colors (redness-greenness and yellowness-blueness), as has been done in previous works using the CIELAB color space [87][89] [115]. In this work, the IPT color space was used, which is perceptually more uniform than CIELAB, particularly in the blue region of the color space [136]. The color in the test field could be adjusted in three dimensions of chroma, hue and lightness, derived in the IPT color space.

To make the color matching task less daunting for the observer, the starting color in the matching field (CRT) was set to hue and lightness values of the test field (LCD) as predicted by display characterization (except in the pilot test, as explained later). However the initial matching field chroma was randomly varied between 75% and 90% of the test field chroma. This was done because preliminary tests revealed that for observers unfamiliar with color, the task of matching was more difficult when both hue and chroma were completely different in the two fields. However, the observers generally made an adjustment in all three dimensions, which was expected since a display characterization is essentially based on an average, standard observer data (in this case, CIE 10° standard colorimetric observer) and does not conform to individual observer characteristics. In addition to setting the initial color, the hue angle range was set to ±30° of the initial value to prevent the observer from deviating too far from the region where a match could be located. The smallest possible changes in the chroma, hue and lightness dimensions were set to 0.001, 0.1° and 0.0001, respectively.

5.2.2.2 How to adjust?

A ShuttleXpress® multimedia control by Contour Design was used in this experiment for color adjustment. This control has five buttons, one wheel and a jog that were programmed to specific functionalities (Fig. 5-33), and was connected to the computer through USB interface. The Chroma/Hue/Lightness button allowed switching from one dimension to the other by subsequent pressing. The jog and the shuttle allowed changing the value of the current dimension. Two additional features that were found to be quite helpful in better executing the color matching task were also implemented. The first was a Save-Undo feature that allowed the observer to temporarily save the matching field color before adjusting it further to refine the match, and to go back to the saved version if needed. The second feature was a Reset functionality, which allowed the observer to go back to the initial setting of the current dimension (Chroma, Hue or Lightness) if encountered with the difficulty in getting closer to a match. The Commit button confirmed observer's match and saved the current display RGB and IPT values for both fields. Radiometric measurements were launched by a separate command once the match was confirmed.

5.2.2.3 To fixate or not to fixate?

No head restraint was used in the experiment. White adapting stimuli were presented in both fields for a couple of seconds before launching a new trial. During the course of the trial, the observer was encouraged to move his/her head sideways from time to time, or to look away, in order to reduce the effect of local adaptation. When test and matching field luminance is greater than the surround, adaptation to the bipartite field is likely. The effect of this adaptation is to reduce the perceived difference between the two halves of the bipartite field after viewing them for several seconds. Another way to avoid the adaptation to the bipartite field stimuli is to present the fields for a small percent (e.g. 20%) of the duty cycle, and replace them by the surround chromaticity for the rest of the time [137]. However, this method is more cumbersome and time-consuming, and may cause annoyance to the observer.

The other issue occasionally encountered by the observers was a halo effect, wherein the peripheral part of the bipartite field appeared to be lighter than the rest of the field. This was likely due to simultaneous contrast induced at the border of the field when dark surround was used. Sideways movement of the head or looking away from the field for a couple of seconds significantly helped in reducing both adaptation and contrast effects. However, it must be emphasized that the final match was always made while focusing on the bipartite field, and not through peripheral vision.

For some stimuli, a color inhomogeneity in the center of the field, commonly known as the Maxwell spot, was noticed by some observers. This is a well-documented effect due to higher density of macular pigment in the central fovea and gradually diminishing outward [9] (page 133) (see also Chapter 2). The observers were asked to ignore this non-uniformity.

5.2.2.4 What about adaptation and surround?

For a small field, the surround serves to maintain a reasonably steady-state of adaptation for the observer [16] (page 137). Note that the term adaptation here refers to the luminance adaptation and not the chromatic adaptation. The effect of a chromatic surround on color matching was outside the scope of current study. To study the effect of adaptation on large-field display color matches, observers were asked to perform color matching in two separate experiments, one in dark surround and the other with an achromatic surround with roughly uniform luminance. For the surround test, a diffuse white mask was used instead of a black mask. A projector (Optoma EP747 with DLP™ technology) placed behind the observer overhead was used to uniformly illuminate the mask. A black circle in the middle of the projected image overlapped with the 10° bipartite field on the mask, so that light from the projector passing through the hole could be minimized. The projector was carefully positioned such that the observer's head did not cast a shadow on the mask, and the small amount of light passing through the hole fell on the black cover on the table in front of the displays, and not on the mirror or the displays themselves. The luminance of the surround was 15 cd/m² in the middle, and

had a horizontal fall-off of about 10% on the far end of both sides. The correlated color temperature of the surround was close to 7400K. The 102cm x 60cm surround field formed an angle of 73° horizontally and 47° vertically in the observer's eyes.

Table 5-7 lists the full-white chromaticities, luminance values and the Correlated Color Temperatures (CCTs) of CRT, LCD and the projector as measured by the spectroradiometer.

Table 5-7. Chromaticities, luminance values and Correlated Color Temperatures of the two displays and the projector

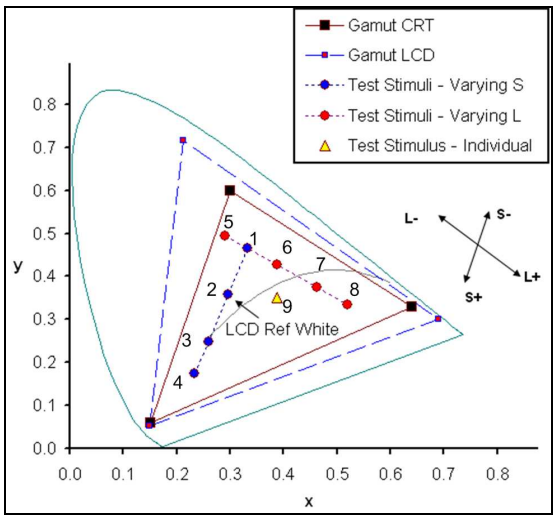
	CRT	**LCD**	**Projector**
x	0.3074	0.306	0.2958
y	0.3255	0.3245	0.3359
Y(cd/m²)	96.04	96.69	14.98
CCT (K)	6828	6919	7363

5.2.3 *Selection of test stimuli*

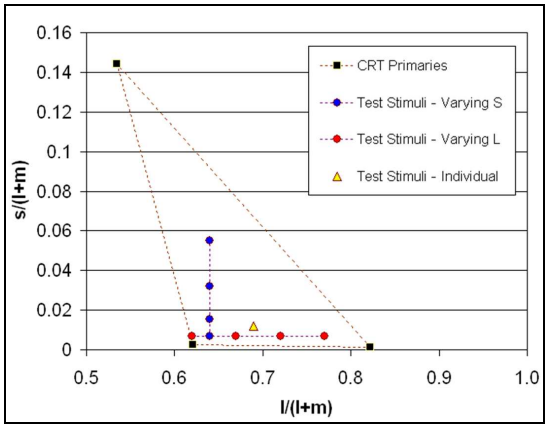
The basis of stimuli selection in the current work differs from previous studies with similar experimental setups, where either the primary or secondary colors were selected as stimuli [87][89], or the color space was sampled in equal hue angle steps [115]. Such choices are useful in comparing observer variability in color matching in different regions of the color space. However, they do not have a physiological basis, and do not consider how the stimuli may affect the long-, medium- and short- wavelength sensitive cone excitations (hereafter referred to as LWS, MWS and SWS respectively), which is an issue of fundamental importance in color matching. Since a major goal of the current study is to evaluate the merits of various color matching functions and cone fundamentals in the context of modern display colorimetry, it was of interest to select the test stimuli for the experiments in such a way that they varied along physiologically significant axes. Thus, MacLeod-Boynton chromaticity diagram [35] was used for specifying the chromaticity coordinates of nine test stimuli. In this diagram, the cone spectral sensitivities form rectangular axes in a constant luminance plane. The abscissa represents the equal and opposite change in LWS and MWS cone excitations (such that the sum is unity), and the ordinate represents the level of SWS cone excitation. It is possible to derive the MacLeod-Boynton chromaticity coordinates from Stockman-Sharpe 10° cone fundamentals [22], on which much of this thesis work and CIE 2006 cone fundamentals [14] are based. However, transforming MacLeod-Boynton chromaticities of a test stimulus into 10° XYZ tristimulus values is not straightforward (see Chapter 4). This transformation is relatively simple using MacLeod-Boynton chromaticities based on Smith-Pokorny 2° cone fundamentals [37], since LWS and MWS are appropriately scaled so that (L+M) gives luminous efficiency function Y. The transformation has been described elsewhere [16] (page 118). However, the (x, y) chromaticity values

so obtained correspond to 1951 Judd modified CIE 2° observer. Thus, for the purpose of selecting the test stimuli, the Judd-revised observer was used to perform display characterization computations and to derive the RGB digital counts for both displays that would result in the specific MacLeod-Boynton chromaticities. Note that, with the exception of stimuli selection, 1964 CIE 10° standard colorimetric observer was used for all colorimetric computations.

Four of the nine selected stimuli varied along s-, with l- being constant (l = 0.64) in the MacLeod-Boynton chromaticity diagram, while four others varied along l- axis with constant s- (s = 0.007). The ninth test color was an isolated point close to skin tone. Fig. 5-34 shows the stimuli in Judd chromaticity diagram (top) and the MacLeod-Boynton diagram (bottom). All nine stimuli had a luminance close to 25 cd/m². The luminance could not be increased any further since it caused some boundary points (e.g. stimulus #5 in Fig. 5-34a) to fall outside the gamut of the CRT.



(a)



(b)

Fig. 5-34. Nine test stimuli in (a) Judd chromaticity diagram, and (b) MacLeod-Boynton chromaticity diagram based on Smith-Pokorny 2° cone fundamentals (spectrum locus not shown)

5.2.4 *Experiments*

Three experiments were conducted, of which the first was a pilot test. All experiments were conducted in a dark room, with all visible surfaces being covered by black paper/cloth. In each test, there were nine test stimuli as described before, and each observer performed three repetitions. Thus, there were 27 trials in each test. Each repetition lasted 45 min – 1 hour, between which, and between two consecutive matches, the observers took a break for several minutes. Each observer participated in the three tests within a span of two weeks.

Specific details of the three experiments follow.

5.2.4.1 Pilot test using only one display

In this test, only the LCD was used for color matching. A window with two rectangles separated by a thin black strip (simulating the mirror edge in the actual experiment) filled the full screen of the LCD. The right rectangle formed the test field, and the left rectangle, whose color could be adjusted by the observer, formed the matching field. When seen through the 10° mask, the visual appearance of the 10° bipartite field was exactly the same as in case of the tests involving two displays. The test was performed in the dark surround condition. The observer task has already been described in Section 5.2.2.

Comparing the results of intra- and inter-observer variability in this pilot test, the validity of the experimental protocol could be ascertained. For example, if for the majority of the observers, the intra-observer variability is more than the inter-observer variability, this would mean the experimental setup is unsuitable for acquiring color matching data, as the uncertainty of observer color matches would not be within acceptable range. In fact, given that the experiment was being performed on a single display, the observer metamerism aspects would not apply, so inter-observer variability should be of the same order as intra-observer variability. On the other hand, if the intra-observer variability is high only for a limited number of observers, we can conclude that these observers are not adept at using the experimental tool for obtaining color matches with adequate certainty, either because of their lower color matching precision (i.e. higher tolerance), or because of their unfamiliarity with the color matching task.

This test offered an advantage over the previous studies [89][115][111], in which it was not straightforward to ascertain whether and to what extent observer variability was influenced by the experimental protocol itself.

In this pilot test, the initial lightness, chroma and hue values of the matching field were randomly set to values significantly different from those of the test field. Note that this was not the case in the other

two experiments (see Section 5.2.2.1). Also, when a match was confirmed by the observers, the lightness, chroma and hue control settings were recorded, but the spectral measurement was not performed, unlike in the other two experiments.

5.2.4.2 Experiment with dark surround

This test was performed using the two displays as discussed before, in the dark surround conditions. No light source other than the bipartite field was present.

5.2.4.3 Experiment with white surround

This test was conducted with white surround condition, as described in Section 5.2.2.4. Comparing the results of this experiment with those of dark surround experiment would enable us to assess the potential role of steady-state, luminance adaptation on display color matches. This is of interest since in practical, real-life situation, the display viewing condition generally includes a lit surround.

Ten observers participated in each of the three experiments. The observers were in the age range of 30 – 50, and all were color normal, as confirmed by Ishihara pseudo-isochromatic plates and a Farnsworth-Munsell 100 hue test.

5.2.5 *Results and Discussion*

5.2.5.1 Intra- and inter- observer variability in pilot test

As explained before, a comparison of the intra- and inter-observer variability in the data from the pilot test with single display will indicate the suitability of the experimental setup for conducting color matching experiments and the ability of observers to perform the color matching task. The intra-observer variability refers to the deviations in matches for a given test color made by a single observer during different trials. The inter-observer variability refers to the deviations in mean observer matches (averaged over several repetitions) for a given test color from one observer to the other. To determine the intra- and inter-observer variability in the pilot test data, the root-mean-square (RMS) errors were computed for the color matches in the lightness, chroma and hue dimensions in the IPT space. These were the original dimensions adjusted by the observers. The display used in this experiment was stable enough so that it can be assumed the test colors presented to the observers were reasonably constant across different sessions. This issue is further clarified afterward.

For computing the intra- observer variability, first the RMS values of the differences between all match repetitions by a given observer and the mean match for that observer were obtained for each test stimulus. For each test stimulus, the mean of these RMS values over all observers gives the

average intra-observer RMS error. Similarly for inter- observer variability, RMS errors were computed between the mean of all observer color matches for each test stimulus, and the mean of each observer matches is computed over all repetitions.

Results of one of the ten observers showed significantly higher (2.5 times) intra-observer variations than the others for the test stimulus #8 (red), and to a lesser extent for test stimulus #4 (blue). This observer was excluded from the analysis of the pilot test data. Fig. 5-35 shows the plots of intra- and inter-observer variability in three color space dimensions for the rest of the nine observers. On an average, the intra- and inter-observer RMS errors are low. Mean intra-observer RMS errors were 1.2% in lightness, 3.0% in chroma and 1.2% in hue, averaged over all test colors. For mean inter-observer RMS error, these values were 1.8%, 4.1% and 1.3% respectively. As explained earlier, in this particular experiment the main difference between intra- and inter- observer variations come from the differences among observers in their precision and repeatability, not from any physiological reasons related to observer metamerism. Thus, the intra- and inter- observer variability are understandably similar. We can expect that in this experiment, the uncertainty of color matches contributed by the experimental setup itself does not exceed the mean intra-observer RMS errors.

Test stimulus #4 (blue) shows significantly higher inter-observer variability in lightness and chroma compared to other stimuli. This indicates that the cyan/blue region is particularly susceptible to lack of precision in observer color matches. It is possible that the Maxwell spot [9] (page 133) plays a role in this, since the effect of macular pigment absorption is likely to be more pronounced for this test stimulus.

The hue in case of test stimulus #2, which was an achromatic color close to the LCD white point, shows relatively high intra- (4.5°) and inter-observer (3.9°) RMS errors in hue. However, for five out of ten observers, the mean intra-observer RMS error was only 2.7°, indicating that the high error resulted from individual observer uncertainty in matching achromatic colors, and was not caused by the experimental setup itself.

Overall, the results from Pilot Test 1 indicate that all observers were able to adjust the matching field to get satisfactorily close to the test field color. All observers expressed satisfaction over their matches, and over the method of adjustment. Thus, we can conclude that the experimental setup is suitable for acquiring valid metameric color matching data. The observer who showed higher variations for test stimuli #4 and #8 was able to achieve satisfactory results for other stimuli, and thus, although excluded from the above analysis, he was not excluded in subsequent experiments.

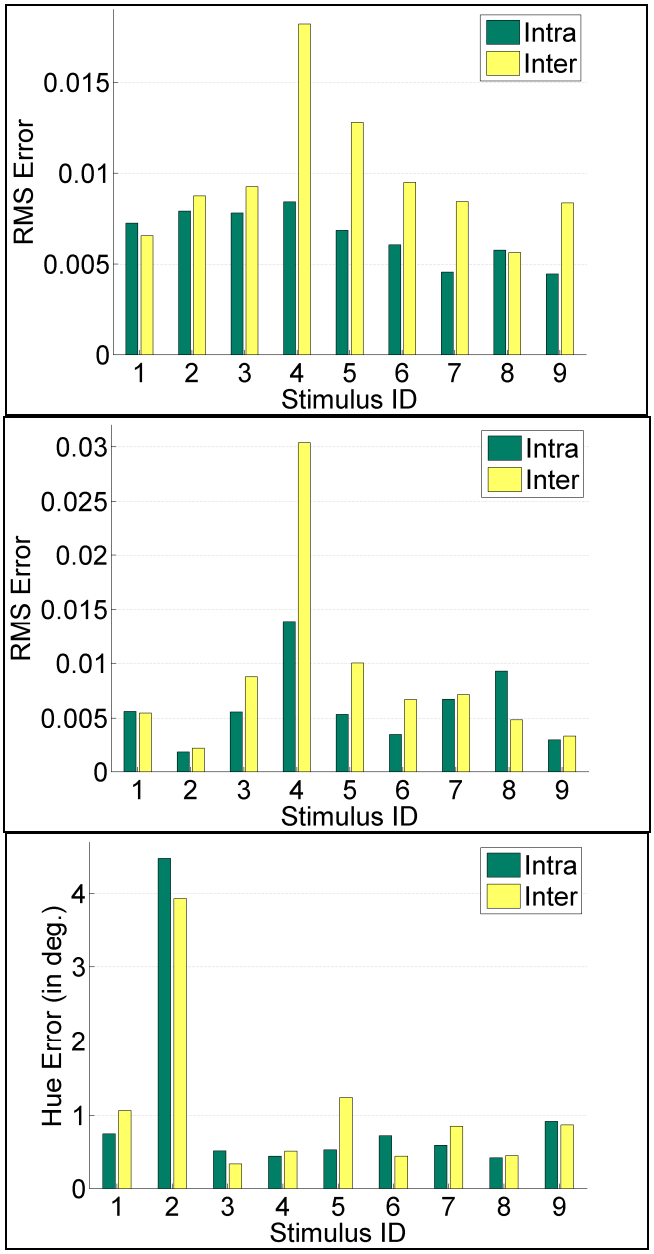


Fig. 5-35. Results from the pilot test with single display: mean intra- and inter-observer RMS errors in lightness (top), chroma (middle) and hue (bottom), computed in IPT color space

5.2.5.2 Intra- and inter- observer variability in two experiments

The intra- and inter-observer variability was also determined for the experiments with dark surround and white surround. In both cases, measured spectral power distributions of the matching field for each observer match were used. Note that in this case, we do not compare the LCD and CRT colors, but rather inspect the variability in the CRT color matches, assuming the test colors on the LCD stayed approximately constant during the experiment. Average color difference on the LCD side across all trials was less than 0.1 ΔE_{00} (CIE 2000 advanced color difference metric [138]), so the

assumption is acceptable. From the spectral data, XYZ tristimulus values and CIELAB coordinates were calculated using the 1964 CIE 10° standard colorimetric observer and display white points. Mean Color Difference from the Mean (MCDM) [10] was computed across three repetitions for each observer in case of intra- observer variability, and across the mean matches of all observers in case of inter- observer variability.

Table 5-8 lists the MCDM values for all nine stimuli, for both tests, calculated based on ΔE_{00}. As expected, inter-observer variability is larger than the intra-observer variability, on an average 1.75 times in case of dark surround experiment and 2.2 times in case of white surround experiment. Inter-observer variability is the higher for test stimuli #2, #6 and #9 compared to other colors, for both experiments.

The surround has the effect of a steady-state adaptation during the color matching. Intra-observer variability slightly reduced on the introduction of a white surround, but the effect on the inter-observer variability is less apparent. The average reduction is 0.13 ΔE_{00} for the intra-observer variability, and 0.02 ΔE_{00} for the inter-observer variability. Overall, no strong effect of surround on the observer color matches was observed.

Table 5-8. Mean Color Differences from the Mean (MCDM) for intra- and inter-observer data from the experiments with dark surround and white surround

Stimulus ID	Dark Surround		White Surround	
	Intra-	Inter-	Intra-	Inter-
1	0.61	0.93	0.53	1.16
2	0.68	1.48	0.55	1.67
3	0.60	1.02	0.51	0.99
4	0.63	1.01	0.48	0.85
5	0.58	0.79	0.53	0.98
6	0.58	1.48	0.44	1.30
7	0.51	0.99	0.39	0.90
8	0.47	0.75	0.46	0.53
9	0.94	1.36	0.53	1.26

5.2.5.3 Color match prediction error with CIE 10° standard colorimetric observer

As mentioned before, spectral data were collected for both displays after each trial in which the observer performed a color match. This allowed computation of chromaticities of colors on two sides of the bipartite field that matched for individual observers. Two different methods were used to compare CIE 10° standard colorimetric observer predictions with individual color matches. In the first method, display characterization data were used to predict a CRT color match of the LCD test color. For each trial, XYZ tristimulus values were computed from the spectral data of the LCD test colors,

using CIE 10° standard colorimetric observer. The XYZ values were averaged over all repetitions for a given observer. These are the XYZ values to be reproduced on the CRT. The CRT inverse model predicted the digital counts that would generate similar XYZ values. For better accuracy, and as verification, the CRT forward model was then used to compute the XYZ values that could actually be reproduced on the CRT. Thus, these XYZ values corresponded to a "standard observer" color match on the CRT, as predicted by the 10° standard colorimetric observer. XYZ values were also computed from the spectral data of the observer color matches on the CRT. These two sets of XYZ values were converted to CIELAB, and ΔE_{00} color difference values were computed. The second and third columns of Table 5-9 list the 90^{th} percentile of these ΔE_{00} values between the predicted and actual observer matches on the CRT side for each of the nine stimulus, averaged over all observers.

The second method was more straightforward. As before, XYZ values were computed from the spectral data for both the LCD test colors and the CRT matching colors, using CIE 10° standard colorimetric observer. For each observer, the XYZ values over all repetitions were averaged, and then were converted to CIELAB values. Finally, ΔE_{00} color difference between the LCD and CRT sets of CIELAB values were computed. These ΔE_{00} values signify the differences that would be perceived by a "standard observer" between the LCD and CRT color matches of individual observers. The last two columns of Table 5-9 list the 90^{th} percentile of these ΔE_{00} values.

Table 5-9. 90^{th} percentile color difference (ΔE_{00}) values computed between i) the CIE 10° standard colorimetric observer predicted matches and observer color matches on the CRT side, and ii) the test colors on LCD and observer matches on CRT

Stimulus ID	Prediction and Observer Matches (on CRT)		Observer Matches (LCD and CRT)	
	Dark Surround	White Surround	Dark Surround	White Surround
1	2.36	3.00	2.05	2.81
2	3.21	3.15	2.81	3.08
3	2.17	2.12	2.20	2.50
4	2.87	2.59	3.16	3.07
5	2.30	2.40	2.16	2.36
6	3.62	3.45	3.26	2.89
7	1.70	1.73	1.63	1.75
8	1.38	1.01	1.42	1.23
9	3.25	2.36	2.82	2.36

Computed ΔE_{00} color difference between the CIE 10° standard colorimetric observer predictions and observer color matches on the CRT are generally higher than the color difference between the actual observer matches on the two displays. This is not surprising since the former is affected by the computational approximations of display modeling, and is dependent on the assumptions of display additivity and proportionality.

Fig. 5-36 plots the ΔE_{00} color difference values corresponding to individual observer matches on LCD and CRT, and the predicted and real observer color matches on the CRT side, both for the dark surround experiment. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles (q_1 and q_3 respectively), the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as red circles. Points are drawn as outliers if they are larger than [q_3 + 0.5 * (q_3 - q_1)] or smaller than [q_1 - 0.5 * (q_3 - q_1)]. Thus for a given stimulus, the size of the box and the length of the whisker indicates inter-observer variability for a given test color, while together with the red circles indicates the range of variability among observers for a given test color. Fig. 5-37 shows the same plots for white surround.

From the data in Table 5-9 and the plots in Figs. 5-36 and 5-37, it is clear that for some observers, some of the colors on the two displays that match for individual observers are predicted by the CIE 10° standard colorimetric observer as having a significant color difference, and similarly, the colors that are predicted by the standard colorimetric observer to be a match when shown on the two displays are sometimes unacceptable to individual observers. This discrepancy is the highest for the test color #2 and #4, an achromatic color and a saturated blue respectively (Table 5-9). In case of dark surround experiment, the mean, maximum and the 90th percentile ΔE_{00} values between individual observer matches on LCD and CRT, across all stimuli and all observers, are 1.4, 3.4 and 2.6 respectively (1.4, 3.5 and 2.7 respectively for the white surround experiment).

In case of Alfvin and Fairchild's [89] experiment, the mean color difference from mean (MCDM) for inter-observer variation was 2.5 CIELAB units. Oicherman et al. [107] on the other hand reported a mean variability of observer matches of around 3 CIELAB units (note that it was not a split-field color-matching). Compared to previous studies, the mean variability among observers might seem relatively low, but there are several caveats in using computed ΔE_{00} color difference values for making that inference.

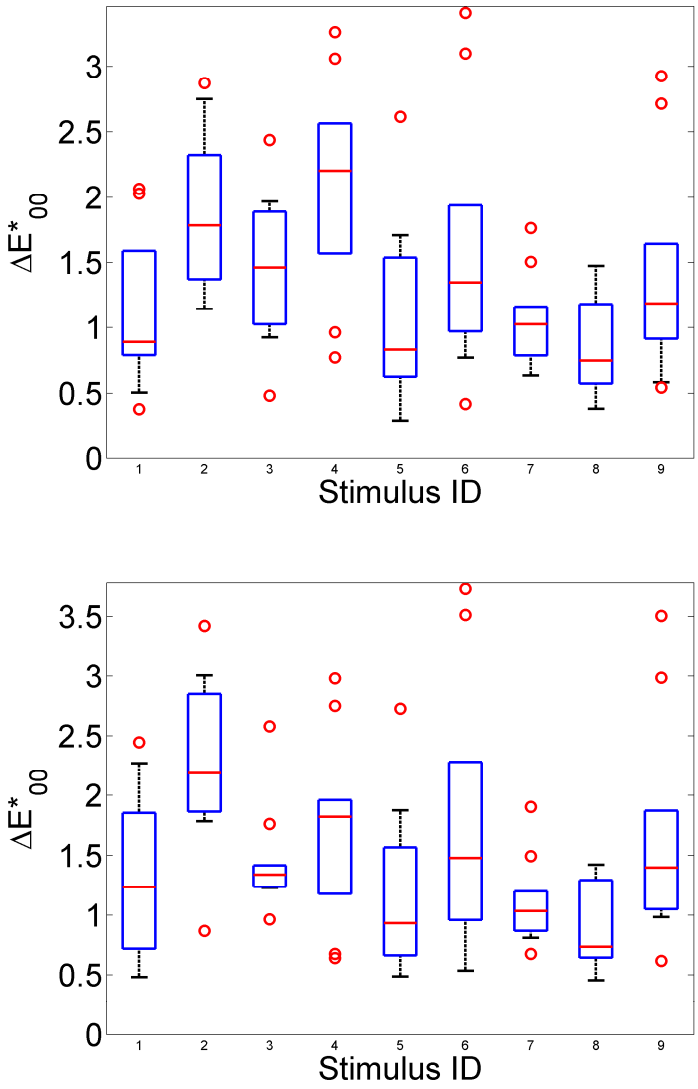


Fig. 5-36. Box plot of inter-observer variability in the results from the dark surround color matching experiment. Top figure shows the ΔE_{00} color difference between CRT and LCD observer matches as predicted by the CIE 10° standard colorimetric observer, and the bottom figure shows the ΔE_{00} color difference between CRT observer matches and corresponding CRT match predictions by the CIE 10° standard colorimetric observer.

The significance of the ΔE_{00} values depends on the context, viewing conditions and the observer. The values reported here are possibly low for complex images and surrounds, and even cross-media color matching. However, in the experimental setup implemented in this study uniform color stimuli are matched by non-novice observers under strictly controlled viewing conditions. In such a scenario, a ΔE_{00} color difference much larger than 1.0 is likely to be perceptible. An average color match prediction error of 1.4 ΔE_{00} over all colors and all observers is likely to be acceptable in most application contexts, but the maximum ΔE_{00} value of 3.4, and the 90th percentile ΔE_{00} value of 2.6 between individual observer matches predicted by the CIE 10° standard colorimetric observer are rather high, particularly in applications that require stringent color matches. This indicates that for

some colors, color match prediction by an average observer results in significant color match errors for many individual observers. In color critical applications involving modern displays, expert observers will likely find such differences unacceptable. The degree of the prediction error is dependent on the spectral characteristics of the display, and also on how close an individual's color vision characteristics are to an average. This disagreement was apparent during informal visual tests prior to the experiment reported here, when color matches obtained by some observers were rejected by others, and vice versa.

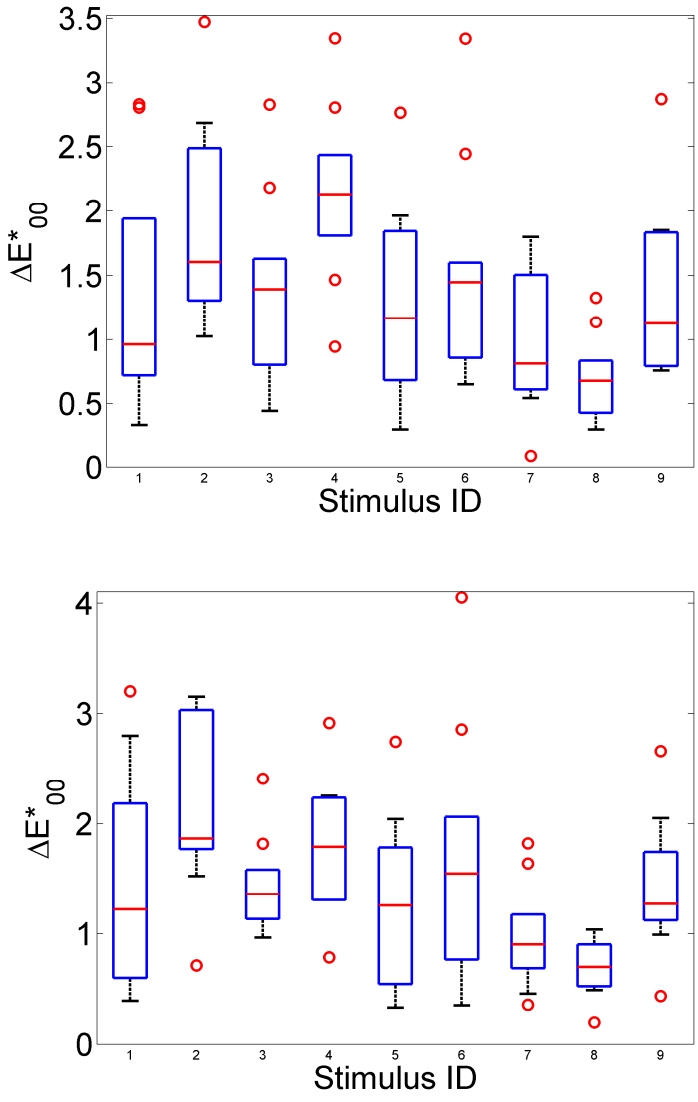


Fig. 5-37. Box plot of inter-observer variability in the results from the white surround color matching experiment. Top figure shows the ΔE_{00} color difference between CRT and LCD observer matches as predicted by the CIE 10° standard colorimetric observer, and the bottom figure shows the ΔE_{00} color difference between CRT observer matches and corresponding CRT match predictions by the CIE 10° standard colorimetric observer.

5.3 Conclusions

This chapter reviewed some of the most significant studies on observer variability and metamerism conducted in the last two or three decades. Such variability arises from the differences in individual's color vision, and thus is of fundamental nature. The effect of observer variability in color perception has been investigated over the years both through classical color matching experiments and through more applied studies, as reviewed in this chapter. Latest addition to this body of scientific studies is the display-based color matching experiment conducted as part of this thesis work. The experimental design took into account several important aspects of large-field color matching. The results obtained from the experiments involving ten observers showed that while average prediction errors for all observers and all stimuli was lower than some of the similar studies performed in the past, the differences were significant for some stimuli. The maximum color difference between the predictions of CIE 10° standard colorimetric observer and real observer matches was 3.4 ΔE_{00}, and the 90th percentile value was 2.6 ΔE_{00}. For a color critical application like in color grading for post-production, such kind of color differences would be considered high.

Unfortunately, an evaluation of observer variability based on computed ΔE_{00} values is not objective enough. In applications involving highly metameric color matches (for example, those involving narrow-band stimuli obtained from LEDs), assessing acceptability of color matches based on ΔE_{00} values may not always be realistic. This arises out of an inherent limitation of an average observer model like the CIE 10° standard colorimetric observer in representing individuals widely differing in their color vision. Since all color difference metrics are essentially based on an average observer, they do not represent Euclidean distances in a perceptual sense for observers sufficiently different from the average. As a result, they fail as a quantitative measure of perceived difference for highly metameric color matches, posing serious restrictions on colorimetric computations and analysis.

An advanced colorimetric system might be imperative for studies and industrial applications involving highly metameric color stimuli. The ground work for such a system is already underway [14] [116]. It is hoped that this thesis research will contribute to that end.

You see things; and you say "Why?" But I dream things that never were; and I say "Why not?" ~ George Bernard Shaw, Back to Methuselah (1921)

6. Colorimetric Observer Categories

6.1 Introduction

Conventional color reproduction relies on colorimetric data for a single “standard colorimetric observer”, representing an average colorimetric observer with normal color vision. The 1931 CIE 2° standard colorimetric observer and 1964 CIE 10° standard colorimetric observer (see Chapter 4) are widely used in the industry. 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 colorimetric observer model, defined 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 [87], but the model significantly under-predicted inter-observer variability [89], and was never adopted by the industry. Thus, applied colorimetry in its current form does not have any provision for incorporating observer variability (commonly termed as observer metamerism, see Chapter 5) into the computations. The limitation, as explained in Chapter 5, has become non-trivial with the advent and wide-spread adoption of modern wide-gamut consumer displays with narrow-band primaries. Light Emitting Diode (LED) based applications are similarly affected. Thus, it is important to find a practical solution to this problem that can be effectively implemented in industrial applications.

A principal hypothesis of this work is that human observers with normal color vision can be classified into a small number of categories based on their color vision. Based on such categorization of the whole observer population, multiple colorimetric observer models can be established for use in applied colorimetry. However, such categories must be appropriately identified and universally agreed upon, and there must be a means to determine which categories should be used under a given circumstance.

There could be multiple ways through which a set of representative colorimetric observer categories can be derived. In this work, a two-step method was developed for deriving these categories. In the first step, 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. A squared Euclidean distance measure (in cone fundamental space) was used in this

analysis, and thus was fundamental in nature. In the second step, a reduced set of representative observer models were derived from the 125 combinations through an iterative algorithm. For this, several predefined criteria on perceptual color differences delta E 2000 (ΔE_{00}) with respect to actual color matching functions of the 47 Stiles-Burch observers were used. Color differences were computed for the 240 Colorchecker DC™ samples viewed under D65 illumination. Thus the goal was to come up with a minimum set of observer models that would satisfy all predefined color difference criteria for each Stiles-Burch observer. The derivation of the reduced set of observer models is more applied in nature in comparison to the model cone fundamentals derived in the first step. Sections 6.2 and 6.3 describe the two-step method in more detail.

An experimental method was also developed in order to assign colorimetric observer categories to individual observers. The method was first implemented on a test setup with two displays, which was later replaced by a proof-of-concept prototype based on LEDs. Sections 6.4 and 6.5 discuss these two implementations respectively.

Finally, Section 6.6 presents results from a collaborative experiment aimed at validating the observer classification method, followed by some concluding remarks on standard and deviate colorimetric observers in Section **Error! Reference source not found.**.

6.2 Deriving colorimetric observer categories – Step-1: cluster analysis

An assumption of this work is that the CIE06 model predictions and the experimentally obtained visual color matching data from the 1959 Stiles-Burch study, when combined together, incorporate most of the variability that can be found among the color normal population. The combined data set used in this study thus 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 [139]. The purpose of the analysis is to arrange the functions into relatively homogeneous groups based on multivariate observations. In this analysis, the total number of variables was 35 (normalized values at 35 wavelengths) and total number of observations was 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. K-means method was chosen as it is one of the more popular nonhierarchical clustering methods, and is capable of handling a large amount of data [139]. Other clustering methods were not investigated, since there is no direct way to ascertain which method produces best results.

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. Cluster centroids were recomputed after each reassignment. Each cluster in the partition was defined by its member objects and by its centroid, or center. As explained later, suitable wavelength ranges (i.e. the number of variables) were chosen for long-wave sensitive (LWS), medium-wave sensitive (MWS) and short-wave sensitive (SWS) cone fundamentals to avoid the influence of variations where functions had low amplitudes. In this regard, it is worthwhile to quote Paul A. Gore [140]: "Researchers are encouraged to select variables based on sound theoretical grounds, to select variables that will maximally discriminate among objects, and to avoid the indiscriminate inclusion of variables".

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. At first, $\bar{r}$ $\bar{g}$ $\bar{b}$ CMFs of Stiles-Burch observers were converted to $\bar{l}$ $\bar{m}$ $\bar{s}$ cone fundamentals through a linear transformation. An approximate 3x3 RGB-to-LMS transformation matrix (Eq. 6-25) was computed from the available average $\bar{r}$ $\bar{g}$ $\bar{b}$ CMFs and average $\bar{l}$ $\bar{m}$ $\bar{s}$ cone fundamentals of 47 Stiles-Burch observers.

$$\begin{bmatrix} \bar{l}_{10}(\lambda) \\ \bar{m}_{10}(\lambda) \\ \bar{s}_{10}(\lambda) \end{bmatrix} = \begin{bmatrix} 0.192325 & 0.749549 & 0.067573 \\ 0.019229 & 0.940909 & 0.113830 \\ 0.000054 & 0.010303 & 0.991441 \end{bmatrix} \begin{bmatrix} \bar{r}_{10}(\lambda) \\ \bar{g}_{10}(\lambda) \\ \bar{b}_{10}(\lambda) \end{bmatrix} \tag{6-25}$$

The cluster analysis was performed on the cone fundamentals, and the model cone fundamental functions were then converted into CIE 10° standard colorimetric observer equivalent CMFs through a 3x3 transformation. Again, an approximate 3x3 LMS-to-XYZ transformation matrix (Eq. 6-26) was computed from the available 1964 10° $\bar{x}$ $\bar{y}$ $\bar{z}$ standard colorimetric observer functions and the average $\bar{l}$ $\bar{m}$ $\bar{s}$ cone fundamentals of 47 Stiles-Burch observers.

$$\begin{bmatrix} \bar{x}_{10}(\lambda) \\ \bar{y}_{10}(\lambda) \\ \bar{z}_{10}(\lambda) \end{bmatrix} = \begin{bmatrix} 1.905378 & -1.321620 & 0.419512 \\ 0.698648 & 0.333043 & -0.013601 \\ -0.024300 & 0.040453 & 2.073582 \end{bmatrix} \begin{bmatrix} \bar{l}_{10}(\lambda) \\ \bar{m}_{10}(\lambda) \\ \bar{s}_{10}(\lambda) \end{bmatrix} \quad (6\text{-}26)$$

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 colorimetric 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} was 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 6-10. 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 five 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. Fig. 6-38 through 6-40 show the five cone fundamentals.

Note that the criterion of one unit of maximum color difference is somewhat arbitrary. A different criterion will likely result in a different number of model sets. Generally speaking, a color difference of less than 1 unit ΔE_{00} is not perceived as significant in most situations, so it is a reasonable criterion for the purpose of choosing the model sets. Further, increasing the number of model sets rapidly increases the total number of possible combinations of model functions, resulting in too many CMFs.

It should be pointed out here that for each cone fundamental, the cluster analysis was performed on data points in a restricted wavelength range that excluded the lower 10% spectral sensitivities or more. This was to ensure higher noise level in the observer data in either end of wavelength range did not affect the final clusters. For LWS cone fundamentals, the wavelength range of 520 nm - 650 nm was used, for MWS cone fundamentals, the wavelength range of 470 nm - 610 nm was used and for SWS cone fundamentals, the wavelength range of 410 nm - 490 nm was used.

Table 6-10. 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 colorimetric 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

Further, SWS spectral sensitivity values of the original Stiles-Burch observer data have poor accuracy at the wavelengths beyond 505 nm [14], which resulted in non-monotonic SWS model functions after cluster formation. To avoid this issue, SWS values of Stiles-Burch observer data below 0.005 in magnitude were ignored in the cluster analysis.

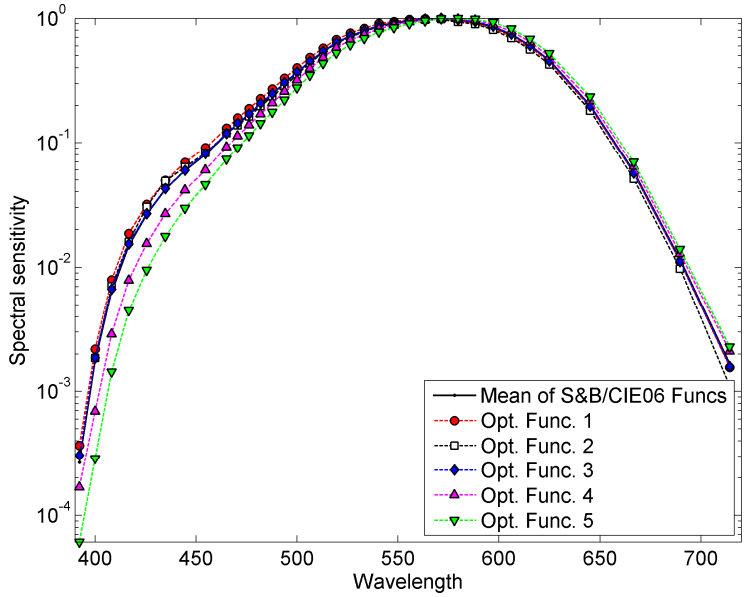


Fig. 6-38. Optimal set of five LWS cone fundamentals

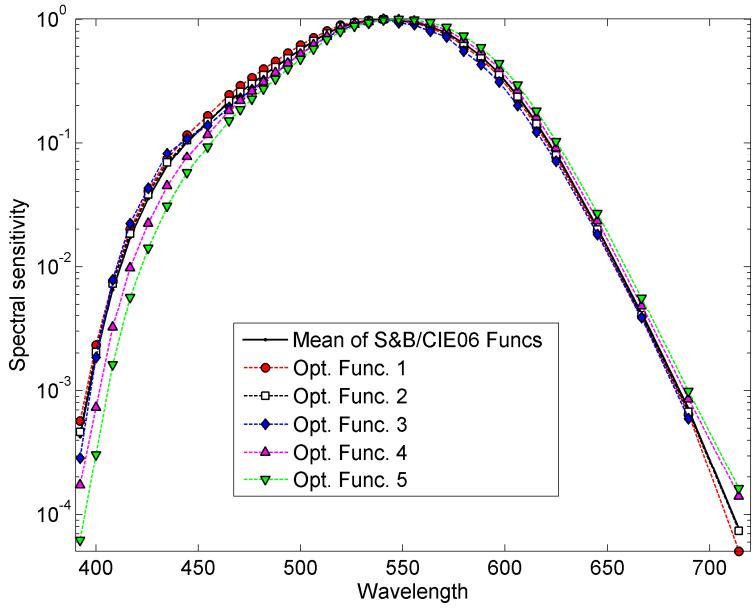


Fig. 6-39. Optimal set of five MWS cone fundamentals

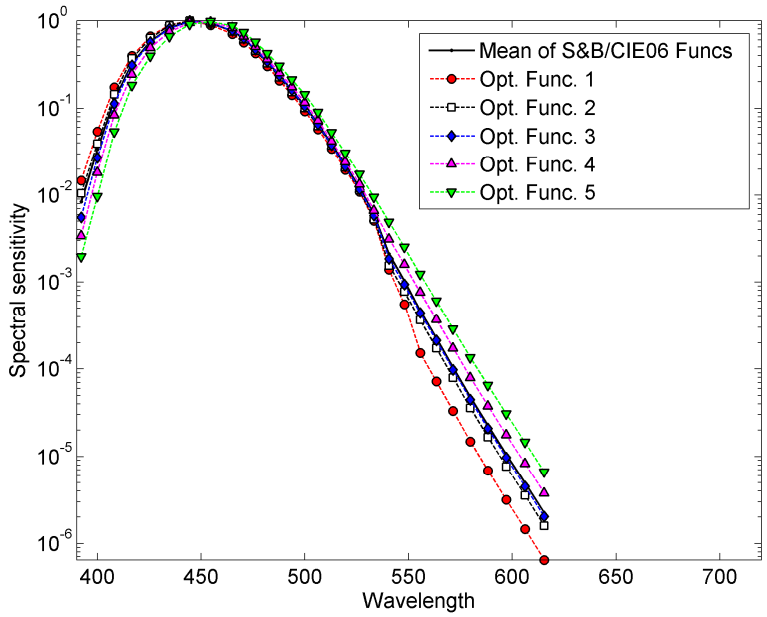


Fig. 6-40. Optimal set of five SWS cone fundamentals

6.3 Deriving Colorimetric Observer Categories – Step-2: identifying reduced sets of model CMFs

6.3.1 *Preliminary reduced set of seven CMFs*

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 set of color difference (ΔE_{00}) criteria for a given observer. Thus, for the said constraints, 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} ~1 criteria for any Stiles-Burch observer. The derivation of such reduced sets does not result in a unique solution, but is dependent on the color data set and the color difference criteria. 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, ΔE_{00} color difference equations have been used for deriving the reduced sets of model CMFs. While these ΔE_{00} equations correspond to CIE 10° standard colorimetric observer and do not fully hold for other observer models, it is hypothesized 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, corresponding to 125 possible observer CMF combinations. We must take into consideration that for some observers with atypical color vision characteristics, a given ΔE_{00} 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 ΔE_{00} threshold was computed using the 10^{th} or the 5^{th} percentile of the 125 ΔE_{00} values, whichever was below 1.2. This meant the worst 5% or 10% ΔE_{00} values would not be considered while deciding which observer categories could be assigned to a given Stiles-Burch observer. For six observers, the ΔE_{00} threshold computed this way was more than 2.0. However, these thresholds were still less than the ΔE_{00} values computed similarly with the CIE 10° standard

colorimetric observer, indicating that these specific Stiles-Burch observers were far away from the average of the population.

The suitability of a given CMF combination for any Stiles-Burch observer was determined by a "CMF Performance Index" (PI), based on the average percent deviation from the ΔE_{00} threshold (a positive PI indicated average ΔE_{00} 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 6-11 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 ΔE_{00} threshold for 17 observers, which is 36.2% of Stiles-Burch observer pool. Combination 58 met the ΔE_{00} threshold for another 14 observers, so combinations 2 and 58 together satisfied 66% of the Stiles-Burch 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.

Table 6-11. The reduced set of seven observer categories, their constituent average CMFs, and the total number of Stiles-Burch observers assigned to 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

6.3.2 *Updated reduced set of eight CMFs*

The spectral power distributions of the Colorchecker samples under D65 are broadband in nature, and so are unlikely to manifest significant observer variability. For deriving the reduced set of observer models with more precision, a better dataset was sought, and obtained from a color system with narrow-band primaries. The new Observer Calibrator prototype, described in Section 6.5, is an appropriate device for this purpose, since it is capable of producing highly metameric color signals. Accordingly, the 240 stimuli used in the second step were replaced by 5832 estimated spectral power

distributions obtained by using the LED primaries in the right half of the bipartite field of the prototype. These colors are characterized by high observer variability. As shown in Fig. 6-41, these color samples cover a wide color gamut formed by the prototype primaries. Rest of the method to derive the reduced set of observer models was the same as before [4].

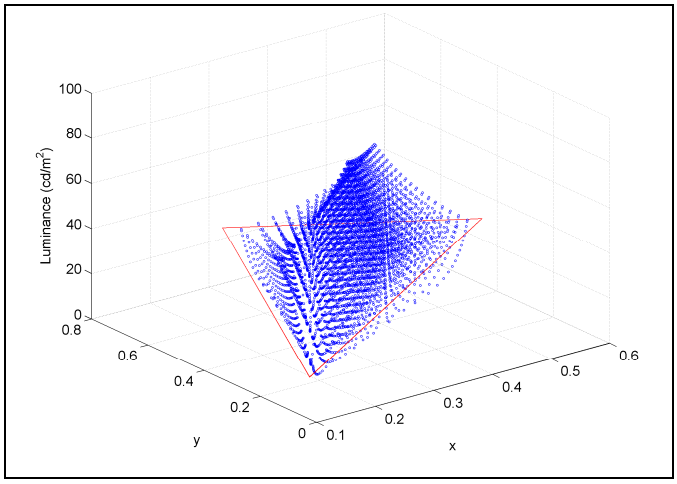


Fig. 6-41. Test color set obtained from Observer Calibrator Prototype

Using the new set of test colors, a total of eight colorimetric observer categories were obtained. In these categories, there are four unique x- functions (Fig. 6-42), three unique y- functions (Fig. 6-43) and four unique z- functions (Fig. 6-44), with more variability in the x- functions than in others.

As before, Table 6-12 shows which of the 125 combinations, and their constituent x-, y-, z- functions were picked for the reduced sets of 8 observer classes. Note that the 4th z- CMF from the cluster analysis was not included in this reduced set.

Table 6-12. The reduced set of eight observer categories, their constituent average CMFs, and the total number of Stiles-Burch observers assigned to various categories

Iteration	Combination	x-	y-	z-	Total Obs	%Obs Covered
1	10	1	2	5	14	29.8
2	62	3	3	2	14	59.6
3	53	3	1	3	6	72.3
4	27	2	1	2	6	85.1
5	36	2	3	1	3	91.5
6	11	1	3	1	2	95.7
7	83	4	2	3	1	97.9
8	76	4	1	1	1	100

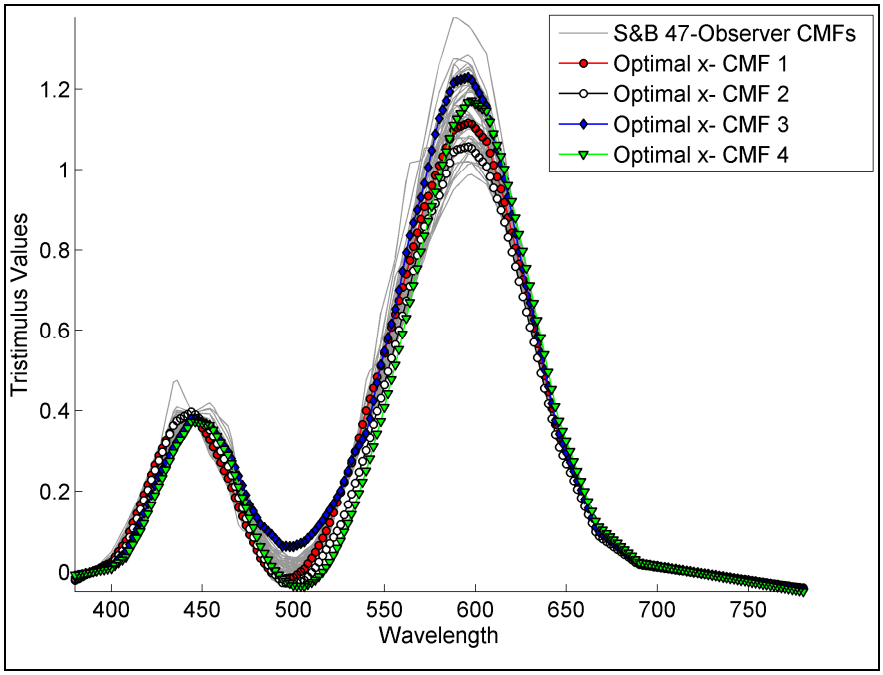


Fig. 6-42. Reduced set of four x- color matching functions plotted on Stiles & Burch 47-observer data.

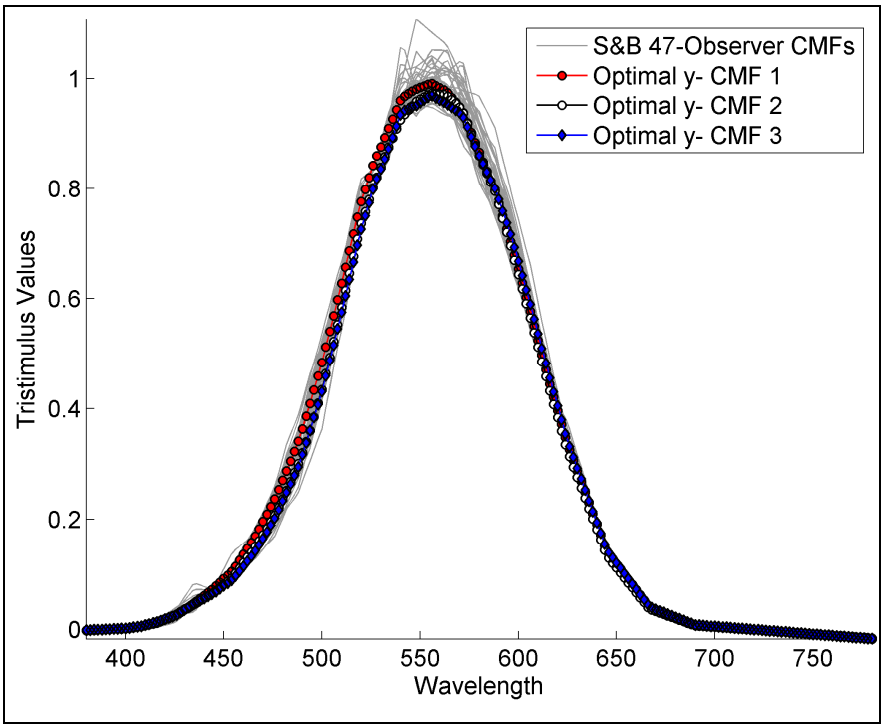


Fig. 6-43. Reduced set of three y- color matching functions plotted on Stiles & Burch 47-observer data.

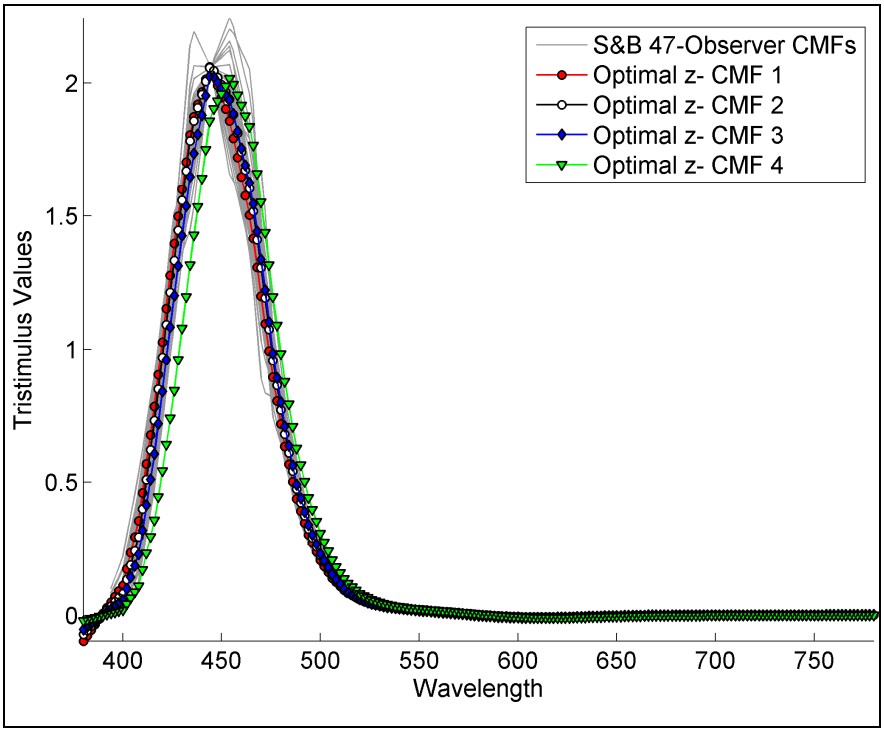


Fig. 6-44. Reduced set of three z- color matching functions plotted on Stiles & Burch 47-observer data.

6.4 Experimental method for classifying color-normal observers using displays

6.4.1 *The setup*

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. 6-45. 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 of Chapter 5 was used, shown in Fig. 6-46 and described in detail in the previous chapter.

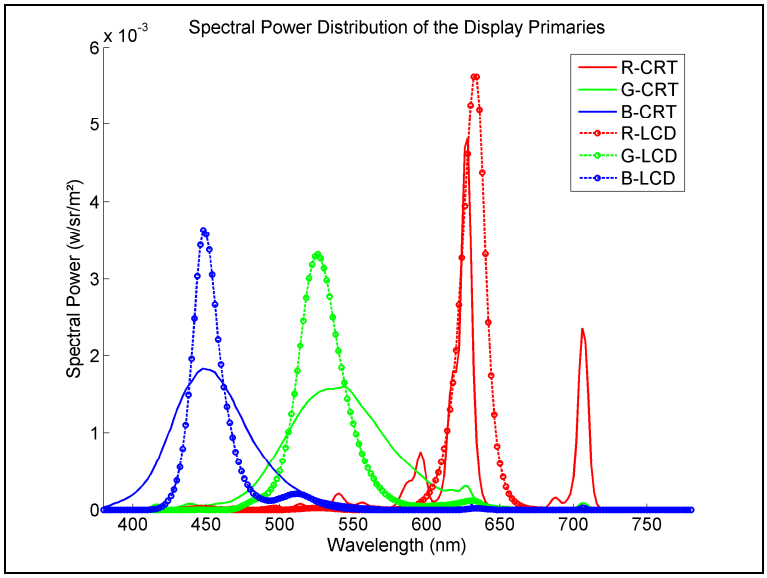


Fig. 6-45. Spectral power distributions of the primaries of the CRT display and LCD

The two displays were characterized using CIE 10° standard colorimetric observer and each of the seven observer categories. Thus, corresponding to each of the eight sets of CMFs (CIE 10° standard colorimetric observer + 7 new categories), a display forward and reverse model were determined.

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 color pairs 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 representation spaces, 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.

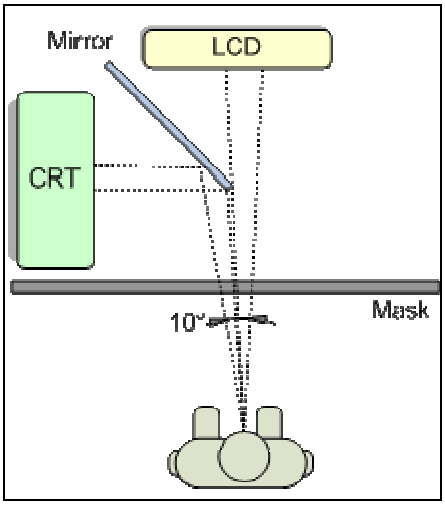


Fig. 6-46. Experimental setup

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.

Thirty observers took part in the observer classification experiment, including the ten observers who participated in the preliminary color matching experiments described in Chapter 5. 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 colorimetric 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 various categories into one of three groups, namely, *unacceptable*, *acceptable* and *satisfactory*. This was accomplished in several steps, by: i) going through various category-specific colors 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.

6.4.2 *Results*

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 6-13 shows two examples of such table for observers #1 and #8. Note that the category 1 is the CIE 10° standard colorimetric 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. 6-27, and included in Table 6-13. 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.

$$\begin{aligned} R_i &= 80S + 20A - 100U \\ U &= 1 - S - A \\ \Rightarrow R_i &= 180S + 120A - 100 \\ R_i' &= \frac{100R_i}{\sum_i R_i} \end{aligned} \tag{6-27}$$

Table 6-13. 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 colorimetric observer)

Ranking	Category							
	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
R'_i Score	36	-179	100	-93	86	-121	-164	29

Ranking	Category							
	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
R'_i Score	82	88	40	100	88	-92	94	-32

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. 6-47, where each bubble corresponds to a category, and the area of a bubble is proportional to its relative score Ri'. 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.

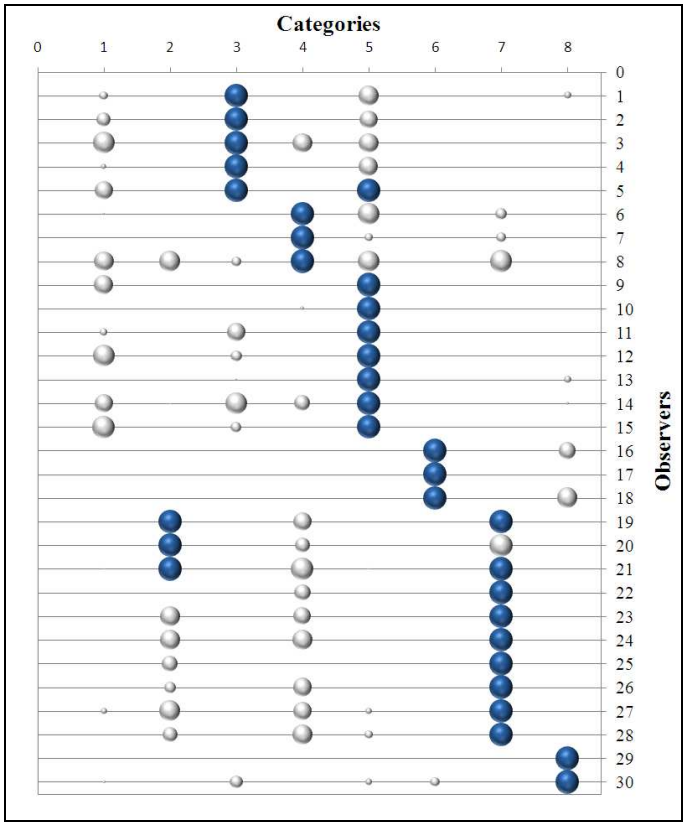


Fig. 6-47. Observer categories as determined through the observer classification experiment (category 1: CIE 10° standard colorimetric observer)

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 (not shown), 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. 6-47, 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 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 colorimetric 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 colorimetric 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 in this experiment, the CIE 10° standard colorimetric observer model is definitely outperformed by other categories for these observers.

When considered alone, the CIE standard colorimetric 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 colorimetric observer. It is possible that given a choice, many observers would prefer a category different from the CIE 10° standard colorimetric observer. A possible explanation for the low preference for the CIE 10° standard colorimetric 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 colorimetric observer as per the observer classification experiment. Category 7 is quite close to category 2, which, as per previous analysis, was the dominant category for the Stiles-Burch observers.

These results raise two fundamental questions: 1) should the standard colorimetric 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 colorimetric observer continue to satisfy our needs today, or is it time to have a provision for multiple observer models in applied colorimetry, and if so, how?

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 colorimetric observer will probably still be reasonably good (see Chapter 5), but is it really the best possible representation of the color-normal population?

This thesis attempts 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.

The results from this first phase of observer classification experiment definitively confirmed the existence of observer metamerism issue in modern displays with narrow-band primaries. But more importantly, they also showed that such display systems could be exploited to better predict the variability in individual observers.

6.5 Observer Classification using Observer Calibrator Prototype

6.5.1 *The prototype*

The display-based setup was not convenient enough to be used in industrial applications. Thus a portable, LED-based instrument prototype for observer classification was conceived. This prototype replicated the observer classification experimental setup based on two displays (one with broadband primaries and the other with narrow-band primaries), described in the previous section.

The prototype configuration is shown in Fig. 6-48. The actual prototype is shown in Fig. 6-49. The illumination system in the prototype is primarily composed of two clusters of four LEDs. Out of the four LEDs in each half-field, one is a white LED and is used only for generating an adaptation field. Two adjacent integrating boxes (IB1, IB2) are designed for light mixing so that uniform colors can be obtained in each half-field. The colors can be viewed monocularly in the 10° bipartite field (F1, F2).

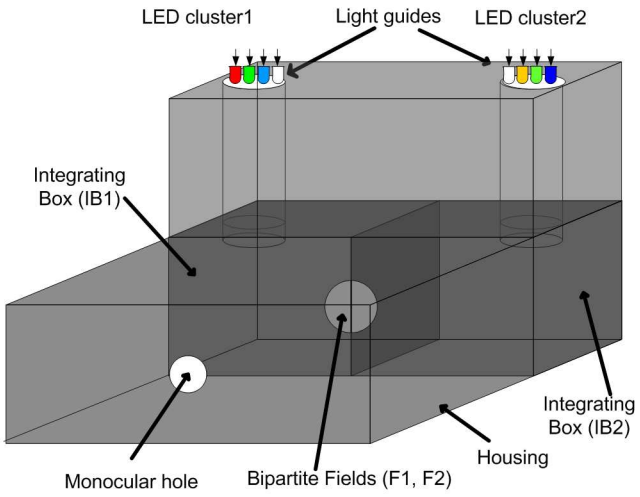


Fig. 6-48. Configuration of the Observer Calibrator Prototype

This prototype is made of different materials. There are two hollow cubical integrating chambers. Each cube is made of three layers: the outer layer of black paper, the middle insulating layer of Mylar and the inner layer of white paper with 85% reflectance. On the top of each integrating chamber, there is a nozzle made of Spectralon (highly diffusing fluoro-polymer) which is surrounded outside by Mylar. LEDs are placed inside the nozzles. The remainder of the prototype is made of black paper and cardboard.

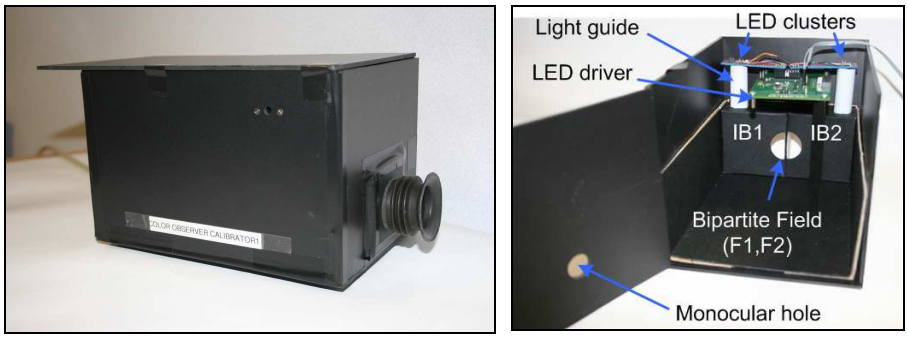


Fig. 6-49. External and internal views of the Observer Calibrator Prototype

A key consideration in the design of the prototype was to achieve high luminance uniformity for both halves of the bipartite field. Luminance was measured on four points along the periphery of each field, as shown in Fig. 6-50. The sizes of the integrating chambers and the LED positions were adjusted until satisfactory luminance uniformity could be obtained. The final dimensions of the prototype are given in Fig. 6-51. A luminance uniformity of 6% was achieved, in other words the max luminance difference between the four points in Fig. 6-50 was 6%.

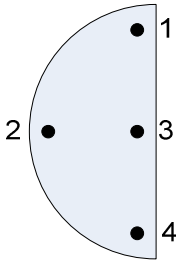


Fig. 6-50. Luminance uniformity measurement points in each half of the bipartite field

The LEDs for the two fields needed to be selected in such a way that the peak wavelengths of the LEDs in one field (F2) fell in the region of high variability in the observer categories, while those of the LEDs in the other field (F1) coincided with region of low variability in the observer categories. This would ensure that an observer looking at different versions of color matches in the prototype would find the left half of the bipartite field relatively constant, while the right half would tend to change. An additional analysis was performed to identify the wavelength regions of x-, y- and z-functions with highest variability among the different observer classes. Fig. 6-52 shows the x-, y- and z- functions of different observer categories and the CIE 10° standard colorimetric 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 426 nm have high variability in case of x-, y- and z- CMFs respectively.

Results from this analysis can be compared with Thornton's [103] "prime-color" and "antiprime" spectral regions of 452-533-607 nm and 497-579-653 nm, respectively. Thornton observed that the CIE 10° standard colorimetric observer performed well when the incoming light composed of the spectral regions near the prime colors [104]. In contrast, it performed poorly in presence of the spectral content in the antiprime region. The prime-color spectral regions fall near the peak wavelengths of the CMFs in Fig. 6-52. Out of the three prime-color wavelengths, 607 nm falls in a zone where x- CMF has high variability. Out of the three antiprime regions, only 579 nm seems to coincide with a region of high observer variability. A primary at 497 nm might have been problematic not because of observer variability, but because of very low contribution of x- CMF. On the other hand, 653 nm was chosen as an antiprime wavelength to avoid other regions already chosen as prime-color and antiprime regions, while retaining reasonable visual response [103]. Thus, the Thornton's antiprime wavelengths either fall in the regions of high observer variability or in the regions of low spectral sensitivity, which can be an explanation of Thornton's observations.

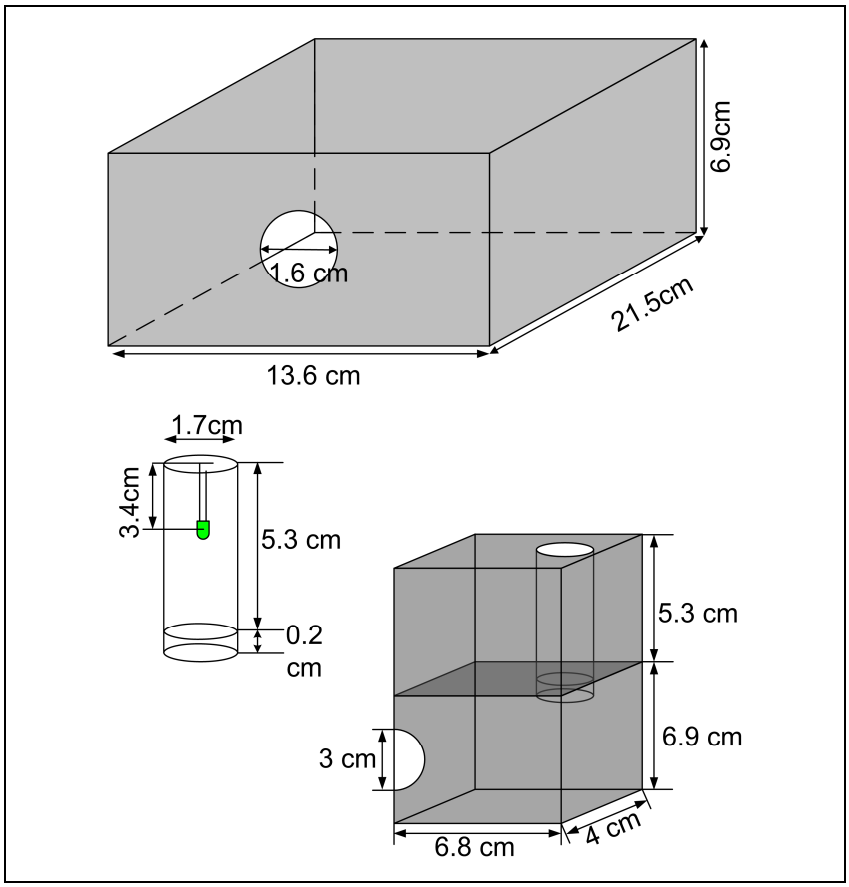


Fig. 6-51. Dimensions of various parts of the Observer Calibrator Prototype

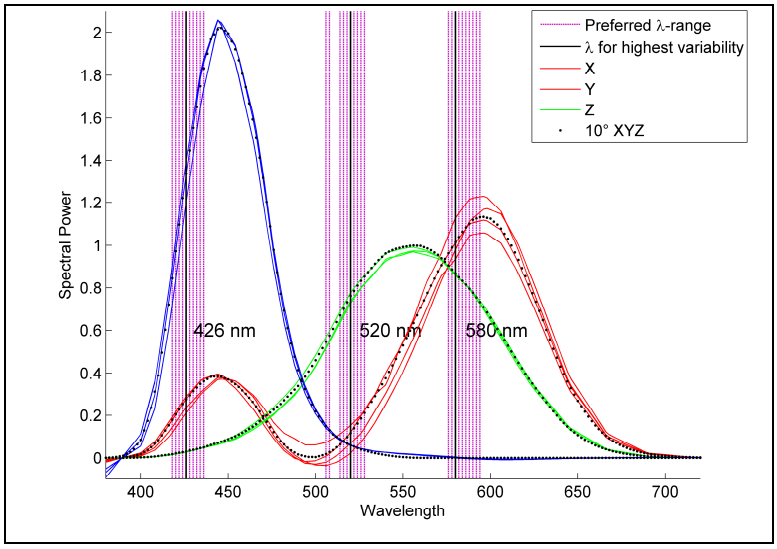


Fig. 6-52. Wavelength regions of x-, y- and z- functions with high variability among various observer categories

However, selection of LEDs for either of the two fields (F1 and F2) had two additional constraints. The first constraint was related to the common gamut of the two fields. The spectral power distributions of the three colored LEDs in a given field at full power determined the chromaticities of primaries, which in turn defined the color gamut achievable for that field. It was important that the two fields had significant amount of common color gamut, otherwise it would be impossible to find a color match between the two fields for different base colors.

The second constraint was related to the luminance level parity between the two fields. The visual task of color matching dictated that there be equivalence in the luminance levels on the two halves of the bipartite field, which required the total peak luminance due to all LEDs on both fields be similar. This constraint was partly overcome at the hardware level where current flowing to individual LEDs could be halved (from 20 mA to 10 mA) by setting appropriate registers, allowing some level of control over the peak luminance of individual LEDs. Still, this imposed a restriction on the choice of LEDs.

Because of the above two constraints and because of unavailability of LEDs with certain peak wavelength, not all LEDs matched the desired characteristics. For example, no blue LED was found with a peak wavelength of around 420 nm and with appropriate power level. Thus, some compromise had to be made in LED selection. Table 6-14 gives colorimetric specifications of the LEDs chosen for the two fields. Note that the luminance values listed in this table are post-hardware adjustment.

Table 6-14. Colorimetric specifications of LEDs chosen for the two fields

	LED	Peak Measured λ (nm)	Peak Luminance (cd/m²)	Chromaticities	
				x	y
Field 1	Blue	470	61.4	0.1186	0.1549
	Green	506	69.8	0.1270	0.6665
	Red	644	26.2	0.6990	0.3010
	White	-	124.3	0.3369	0.3381
Field 2	Blue	462	23.4	0.1349	0.0982
	Green	518	70.2	0.2315	0.6870
	Orange	594	59.7	0.5925	0.4075
	White	-	124.3	0.3369	0.3381

Fig. 6-53 shows the spectral power distributions of the observer calibrator primaries in the two fields, and the color gamuts obtained from them is shown in Fig. 6-54. The central points in the gamut are obtained by adding up the response of the three LEDs at peak power. The color gamuts are significantly larger than typical display gamuts or Rec. 709. Note that the white in Fig. 6-53 is the white LED used for generating the adaptation stimulus for both half-fields.

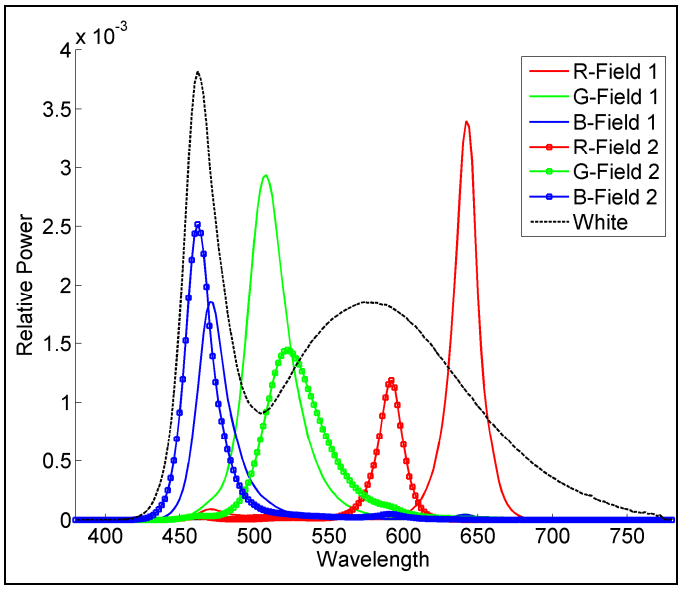


Fig. 6-53. Spectral power distributions of the LEDs used in the Observer Calibrator Prototype

The white LED was more powerful compared to others, so for the adaptation field only about 50% power was used. The luminance values of the adaptation field on the narrow-band and broad-band side were close to each other.

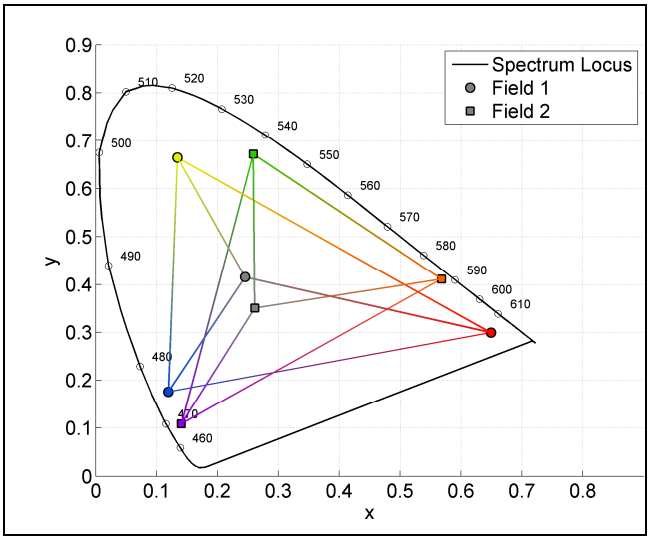


Fig. 6-54. Gamuts of the LED primaries in the Observer Calibrator Prototype

The schematic of the prototype along with the computer interface is shown in Fig. 6-55. The prototype has an LED driver that controls the LEDs. The LED driver is interfaced to a computer. A software application (very similar to the one described in Chapter 5) 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. There is a user control device connected to the computer through Universal Serial Bus (USB) that allows the observer to browse through various versions of a color match corresponding to individual categories.

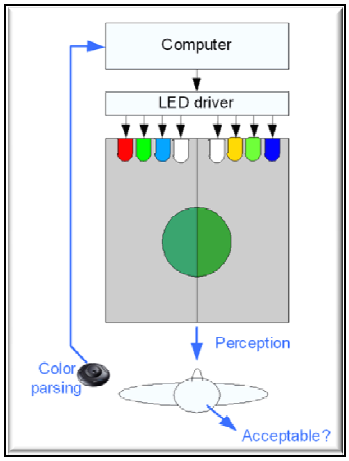


Fig. 6-55. Schematic of the Observer Calibrator Prototype with computer interface

6.5.2 *Observer Classification Method*

Nine matching colors were produced in each half of the 10° bipartite field corresponding to nine observer models, namely the CIE 10° standard colorimetric observer and the eight reduced set of observer categories (as determined in Section 6.3.2). As in the display-based setup, the test software

allowed the nine versions of color matches to be presented in a random order in each trial, allowing the observer to browse through them with the help of the user control. His or her task was then to follow a multi-step method and classify these nine versions of color matches into *Superior*, *Average* or *Inferior* categories. The names of the categories were changed from the earlier *Unacceptable*, *Acceptable* and *Satisfactory* since an acceptability judgment was more subjective compared to a superiority judgment. The latter can be thought of as a relative ranking among the available color matches. Based on several such trials (for different base colors), the category that most often produces the best match is identified, and is the category assigned to the given observer. Eight base colors, shown in Fig. 6-56, were selected for the experiment using the same method as described in Section 6.4.1.

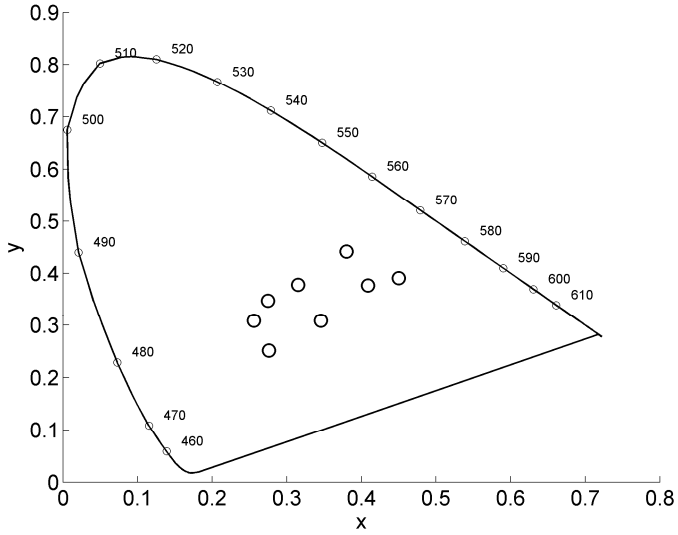


Fig. 6-56. Eight test stimuli used in Observer Classification test performed with the Observer Calibrator Prototype

6.5.3 *Two experiments with the Observer Calibrator*

Two observer classification experiments were performed using the Observer Calibrator prototype. The experiments were performed with collaboration with two universities, one in Germany and the other in Hungary.

The first experiment was performed at *the Institute of Printing Science and Technology (IDD), Technische Universität Darmstadt* in Darmstadt, Germany. Twenty-seven observers (10 female and 17 male observers with an average age of 34.5 years) with normal color vision participated in this experiment. The second experiment was performed at the *University of Veszprém*, in Veszprém, Hungary. Twenty-two observers (5 female and 17 male with average age of 28.8 years) with normal

color vision participated in this experiment. For each observer, the experiment took 40-45 minutes on an average.

6.5.4 *Results*

Fig. 6-57 shows the observer categories for all 49 observers from the two tests obtained using the same method as outlined in Section 6.4.2. Results from the Darmstadt experiment are shown on the left, and those from the Veszprém experiment is shown on the right. Note that these categories are not the same as in display-based experiment, shown in Fig. 6-47.

Table 6-15 summarizes the results obtained from the two experiments, involving a total of 49 observers. Categories 4, 5 and 6 are the most populated. Together they cover 63% of this observer panel. Categories 3, 7 and 9 are the least populated.

As before, category 1 (CIE 10° standard colorimetric observer) was the assigned category for a minority of observers, only 4 out of 49 (around 8%).

Categories 8 and 9, known to be quite distinct from the other categories, were assigned to 6 out of 49 observers (around 12%). A 22-year old observer from Veszprém was assigned to category 8, but for all other observers belonging to these two categories the average age was 52.

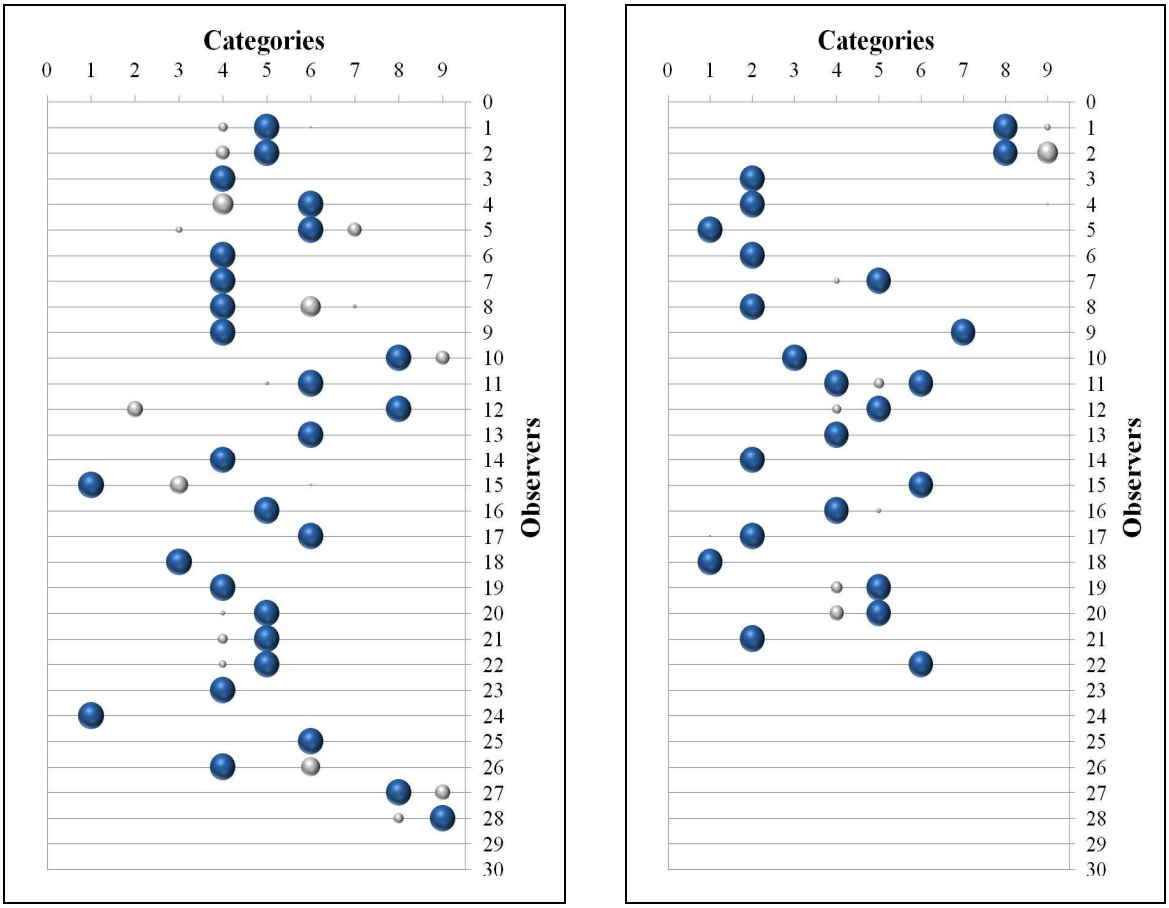


Fig. 6-57. Results from observer classification experiment in Darmstadt, Germany (left) and Veszprem, Hungary (right) using the observer calibrator (category 1: CIE 10° standard colorimetric observer)

Table 6-15. Observer classification result summary based on 49 observers

Observer Categories	No of Observers	% Observers
Cat-1 (CIE 10° standard)	4	8.2
Cat-2	7	14.3
Cat-3	2	4.1
Cat-4	12	24.5
Cat-5	10	20.4
Cat-6	9	18.4
Cat-7	1	2
Cat-8	5	10.2
Cat-9	1	2

6.5.5 *Limitations of current prototype*

The current version of the prototype has two main limitations. First, the LED selection for the two fields was constrained by the availability of LEDs with specific wavelengths. Thus, not all LEDs could satisfy the design criteria.

The second limitation was related to the hardware. The LED driver being 8-bit could provide only 8-bit resolution for controlling the LED input power. This imposed a constraint on precise color reproduction. A 10-bit LED driver would be more appropriate for this prototype.

6.6 Validation of Observer Classification method

An observer classification experiment was planned in collaboration with the IDD in Darmstadt, Germany with two principal aims. First, to validate the observer classification method with the help of the results obtained from an independent visual experiment, and second, to probe an interesting but unanswered question: do color matching functions (CMFs) influence our perception of small suprathreshold color differences? If so, what is the extent of this influence? The threshold discrimination (also referred to as just noticeable distance or JND), which is a measure of uncertainty and variability, is typically determined by color matching, and thus by individual color matching functions. For larger color difference (suprathreshold), it is assumed that the influence of the color matching function on the perceived color difference decreases continuously. Thus, we can assume that there is some impact of CMFs on perceived small suprathreshold color differences in addition to higher order processes. Under this hypothesis, small color difference judgments viewed on a display with narrow-band primaries should be significantly influenced by individual variability in color matching functions. Such an experiment was performed earlier by Urban et al [141]. A significant correlation between the small color difference judgments and the observer categories would help validate the observer classification method, and also provide support to hypothesis that there exists a relationship between small color difference judgments and color matching functions.

Prior to describing the correlation analysis of the results obtained from the two experiments, it would be useful to briefly discuss the setup for the color difference experiment.

6.6.1 *Setup for the color difference experiment by Urban et al [141]*

A color-difference experiment was performed on a liquid crystal display (LCD) prior to conducting the collaborative observer classification experiment. The method of constant stimuli [142] was used to determine color differences around five CIE color centers (CIE Gray, CIE Red, CIE Yellow, CIE Green, and CIE Blue [143]). These color differences were perceived equally to the color difference in the anchor pair consisting of two neutral gray tones with a color difference of $\Delta E_{ab} = \Delta L^* = 2.2$. 14

directions around each color center were investigated (Fig. 6-58). Along each direction five test colors were chosen resulting in 14x5 = 70 color comparisons for each color center and a total of 350 comparisons for the whole experiment.

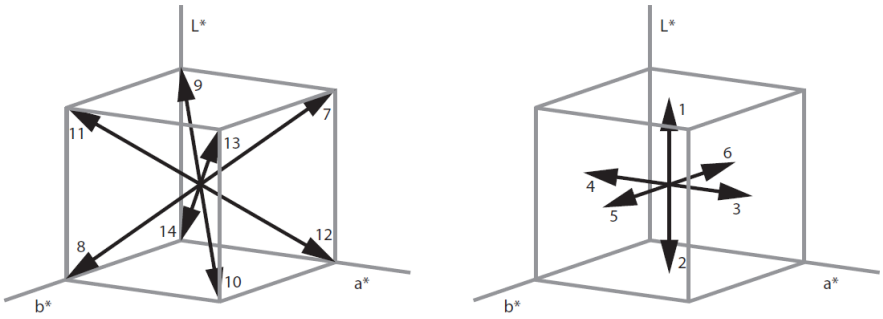


Fig. 6-58. Investigated directions around the color centers [141]

The arrangement of patches is shown in Fig. 6-59. Each test pair was composed of a color center and one of the test colors. The patches covered approx. 10° of the visual field. The positions of the displayed test and anchor pairs were switched randomly as well as the color positions within the pairs. Observers were asked to choose the color pair (anchor or test pair) with the largest perceived color difference. A detailed description of the experiment and results can be found in [141].

Details of the observer classification experiment have already been explained in the previous sections.

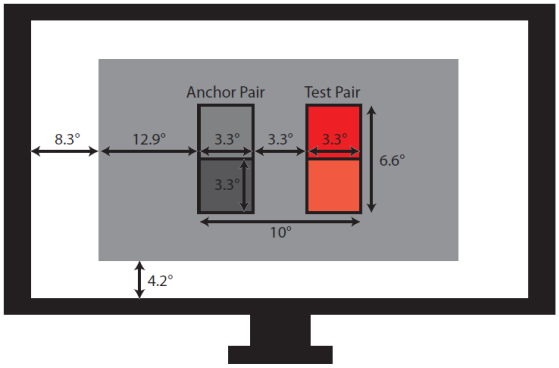


Fig. 6-59. Experimental setup on LCD - degree of visual field [141]

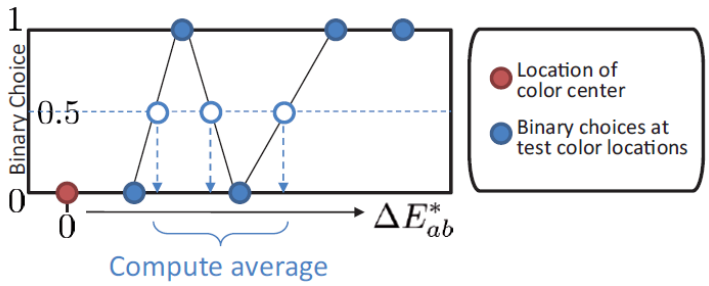


Fig. 6-60. Calculation of the individual threshold for a single observer. If there are multiple crossings of the 0.5 line, the corresponding ΔE^*_{ab} distances are averaged for thresholding [141]

6.6.2 *Correlation of color difference data with colorimetric observer categories*

Results from the observer classification experiment have been presented in Section 6.5.4. In analyzing the data from the color difference experiment [141], an individual threshold was calculated based on the binary choices of each single observer as shown in Fig. 6-60. However, note that the computed individual color difference thresholds are biased by quantization errors due to the small number of binary choices. The color difference between the color center and the color, indicated by this threshold, is perceived by the current observer similar to the color difference of the anchor pair. Based on the individual thresholds an average observer was calculated for the observer panel. Fig. 6-61 shows the mean deviation of individual observer thresholds from the average observer threshold, where individual data points are marked by the assigned colorimetric observer categories for each observer. Note that here average thresholds and individual thresholds are calculated in the CIELAB color space. As shown in the diagram, two observers belonged to category 1 (CIE 10° standard colorimetric observer), one to category 3, eight to category 4, six to category 5, six to category 6, three to category 8 and one in category 9 (see also Fig. 6-57 left panel). No observer belonged to categories 2 and 7. Thus categories 4, 5 and 6 were most popular. The two observers belonging to category 1 are closer to the standard colorimetric observer than others in this observer population.

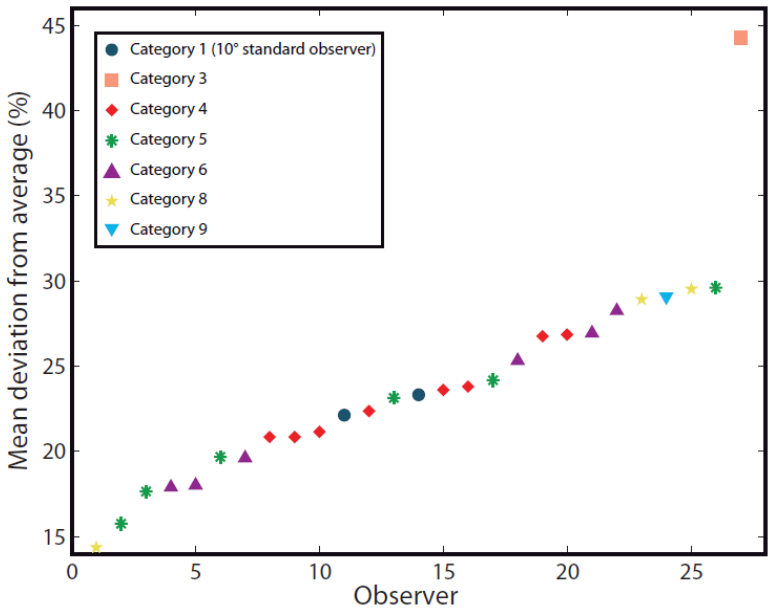


Fig. 6-61. Mean deviation of individual observer thresholds from the average observer threshold

An interesting observation about Fig. 6-61 is if we assume the two individual points belonging to category 1 as references, other categories seem to be symmetrically placed above and below these two points. For example, category 4 is tightly spread above and below the category 1 points, while category 6 points are further away beyond the category 4 points both above and below, and category 5 points are distributed over a wider range. Category 3, 8 and 9 are farthest away from category 1 points, either above, below or both. This indicates that there exists a link between the observer categories and individual observer's color difference perception.

Separate analysis was conducted for investigating the correlation between average observer color difference thresholds and observer categories. This analysis involved the use of CMFs for various categories, thus CIEXYZ color space was preferred over CIELAB since the conversion of CIEXYZ to CIELAB is valid only for 2° or 10° standard colorimetric observer. The XYZ coordinate system on the other hand is purely computational, and can be defined for any specific CMF. The xyY chromaticity diagram is defined by the specific monochromatic primaries used in obtaining the original color matching functions. Since all observer categories are essentially based on Stiles-Burch 10° CMFs, chromaticity coordinates obtained by using individual categories can be compared and even plotted on the same diagram. However, it is important to note that the distances in CIEXYZ color space are not representative of color differences perception, and the scale is not uniform in different areas of color space.

At this point, a perceptual space for these categories does not exist, and so there is no appropriate perceptual metric available to us. From the spectral power distributions (SPDs) of all test stimuli and the color matching functions (CMFs) for each of the nine categories (category 1 being CIE 10° standard colorimetric observer), CIEXYZ and category-specific XYZ (henceforth CatXYZ) values were computed. Since all observer thresholds were originally computed in CIELAB using the CIE 10° standard colorimetric observer, these needed to be converted to catXYZ. For each category, a 3x3 transformation matrix (M_{Cat}) was computed in a least square sense from CIEXYZ ($XYZ_{stimuli,Std}$) and CatXYZ ($XYZ_{stimuli,Cat}$) data of all color stimuli obtained earlier. Observers' average color difference threshold data were then converted from CIELAB ($Lab_{stimuli,Std}$) to CIEXYZ, which were then converted to CatXYZ by multiplying with the transformation matrices. Eq. 6-28 explains these two steps.

$$M_{Cat} = XYZ_{stimuli,Cat}^{-1} \cdot XYZ_{stimuli,Std}$$
$$LAB_{stimuli,Std} \rightarrow XYZ_{obsAv,Std} \xrightarrow{M_{Cat}} XYZ_{obsAv,Cat} \quad (6\text{-}28)$$

These computations allow us to plot the observer data organized by categories with coherence in scale. In this analysis, root-mean-square (RMS) distances between XYZ coordinates of observer color difference thresholds and color centers have been considered, with the hypothesis that around a given color center, small color differences in a given direction can be assumed to be Euclidean. Figs. 6-63 through 6-67 show the RMS distances between the test colors along various directions and different color centers. Fig. 6-62 explains the symbols used in these figures, showing the five test colors for a single direction represented in terms of RMS distances of various category-specific XYZ values.

In each figure, the black central line is the color center. All XYZ RMS distances are measured from this color center and represented on two sides of the central line. The fourteen directions (see Fig. 6-58) are organized in pairs along the ordinate, with each direction having seven colored lines corresponding to various categories. Note that the first line is for CIE 10° standard colorimetric observer, and the rest are for categories 3, 4, 5, 6, 8 and 9, as shown in Fig. 6-62. Categories 2 and 7 are not present since no observer was assigned these categories in this experiment. Each colored line joins the five test stimuli in a given direction for any given category, shown as black dots. As mentioned before, the RMS distances in this figure are not perceptual. But conveniently, comparing the lengths of these lines gives an idea of the relative distance scales in various directions and categories for a given color center, thus allowing us to compare the RMS distances of observer thresholds. For the first colored line in each figure, all circles and the star represent CIEXYZ RMS distances (Cat. 1), while for the rest the circles represent the RMS distances in respective CatXYZ spaces. Note that while the RMS distances corresponding to different categories along a given direction can be compared, distances along various directions should not be compared to each other.

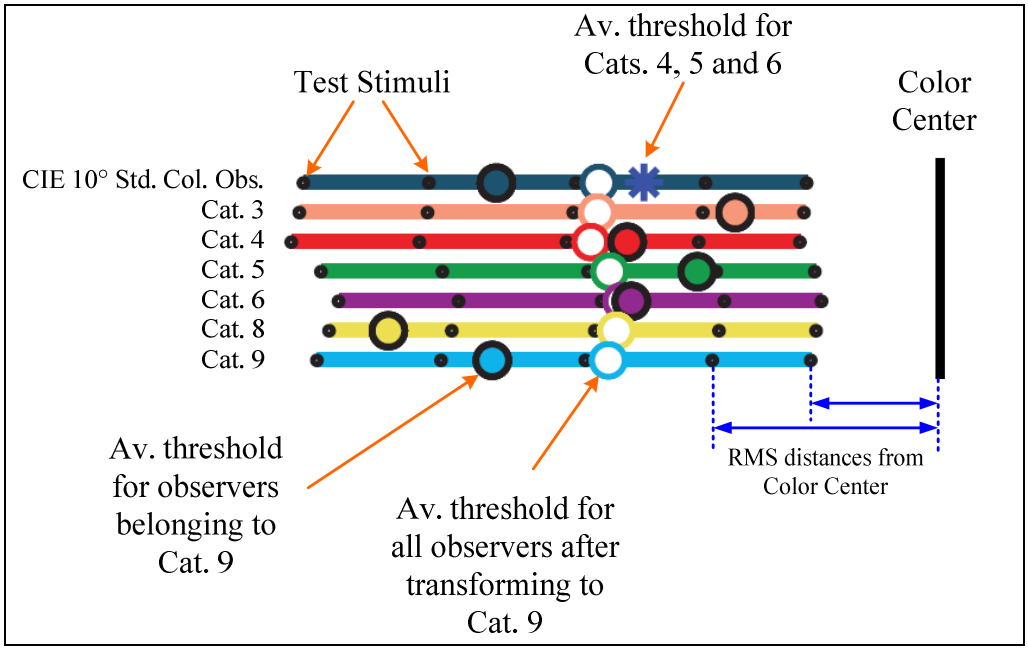


Fig. 6-62. Explanation of symbols used in Figs. 6-63 through 6-67. Each of the 14 directions has similar representation. All distances are RMS distances in CIEXYZ or CatXYZ coordinate systems.

The empty circles represent the color difference threshold (see Fig. 6-60) along a given direction and a given color center, averaged over all observers. To compute these color difference thresholds (T_{catX}) for any given category (CatX), Eq. 6-29 below was used. First, all intra-category average thresholds (T_{cat1}, T_{cat2} etc) were converted to corresponding CIEXYZ thresholds for the given category, using transformations given in Eq. 6-28. The resulting thresholds were then multiplied by the number of observers belonging to the respective categories (N_{cat1}, N_{cat2} etc), then summed, and then divided by the total number of observers. Such category-wise weighting takes into account the fact that the categories were not equally populated, so more weights were assigned to average thresholds coming from more populated categories. This weighted threshold corresponded to average color difference threshold in CIEXYZ, which was then multiplied by the transformation matrix (M_{catX}) from Eq. 6-28 to obtain the thresholds (T_{catX}) for the given category CatX.

$$T_{CatX} = \frac{\sum_j \left[\left(M_{Catj}^{-1} \cdot T_{Catj} \right) \cdot N_{catj} \right]}{\sum_j N_j} \cdot M_{CatX} \qquad (6\text{-}29)$$

The filled circles represent similar RMS distances where color difference thresholds are computed for observers grouped by their assigned categories (T_{cat1}, T_{cat2} etc in Eq. 6-29). So the filled circles in the 1st line are for observers belonging to Category 1 (CIE 10° standard colorimetric observer), filled circles in the 2nd line are for observers belonging to Category 2, so on and so forth.

Finally, the blue star on the first line are for color difference thresholds computed only for observers belonging to the three dominant categories, namely 4, 5 and 6. As before, weightings based on number of observers were applied.

To summarize the foregoing discussion, the empty circles represent the global average observer thresholds obtained by transforming all intra-category average threshold values to a given category, and the filled circles simply represent the intra-category average observer threshold only for a given category. The distances between the empty circles and the filled circles for any category indicate how different this category is from the averaged observer data.

Typically, categories 3, 8 and 9 have the largest distances between the empty and filled circles. In the observer classification experiment, observers of categories 8 and 9 rejected color matches corresponding to the CIE standard colorimetric observer with high certainty, for all seven test colors. This bolsters the inference that these categories are indeed quite different from the standard colorimetric observer. In this experiment, all observers belonging to these two categories were in the highest age-group, but other experiments (see Section 6.5.3) have indicated that some young observers can also belong to these categories.

Only one observer was assigned category 3. For this category, distances from the color center are often less than that in case of other categories, which may indicate the observer had better color discrimination than average observers in other categories. However, as per Fig. 6-61, this observer had the highest deviation from the mean color difference threshold, which is also consistent with Figs. 6-63 through 6-67. Because of statistical insufficiency of the data, these observations cannot be considered as general inferences with regard to category 3.

In many cases, RMS distances between global average thresholds (empty circles) and intra-category average thresholds (filled circles) for category 1 (CIE 10° standard colorimetric observer) are larger than those in case of categories 4, 5 and 6, which indicates observers belonging to category 1 are relatively further away from the average observer data. This indicates that the perception of such individuals can still be strongly distinct from the statistical mean of a certain observer population. On the other hand, color difference thresholds averaged for observers in categories 4, 5 and 6 (blue stars), are in general significantly closer to the global average. Over 70% observers belonged to these three categories.

In a perceptual color space optimized for each category the average color difference thresholds (filled circles) would ideally form a vertical line, all thresholds being at the same distance from the color center. But this is not the case here. The transformations between CIEXYZ and CatXYZ are approximate. As explained before, the average color difference threshold computations were

performed in the CIELAB space, and was then converted to CIEXYZ and then to CatXYZ. Thus, transformed threshold points for some categories do not always fall on the colored lines, implying the RMS distances in CatXYZ space can in some cases exceed the distance of farthest or nearest test stimulus.

Table 6-16. Absolute difference between global observer average thresholds and intra-category average thresholds for Color Center 5 (Blue)

Direction	Category						
	1	3	4	5	6	8	9
1	0.20	0.20	0.20	0.26	0.35	0.03	2.65
2	0.71	0.29	0.16	0.29	0.28	0.69	1.45
3	0.02	0.10	0.02	0.00	0.02	0.04	0.04
4	0.03	0.07	0.01	0.00	0.00	0.01	0.04
5	0.02	0.90	0.20	0.17	0.44	0.80	0.47
6	0.30	0.06	0.17	0.18	0.06	0.34	0.39
7	0.21	2.24	0.28	0.11	0.05	0.45	1.71
8	0.10	1.66	0.39	0.83	0.09	0.66	0.70
9	0.08	0.11	0.08	0.11	0.08	0.11	0.18
10	0.28	0.24	0.24	0.11	0.24	0.63	0.52
11	0.12	0.47	0.26	0.20	0.77	0.13	0.46
12	0.44	0.92	0.11	0.14	0.44	0.10	0.80
13	0.47	0.47	0.47	0.53	0.06	0.43	0.94
14	0.27	0.52	0.06	0.10	0.12	0.34	0.37

The distances between global average thresholds and average thresholds within categories are typically larger for categories 3, 8 and 9 compared to other categories. This implies that there exists possibility to improve average color difference prediction for observers belonging to these categories, if we can use color matching functions that are more appropriate than the standard colorimetric observer. These observers stand to gain the most by a practical implementation of the concept of observer classification. Results for the blue color center (color center 5) are particularly interesting in this regard. The absolute difference between global average thresholds and average thresholds within categories for this color center are shown in Table 6-16 (see also Fig. 6-67). The shaded values indicate category-specific average thresholds having large differences with respect to the global average thresholds. For categories 3, 8 and 9 these distances are relatively large along several directions, which is an indication that observers in these categories will tend to have high disagreement in color difference judgment in blues (color center 5) with the rest of the population.

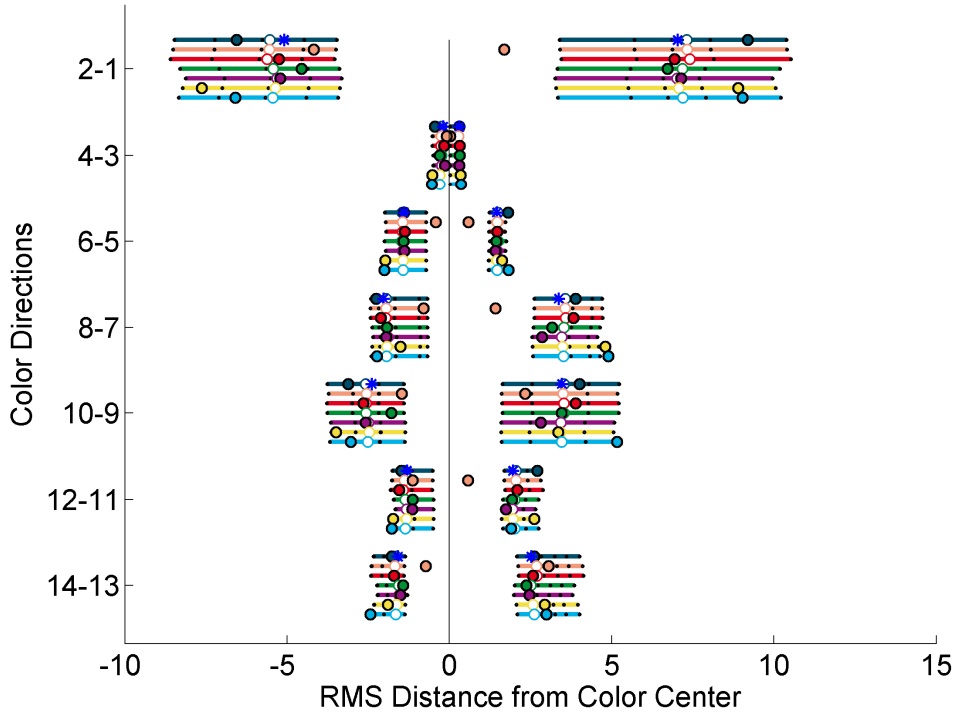


Fig. 6-63. RMS Distances in XYZ coordinate system for Color Center 1 (Gray)

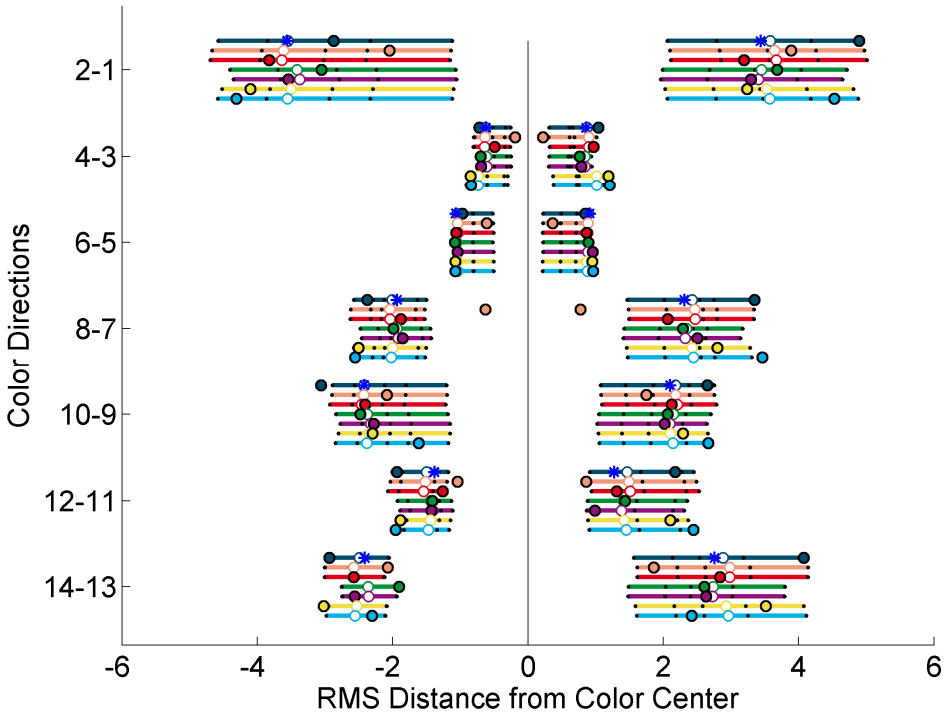


Fig. 6-64. RMS Distances in XYZ coordinate system for Color Center 2 (Red)

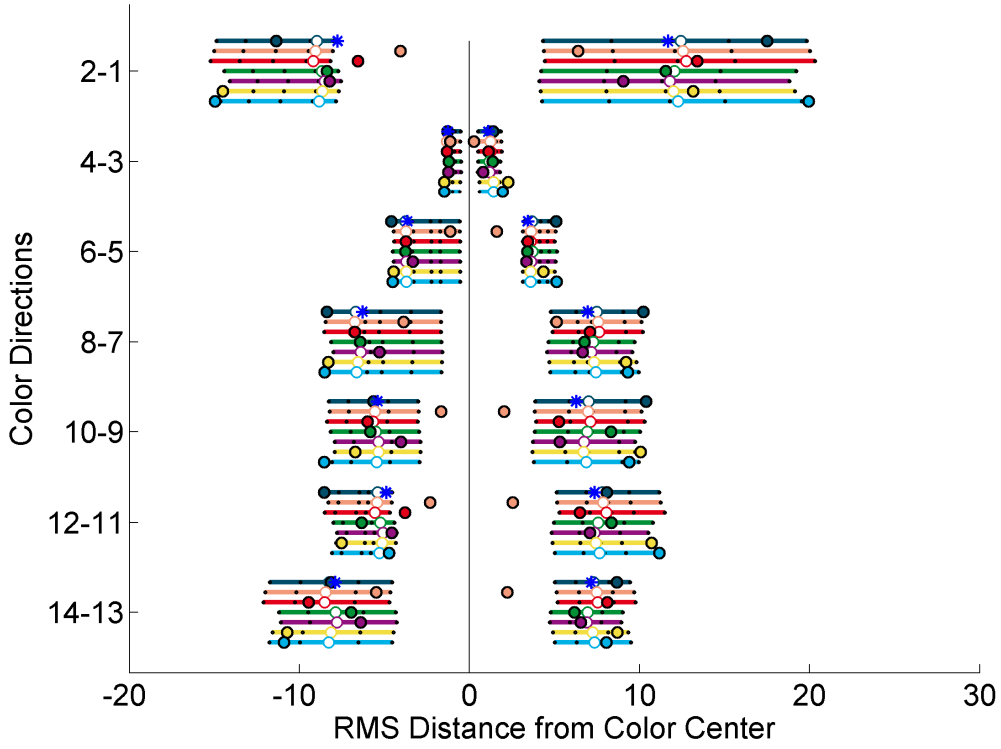


Fig. 6-65. RMS Distances in XYZ coordinate system for Color Center 3 (Yellow)

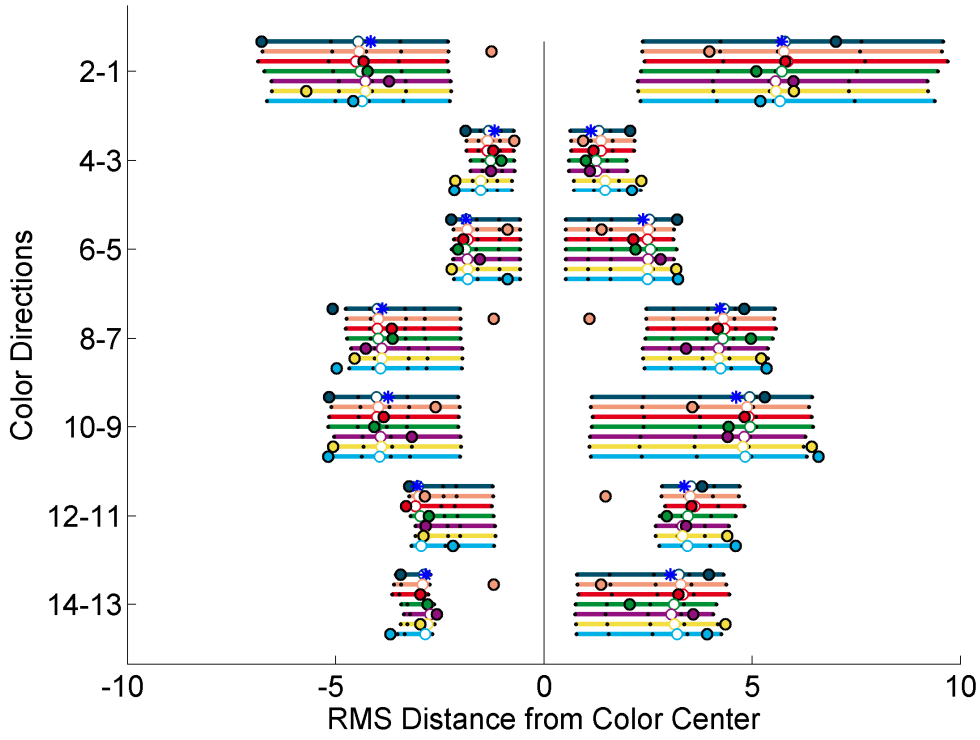


Fig. 6-66. RMS Distances in XYZ coordinate system for Color Center 4 (Green)

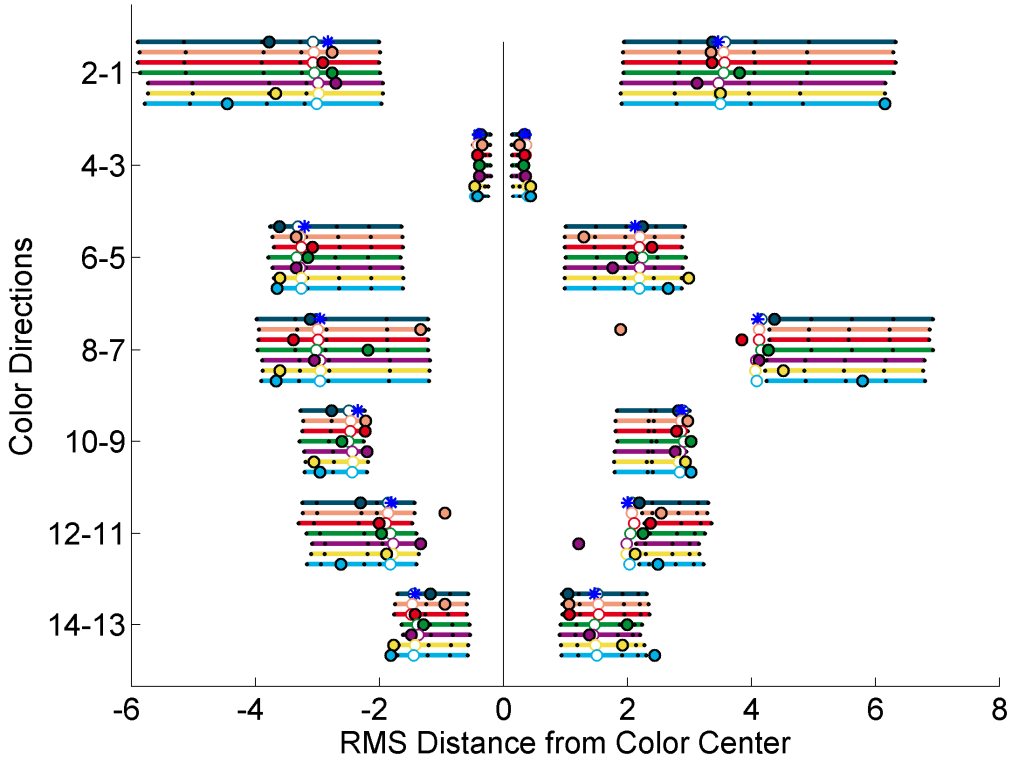


Fig. 6-67. RMS Distances in XYZ coordinate system for Color Center 5 (Blue)

6.6.3 *Conclusions from correlation analysis*

As the observer classification experiment conducted with the help of Observer Calibrator prototype suggest, the CIE 10° standard colorimetric observer was assigned to only 4 out of 49 observers. This means only around 8% of the observers conform to the current standard colorimetric observer.

With regard to the correlation analysis of observer classification data and color difference judgments, two main inferences emerge. Firstly, color difference thresholds for categories that are very different from the CIE standard colorimetric observer, as indicated by the observer classification results, have large differences from the global average thresholds. Secondly, average thresholds for observers belonging to dominant categories are generally very close to the global average thresholds. The consistency between observer categories and color difference data give an indirect validation of the observer classification method. The results also lead us to conclude that colorimetric observer categories, derived from classical color matching data, can help in the prediction of average suprathreshold color difference perception for a given observer population. Determining the extent to which the results can be improved needs further investigation, requiring additional visual data and appropriate metrics.

In his paper on the variability of small suprathreshold color difference perception, Kuehni [144] comments that "*the results indicate that assessments of the magnitude of small color differences have*

considerable variability within, but particularly between observers. …From this it is evident that mean observer data depend to a significant degree on the composition of the observer panel. It is also evident that color difference formulas fitted to mean data can predict perceived color differences accurately only for a minority percentage of color-normal observers."

Long ago, Rich and Jalijali [74] also talked about the possibility that perception of small color difference could be observer dependent. They noted: "*Unfortunately, it thus appears that the perception of color-differences are not observer independent. This implies that color difference or color acceptance formulas based on single observers…are risky ventures. This also implies that scaling or ranking of color or color-differences will be influenced or affected by observer differences. …The result of scalings by an unreliable observer may be nonlinear, distorted, or just very noisy.*"

Results obtained from the collaborative experiment reinforce these assertions, but more importantly, opens up two important possibilities for future discussion. Firstly, is it possible to customize color difference equations for individual observer categories, and even derive more uniform color spaces for these categories? And secondly, can we use our knowledge of observer categories to derive a better standard colorimetric observer from a limited amount of visual data, so that we can achieve a more uniform color space, and simplified color difference equations? The relevance of these questions for the color imaging industry can be better appreciated in the context of Kuehni's plea [145] for an industry-wide, systematic effort to address the existing issues with estimation of color differences: "*It seems appropriate and useful to color-related industries to make a concerted effort at the beginning of the new century to resolve the issues around an objective method of color control to the degree that the biggest variable, the observer, allows. Only a widely controlled and comprehensive effort will make this possible.*"

6.7 Final words on standard and "deviate" colorimetric observers

The fact that the CIE 10° standard colorimetric observer was the chosen category for only around 8% of all observers tested, a result in congruence with earlier findings (see Chapter 5 and Section 6.4.2), 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 observer CMFs, which results in a mathematical model of an average observer that does not quite correspond to the observers who participated in the observer classification experiments. In other words, observers who are sufficiently different from the average unduly skew the results of the mean.

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 an 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”. While the terms “standard” and “deviate” were likely used by the scientific community in purely mathematical context, the term “deviate observer” is often interpreted by non-experts 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 a color vision problem. In a way, the terminology used traditionally in color science community gives way to this wrong understanding. It is important to acknowledge that it is perfectly natural for individual color normal observers to be different from each other. As has been shown in this chapter, 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 to warrant a revisit to the aspects of 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.

A more appropriate way of defining the colorimetric observer models could be similar to what CIE did to define the “standard illuminants”, by using terms like CIE standard illuminant A, B, C, 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, “colorimetric observer model A” under normal circumstances. One advantage of this method is no model is claimed as “standard” or “deviate”, just like the CIE illuminants. The other advantage is that for a restricted population, a color researcher or engineer can choose to use a more appropriate model. For example, with respect to observer categories introduced in this chapter, categories 8 and 9 (or their updated or improved future versions) will be more appropriate for an elderly population than the categories 4 and 5. Of course such colorimetric observer models 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.

It can also be hoped that an existence of multiple categories will encourage the color engineers to be more attentive to the specific observer model being used in their colorimetric computations. At present, many professionals working in the area of color technology tend to use the CIE 2° standard colorimetric observer and the CIE 10° standard colorimetric observer interchangeably, failing to acknowledge the significant impact of the choice of the colorimetric observer model on their

computations and designs. For example, many measurement instruments use CIE 2° standard colorimetric observer by default, while the small-field color matching may not be appropriate in the context in which such measurements are being performed. Such discrepancies often go unnoticed.

Finally, it must to emphasized once more that a very large observer population must be tested using the observer classification method before the set of representative observer categories can be finalized. Such task can best be handled by a standardization body like CIE, based on the findings of this work and of other researchers in the domain.

Discovery consists of seeing what everybody has seen and thinking what nobody has thought. ~ Albert von Szent-György. The Scientist Speculates: An Anthology of Partly-Baked Ideas

7. Observer-dependent color imaging: workflow, implementation and benefits

7.1 Introduction

In Chapter 6, a practical method was implemented to classify observers into one of several categories. This chapter proposes a workflow for color critical industrial applications in order to exploit the knowledge of observer categories to obtain observer-specific color matches.

7.2 Colorimetrically accurate imaging workflow

As explained in the beginning of Chapter 5, observer variability and metamerism can be a nontrivial issue in post-production applications, which involve critical color matching tasks, for example the color grading of the raw movie content at the post-shooting stage. The main goal of color reproduction in an application such as this is markedly different from that in typical consumer applications. To appreciate this fact, we should consider the set of five objectives of color reproduction put forth by Fairchild [146]. These objectives are an updated version of Hunt's original proposition of six objectives [147], and are as follows:

i) Color reproduction: basic ability of devices to reproduce colors.

ii) Pleasing color reproduction: ability of devices to reproduce acceptable colors, where observers have no knowledge of original scene, and so no expectation beyond a pleasing image.

iii) Colorimetric color reproduction: ability of devices to produce colorimetrically accurate colors. Involves reproduction on calibrated and characterized devices, allowing the CIE tristimulus values of the original image to be accurately reproduced on any given output device. Useful only when viewing conditions for the original and reproduced images are identical.

iv) Color appearance reproduction: ability of devices to maintain appearance attributes. Reproduction involves calibrated and characterized devices, requires a color appearance

model, and also information about viewing conditions for the original and reproduced images. Poses many challenges to be realized in commercial applications

v) Color preference reproduction: ability of devices to manipulate colors to ensure subjective preference of the user for a given medium and subject.

Color reproduction in post-production applications like color grading does not quite match any of the above, except probably to some extent the colorimetric color reproduction or color appearance reproduction. In most consumer applications, color reproduction focuses on user preferences (the last objective above), but in post-production, the main goal is generally to preserve the artistic intent of the Director of Photography (DP) and the Colorist, irrespective of the ultimate consumer's personal preferences. Doing so for a variety of media, for example, large-screen content (film and/or digital), digital mastered content (television and/or DVD) etc poses a great color reproduction challenge, even more with wide-gamut displays introduction. At the very least, it is critical that throughout the post-production workflow the colors are represented accurately.

Color imaging workflow in any practical application can be organized in three steps [148]: i) device-dependent representation where colors are specified for a given imaging device only, ii) device-independent representation where colors are specified in terms of colorimetric coordinates such as CIEXYZ or CIELAB, and iii) viewing-condition-independent representation that take into account the color appearance of any given scene with specific viewing conditions such as luminance level, surround, chromatic adaptation, etc. and attempt to specify the final image appearance. Even before considering the appearance attributes of the content, the device-independent colorimetric representation, which is quite fundamental from color science point of view, need to be perfected in order to achieve a colorimetrically accurate imaging workflow.

In this regard, the choice of the color space is critical. Any color space specification essentially uses an average or a standard colorimetric observer. Thus, individual variability is an issue that cannot be overlooked, particularly when this variability is significant, either because of the spectral characteristics of the imaging device employed, or because of the observer's color vision. In fact, many colorists in the entertainment industry prefer to work on raw colors (device RGB) instead of device-independent color specifications like CIELAB. Thus the prost-production workflow can potentially benefit from an improved, personalized color space that is perceptually more uniform for a certain individual observer (a colorist or a DP) than what is currently available. This is the ultimate goal of the workflow discussed in the next section.

7.3 Observer-dependent color imaging workflow

In a typical color image processing pipeline, a significant part of the processing is device independent, irrespective of the devices involved in the input or output side. However, all the processing, whether device dependent or device independent, is based on a single CIE standard colorimetric observer. In this thesis work, a new term is hereby introduced: *observer dependent color imaging* (ODCI). The concept is illustrated in Fig. 7-68, applied to some typical color imaging workflows. However, other embodiments/applications are also possible. Note that the concept of *observer dependent color imaging* applies only to a small part of the imaging workflow, at the acquisition (input) or rendering (output) level, keeping the rest of the chain unaffected.

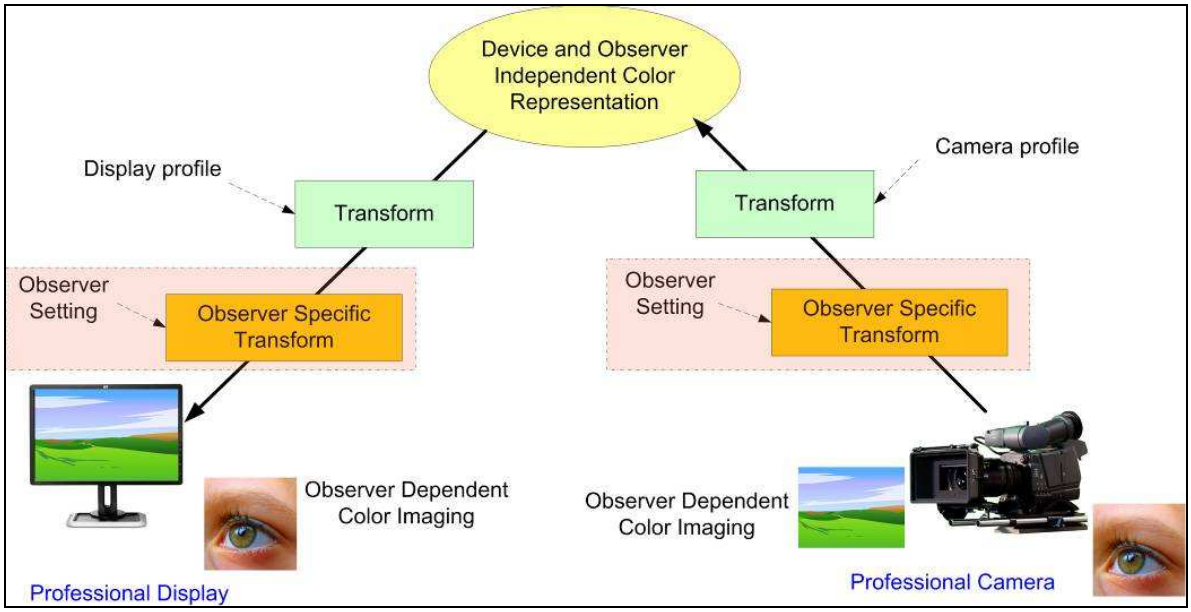


Fig. 7-68. Observer dependent color imaging workflow

ODCI workflow will typically be implemented at the output side, for example, for display processing. A display profile-specific transform is currently applied to device independent color representation, to obtain display color codes (also described as display channel values or digital counts). The proposed workflow will introduce an additional step (the orange block in the figure) where a further transform will be applied based on specific observer setting on the device, and will result in modified display color codes, customized for a specific observer. This will ensure that the colors perceived by this observer on the display are approximately identical to the perception intended for a CIE standard colorimetric observer, which is the underlying assumption in the whole color reproduction chain.

The observer specific transform described above could be implemented in several ways. In a more device-specific implementation, it could be in the form of observer-specific display Lookup Tables (LUT). Such a LUT would convert digital counts corresponding to CIE standard colorimetric observer

specific colors directly to digital counts corresponding to a given observer category specific colors. In a more generic, two-step implementation, observer-specific colorimetric transformation can be applied to convert CIE standard colorimetric observer specific values XYZ (henceforth CIEXYZ) to observer category specific XYZ values (henceforth CatXYZ) , and then in the next step convert the CatXYZ values to corresponding digital counts through appropriate display LUTs.

ODCI can also be applied on the input side, on professional camera system. Colors seen by the photographer can be converted to corresponding colors that would have been seen by a CIE standard colorimetric observer, using a transform similar to the one described above. In this case, display primaries are replaced by camera spectral sensitivities. Rest of the chain remains the same as conventional processing.

However, this workflow may not be practical in the context of some typical color imaging applications, for example, for input devices like consumer digital cameras and scanners, and for output devices like generic printers. These devices, and/or the content generated by them, are likely to be viewed by many different users under uncontrolled viewing conditions. In these application contexts, a precise, observer-dependent color reproduction (and perception) is neither practical nor useful. A standard, average colorimetric observer seems more appropriate in these cases.

7.4 Implementation - derivation of colorimetric transformations between the CIE 10° standard colorimetric observer and eight categories

Before discussing the implementation aspects, it must be reiterated that there is an assumption involved in the computation of the reduced set of color matching functions. This assumption pertains to the conversion from the cone fundamentals of chosen observer categories to the corresponding color matching functions equivalent to CIE XYZ system. As described in Chapter 6, a linear transformation matrix was computed to convert the $\bar{r}$ $\bar{g}$ $\bar{b}$ color matching functions (CMFs) of Stiles-Burch observers first to corresponding cone fundamentals, and then to CIE 10° standard colorimetric observer equivalent CMFs. For this purpose, approximate transformation matrices were computed from the average observer data, which were then used on individual observer data.

In reality, deriving an XYZ tristimulus space from a given set of $\bar{r}$ $\bar{g}$ $\bar{b}$ CMFs is not so straightforward. Recommending a standard procedure for this derivation is within the scope of CIE TC 1-36 that published the first part of its report [14] on physiologically-based CMFs. The official CIE recommendation for the derivation of tristimulus space will be based on Wold and Valberg's method [149], which uses the same principles as used in developing CIE 1931[41] standard, while imposing additional restrictions. At the time this thesis research was conducted, this recommendation

was not available, so an approximate linear transformation was used instead. Note that all observer categories are based on the Stiles-Burch experiment [44] involving monochromatic stimuli of unit radiance and wavelengths 645.2 nm, 526.3 nm and 444.4 nm as primaries, which define the axes of the chromaticity space. Thus, we can assume that a single average transformation between the RGB and XYZ chromaticity spaces for all observer categories is an acceptable approximation.

7.4.1 *Method of transformation*

Two methods were used for the transformations of CIEXYZ values corresponding to the CIE 10° standard colorimetric observer ($X_{Cat\text{-}1}\ Y_{Cat\text{-}1}\ Z_{Cat\text{-}1}$) into the CatXYZ values corresponding to various observer categories ($X_{Cat\text{-}A}\ Y_{Cat\text{-}A}\ Z_{Cat\text{-}A}$, where A denotes one of the eight categories and varies from 2 through 9). The first method used a linear transformation, while the second used a nonlinear spline-based 3D interpolation. The transformations were computed using Eq. 7-30, where $T_{Cat\text{-}A}$ represents either a linear 3x3 matrix or a three-dimensional lookup table (LUT), and '*' represents a matrix product or a LUT application.

$$\begin{bmatrix} X_{\text{Cat-A}} \\ Y_{\text{Cat-A}} \\ Z_{\text{Cat-A}} \end{bmatrix} = T_{Cat-A} * \begin{bmatrix} X_{\text{Cat-1}} \\ Y_{\text{Cat-1}} \\ Z_{\text{Cat-1}} \end{bmatrix} \tag{7-30}$$

Thus, both methods required that the tristimulus values be computed for a given set of spectral data. For this, estimated spectral power distributions of a large set of stimuli were used in each case. These spectral power distributions were obtained by using the LED primaries in the right half of the bipartite field of the prototype. These colors are characterized by high observer variability. From the spectral power distributions, CIEXYZ and CatXYZ values were computed for each CMF. From here onward, this dataset is referred to as modeling dataset. The modeling datasets were slightly different in case of linear transformation and in case of 3D interpolation, as described later.

For each of the two methods, an independent set of eight tristimulus values (hereafter referred to as verification dataset) was used for the verification of the accuracy of the transformation. From here onward, this dataset is referred to as verification dataset. Slightly different verification datasets were used in the two methods, as clarified later.

Using [$X_{10}\ Y_{10}\ Z_{10}$] values from the modeling dataset in Eq. 7-30, [$X_{Cat\text{-}A}\ Y_{Cat\text{-}A}\ Z_{Cat\text{-}A}$] values were predicted using a linear (3x3 matrix) or nonlinear transformation (3D-LUT), which were then compared to corresponding tristimulus values computed from the spectral data. The errors between the predicted and actual values would indicate the accuracy of the transformations.

In the following two sections, details of the modeling dataset and the verification dataset are included, followed by a discussion of the results obtained from each method.

7.4.2 *Using linear transformation with 3x3 matrices*

For computing the transformations, the same set of 5832 estimated spectral power distributions used earlier for deriving the reduced set of CMFs (see Chapter 6) was selected as the modeling dataset.

Spectral power distributions of the eight test stimuli (Fig. 6-19) from the observer classification experiment were used as verification dataset. From these spectral data, tristimulus values corresponding to the CIE 10° standard colorimetric observer and the eight observer categories were computed, and are shown in Table 7-17. The standard observer is marked as category 1 as in Chapter 6. Note that different normalization factors were used in the tristimulus value computation for different CMFs so that in each case the luminance of white obtained by setting the LED field primaries to their maximum powers equals to 100.

Table 7-17. Modeling dataset for the linear transformation method. Tristimulus values of eight test stimuli used in the Observer Calibrator, corresponding to CIE 10° standard colorimetric observer (category 1) and the reduced set of eight observer categories

CMF	*Tristim. Values*	*Test Stimuli*							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
CIE 10° Std. Col. Obs. (Cat-1)	X_{10}	13.7463	13.8509	10.9885	13.2999	18.7832	15.7506	13.86	17.3884
	Y_{10}	12.6658	12.6201	9.5445	16.8156	16.7856	18.3087	16.7239	20.8577
	Z_{10}	7.2181	23.7031	3.8872	18.2753	18.7971	7.3985	23.5946	16.9055
Obs. Cat. - 2	X_{Cat2}	14.0023	13.8334	11.2165	13.4224	18.9771	16.1107	13.901	17.6494
	Y_{Cat2}	12.8051	12.6528	9.6731	16.8372	16.9211	18.4421	16.723	20.9417
	Z_{Cat2}	8.3687	27.2074	4.5135	21.1926	21.626	8.7488	27.2527	19.6731
Obs. Cat. - 3	X_{Cat3}	15.6041	15.8283	12.4576	15.2	21.3762	17.8982	15.8645	19.8201
	Y_{Cat3}	12.9308	12.6846	9.7894	16.8553	17.0449	18.5601	16.7219	21.0156
	Z_{Cat3}	7.1404	23.4214	3.8416	18.1131	18.5753	7.3631	23.3607	16.7665
Obs. Cat. - 4	X_{Cat4}	14.8566	15.0701	11.8609	14.4719	20.3522	17.0409	15.1046	18.8707
	Y_{Cat4}	12.5697	12.6316	9.4546	16.802	16.7126	18.1969	16.739	20.7927
	Z_{Cat4}	7.2027	23.6955	3.8669	18.3211	18.7751	7.4253	23.6355	16.9479
Obs. Cat. - 5	X_{Cat5}	12.5712	12.7893	10.0738	11.9451	17.2729	14.1521	12.5314	15.6421
	Y_{Cat5}	12.5697	12.6316	9.4546	16.802	16.7126	18.1969	16.739	20.7927
	Z_{Cat5}	7.1859	23.539	3.8696	18.2072	18.6764	7.412	23.4785	16.8587
Obs. Cat. - 6	X_{Cat6}	13.2037	13.4328	10.5807	12.5461	18.1419	14.8641	13.1619	16.4291
	Y_{Cat6}	12.9308	12.6846	9.7894	16.8553	17.0449	18.5601	16.7219	21.0156
	Z_{Cat6}	7.1404	23.4214	3.8416	18.1131	18.5753	7.3631	23.3607	16.7665
Obs. Cat. - 7	X_{Cat7}	14.0676	13.898	11.2688	13.485	19.0656	16.1858	13.9659	17.7318
	Y_{Cat7}	12.9308	12.6846	9.7894	16.8553	17.0449	18.5601	16.7219	21.0156
	Z_{CaT7}	7.1404	23.4214	3.8416	18.1131	18.5753	7.3631	23.3607	16.7665
Obs. Cat. - 8	X_{Cat8}	13.8802	14.1044	11.1807	12.7397	19.1041	15.2706	13.435	16.8022
	Y_{Cat8}	12.8051	12.6528	9.6731	16.8372	16.9211	18.4421	16.723	20.9417
	Z_{Cat8}	7.53	24.7721	4.0426	19.1535	19.6281	7.7627	24.7094	17.7179
Obs. Cat. - 9	X_{Cat9}	13.2769	13.4914	10.6948	12.186	18.2738	14.6069	12.8511	16.072
	Y_{Cat9}	12.5697	12.6316	9.4546	16.802	16.7126	18.1969	16.739	20.7927
	Z_{Cat9}	7.2027	23.6955	3.8669	18.3211	18.7751	7.4253	23.6355	16.9479

As outlined in the previous section, 3x3 matrices for obtaining CatXYZ values from the CIEXYZ values were computed using the modeling dataset. Eq. 7-30 was solved for each $T_{Cat\text{-}A}$ in the least square sense to obtain transformation matrices for various categories ($M_{Cat\text{-}2}$, $M_{Cat\text{-}3}$ etc). Eqs. 7-31 through 7-38 give these matrices.

$$M_{Cat-2} = \begin{bmatrix} 0.980139 & 0.020211 & -0.028272 \\ 0.013582 & 0.958319 & 0.039481 \\ -0.016043 & -0.006330 & 1.104520 \end{bmatrix} \tag{7-31}$$

$$M_{Cat-3} = \begin{bmatrix} 1.082740 & 0.039545 & -0.009477 \\ 0.005694 & 0.945804 & 0.011083 \\ 0.006266 & -0.011986 & 0.949995 \end{bmatrix} \quad (7\text{-}32)$$

$$M_{Cat-4} = \begin{bmatrix} 1.082740 & -0.011363 & -0.012619 \\ 0.005694 & 1.011010 & 0.012059 \\ 0.006266 & 0.006649 & 1.010796 \end{bmatrix} \quad (7\text{-}33)$$

$$M_{Cat-5} = \begin{bmatrix} 0.979985 & -0.011363 & -0.009032 \\ -0.065248 & 1.011010 & 0.011730 \\ 0.007037 & 0.006649 & 1.002210 \end{bmatrix} \quad (7\text{-}34)$$

$$M_{Cat-6} = \begin{bmatrix} 0.979985 & 0.039545 & -0.009477 \\ -0.065248 & 0.945804 & 0.011083 \\ 0.007037 & -0.011986 & 0.949995 \end{bmatrix} \quad (7\text{-}35)$$

$$M_{Cat-7} = \begin{bmatrix} 0.980139 & 0.039545 & -0.009477 \\ 0.013582 & 0.945804 & 0.011083 \\ -0.016043 & -0.011986 & 0.949995 \end{bmatrix} \quad (7\text{-}36)$$

$$M_{Cat-8} = \begin{bmatrix} 1.121113 & 0.020211 & -0.012619 \\ -0.160701 & 0.958319 & 0.012059 \\ 0.005468 & -0.006330 & 1.010796 \end{bmatrix} \quad (7\text{-}37)$$

$$M_{Cat-9} = \begin{bmatrix} 1.121113 & -0.011363 & -0.012619 \\ -0.160701 & 1.011010 & 0.012059 \\ 0.005468 & 0.006649 & 1.010796 \end{bmatrix} \quad (7\text{-}38)$$

Using the linear transformations on the verification dataset, predicted CIEXYZ and CatXYZ values were computed. Fig. 7-69 plots the chromaticity coordinates for the eight test stimuli obtained from these CIEXYZ and CatXYZ values. Values obtained directly from the spectral data are plotted as squares while the values obtained through a linear transformation of the CIEXYZ values are shown as triangles.

Chromaticity errors between the predicted tristimulus values and those computed earlier from the spectral data are given in Fig. 7-70. The errors were computed in terms of Euclidean distances for

each of the eight stimuli in xy-chromaticity diagram. On each box, the central mark is the median of eight distances, the edges of the box are the 25th and 75th percentiles, and the dotted error bars extend to the most extreme data points not considered outliers. Note that as in Chapter 6, these categories are marked as 2 through 9 and the CIE 10° standard colorimetric observer is marked as category 1.

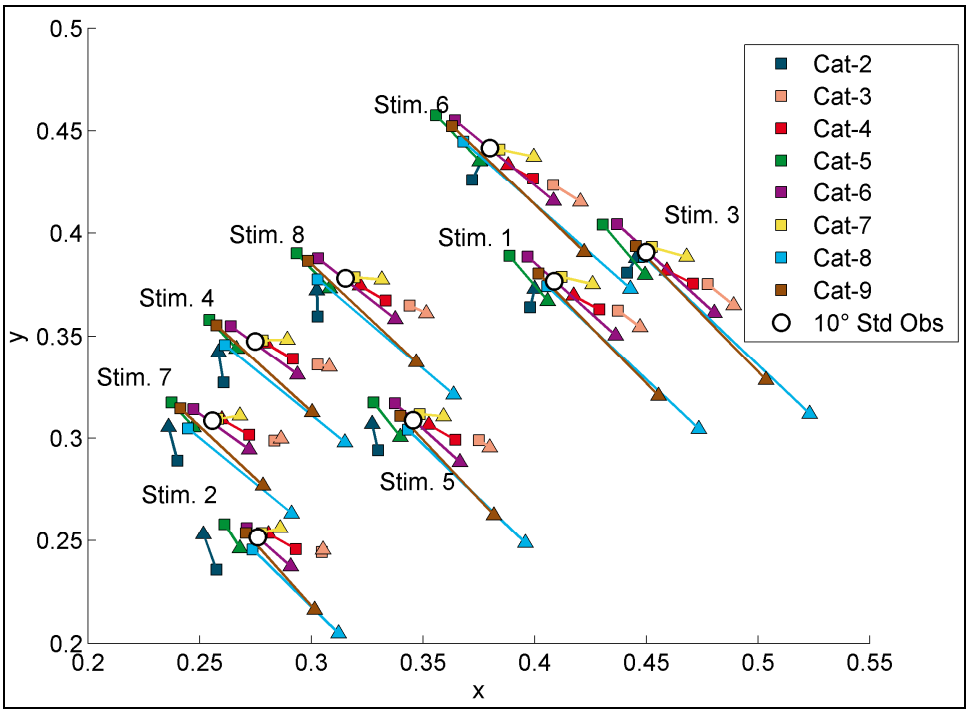


Fig. 7-69. xy-chromaticities of eight test stimuli corresponding to various observer categories and CIE 10° standard colorimetric observer. Squares: coordinates obtained from spectral data. Triangles: coordinates obtained through linear transformation of CIEXYZ values. The test stimuli were used in the Observer Calibrator prototype described in Chapter 6

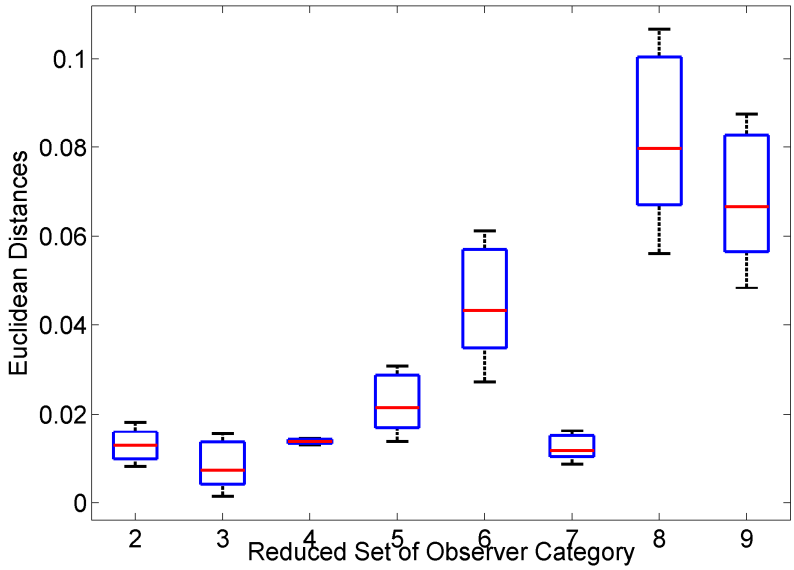


Fig. 7-70. Chromaticity prediction errors due to linear transformation of observer-specific tristimulus values

As evident in Figs. 7-69 and 7-70, high prediction errors exist for the observer categories 7 and 8, indicating the linear transformation from CIEXYZ used in this computation may not be adequately accurate for these two categories. Colors in green and red have higher prediction errors for these categories. The prediction errors for categories 2, 3, 4 and 7 are relatively low. Category 4 is unique in the sense that there is almost no difference in the prediction errors among the eight stimuli as indicated by the similar distances between the squares and the triangles for all test stimuli in Fig. 7-69. This results in category 4 having a flat line instead of a box in Fig. 7-70. What does it mean for the observers of this category? It is likely that the color matches obtained through a linear transformation of CIEXYZ colors as per Eq. 7-33 would all seem to be equally good or bad for the observers of category 4. If we ask such an observer to participate in a color matching experiment involving just one test color, and then derive a linear transformation between the matched color and the match predicted by the CIE 10° standard colorimetric observer, it is likely that such a transformation will give a reasonably accurate result for other colors. Recall that according to observer classification experiments, category 4 is the most popular category with 24.5% observers belonging to this category (see Table 6-6 and Fig. 6-20).

It is also interesting to note that a linear transformation for category 3 results in accurate prediction of chromaticities in the blue regions of the color space (Fig. 7-69), while other categories show much more variations. For this category, the errors increase as we move toward the red and green. Note that there was only one observer in this category, so the above observations may not be generalized for all observers belonging to category 3, at least until more data specific to this category are available.

Finally, if we compare the distances between white circles (CIE 10° standard colorimetric observer) and the squares corresponding to different categories (Fig. 7-69), the distances for a given category are similar for different test stimuli, but vary from one category to the other. This is expected since all these points are obtained through the fundamental colorimetric equations using the spectral data and corresponding CMFs. So while the Euclidean distances between the white circles and the squares for a given category do not vary appreciably over the different parts of the color space, the differences between various CMFs dictates that the squares be located differently in the color space.

In general for all test stimuli, category 7 is the closest point to white circles, which indicates this category results in similar results as the current standard colorimetric observer in the xy-chromaticity space. As per the observer classification experiments, only 2% of total observers (1 out of 49) belonged to this category (see Table 6-6). This category might thus be redundant. Further experiments can confirm this assumption.

7.4.3 *Using three-dimensional lookup tables obtained from spline-based 3D interpolation*

For computing the three-dimensional lookup tables, a set of 4913 estimated spectral power distributions (similar to the larger set of 5832 used in case of linear transformations) was selected as the modeling dataset.

As for the verification dataset, a slightly different set of chromaticities were used compared to the case of linear transformations. This new set, shown in Table 7-18, was a result of an inadvertent computational error, which was detected after 3D interpolations were completed for all observer categories. Since the computations were highly time-consuming (4 hours for each category), the verification was not rerun in the interest of time. Note however that the chromaticities are close to the original set and the results presented in this section are valid and accurate.

Using the 3D interpolations on the verification dataset of CIEXYZ values, predicted CatXYZ values were computed. Fig. 7-71 plots the chromaticity coordinates for the eight test stimuli obtained from these CIEXYZ and CatXYZ values. Values obtained directly from the spectral data are plotted as squares while the values obtained through a linear transformation of the CIEXYZ values are shown as triangles. The triangles and the squares are superimposed, confirming 3D interpolation method accurately predicts the CatXYZ values.

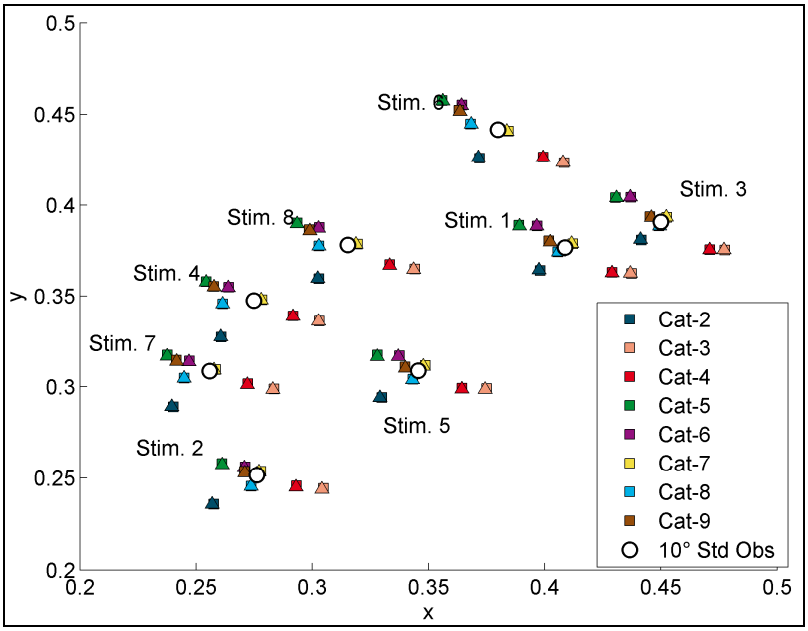


Fig. 7-71. xy-chromaticities of eight test stimuli corresponding to various observer categories and CIE 10° standard colorimetric observer. Squares: coordinates obtained from spectral data. Triangles: coordinates obtained through 3D interpolation of CIEXYZ values (superimposed on squares due to low prediction errors). The test stimuli were similar to those used for linear transformation

Table 7-18. Modeling dataset for the 3D interpolation method. Tristimulus values of eight test stimuli, corresponding to CIE 10° standard colorimetric observer (category 1) and the reduced set of eight observer categories

CMF	*Tristimulus Values*	*Test Stimuli*							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
CIE 1964 10° Std. Col. Obs.	X	16.6255	16.7520	13.2901	16.0855	22.7173	19.0496	16.7630	21.0305
	Y	15.3187	15.2634	11.5436	20.3376	20.3013	22.1435	20.2267	25.2264
	Z	8.7300	28.6677	4.7014	22.1030	22.7342	8.9481	28.5365	20.4463
Red. Cat-1	Cat-X	16.3638	16.1665	13.1082	15.6861	22.1776	18.8278	16.2454	20.6260
	Cat-Y	14.9647	14.7867	11.3045	19.6768	19.7748	21.5525	19.5434	24.4736
	Cat-Z	9.7801	31.7960	5.2747	24.7668	25.2733	10.2243	31.8490	22.9911
Red. Cat-2	Cat-X	18.1511	18.4119	14.4910	17.6811	24.8653	20.8197	18.4540	23.0553
	Cat-Y	15.0415	14.7550	11.3873	19.6065	19.8271	21.5896	19.4513	24.4459
	Cat-Z	8.3059	27.2444	4.4686	21.0696	21.6072	8.5650	27.1738	19.5032
Red. Cat-3	Cat-X	18.1511	18.4119	14.4910	17.6811	24.8653	20.8197	18.4540	23.0553
	Cat-Y	15.3570	15.4327	11.5511	20.5279	20.4187	22.2321	20.4509	25.4035
	Cat-Z	8.7999	28.9500	4.7243	22.3838	22.9385	9.0719	28.8767	20.7061
Red. Cat-4	Cat-X	15.3589	15.6254	12.3077	14.5939	21.1032	17.2904	15.3103	19.1107
	Cat-Y	15.3570	15.4327	11.5511	20.5279	20.4187	22.2321	20.4509	25.4035
	Cat-Z	8.7793	28.7588	4.7277	22.2446	22.8179	9.0557	28.6848	20.5972
Red. Cat-5	Cat-X	15.3589	15.6254	12.3077	14.5939	21.1032	17.2904	15.3103	19.1107
	Cat-Y	15.0415	14.7550	11.3873	19.6065	19.8271	21.5896	19.4513	24.4459
	Cat-Z	8.3059	27.2444	4.4686	21.0696	21.6072	8.5650	27.1738	19.5032
Red. Cat-6	Cat-X	16.3638	16.1665	13.1082	15.6861	22.1776	18.8278	16.2454	20.6260
	Cat-Y	15.0415	14.7550	11.3873	19.6065	19.8271	21.5896	19.4513	24.4459
	Cat-Z	8.3059	27.2444	4.4686	21.0696	21.6072	8.5650	27.1738	19.5032
Red. Cat-7	Cat-X	16.2211	16.4831	13.0664	14.8883	22.3260	17.8460	15.7009	19.6360
	Cat-Y	14.9647	14.7867	11.3045	19.6768	19.7748	21.5525	19.5434	24.4736
	Cat-Z	8.7999	28.9500	4.7243	22.3838	22.9385	9.0719	28.8767	20.7061
Red. Cat-8	Cat-X	16.2211	16.4831	13.0664	14.8883	22.3260	17.8460	15.7009	19.6360
	Cat-Y	15.3570	15.4327	11.5511	20.5279	20.4187	22.2321	20.4509	25.4035
	Cat-Z	8.7999	28.9500	4.7243	22.3838	22.9385	9.0719	28.8767	20.7061

As before, chromaticity errors between the predicted and actual tristimulus values (CIEXYZ and CatXYZ), computed in terms of Euclidean distances for each of the eight stimuli in xy-chromaticity diagram, are shown in Fig. 7-72. The absolute errors are less than 0.001, and thus negligible.

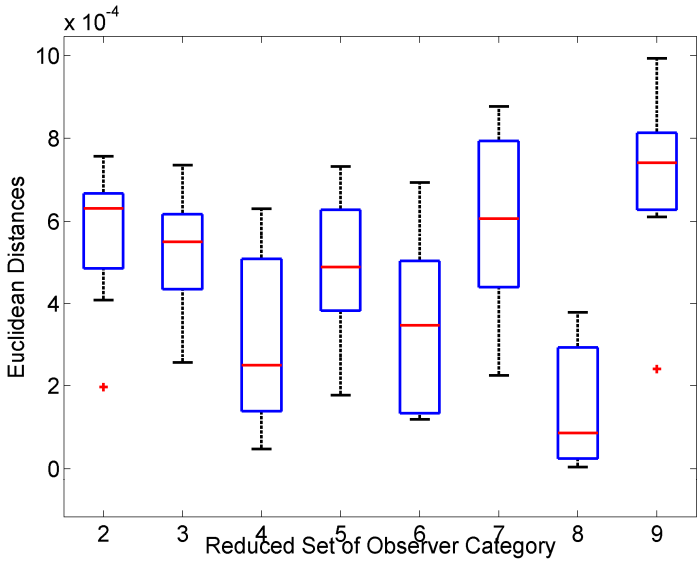


Fig. 7-72. Chromaticity prediction errors due to 3D interpolation of observer-specific tristimulus values

7.5 Advantages of ODCI in an applied context

As explained in the previous chapters, the issue of observer metamerism has become non-trivial with the advent and wide-spread adoption of modern wide-gamut consumer displays. 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 [150] [151] (see also Chapter 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. However, the future of televisions and consumer displays lie in these wide-gamut displays. Even many latest professional displays are equipped with such narrow-band primaries. The potential advantages of the ODCI workflow should be assessed in this context.

The practical advantage of an ODCI workflow is that in an imaging device, e.g. a display, the user can have a control - just like brightness or saturation control typical in today's displays, which will allow him/her to select a specific observer setting. This setting can be selected based on the observer classification test described in the previous chapter, which will make the colors appear to him/her close to what would have appeared to a standard colorimetric observer (i.e. observers with characteristics identical to the CIE standard observer functions). Or, the setting can be based on the default dominant category.

Thus, by selecting an appropriate observer setting for each observer, the variability in the color perception from one observer to the other can be minimized. This will significantly reduce the uncertainty in color critical tasks introduced by observer variability. An example could be color correction by a colorist during post-production, where any potential disagreement between the colorist and the Director of Photography can be minimized by selecting an appropriate set of CMFs for each person (through user control).

As an extension, the ODCI workflow can be used to customize a device not only for the color normal observers, but also for the anomalous trichromats, who are currently not able to have the same color experience as a color normal observer, in spite of not being color blind. An appropriate ODCI implementation can aid to meet the needs for this kind of special group of consumers, even though appropriate observer categories would first need to be established. The ODCI workflow can, in principle, make it possible to allow every observer to perceive a given color in the way it was originally intended by the content creator, irrespective of individual observer variability, as long as the observer is a trichromat.

The concept can be applied to any application of Digital Image Processing/Digital Video Processing. It could, in principle, be applied to any industrial application involving color management and reproduction.

Specific to the application contexts relevant for content processing, ODCI has potential to help develop technologies for observer-dependent color correction method in post-production workflow. More generally, graphics arts and the use of creative computer software can benefit from the workflow proposed here. It is also applicable to high quality color reproduction for TV/PC end users as observer dependent calibration can easily be implemented in the form of a Look-Up Table transform in personal computers, set-top boxes or gateways.

The ODCI workflow is likely to be most useful for professional and high-end consumer display applications.

7.6 Conclusions

This chapter presented the first set of tools for implementing observer categories in a practical application context. The *observer dependent color imaging* (ODCI) workflow was described. Implementation of this workflow in a display application necessitates that colors corresponding to CIE 10° standard colorimetric observer be converted into category-specific colors. Simpler linear transformations as well as three-dimensional Lookup Tables obtained from nonlinear interpolations were computed and preliminary analysis of the results was presented. These transformations are in the

form of very large Lookup Tables, and so could not be included in the thesis manuscript. However, they will be made available to the research community through the author's personal website www.abhijitsarkar.com.

These results show that for some observer categories, a linear transformation from tristimulus values corresponding to the CIE 10° standard colorimetric observer to those corresponding to various categories will result in relatively low chromaticity computation errors, while for other categories the error will be more significant. This raises a fundamentally important question. Should CIE XYZ tristimulus computation be adapted differently for observers whose color matching functions are known to be very different, so that for a given color stimulus, similar tristimulus values can be obtained for these two observers? This question has so far been redundant since colorimetry is based on a single observer model. However, if we decide to expand colorimetric computation to provide the option of multiple-observer models, the most basic equation in color science, computation of the CIE tristimulus values may need to be modified to make room for such expansion.

In a more applied context, when it comes to the implementation of ODCI workflow, the transformation from color specification based on the CIE 10° standard colorimetric observer to those based on different observer categories must be implemented and tested in visual experiments. This chapter set forth the preliminary approach for achieving that goal.

We shall not cease from exploration, and the end of all our exploring will be to arrive where we started and know the place for the first time. ~ T. S. Eliot, Little Gidding (1942)

8. Conclusions

8.1 Contributions

As described in the introduction, the main motivation behind this thesis was to find a practical solution to the problem of observer metamerism in industrial applications. However, the contributions of this thesis turn out to be relevant not only for industrial applications where observer metamerism is an issue, but also for more fundamental studies in the domain of color and vision sciences.

The most important contribution of this thesis has been to prove the main hypothesis, that human observers with normal color vision can be classified into a small number of categories based on their color vision. This work proposes a set of eight colorimetric observer categories for use in colorimetry. However, it is important to note that there is no unique way to derive these categories. It is expected that the proposed categories will be further tested, and updated as needed, in future research works of various color and vision scientists. Establishing the most appropriate set of categories was not the main purpose of this work, and this is a task that is better left for a standardization body like Commission Internationale de l'Éclairage (CIE).

Another key contribution of this work is the development of the observer classification method as well as the proof-of-concept prototype, described as the *Observer Calibrator* throughout this thesis. This observer classification method, together with the compact and economical prototype, is the enabling factor for the practical implementation of observer categories in industrial applications. Moreover, the *Observer Calibrator* can be an immensely helpful research tool in all scientific studies in color science and color vision that employ visual psychophysics. This tool, possibly the first of its kind, can help in the selection of a small group of observers, such that this group is representative of a large population of color normal observers. This can possibly be achieved by selecting observers belonging to various categories. It is also possible to select observers only from the categories prevalent or dominant among large population of color normal observers. At present, there is no easy and effective means to do such "observer profiling" of color normal observers, even though multiple color deficiency tests exist, for example, Ishihara PseudoIsochromatic Plates or Anomaloscopes employing Rayleigh matches (for red-green color deficiency test) or Moreland matches (for blue-green color deficiency test). Most scientific studies ignore this very important aspect of experimental design to ensure representativeness of the observer pool participating in psychophysical experiments.

Nonetheless, the outcome of some of these experiments can be critically affected by the choice of observers.

Final contribution of this thesis is to provide a first step toward an implementation of colorimetric observer categories in a practical color imaging workflow. This workflow, described in this thesis as the *observer dependent color imaging* (ODCI), will typically be implemented at the output side, for example, for display processing. The basic idea is to convert color specifications corresponding to CIE 10° standard colorimetric observer (as they generally are) into color specifications corresponding to individual observer categories. The display device must be characterized to derive forward and inverse display models separately for each of these observer categories. Nonlinear transformations that result in accurate color transformations are derived. These transformations will be made available to the research community through the personal website of the author: www.abhijitsarkar.com.

In the next section, various key achievements of the thesis are described in more detail.

8.2 Achievements

8.2.1 *Theoretical analysis of CIE TC 1-36 (CIEPO06) physiologically-based observer model*

A comprehensive theoretical analysis was conducted on various aspects of the physiologically-based observer model proposed by the Technical Committee TC 1-36 of the Commission Internationale de l'Éclairage (CIE). In the context of color perception on modern narrow-band displays, the performance of the CIEPO06 model in predicting the average observer data corresponding to various age-groups was evaluated, and the results were compared with those from the CIE 10° standard colorimetric observer. This analysis used a comprehensive, well recognized color-matching dataset for 47 observers obtained through classical color-matching experiment.

The CIEPO06 model performance was improved significantly upon a nonlinear optimization of the model. It was proposed that one of the physiological factors, namely the photopigment optical density, be made age-dependent.

8.2.2 *Color-matching experiment with two displays to study observer metamerism in narrow-band displays*

The effect of observer metamerism in modern display applications was investigated through color-matching experiments. This involved two displays, one was a Cathode Ray Tube (CRT) display with broad-band primaries, and the other was a Liquid Crystal Display (LCD) with narrow-band primaries. The experimental design took into account several important aspects of large-field color matching.

The results obtained from the experiments involving ten observers showed that while using the CIE standard colorimetric observer the average prediction errors for all observers and all stimuli was lower than some of the similar studies performed in the past, the differences were significant for some stimuli.

8.2.3 *Derivation of eight colorimetric observer categories through statistical analysis*

A two-step method was developed for deriving a minimal set of colorimetric observer categories meeting several predefined requirements. In the first step, five 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 set of 47-observer data from 1959 Stiles-Burch study, and 61 color matching functions derived from the CIEPO06 model corresponding to 20-80 age parameter range. Squared Euclidean distance measure (in cone fundamental space) was used in this analysis. In the second step, a reduced set of representative observer models (or categories) were 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 is that both spectral and colorimetric features of the color-matching functions were considered to minimize model redundancy and ensure uniqueness of the selected categories.

8.2.4 *Development of an observer classification method and implementation using two displays*

An experimental method was developed in order to assign colorimetric observer categories to individual observers. The two displays used in the color matching experiment described before were used. They were characterized using 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.

8.2.5 *Development and testing of Observer Calibrator prototype*

A portable, LED-based instrument prototype for observer classification was conceived. This prototype replicated the observer classification experimental setup based on two displays described in the

previous section. The prototype has an LED driver that controls 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. A correlation analysis was 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. The consistency between observer categories and color difference data gave an indirect validation of the observer classification method. The results also led to the conclusion that colorimetric observer categories, derived from classical color matching data, can help in the prediction of average suprathreshold color difference perception for a given observer population. If this observation is further validated in future research, colorimetric observer categories will have a significant impact on the formulations of color difference metrics and perceptual color spaces.

The *Observer Calibrator* can be an immensely helpful research tool in all scientific studies in color science and color vision that employ visual psychophysics. This tool can help in the selection of a small group of observers, such that this group is representative of a large population of color normal observers. Of course, it can as well be a tool to further study color vision variability.

8.2.6 *Observer-Dependent Color Imaging (ODCI) workflow*

This thesis work provided a first step toward an implementation of colorimetric observer categories in a practical color imaging workflow. Implementation of this workflow in a display application necessitates that colors corresponding to CIE 10° standard colorimetric observer be converted into category-specific colors. Simpler linear transformations as well as three-dimensional Lookup Tables obtained from nonlinear interpolations were computed and preliminary analysis of the results was presented.

8.3 Perspectives

The use of a standard observer in colorimetric computations is essentially based on the assumption that the whole population of color normal observers can reasonably be represented by a single colorimetric observer model, defined by a set of three Color Matching Functions (CMFs). This assumption is arguably the greatest weakness in the formulation of colorimetry. The constraints of this assumption were known to the color science community ever since CIE colorimetry was established,

and adopted universally. Indeed, such an approximation did not pose much problem in any of the conventional industrial applications, until recently. Its weakness has become non-trivial with the advent and wide-spread adoption of modern wide-gamut consumer displays with narrow-band primaries, facilitated by the Light Emitting Diode (LED) technology. Recent studies show that individual variability in color vision characteristics often lead to disagreement among observers over color matches, and overall color experience on such devices is adversely affected. This observation has been reaffirmed in this work. Thus, more than ever before, there is a need to find a practical solution to this issue of observer variability (i.e. observer metamerism). We need a solution that can be effectively implemented in industrial applications.

It is hoped that the concept of colorimetric observer categories, the method of observer classification, and the Observer Calibrator prototype, all developed during the course of this thesis, will contribute toward this goal. The *observer dependent color imaging* (ODCI) workflow proposed in this thesis is an embodiment of the envisaged solution, which once developed further, it can be hoped, will significantly reduce the problem of observer metamerism for color critical applications.

As pointed out in the previous section, a correlation analysis of observer categories and suprathreshold color difference judgment data obtained for the same group of observers showed interesting consistency. This raises a question for the future researchers of this topic - can colorimetric observer categories have a fundamental impact on how we use visual data to derive color difference formulae and perceptually uniform color spaces? It seems doing further research on observer categories will not only be interesting for color imaging applications, but could also prove to be highly relevant for basic research in color science and vision.

This thesis has exploited some of the latest advances made in the field of color vision in the past two decades, and has made a systematic effort to offer a practical and scientifically sound solution to the issue of observer variability. In doing so, this work has attempted to bridge the gap that currently exists between the scientific community of vision researchers, and the professional community of color science specialists.

In perspective, this work tries to move away from conventional wisdom of “standard” and “deviate” observers that has dominated colorimetry for many decades. During the course of this thesis it was observed that the terminology used traditionally in color science community gives way to a misunderstanding, particularly among non-experts. The numerical constructs of a “standard observer” and a “standard deviate observer” have facilitated, though unintentionally, a conception in the general population that a human observer should have the same or similar color vision as represented by the “standard”; otherwise he or she is second-rate or unqualified as an observer. Many people showed a reluctance to participate in the observer classification test, fearing he or she would “fail” the test. The

message that needs to be sent across to the users of color is that it is perfectly fine to have a color vision different from that of an average observer. In fact, it is normal. Based on the results obtained from the observer classification experiments involving 49 observers, no single observer category was assigned to more than a quarter of the whole population. This of course depends on the selection of categories. Also noteworthy is the fact that only around 8% of these observers conformed to the CIE 10° standard colorimetric observer.

This thesis research makes a case for an alternate approach in which the colorimetric observer models would be defined in a way similar to what CIE did to define the illuminant models, by using terms like CIE standard illuminant A, B, C, 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, "colorimetric observer model A" under normal circumstances. One advantage of this method is no 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 specialist can choose to use a more appropriate model. It is also hoped that this approach will encourage the color engineers to be more attentive to the specific observer model being used in their colorimetric computations. Unfortunately, to this date, the observer model appears to be the most neglected aspect of applied colorimetry.

The proposed paradigm shift in the treatment of observer models in colorimetry is easier said than done. As this three years' of research is coming to an end, a humbling realization is setting in that this work is but a stepping stone. Resolving all the standing issues with regard to colorimetric observers is not a matter of one doctoral thesis, and not even one isolated research initiative. A community-wide, concerted effort is needed to take this work to the next level.

Following could be a rough guideline for future work:

1. **Revisit the observer category derivation:** the original dataset of Stiles and Burch observers did not have sufficient representation of higher age-group observers above the age of 50. Also, the effect of genetic polymorphism due to which long- and/or medium-wave sensitive cone fundamentals undergo a peak shift is not likely to be present in the combined dataset. These weaknesses in the underlying dataset can possibly affect the derivation of the observer categories. Further theoretical analysis could be conducted to ascertain the most appropriate mathematical process and statistical data.

2. **Finalize a first set of candidate colorimetric observer categories:** This could be an updated version of eight observer models proposed in this thesis.

3. **Improve upon and standardize the Observer Calibrator instrument:** The prototype demonstrated a strong Maxwell-spot effect. It could possibly be reduced by changing one or more LED primaries. However, this would affect the current balance of luminance, and the common color gamut on two sides of the bipartite field. This aspect needs investigation. Further, it will be important to improve the hardware (LED driver) from the current 8-bit to 10-bit so that color reproduction can be more precise.

4. **Collect a very large amount of observer classification data:** It will be critical to have observer classification data from hundreds of observers around the world, obtained by using the Observer Calibrator and the first set of candidate observer categories.

5. **Finalize the set of standard colorimetric observer categories:** Based on the observer classification data collected in the previous step, we can determine which of the candidate categories are most appropriate to be selected for the final set. From practical point of view, the final set should have minimal number of categories that would satisfy one or more pre-determined criteria with respect to the observer population tested. One of these categories (probably the most dominant one) would be the reference category, the basis of all generic colorimetric computations. This could also be the current CIE 10° standard colorimetric observer (for large-field applications), or preferably, its improved version.

6. **Establish transformations between categories:** For applications that need to account for observer variability, generic color representations need to be converted to observer category-specific color representations, implementing *Observer-Dependent Color Imaging* workflow. Nonlinear transformations similar to those established here could be used, or nonlinear color conversion equations can be developed.

This clearly involves a lot of efforts. Is it worth all that efforts? For many applications, it might be sufficient to have a single observer model. But for those novel applications for which a single observer model is insufficient, or for which an average observer model is inappropriate, it is important that there is a practical, scalable solution. It is the responsibility of us, the color scientists and researchers, to ensure that modern colorimetry has such adaptability. This thesis concludes with this vision for our field in the 21st century.

References

[1] R. A. Young, Oh say, can you see? the physiology of vision. In Proceedings of SPIE Human Vision, Visual Processing and Digital Display, San Jose, CA. **1453**, 92—123 (1991).

[2] S. E. Palmer, *Vision Science – Photons to Phenomenology* (MIT Press, Cambridge, MA., 1999).

[3] D. Hubel, *Eye, brain and vision* (New York: Scientific American Books, 1995).

[4] R. W. Rodieck, *The first steps in seeing* (Sinauer Associates, Sunderland, MA, US, 1998), xi ed.

[5] N. K. Logothetis, Vision: A window on consciousness, Scientific American **281**, 68–75 (1999).

[6] B. A. Wandell, *Foundations of vision* (Sinauer Associates., Sunderland, MA, US, 1995), xvi ed.

[7] O. Packer and D. Williams, *The Science of Color* (Elsevier, Oxford, 2003), vol. 2, chap. Light, the retinal image, and photoreceptors, pp. 41–102, 2nd ed.

[8] P. Lennie, *The Science of Color* (Elsevier, Oxford, 2003), vol. 2, chap. The Physiology of Color Vision, pp. 217–246, 2nd ed.

[9] G. Wyszecki and W. Stiles, *Color science: concepts and methods, quantitative data and formulae* (John Wiley & Sons, New York, NY, 1982), 2nd ed.

[10] R. S. Berns, *Billmeyer and Saltzman's principles of color technology* (John Wiley & Sons, 2000), 3rd ed.

[11] A. Stockman and L. Sharpe, *Color Vision: From Genes to Perception* (Cambridge University Press, Cambridge, 2001), chap. Cone spectral sensitivities and color matching, pp. 53–88.

[12] J. Mollon, *The Science of Color* (Elsevier, Oxford, 2003), vol. 2, chap. The origins of modern color science, pp. 1–39, 2nd ed.

[13] L. M. Hurvich and D. Jameson, An opponent-process theory of color vision, Psychological Review **64**, 384–404 (1957).

[14] CIE, Fundamental Chromaticity Diagram with Physiological Axes – Part I, CIE Technical Report pp. 170–171 (2006).

[15] G. S. Brindley, *Physiology of the retina and the visual pathway* (Oxford, England: Williams and Wilkins., 1960), 2nd ed.

[16] V. C. Smith and J. Pokorny, *The Science of Color* (Elsevier, Oxford, 2003), chap. Color matching and color discrimination, 2nd ed.

[17] J. Pokorny, V. Smith, G. Verriest, and A. Pinckers, eds., *Congenital and acquired color vision defects* (Grune and Stratton, New York, 1979).

[18] A. Stockman, D. I. A. MacLeod, and J. A. Vivien, Isolation of the middle- and long-wavelength-sensitive cones in normal trichromats, J. Opt. Soc. Am. A **10**, 2471–2490 (1993).

[19] A. König and C. Dieterici, Die Grundempfindungen in normalen und anomalen Farben Systemen und ihre Intensitats-Verthielung im Spectrum, Zeitschrift für Psychologie und Physiologie der Sinnesorgane **4**, 241—347 (1893).

[20] G. Meyer and D. Greenberg, Color-defective vision and computer graphics displays, IEEE Computer Graphics and Applications **8**, 28–40 (1988).

[21] H. Brettel, F. Viénot, and J. D. Mollon, Computerized simulation of color appearance for dichromats, J. Opt. Soc. Am. A **14**, 2647–2655 (1997).

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

[23] M. Neitz and J. Neitz, Molecular genetics of color vision and color vision defects, Arch. Ophthalmol. **118**, 691–700 (2000).

[24] L. Sharpe, A. Stockman, H. Jägle, and J. Nathans, *Color Vision: From Genes to Perception* (Cambridge University Press, Cambridge, 2001), chap. Opsin genes, cone photopigments, color vision, and color blindness, pp. 3–52.

[25] W. Wright, *Handbook of Sensory Physiology Vol. VII/4: Visual Psychophysics* (Springer-Verlag, Berlin, 1972), chap. Colour mixture, pp. 435–469.

[26] H. G. Grassmann, "Theory of compound colors, Philosophic Magazine 4(7) 254–264 (1854)", in Sources of Color Science, , D. L. MacAdam, ed. (Cambridge, MA, USA, The MIT press., 1970).

[27] B. Oicherman, Effects of colorimetric additivity failure and of observer metamerism on cross-media colour matching, Ph.D. thesis, Department of Colour Science, The University of Leeds (2007).

[28] CIE, Reappraisal of colour matching and Grassmann's laws (Final draft for ballot), CIE Technical Report (2009).

[29] F. Blottiau, Les défauts d'additivité de la colorimétrie trichromatique, Rev. d'Opt. **26**, 193 (1947).

[30] P. W. Trezona, Additivity of colour equations, Proceedings of the Physical Society. Section B **66**, 548 (1953).

[31] P. W. Trezona, Additivity of Colour Equations: II, Proceedings of the Physical Society. Section B **67**, 513 (1954).

[32] B. Crawford, Colour matching and adaptation, Vision Research **5**, 71–78 (1965).

[33] R. D. Lozano and D. A. Palmer, Large-field color matching and adaptation, J. Opt. Soc. Am. A **58**, 1653–1656 (1968).

[34] P. W. Trezona, Maxwell and maximum saturation methods of colour matching: some sources of error, CIE Symposium on advanced colorimetry, CIE publication x007, Central bureau of the CIE, Vienna, Austria pp. 88–90 (1993).

[35] D. MacLeod and R. Boynton, Chromaticity diagram showing cone excitation by stimuli of equal luminance, J. Opt. Soc. Am. A **69**, 1183–1186 (1979).

[36] R. Boynton and N. Kambe, Chromatic difference steps of moderate size measured along theoretically critical axes, Color Research & Application **5**, 13–23 (1980).

[37] V. Smith and J. Pokorny, Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm, Vision Research **15**, 161–171 (1975).

[38] V. Smith and J. Pokorny, The design and use of a cone chromaticity space: a tutorial, Color Research & Application **21**, 375–383 (1996).

[39] W. D. Wright, A re-determination of the trichromatic coefficients of the spectral colors, Trans. Opt. Soc. **30**, 141–164 (1929).

[40] J. Guild, The colorimetric properties of the spectrum, Philosophical Trans. Royal Soc. London A **230**, 149–187 (1932).

[41] CIE, Recueil des travaux et compte rendu de séances, Sixième Session, Genève, Juillet, 1924 pp. 67–69 (1926).

[42] W. W. Coblentz and W. B. Emerson, Relative sensibility of the average eye to light of different colors and some practical applications of radiation problems, U. S. Bureau of Standards Bulletin **14**, 167 (1918).

[43] K. S. Gibson and E. P. T. Tyndall, Visibility of radiant energy, Bulletin Bureau of Standards **19**, 131 (1923).

[44] W. Stiles and J. Burch, NPL colour-matching investigation: final report (1958), Journal of Modern Optics **6**, 1–26 (1959).

[45] N. I. Speranskaya, Determination of spectral color co-ordinates for twenty-seven normal observers, Optics and Spectroscopy **7**, 424–428 (1959).

[46] CIE, Proceedings 1959 (Brussels), Bureau Central de la CIE, Paris (1960).

[47] D. B. Judd, Judd's method for calculating the tristimulus values of the cie 10° observer, CIE symposium on advanced colorimetry, CIE publication x007, Central bureau of the CIE, Vienna, Austria pp. 107–114 (1993).

[48] CIE, Improvement to industrial colour-difference evaluation, CIE Publication No. 142, CIE Central Bureau, Vienna, (2001).

[49] CIE, Industrial colour-difference evaluation, CIE Publication No. 116, CIE Central Bureau, Vienna (1995).

[50] J. Pokorny, V. Smith, and M. Lutze, Aging of the human lens, Applied Optics **26**, 1437–1440 (1987).

[51] D. van Norren and J. Vos, Spectral transmission of the human ocular media, Vision Research **14**, 1237—44 (1974).

[52] B. Crawford, The scotopic visibility function, Proceedings of the Physical Society, B **62**, 321—34 (1949).

[53] K. H. Ruddock, Evidence for macular pigmentation from colour matching data, Vision Research **3**, 417 – 429 (1963).

[54] F. Viénot, *Colour Vision* (Academic Press, New York, 1983), chap. Can variation in macular pigment account for the variation of colour matches with retinal position?, pp. 107—16.

[55] B. R. J. Hammond, B. R. Wooten, and D. M. Snodderly, Individual variations in the spatial profile of human macular pigment, J. Opt. Soc. Am. A **14**, 1187–1196 (1997).

[56] F. Viénot, Color-match changes between the fovea and the perifovea, Color Research & Application **8**, 215–220 (1983).

[57] J. Moreland, Analysis of variance in anomaloscope equations, Documenta Opthalmologica Proceedings Series **39**, 111–119 (1984).

[58] J. Moreland, Moreland match revisited, Visual Neuroscience **21**, 471–476 (2004).

[59] S. S. Miller, Psychophysical estimates of visual pigment densities in red-green dichromats, The Journal of Physiology **223**, 89–107 (1972).

[60] V. Smith and J. Pokorny, Psychophysical estimates of optical density in human cones, Vision Research **13**, 1199–1202 (1973).

[61] A. Renner, H. Knau, M. Neitz, J. Neitz, and J. Werner, Photopigment optical density of the human foveola and a paradoxical senescent increase outside the fovea, Visual neuroscience **21**, 827–834 (2004).

[62] W. Swanson and G. Fish, Age-related changes in the color-match–area effect, Vision research **36**, 2079–2085 (1996).

[63] V. Smith, J. Pokorny, and S. Starr, Variability of color mixture data–I. Interobserver variability in the unit coordinates, Vision Research **16**, 1087–1094 (1976).

[64] J. Neitz and G. Jacobs, Polymorphism in normal human color vision and its mechanism, Vision research **30**, 621–636 (1990).

[65] A. B. Asenjo, J. Rim, and D. D. Oprian, Molecular determinants of human red/green color discrimination, Neuron **12**, 1131 – 1138 (1994).

[66] M. Neitz, J. Neitz, and G. Jacobs, Spectral tuning of pigments underlying red-green color vision, Science **252**, 971–974 (1991).

[67] S. Merbs and J. Nathans, Absorption spectra of human cone pigments, Nature **356**, 433—5 (1992).

[68] L. T. Sharpe, A. Stockman, H. Jägle, H. Knau, G. Klausen, A. Reitner, and J. Nathans, Red, green, and red-green hybrid pigments in the human retina: Correlations between deduced protein sequences and psychophysically measured spectral sensitivities, The Journal of Neuroscience **18**, 10053–10069 (1998).

[69] F. Viénot, Muséum National d'Histoire Naturelle, Centre de Recherche sur la Conservation des Collections, 36 rue Geoffroy Saint-Hilaire, 75005 Paris, (personal communication, 2011).

[70] V. C. Smith and J. Pokorny, Large-field trichromacy in protanopes and deuteranopes, J. Opt. Soc. Am. A **67**, 213–220 (1977).

[71] D. A. PALMER, Tetrachromatic matches, J. Opt. Soc. Am. **62**, 828–830 (1972).

[72] P. Trezona, The tetrachromatic colour match as a colorimetric technique, Vision Research **13**, 9 – 25 (1973).

[73] P. Trezona, Additivity in the tetrachromatic colour matching system, Vision Research **14**, 1291 – 1303 (1974).

[74] D. C. Rich and J. Jalijali, Effects of observer metamerism in the determination of human color-matching functions, Color Research & Application **20**, 29–35 (1995).

[75] P. K. Kaiser and H. Hemmendinger, The color rule: A device for color-vision testing, Color Research & Application **5**, 65–71 (1980).

[76] W. Stiles and G. Wyszecki, Intersections of the spectral reflectance curves of metameric object colors, J. Opt. Soc. Am. A **58**, 32–38 (1968).

[77] N. Ohta and G. Wyszecki, Location of the nodes of metameric color stimuli, Color Research & Application **2**, 183–186 (1977).

[78] R. Kuehni, Intersection nodes of metameric matches, Color Research and Application **4**, 101–102 (1979).

[79] W. Thornton, Matching lights, metamers, and human visual response, J. Color Appearance **2**, 23–29 (1973).

[80] R. Berns and R. Kuehni, What determines crossover wavelengths of metameric pairs with three crossovers? Color Research & Application **15**, 23–28 (1990).

[81] F. Viénot, Muséum National d'Histoire Naturelle, Centre de Recherche sur la Conservation des Collections, 36 rue Geoffroy Saint-Hilaire, 75005 Paris, (personal communication, 2011).

[82] R. Kuehni and R. Ramanath, Comparing observers, Color Research & Application **29**, 183–186 (2004).

[83] E. Allen, An index of metamerism for observer differences : . Proceedings of the 1st AIC congress, Color 69, Musterschmidt, Göttingen pp. 771–784 (1970).

[84] Y. Nayatani, K. Hashimoto, K. Takahama, and H. Sobagaki, Comparison of methods for assessing observer metamerism, Color Research & Application **10**, 147–155 (1985).

[85] K. Takahama, H. Sobagaki, and Y. Nayatani, Prediction of observer variation in estimating colorimetric values, Color Research & Application **10**, 106–117 (1985).

[86] N. Ohta, Formulation of a standard deviate observer by a nonlinear optimization technique, Color Research & Application **10**, 156–164 (1985).

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

[88] A. North and M. Fairchild, Measuring color-matching functions. Part II. New data for assessing observer metamerism, Color Research & Application **18**, 163–170 (1993).

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

[90] B. Oicherman, M. Luo, A. Tarrant, and A. Robertson, The uncertainty of colour-matching data, in Proceedings of IS&T/SID 13th Color Imaging Conference, (2005), pp. 326–332.

[91] K. Katori and M. Fuwa, Spectral luminous efficiency function derived from color matching functions of 10 degree field, Acta Chromatica **3**, 129–140 (1979).

[92] Y. Nayatani, T. Tanaka, H. Sobagaki, K. Takahama, and K. Hashimoto, Field trials for assessing the method of observer metamerism adopted by cie, Color Research & Application **16**, 97–107 (1991).

[93] W. Wright, A re-determination of the mixture curves of the spectrum, Trans Opt Soc London **31**, 201–211 (1929–30).

[94] H. Fairman, M. Brill, and H. Hemmendinger, How the cie1931 color-matching functions were derived from the wright–guild data, Color Research & Application **22**, 11–23 (1997).

[95] A. D. Broadbent, A critical review of the development of the cie1931 rgb color-matching functions, Color Research & Application **29**, 267 (2004).

[96] I. Nimeroff, J. R. Rosenblatt, and M. C. Dannemiller, Variability of spectral tristimulus values, J. Opt. Soc. Am. A **52**, 685–691 (1962).

[97] M. Shaw and M. Fairchild, Evaluating the 1931 cie color-matching functions, Color Research & Application **27**, 316–329 (2002).

[98] W. Stiles and J. Burch, Interim report on the commission internationale de l'eclairage, zurich, on the national physical laboratory's investigation of colour matching, Optica Acta **2**, 168–181 (1955).

[99] A. Stockman, Colour & vision research laboratory website, http://www.cvrl.org/, (2011).

[100] F. Viénot, New equipment for the measurement of color-matching functions, Color Research & Application **2**, 165–170 (1977).

[101] F. Viénot, Relations between inter- and intra-individual variability of color-matching functions. experimental results, J. Opt. Soc. Am. A **70**, 1476–1483 (1980).

[102] J. Pokorny, V. C. Smith, and S. J. Starr, Variability of color mixture data–ii. the effect of viewing field size on the unit coordinates, Vision Research **16**, 1095 – 1098 (1976).

[103] W. Thornton, Toward a more accurate and extensible colorimetry. Part I. Introduction. the visual colorimeter-spectroradiometer. Experimental results, Color Research & Application **17**, 79–122 (1992).

[104] W. Thornton, Toward a more accurate and extensible colorimetry. Part II. Discussion, Color Research & Application **17**, 162–186 (1992).

[105] W. Thornton, Toward a more accurate and extensible colorimetry. Part III. Discussion (continued), Color Research & Application **17**, 240–262 (1992).

[106] CIE, Symposium on advanced colorimetry, CIE Publication No. x007, Commission Internationale de l'Eclairage, Vienna, Austria (1993).

[107] B. Oicherman, R. Luo, and A. Robertson, Observer metamerism and colorimetric additivity failures in soft proofing, in Proceedings of the 14th Color Imaging Conference (CIC), (2006), pp. 24–30.

[108] C. Oleari and M. Pavesi, Grassmann's laws and individual color-matching functions for non-spectral primaries evaluated by maximum saturation technique in foveal vision, Color Research & Application **33**, 271–281 (2008).

[109] Y. Nakano, M. Kurokami, H. Moriki, K. Suehara, J. Kohda, and T. Yano, Polychromator using digital micro-mirror device and its application to additivity test of color matching, Proceedings of the 25th session of the CIE, Division 1, San Diego, USA. **1**, 56–59 (2003).

[110] R. Lozano and D. Palmer, The additivity of large-field colour matching functions, Vision Research **7**, 929 – 937 (1967).

[111] A. North and M. Fairchild, Measuring color-matching functions. Part I, Color Research & Application **18**, 155–162 (1993).

[112] Y. Nayatani, Comments to the Articles "Measuring Color-Matching Functions, Part I and Part II" by Amy D. North and Mark D. Fairchild, Color Research & Application **19**, 383–389 (1994).

[113] I. Pobboravsky, Effect of small color differences in color vision on the matching of soft and hard proofs, in TAGA Proceedings, (1988), p. 62–79.

[114] J. A. Diaz, A. Chiron, and F. Viénot, Tracing a metameric match to individual variations of color vision, Color Research & Application **23**, 379–389 (1998).

[115] B. Oicherman, R. Luo, B. Rigg, and A. Robertson, Effect of observer metamerism on colour matching of display and surface colours, Color Research & Application **33**, 346–359 (2008).

[116] P. Csuti and J. Schanda, Colour matching experiments with RGB-LEDs, Color Research & Application **33**, 108–112 (2008).

[117] P. Csuti and J. Schanda, A better description of metameric experience of LED clusters, in Proceedings of Light and Lighting Conference with Special Emphasis on LEDs and Solid State Lighting, (Budapest, Hungary, 2009).

[118] Onelight spectra, http://www.onelightcorp.com/, (2011).

[119] How DLP Technology Works, http://www.dlp.com/technology/how-dlp-works/default.aspx, (2011).

[120] D. Judd, *Technical Committee No. 7 Report of Secretariat United States Commission* (International Commission on Illumination, 1951), vol. 1, pp. 1– 60.

[121] J. J. Vos, Colorimetric and photometric properties of a 2° fundamental observer, Color Research & Application **3**, 125–128 (1978).

[122] CIE, Report CIE 1988 2° spectral luminous efficiency function for photopic vision, CIE Publ 86-1990 **86** (1990).

[123] A. Stockman, D. MacLeod, and N. Johnson, Spectral sensitivities of the human cones, J. Opt. Soc. Am. A **10**, 2491–2521 (1993).

[124] D. MacLeod and M. Webster, Factors influencing the color matches of normal observers, Colour Vision: Physiology and Psychophysics, JD Mollon and LT Sharpe, eds.(Academic, London) pp. 81–92 (1983).

[125] V. Smith, J. Pokorny, and Q. Zaidi, How do sets of color-matching functions differ, Colour Vision: Physiology and Psychophysics, JD Mollon and LT Sharpe, eds.(Academic, London) pp. 93–105 (1983).

[126] Y. Nakano, Y. Nakayasu, H. Morita, K. Suehara, J. Kohda, and T. Yano, Individual difference of color matching functions and its cause, ISCC/CIE Expert Symposium, Ottawa, Ontario, Canada (2006).

[127] J. He and S. Shevell, Individual differences in cone photopigments of normal trichromats measured by dual rayleigh-type color matches. Vision research (1994).

[128] F. Viénot, L. Serreault, and P. P. Fernandez, Convergence of experimental multiple rayleigh matches to peak l-and m-photopigment sensitivity estimates, Visual neuroscience **23**, 419–427 (2006).

[129] J. Barbur, M. Rodriguez-Carmona, J. Harlow, K. Mancuso, J. Neitz, and M. Neitz, A study of unusual rayleigh matches in deutan deficiency, Visual neuroscience **25**, 507–516 (2008).

[130] P. Thomas and J. Mollon, Modelling the rayleigh match, Visual neuroscience **21**, 477–482 (2004).

[131] M. A. Webster, Reanalysis of lambda-max variations in the stiles-burch 10-degrees color-matching functions, J. Opt. Soc. Am. A **9**, 1419–1421 (1992).

[132] V. Smith and J. Pokorny, Chromatic-discrimination axes, CRT phosphor spectra, and individual variation in color vision, J. Opt. Soc. Am. A **12**, 27–35 (1995).

[133] E. Miyahara, V. Smith, and J. Pokorny, How surrounds affect chromaticity discrimination, J. Opt. Soc. Am. A **10**, 545–553 (1993).

[134] A. Elsner, L. Berk, S. Burns, and P. Rosenberg, Aging and human cone photopigments, J. Opt. Soc. Am. A **5**, 2106–2112 (1988).

[135] M. Webster and D. MacLeod, Factors underlying individual differences in the color matches of normal observers, J. Opt. Soc. Am. A **5**, 1722–1735 (1988).

[136] F. Ebner and M. Fairchild, Development and testing of a color space (IPT) with improved hue uniformity, in Proceedings of IS&T/SID 6th Color Imaging Conference, (1998), pp. 8–13.

[137] C. Bartleson, *Threshold and Matching* (Academic Press, Orlando, FL, 1984), vol. 5: Visual Measurements of *Optical Radiation Measurement*, chap. 7.

[138] M. Luo, G. Cui, and B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, Color Research & Application **26**, 340–350 (2001).

[139] R. Johnson and D. Wichern, *Handbook of Applied Multivariate Statistical Analysis* (Pearson Education International, NJ, USA, 2007), 6th ed.

[140] P. A. J. Gore, *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (Academic Press, San Diego, CA, 2000), chap. Cluster Analysis, pp. 297–321.

[141] P. Urban, M. Fedutina, and I. Lissner, Analyzing small suprathreshold differences of LCD-generated colors, J. Opt. Soc. Am. A **28**, 1500—1512 (2011).

[142] E. D. Montag and D. C. Wilber, A comparison of constant stimuli and gray-scale methods of color difference scaling, Color Research & Application **28** (2003).

[143] CIE, Parametric effects in colour difference evaluation, CIE Publication No. 101, Technical report, Central Bureau of the CIE, Vienna, Austria (1993).

[144] R. G. Kuehni, Variability in estimation of suprathreshold small color differences, Color Research and Application **34**, 367–374 (2009).

[145] R. G. Kuehni, Color difference formulas: An unsatisfactory state of affairs, Color Research & Application **33**, 324–326 (2008).

[146] M. D. Fairchild, *Color appearance models* (Wiley-IS&T Series in Imaging Science and Technology, Chichester, West Sussex, England, 2005), 2nd ed.

[147] R. Hunt, Objectives in colour reproduction, J. Phot. Sci. **18**, 205–215 (1970).

[148] M. Fairchild, A color scientist looks at video, 3rd International Workshop on Video Processing and Quality Metrics (VPQM), Scottsdale, Invited Paper 1 (2007).

[149] J. H. Wold and A. Valberg, The derivation of xyz tristimulus spaces: A comparison of two alternative methods, Color Research & Application **26**, S222–S224 (2001).

[150] M. Fairchild and D. Wyble, Mean observer metamerism and the selection of display primaries, Final Program and Proceedings-IS&T/SID Color Imaging Conference, Albuquerque, NM, USA pp. 151–156 (2007).

[151] R. Ramanath, Minimizing observer metamerism in display system, Color Research & Application **34**, 391–398 (2009).