Towards Standardizing a Reference White Chromaticity for High Definition Television

Matthew Donato, *Rochester Institute of Technology*, *College of Imaging Arts and Sciences*, *School of Film and Animation Technology*, December 5, 2014

***Abstract*—Emerging high definition displays exhibit great spectral variance with respect to their unique illuminant technologies. As a result, their spectral power distributions differ. When calibrating these spectrally unique displays, calibration instruments claim chromaticity matches while film industry professionals (naturally diverse in their color perception) experience chromaticity mismatches. This was not an issue when early phosphor-based displays in use were spectrally similar to one another. Both a simulation and psychophysical test are developed in an attempt to quantify differences in color perception between classic phosphor-based displays and emerging high definition displays. The proposed method can be used to effectively determine a mean visually-corresponding chromaticity offset from a given standard white point chromaticity for both LED and OLED displays to satisfy a greater population of observers. While this offset may be satisfactory for a greater number of observers, a single observer model cannot accurately predict metameric matches for an entire population of diverse observers. This issue is magnified as three-primary color rendering becomes increasingly monochromatic.**

I. INTRODUCTION

KNOWLEDGE of the fundamentals of human color vision, display-rendered imagery and display calibration standards is necessary to understand why display-rendered imagery may be colorimetrically mismatched between observers.

A. Human Color Vision

Human color perception is derived from a visual integration of the complex spectral reflection information from a particular environment. This integrated stimulus is then processed by the brain, rendering a perception of color. Color is best described as a human-assigned label for perceived stimuli; objects in the world do not inherently have color.

Key to normal human color perception are three individual cone classes unique such that cones in each class vary in sensitivity to different regions of visible electromagnetic energy. For this reason, normal human color perception is labeled trichromatic. These cone classes are labeled with respect to their sensitivities to “short,” “medium,” and “long” wavelengths of visible energy, which vary between individual observers. Such sensitivities are often simplified to the perception of blue, green and red light, respectively. A unique integration of a scene by each cone class is necessary to perceive color; an observer with only a single cone class would perceive a monochromatic environment.

Manuscript received December 5, 2014. This work submitted in partial fulfillment of the requirements for the B.S. degree in Motion Picture Science from the School of Film and Animation, Rochester Institute of Technology.

As a result of this integration of a spectrally complex world, it is possible for a variety of spectrally unique targets to produce identical responses for a given observer. Spectrally variant targets perceived as color matches are known as metamers [1].

B. Human Perception of Display Technologies

In display rendered imagery, the objective is to take advantage of the limitations of integrations performed by the human visual system. By using only a limited subset of emissive illuminants known as primaries, displays can recreate metameric matches of real scenes. If an observer believes that the colors perceived on a display match his/her expectation of real-world scenes, his/her visual system has been “fooled,” and a metameric match has been successfully rendered.

Cathode ray tube (CRT) displays originally dominated the marketplace for commercially available color displays. CRT displays render color by translating input video signal to a corresponding voltage. This voltage is used to fire an electron gun, varying in intensity, upon three unique classes of phosphors. These standardized phosphors are used to render red, green or blue image signals. From a comfortable viewing distance, the array of these small samples of red, green and blue energy are spatially and temporally integrated by the human visual system. CRT displays are capable of rendering metameric matches of real scenes with a limited gamut (i.e. range) of colors.

The specific light-emissive properties of the phosphor elements of a CRT display define its color reproduction qualities and limitations. A spectral power distribution (SPD) defines the intensity (or power per area) of an emissive source at each wavelength of electromagnetic energy. For typical display devices, the complete region of electromagnetic energy is often truncated to the visible light region. Figure 1 illustrates an SPD for the red, green and blue primaries of a typical broadcast CRT display.

C. The Issue With Emerging High Definition Displays

The integrated response of trichromatic human vision can be algebraically rearranged so that it represents chromaticity, defining only the color of a stimulus without respect to its luminance or brightness. It is in this chromaticity space that a pair of chromaticity coordinates are used to describe simply the color of a once complex spectral reflectance of a target for a single observer. In display calibration, standards bodies define pairs of chromaticity coordinates to which the chromaticity of display primaries should be adjusted to ensure consistent color reproduction across devices. An end user is responsible for modifying the intensity of his/her display primaries until a colorimetric measurement made

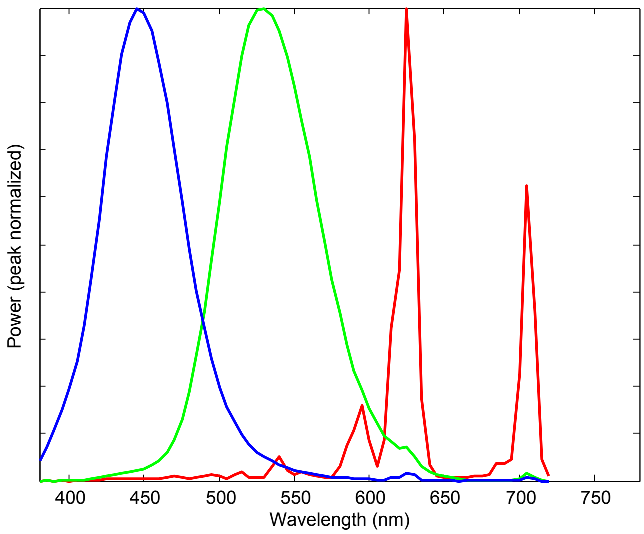


Fig. 1: SPD of the red, green and blue phosphor primaries of a typical broadcast CRT display, peak normalized. Plot line colors correspond to display phosphor color designation.

with a colorimeter matches the chromaticity aim of a given standard.

Given the wide, consistent and standardized primaries of CRT displays, the implementation of a standard chromaticity aim often renders consistent, repeatable calibrations. This results in an increased number of metameric matches and successful colorimetric renderings. Standards for display primary chromaticity aims were written when CRT displays dominated the consumer display market.

Presently, high definition displays dominate the consumer market for viewing motion picture, broadcast and web content. Advancements have been made to display technologies to improve their color gamut, enabling displays to render a wider range of colors otherwise prohibited by the physical limitations of phosphor-based CRT displays. Other advancements in high definition display technologies improve energy efficiency, display dynamic range and physical display shape (i.e., curved displays). These new displays all require the use of different types of display primaries. With these new displays, professionals in the film industry have noticed that colorimeters report identical chromaticity coordinates for varying display technologies that simultaneously appear to be visual mismatches.

Colorimetric instruments have the capacity to quantitatively report colorimetry with respect to a single observer, often mathematically modeled. It is unlikely that a single observer model can predict colorimetry accurately and consistently with respect to every real observer as populations of observers are diverse with respect to their color perception.

Large arrays of new, high definition displays can be found aligned panel-to-panel in consumer electronics stores, and meticulous inspection often reveals significant visual mismatches between these displays. It is extremely likely that a particular display vendor has not calibrated their displays, however, such color differences may be magnified as a result of the great variance in SPDs of primary illuminants of high definition displays.

Included in this group of high definition displays are those illuminated by light emitting diodes (LED), organic light emitting diodes (OLED) and cold cathode fluorescent (CCFL) bulbs as well as plasma displays. Figures 2a-2e illustrate the variety amongst SPDs of the primaries for these display illuminants. For comparison, the SPD of a broadcast CRT display is repeated in Figure 2f.

Compared to the SPDs of the primaries of a CRT display, SPDs of the primaries of high definition display illuminants, particular LED displays and OLED displays, are significantly narrower with sharper peaks and with less spectral overlap.

Amongst LED SPDs alone, spectral primary variations are present with respect to their widths and amount of overlap. These variations exist as a result of the inexpensive manufacturing costs and lack of standardization of LED illuminants.

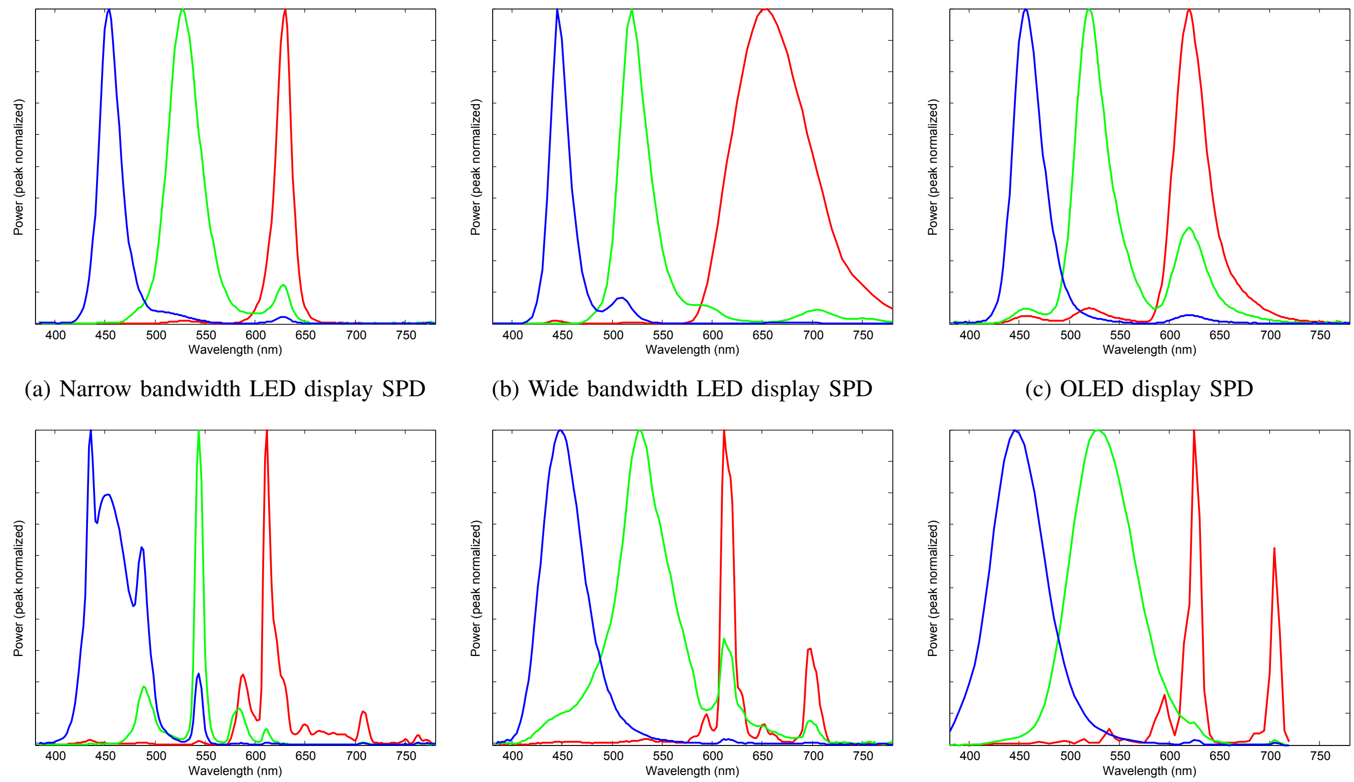
Conversely, the SPD of a plasma display is visually similar to the SPD of a CRT display.

Changes in the spectral properties of display illuminant technologies increase the likelihood that naturally diverse observers will experience different colors for a single rendered image. Displays utilizing narrower bandwidth primaries increase the magnitude of this variance. At the same time, colorimeters used to tune displays can only predict colorimetry with respect to a single mathematical observer model. They cannot, in their simplicity, model matched colorimetry for all observers. Such spectral variations are therefore problematic as artists and advertisers wish to render specific colorimetry for their motion pictures, broadcast content and web content.

II. A Brief History of Broadcast Television Standards

In a 1953 petition to the Federal Communications Commission (FCC), the National Television System Committee (NTSC) defined a series of specifications for CRT displays to be accepted as standards. This petition sets the framework to define colors which should be rendered when various drive signals are sent to a CRT display [2]. Contained within this petition is a standard aim calibration of displayed red, green, blue and white point chromaticities as measured by a colorimeter. Red, green and blue aim chromaticities were defined to be measured when solely red, green or blue signals were sent to a display. A signal which drives equal amounts of the red, green and blue primaries should render an image sans chromaticity coordinates. This rendered signal is defined to be the white point of the display. The selection of the white point standard is a chromaticity aim derived from the full spectrum of Standard Illuminant C, originally defined by the International Commission on Illumination.

In 1979, another standard was set forth by the Society of Motion Picture and Television Engineers (SMPTE). This standard defines a new, yet similar, set of chromaticity coordinates to which the phosphors of CRT displays should be tuned, originally labeled P-22. Most notably, the definition of a white point has been changed in this standard so that its chromaticity matches that of CIE Standard Illuminant D65. Standard Illuminant D65 “is intended to represent average daylight and has a correlated colour temperature of approximately 6500 K” [3]. Like Standard Illuminant C, Standard Illuminant D65 models a complete spectrum of an illuminant. The standard refers only to the chromaticity of



(a) Narrow bandwidth LED display SPD

(b) Wide bandwidth LED display SPD

(c) OLED display SPD

(d) CCFL display SPD

(e) Plasma display SPD

(f) CRT display SPD

Fig. 2: SPDs for six selected display illuminant technologies, peak normalized. These SPDs are derived from peak-normalized average SPDs of displays with identical illuminant technologies. SPDs are measured when fully-driven red, green and blue drive signals are sent to each display in its native state. Some displays employ color matrixing in their native state, as possibly evidenced by the similarity in shape of the smaller of the two green primary SPD peaks and the single red primary SPD peak of the OLED display SPD, Figure 2c. Plot line colors correspond to display primary color designation.

this illuminant viewed on a CRT display with respect to a single observer. The P-22 standard has since been updated; as of 1994, this specification is denoted as SMPTE C [4]. The D65 aim white point chromaticity for display calibration has remained unchanged to this day.

Further standards to define chromaticity aims for broadcast displays have been developed by the International Telecommunication Union (ITU): ITU-R BT.601 [5] and ITU-R BT.709 [6], originally published in 1982 for standard definition displays and 1990 for high definition displays, respectively. Such standards retain the D65 white point used in the SMPTE C phosphor standard, while describing similar chromaticity aims for red, green and blue phosphors. Both ITU standards have also been updated since original publication. The most recent chromaticity aims for the standards mentioned thus far are summarized in Table I.[1]

Using simplified and limited chromaticity coordinates to define display calibration techniques was sufficient to render consistent colorimetry with phosphor-based illuminants as a result of the broad and overlapping SPDs of their primaries. Additionally, phosphors used in CRT displays were standardized, thus making their spectral properties more consistent. This resulted in more metameric matches

TABLE I: Selected standard chromaticity aims to be used with the calibration of broadcast displays.

	White		Red	
	x	y	x	y
NTSC	0.310	0.316	0.67	0.33
ITU-R BT.601-7	0.3127	0.3290	0.630	0.340
ITU-R BT.709-5	0.3127	0.3290	0.640	0.330
SMPTE RP 145-2004	0.3127	0.3290	0.630	0.340

	Green		Blue	
	x	y	x	y
NTSC	0.21	0.71	0.14	0.08
ITU-R BT.601-7	0.310	0.595	0.155	0.070
ITU-R BT.709-5	0.300	0.600	0.150	0.060
SMPTE RP 145-2004	0.310	0.595	0.155	0.070

between diverse observers and calibration equipment.

Given the increase in metameric failures observed when calibrating the white point chromaticity of high definition displays, there are present discussions amongst SMPTE regarding methodologies to overcome this calibration conundrum. The intentions of such discussions are to release updated standards and recommended practices for overcoming metameric failures present for observers of high definition displays.

The objective of this experiment is to devise a method of improving white point calibration standardization practices.

[1]Assuming 525-line North American standard.

The white point chromaticity of a number of high definition displays is visually matched to the white point of a reference CRT display, and the SPD of the white point on both displays is measured. With this matched high definition display spectrum, an analysis of colorimetric integration techniques is completed in order to refine the usage of existing and expensive single-observer model colorimeters and measurement techniques. The intention is that high definition displays can then be calibrated to render colors more satisfactorily for a diverse population of observers.

III. Color Science as it Relates to Display Calibration

A. The CIE 1931 2° Standard Observer

Since the perception of color requires an observer, it is understood that any attempt to standardize simplified colorimetric measurements must be made with respect to a given observer. As a result, the CIE defined a hypothetical observer, the CIE 1931 2° standard observer, which was intended to be used for all colorimetric calculations [7].

Early assessments to quantify the unique sensitivities of each cone class were completed in separate experiments by Guild (1926-1927) and Wright (1928-1929) consisting of a combined 17 individual observers, each with normal trichromatic vision. Understanding the additive nature of light, psychophysical color matching assessments devised by Guild and Wright tasked observers with matching varying amounts of separate red, green and blue lights to a number of monochromatic stimuli. In order for all observers to make successful matches, it was sometimes necessary for additional light to be added to the monochromatic stimuli. This resulted in an empirically assessed negative sensitivity to light for a particular cone class.

The mean results from the experiments performed by Guild and Wright define an average relationship between wavelength of visible electromagnetic energy and sensitivity (also known as cone fundamentals) for each of the three cone classes for the small population of normal trichromatic observers. Since some spectral sensitivity measurements resulted in negative sensitivities which are impractical to integrate, the three curves were linearly recombined into the color matching functions known as the CIE 1931 2° standard observer. These three standard observer functions, denoted as $\bar{x}$, $\bar{y}$, and $\bar{z}$, are illustrated in Figure 3. Importantly, $\bar{y}$ represents the typical human photopic response to luminance or brightness.

Given variations in cone class distributions within the retina and the setup of the experiments performed by Guild and Wright, these color matching functions are only valid for stimuli presented within a 2° field of view.

Based on the physiology of color perception, the standard observer functions are used to predict integrated colorimetry. By multiplying each color matching function with the spectrum of a source or source illuminant, $SPD(\lambda)$, and integrating the result, tristimulus values are computed. These tristimulus values, X, Y, and Z are computed using Equations 1-3.

$$X = \int_{380nm}^{780nm} SPD(\lambda)\bar{x}(\lambda)d\lambda \tag{1}$$

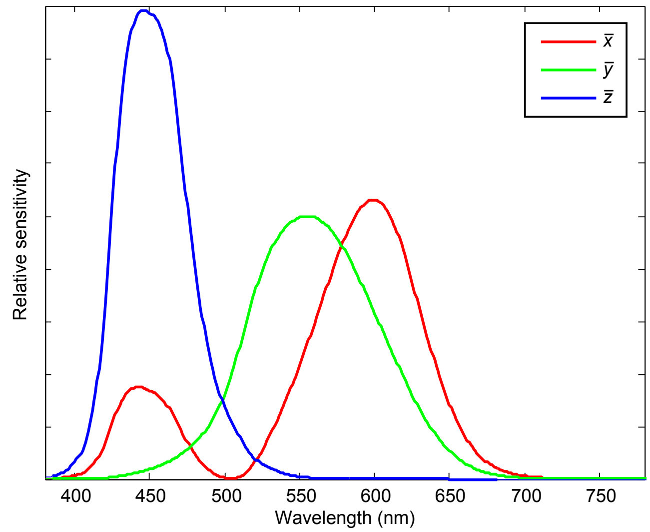


Fig. 3: The CIE 1931 2° standard observer color matching functions.

$$Y = \int_{380nm}^{780nm} SPD(\lambda)\bar{y}(\lambda)d\lambda \tag{2}$$

$$Z = \int_{380nm}^{780nm} SPD(\lambda)\bar{z}(\lambda)d\lambda \tag{3}$$

XYZ tristimulus values are a means to quantify the luminance and chrominance of a given stimulus with respect to the CIE 1931 2° standard observer. The values for each of the three dimensions of this color space bear resemblance to the three unique cone responses of human photopic color vision.

The integrated XYZ tristimulus responses as predicted with the CIE 1931 2° standard observer are computed discretely with increments no smaller than $d\lambda = 5$nm; "no significant difference in the results would follow from employing a smaller interval" [7]. At this point, it is possible to consistently define and repeatably reproduce a color stimulus as perceived by the standard observer under any condition.

Computing XYZ tristimulus values also enables the prediction of metameric stimuli with respect to the CIE 1931 2° standard observer. If Equations 1-3 yield the same XYZ tristimulus as any number of spectrally variant stimuli, those stimuli are all considered metamers. Practically, if a given observer (with his/her own unique color matching function) has experienced the same color appearance for spectrally disparate stimuli, the observer has experienced metamers.

XYZ tristimulus values can be further simplified to represent the color of a stimulus without regard to its luminance or brightness. This space is known as xy-chromaticity, and is computed using tristimulus normalizations as outlined in Equations 4 and 5. It is in this space that the gamut of human-perceivable colors is defined. Its boundary, known as the spectral locus, is illustrated in Figure 4 [8], marked with the location of the chromaticity of D65 white.

$$x = \frac{X}{X+Y+Z} \quad (4)$$

$$y = \frac{Y}{X+Y+Z} \quad (5)$$

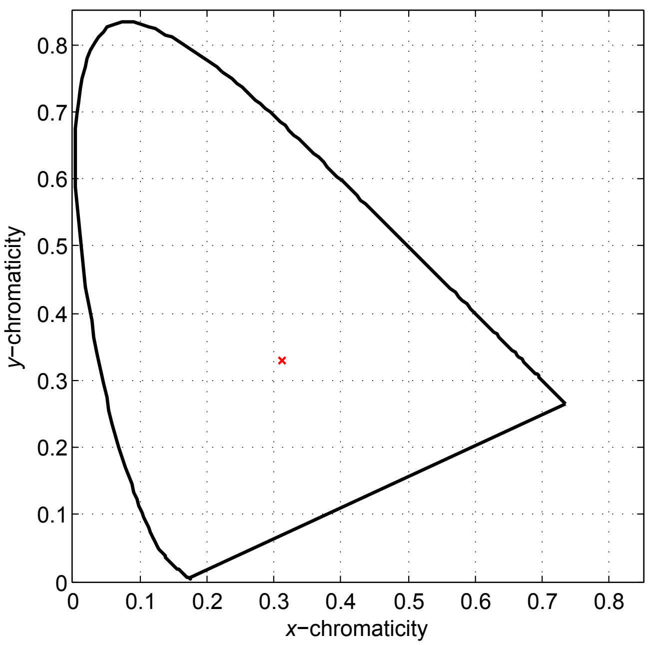


Fig. 4: xy-chromaticity diagram with the chromaticity coordinates of D65 white, (0.3127, 0.3290), in red.

The xy-chromaticity coordinates serve to define a space in which fundamental aim calibration chromaticity values (Table I) are specified. By integrating a combination of the spectrum of a source with $\bar{x}$, $\bar{y}$, and $\bar{z}$, colorimeters can report normalized xy-chromaticity coordinates for a stimulus. In a display calibration environment, professional display hardware controls are modified until the desired chromaticity aims are achieved within an acceptable tolerance.

B. Beyond the CIE 1931 2° Standard Observer

Given that the CIE 1931 2° standard observer is only valid for stimuli presented in a 2° field of view, the CIE published a second color matching function in 1964. This color matching function is appropriately used to predict XYZ tristimulus values for stimuli presented in a 10° field of view and is derived from psychophysical experiments performed by Stiles, Burch and Speranskaya published in 1959 [9].

With respect to the original 1931 color matching function, research by Judd published in 1951 concluded that the CIE 1931 2° standard observer and original luminous efficiency function overestimated sensitivity for short wavelengths of visible energy [10]. As a result, Judd proposed modifications to the $\bar{x}$, $\bar{y}$, and $\bar{z}$ functions which most drastically affected sensitivities to wavelengths below 460nm.

Following the modifications to the standard observer proposed by Judd, Vos proposed additional refinements in 1978 given the availability of improved computational procedures [10]. The original standard observer, Judd-modified standard observer and Judd/Vos-modified standard observer are plotted in Figure 5.

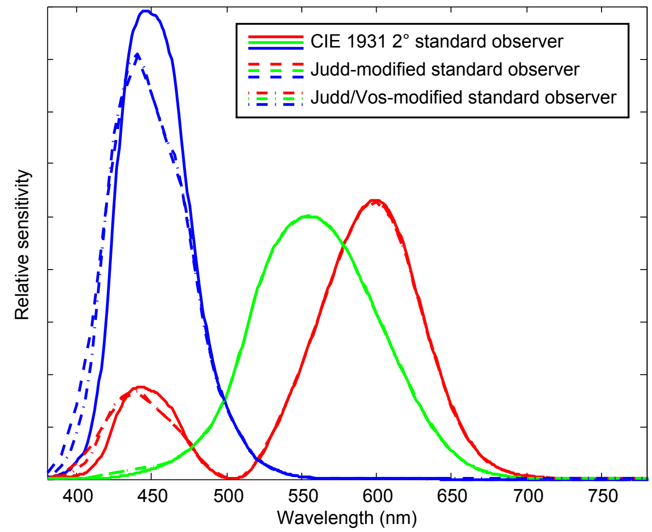


Fig. 5: Comparison of the CIE 1931 2° standard observer, Judd-modified standard observer and Judd/Vos-modified standard observer. Blue, green and red plot line colors correspond to $\bar{x}$, $\bar{y}$, and $\bar{z}$ for each color matching function, respectively.

C. Opponent Color Theory and CIELAB

The basis of the color vision theory discussed thus far is defined as trichromatic theory, supported by the presence three unique cone classes in trichromatic human vision. However, an assessment of a subjective labeling of colors performed by Hering in 1920 identifies color opponents that are never experienced simultaneously for a given stimulus [11]. For example, it is not possible to distinguish a reddish-green or yellowish-blue stimulus. However it is possible to identify stimuli containing all other combinations of red, green, yellow and blue. Further psychophysical testing performed in the 1950s provided quantitative evidence to support that opponent signals are sent to the brain for color processing as opposed to raw trichromatic cone responses.

In 1976, the CIE defined CIELAB, a color space which describes colors with respect to their luminance, L^* and chrominance separated along two axes: a^*, which represents a difference in redness versus greenness and b^*, which represents a difference in yellowness versus blueness. Colorimetry in the CIELAB space is computed with respect to the XYZ tristimulus values of a stimulus as well as the of XYZ tristimulus values of a reference white viewing illuminant, X_n, Y_n, and Z_n. Using Equations 6-8, L^*, a^* and b^* values are computed for XYZ tristimulus values greater than 0.008856 [11].

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16 \quad (6)$$

$$a^* = 500\left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3}\right] \quad (7)$$

$$b^* = 200\left[\left(\frac{Y}{Y_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3}\right] \quad (8)$$

For the average observer who understands color labeling, opponent color theory is sometimes more intuitively understood than trichromatic color theory.

D. Color Differences

To quantify perceptual color differences between stimuli, the CIE developed an elegant Pythagorean color difference formula in 1976, known as ΔE_{76}. Following this, the CIE developed more sophisticated models to quantify color differences. Often used for its computational robustness, ΔE_{94} converts colors in the CIELAB space to colors in a $L^*C^*h^*$ (lightness, chroma, hue) space prior to computing color differences [12]. It is generally accepted that a ΔE_{94} value greater than 1 indicates a noticeable color difference. ΔE_{94} is selected for color difference assessments in this experiment.

E. Recent Developments in Human Color Perception to Quantify and Combat Observer Metamerism

The understanding and significance that a single observer model cannot be used to accurately predict colorimetry for all diverse observers leads to active research towards developing new models for representing human color perception. Recognizing this, Alfvin and Fairchild noted that there will may be quantifiable uncertainties in color matching predictions using a single observer model [13]. In 1996, an experiment was designed where observers were tasked to match the chromaticity of a CRT display to seven adjacent, fixed and illuminated color transparencies. The spectrum of the CRT display was measured and XYZ tristimulus values were computed with respect to the CIE 1931 2°, 1964 10° and Stiles & Burch 2° observer models. For a population of 20 observers each completing the experiment three times, inter- and intra-observer variability was assessed with color differences up to 19 ΔE units and just above 10 ΔE units, respectively. Notably, the CIE standard observer models were still found to predict color within an acceptable tolerance with respect to the matches made by observers.

Given such diversity amongst normal color observers, the CIE published a means of deriving a number of new color matching functions for a given observer based on his/her age and subtended viewing angle [14]. This model, CIE2006, (named after its publication date) acknowledges that changes in age and subtended viewing angle have quantifiable effects on color perception.

In 2010 Sarkar et al. assessed a combination of 108 cone fundamental models in order to more efficiently predict color while reducing observer metamerism for industrial applications [15]. These 108 models consisted of 61 models from the CIE2006 model as well as 47 cone models from the work completed by Stiles and Burch. A statistical cluster analysis was used to separate the group of 108 models into seven most variant models. Through a forced-choice color matching psychophysics experiment, the parameters of the cluster analysis were verified. It was concluded that the subset of seven cone fundamental models could be used to more effectively represent color perception variability amongst a group of diverse observers than a single observer model. The following year, research conducted by Fedutina et al. added an eighth unique observer to this set of cone fundamental models [16]. Classifying a wide population of observers with a subset of cone fundamental models (as opposed to a single model) could be beneficial for reducing metamerism, provided that the cone fundamentals of all real observers are to be assessed.

Using Monte Carlo simulation techniques, Fairchild and Heckaman simulated 1,000 observer models and assessed the color appearance of an X-Rite Color Checker with respect to each observer [17]. Of the 24 color samples on the X-Rite Color Checker, color differences up to a ΔE_{94} value of 33 were computed for the 1,000 observer models, with the mean color difference for each of the 24 patches computed to be a ΔE_{94} value of 9.6.

F. Metamerism Failures and Display Calibration

Variations in observer color prediction reveal the importance of designing displays whose spectral characteristics are capable of rendering metamerically consistent imagery for a population of observers. It is less likely that imagery rendered on displays with narrower primaries will render metameric matches between observers experiencing display-rendered imagery.

With the goal of improving color gamut for their previously popular CRT reference displays, Sony Corporation released an OLED broadcast reference display in 2011 as a replacement. After using colorimeters to calibrate their OLED displays to the D65 white point chromaticity of (0.3127, 0.3290) and visually comparing this white point to a D65-white calibrated CRT display, observers experienced colorimetric mismatches between the two technologies [18].

Sony examined the SPDs of both their CRT reference displays and OLED displays and identified that the integration used to determine xy-chromaticity coordinates with respect to the CIE 1931 2° standard observer did not yield perceptual matches, which is to be expected.

To combat this, Sony asked a number of observers to view a white reference patch CRT display calibrated to a D65 white point with respect to the CIE 1931 2° standard observer and adjust the chromaticity of an OLED display until it rendered a visual match with the CRT display. Then, Sony measured the chromaticity of the visually-matched OLED with respect to the CIE 1931 2° standard observer. This resulted in a cluster of points geometrically displaced from that of D65 white on the chromaticity diagram.

Sony concluded that if the Judd-modified standard observer was used to match the appearance of white on the OLED display to the appearance of D65 white on the CRT display, then the chromaticity of the white point (with respect to the CIE 1931 2° standard observer) would then render more satisfying results for a number of observers. Understanding that colorimeters used in practice today typically measure chromaticity with respect to the CIE 1931 2° standard observer, Sony recommended that, in order to calibrate their OLED displays to D65 white, users should seek an aim chromaticity of (0.3067, 0.3180) as opposed to (0.3127, 0.3290).

In 2014, Long and Fairchild assessed the perception of three-primary, seven-primary and eight-primary displays with respect to the CIE2006, Sarkar/Fedutina et al., and

Fairchild/Heckaman observer models to determine the effects of considering additional display primaries on display perception [19]. Of a collection of three-primary displays, monochromatic three-primary displays were found to reproduce the widest possible gamut but did so at the expense of observer metamerism. Increasing the number of display primaries in a monochromatic system was also found to improve color gamut, so long as the peak wavelength of the selected primaries are empirically assessed to improve observer consistency. This also increases the likelihood of rendering more colorimetrically relevant imagery.

IV. Display Matching Simulation

A. *Constructing the Simulation*

Given that the xy-chromaticity of D65 white (0.3127, 0.3290) as viewed on a CRT reference display is used for calibration, recent research consistently concludes that this simple colorimetric aim used with respect to the CIE 1931 2° standard observer will not render matching white chromaticity for spectrally different displays.

Standards organizations seek to determine a more appropriate set of chromaticity coordinates for new displays. In order to publish standards and recommended practices for use by manufacturers and end users employing colorimeters, a simulation is devised based on the empirical assessment completed by Sony. To represent a diverse population of observers, 13 observer models included in the simulation:

- Eight statistically variant color matching functions as determined by Sarkar, et al. and Fedutina, et al.;
- The CIE 1931 2° standard observer;
- The Judd-modified CIE 1931 2° standard observer;
- The Judd/Vos-modified CIE 1931 2° standard observer;
- An XYZ representation of the CIE 2006 2° cone fundamentals [20]; and
- The Stiles & Burch (1955) 2-deg color matching functions [20].

Given the following variables, the simulation computes predicted corresponding xy-chromaticity offsets from (0.3127, 0.3290) with respect to CIE 1931 2° standard observer "viewing" a high definition display visually matched to a properly calibrated reference display:

- The fully-driven, individually-measured, red, green and blue primary SPDs of a reference display, $[RGB]_{ref}$, a $[3 \times N]$ column vector; specifically, a Sony PVM-14l2 broadcast reference CRT display;
- A reference color matching function, $[CMF]_{1931}$, an $[N \times 3]$ row vector; specifically, the CIE 1931 2° standard observer;
- An XYZ tristimulus aim to which the displays will be matched, a $[1 \times 3]$ row vector; specifically, D65 white, $[95.047, 100, 108.883]$;
- A new color matching function to model an observer, $[CMF]_o$, an $[N \times 3]$ row vector to be selected from the aforementioned list; and
- The fully-driven, individually-measured, red, green and blue primary SPDs of a new display to be matched, $[RGB]_{new}$, a $[3 \times N]$ column vector.

For an illuminated display, $[RGB]$, the XYZ tristimulus values of a given signal can be computed with respect to a particular color matching function, $[CMF]$, as shown in Equation 9. Using matrix mathematics, Equation 9 is algebraically rearranged in simulation to calculate corresponding chromaticity coordinates with respect to a new observer.

$$XYZ = [RGB] \cdot [CMF] \tag{9}$$

The simulation is built in MATLAB 7.12.0 following the mathematics outlined in Figure 6.

With the XYZ tristimulus values of D65 white with respect to the CIE 1931 2° standard observer, $[XYZ]_{D65} = [95.047, 100, 108.883]$, determine a scale factor, S, a $[1 \times 3]$ row vector, to modify the reference display spectral power distribution to a proper calibration.

$$S = [XYZ]_{D65} \cdot [[RGB]_{ref} \cdot [CMF]_{1931}]^{-1}$$

↓

Use scale factor S to simulate an ideally calibrated reference display, $[RGB]_{ref,S}$.

$$[RGB]_{ref,S} = S \cdot [RGB]_{ref}$$

↓

Determine a visual match tristimulus $[XYZ]_m$ using the ideal calibrated reference display, $[RGB]_{ref,S}$.

$$[XYZ]_m = [RGB]_{ref,S} \cdot [CMF]_o$$

↓

Determine a second scale factor, S_{new}, a $[1 \times 3]$ row vector, to match a new display to $[XYZ]_m$ with respect to a unique observer model, $[CMF]_o$.

$$S_{new} = [XYZ]_m \cdot [[RGB]_{new} \cdot [CMF]_o]^{-1}$$

↓

Use scale factor S_{new} to simulate a visual match, $[RGB]_{new,m}$, of the new monitor to the tristimulus values of the reference display, $[XYZ]_m$, computed with respect to $[CMF]_o$.

$$[RGB]_{new,m} = S_{new} \cdot [RGB]_{new}$$

↓

Calculate corresponding XYZ tristimulus values, $[XYZ]_{cc}$, of the visually matched new display, $[RGB]_{new,m}$ with respect to the CIE 1931 2° standard observer, $[CMF]_{1931}$.

$$[XYZ]_{cc} = [RGB]_{new,m} \cdot [CMF]_{1931}$$

↓

Determine corresponding xy-chromaticity coordinates, xy_{cc} for $[XYZ]_{cc}$ using Equations 4 and 5.

↓

Compute the offset chromaticity from D65 white, xy_{off}, by subtracting $xy_{D65} = (0.3127, 0.3290)$ from xy_{cc}.

$$xy_{off} = xy_{cc} - xy_{D65}$$

Fig. 6: Corresponding chromaticity simulation flowchart.

B. Confidence Ellipses

Given the corresponding chromaticity offset for any number of observers for a particular display, a 95% confidence ellipse is computed for the spread of the population. Assuming normal variability amongst a population observers, this ellipse would encompass chromaticity offsets visually assessed by 95% of that population.

C. Using the Simulation to Assess Appropriate Chromaticity Offsets for Different Display Illumination Technologies

Using a Photo Research PR-705 SpectraScan Spectroradiometer as well as data collected by peers, the fully-driven SPDs of the individual red, green and blue primaries of 24 displays with varying illuminant technologies are gathered. Display backlight illuminants include, but are not limited to: CCFL bulbs, LEDs and OLEDs. Corresponding chromaticity offsets are computed for the 24 displays using the procedures outlined in Section IV-A. The offsets, encircled with 95% confidence ellipses are illustrated in Figure 7. The mean offsets for each display are tabulated in Table II.

Generating the 24 offsets for each display technology via unique observer models reveals differences in the magnitude of observer variability as a function of display illuminant type. At the same time, similar display illuminants have similar spreads of corresponding chromaticity offsets. For example, Displays 1-4, 13, and 14 are all CCFL-backlit displays. The simulated spread chromaticity offsets for these six displays are consistent in size and orientation, as are the spread of offsets for displays 16 and 18 which are both plasma displays. The corresponding chromaticity offset simulation for display 9, a laser projector, yields the greatest spread of observer offsets. This is to be expected as laser primaries are nearly monochromatic.

For a more concise assessment of observer variability as a function of display illuminant technology, an average peak-normalized SPD of each illuminant technology is computed. The simulation is run again and the corresponding chromaticity results are shown Figures 8a-8e, with tabulated mean offsets in Table III. These display models correspond to the average SPDs illustrated in Figures 2a-2e, respectively.

TABLE III: Simulated mean xy-chromaticity offsets from D65 white calculated for 5 peak-normalized averaged display illuminant technologies corresponding to Figures 8a-8e.

Display Type	x-offset	y-offset
Narrow bandwidth LED	-0.0014	-0.0036
Wide bandwidth LED	-0.0001	-0.0027
OLED	-0.0009	-0.0075
CCFL	-0.0011	-0.0017
Plasma	-0.0006	-0.0014

Results of this simulation reveal a varying spread of diversity amongst observer models with respect to their simulated visual match of a CRT reference display and an alternate display. Consistent with the hypothesis, there is a noticeable relationship between the bandwidth and spread of the SPDs of the primaries of a display and the spread of observer chromaticity matches; primaries whose SPDs are narrow yield larger spreads of observer chromaticity matches. Based on its design, the corresponding chromaticity offset simulation suggests that real observers may also vary in color perception diversity with respect to their visual matches of reference CRT displays and a variety of other display technologies.

Notably, there are significant differences in the spread of corresponding chromaticity offsets for simulated observer models for two different types of LED displays, illustrated in Figures 8a and 8b. Such diversity is present as a result of varying spectral characteristics due to the lack of standardization of LED illuminants.

As a result of its spectral similarities to CRT displays, the distribution of simulated observer model chromaticity offsets for plasma displays (Figure 8e) is compact and centered closely about D65 white. Unsurprisingly, industry observers have not made remarks with regards to unpleasant color rendering or metameric failures with plasma displays. This is to be expected, as the SPD of a plasma display, Figure 2e, is similar to that of a CRT display, Figure 2f.

D. Using the Simulation to Assess Intra-Display Chromaticity Offset Variability

To quantify the consistency of a single xy-chromaticity offset for a given display model, the fully-driven SPDs of the individual red, green and blue primaries of 40 NEC PA242W LED-backlit LCD displays are measured using a Photo Research PR-655 SpectraScan Spectroradiometer. These SPD measurements are also processed through the simulation described in Section IV-A to yield 40 unique xy-chromaticity offsets for each of the 40 like-modeled displays.

The 40 xy-chromaticity offset plots with 95% tolerance ellipses are illustrated in Figure 9, and a statistical analysis of the 40 mean offsets is tabulated in Table IV. The ΔE_{94} mean color difference from the mean (MCDM) chromaticity offset for each of the 40 unique xy-chromaticity offsets is 0.5618. This color difference is less than a ΔE_{94} of 1, indicating that a chromaticity offset prediction for a particular display is likely valid for all displays of the same model.

TABLE IV: Offset statistics for the 40 mean xy-chromaticity offsets assessed in simulation for each of 40 NEC PA242W LED-backlit LCD displays.

	x-offset	y-offset
Mean	-0.0001	-0.0027
Min.	-0.0008	-0.0038
Max.	0.0007	-0.0007
Std. dev.	0.0003	0.0010

.

V. Psychophysical Display Matching Experiment

A. Building the Experiment

Using MATLAB 7.12.0 with Psychtoolbox-3 [21], a method of adjustments experiment is designed in order to enable an observer to adjust the chromaticity of a display whose illumination technology is spectrally variant from that of a reference CRT display, a Sony PVM-14L2, until the observer believes the two displays are visually matched.

Observers are asked to make color matches by manipulating the chromaticity of a white stimulus presented on the

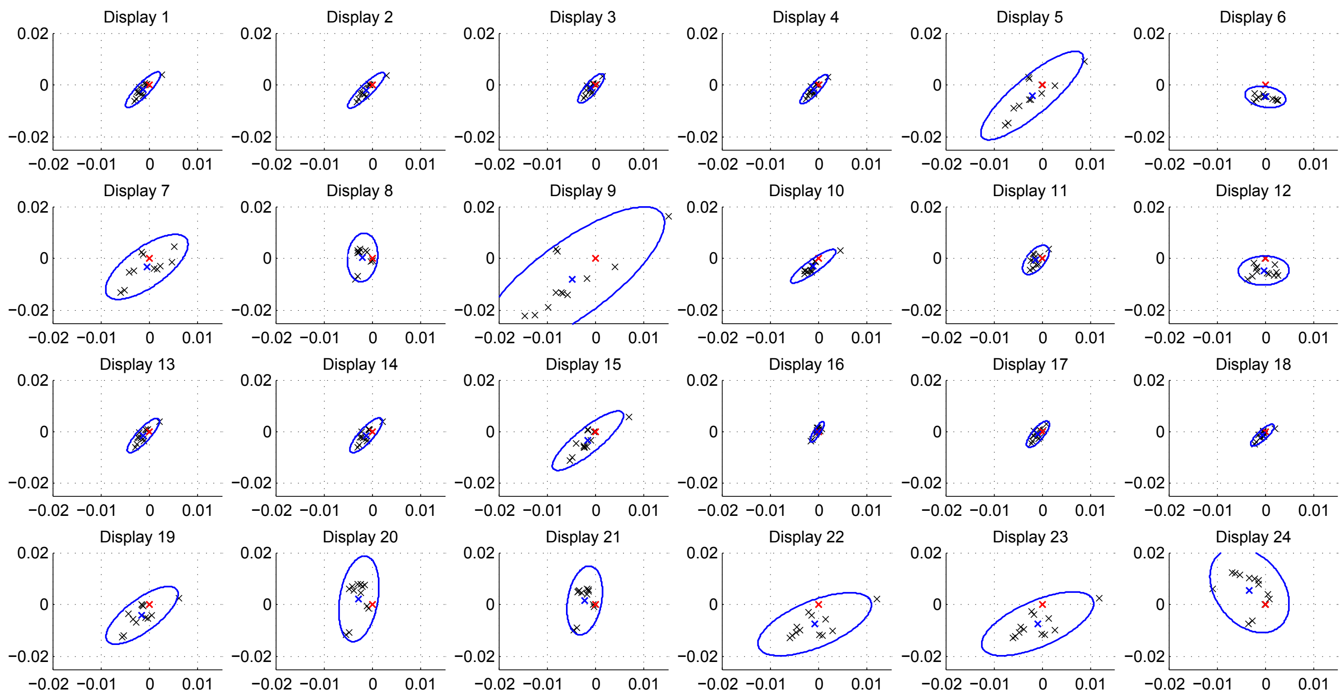
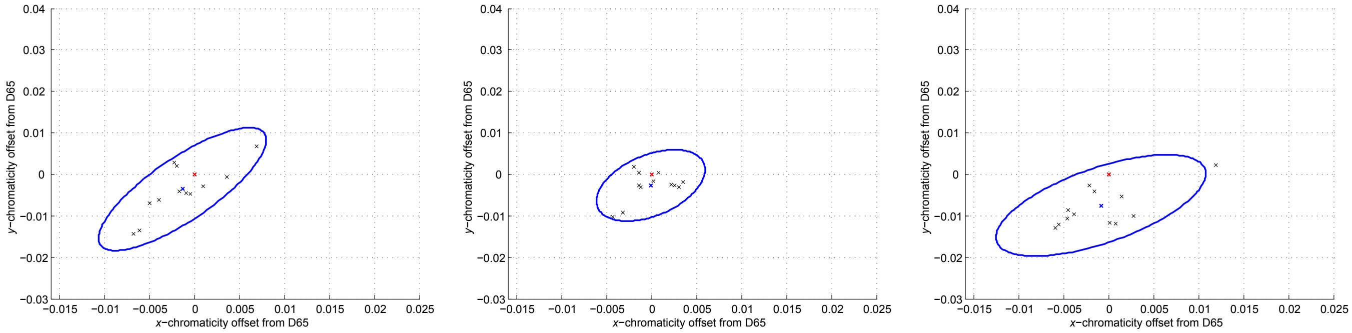


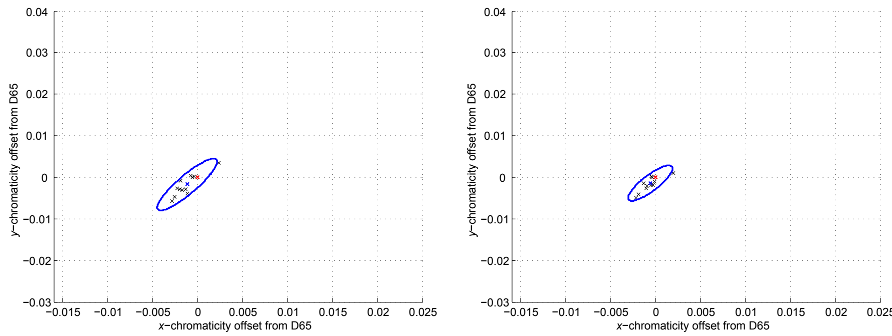
Fig. 7: Assessment of the corresponding chromaticity offsets for 24 displays with various illuminant technologies. The origin, marked with a red ×, represents D65 white chromaticity. Each black × represents the corresponding chromaticity offset for a given observer model, and the blue × indicates the mean corresponding chromaticity offset for a given display. 95% confidence ellipses encircle individual observer model offsets in blue. The x-chromaticity offset from D65 is recorded on the horizontal axis, the y-chromaticity offset from D65 is recorded on the vertical axis.

TABLE II: Simulated mean xy-chromaticity offsets from D65 white calculated for 24 spectrally variant displays corresponding to Figure 7.

Display No.	Display Name	x-offset	y-offset
1	Apple Cinema Display A1082 (1)	-0.0013	-0.0018
2	Apple Cinema Display A1082 (2)	-0.0013	-0.0020
3	Dell 1708FPt	-0.0009	-0.0013
4	Dell 2405FPW	-0.0009	-0.0016
5	Dolby PRM-4200	-0.0021	-0.0041
6	EIZO S2433W	0.0001	-0.0046
7	HP LP2480zx	-0.0006	-0.0033
8	HP ZR2440w	-0.0021	0.0004
9	ITU-R BT.2020-compatible lasers	-0.0049	-0.0080
10	Leader LV5800	-0.0011	-0.0030
11	NEC 3000 3-DLP digital cinema projector	-0.0013	-0.0005
12	NEC LCD3090WQXi-BK	-0.0003	-0.0047
13	Panasonic BT-LH1700WP (1)	-0.0013	-0.0015
14	Panasonic BT-LH1700WP (2)	-0.0013	-0.0015
15	Panasonic PT-AX200U	-0.0016	-0.0035
16	Panasonic TC-P50S30	-0.0003	-0.0001
17	Samsung LN40C630K1FXZA	-0.0009	-0.0010
18	Samsung PN51D8000FFXZA	-0.0006	-0.0014
19	Samsung SCH-I545 (Galaxy S4)	-0.0015	-0.0041
20	Sharp Aquos LC-46LE830U	-0.0028	0.0022
21	Sharp Aquos LC-70UD1U	-0.0023	0.0016
22	Sony BVM-E250	-0.0008	-0.0076
23	Sony BVM-E250A	-0.0009	-0.0074
24	Sony XBR-46HX929	-0.0032	0.0055



(a) Narrow bandwidth LED display offset (b) Wide bandwidth LED display offset (c) OLED display offset



(d) CCFL display offset (e) Plasma display offset

Fig. 8: Chromaticity offset simulation results for five selected display illuminants created from the average SPDs of identical illuminant technologies. Offsets computed with respect to the corresponding average illuminant SPDs defined in Figures 2a-2e, respectively. Plot legend consistent with that of Figure 7.

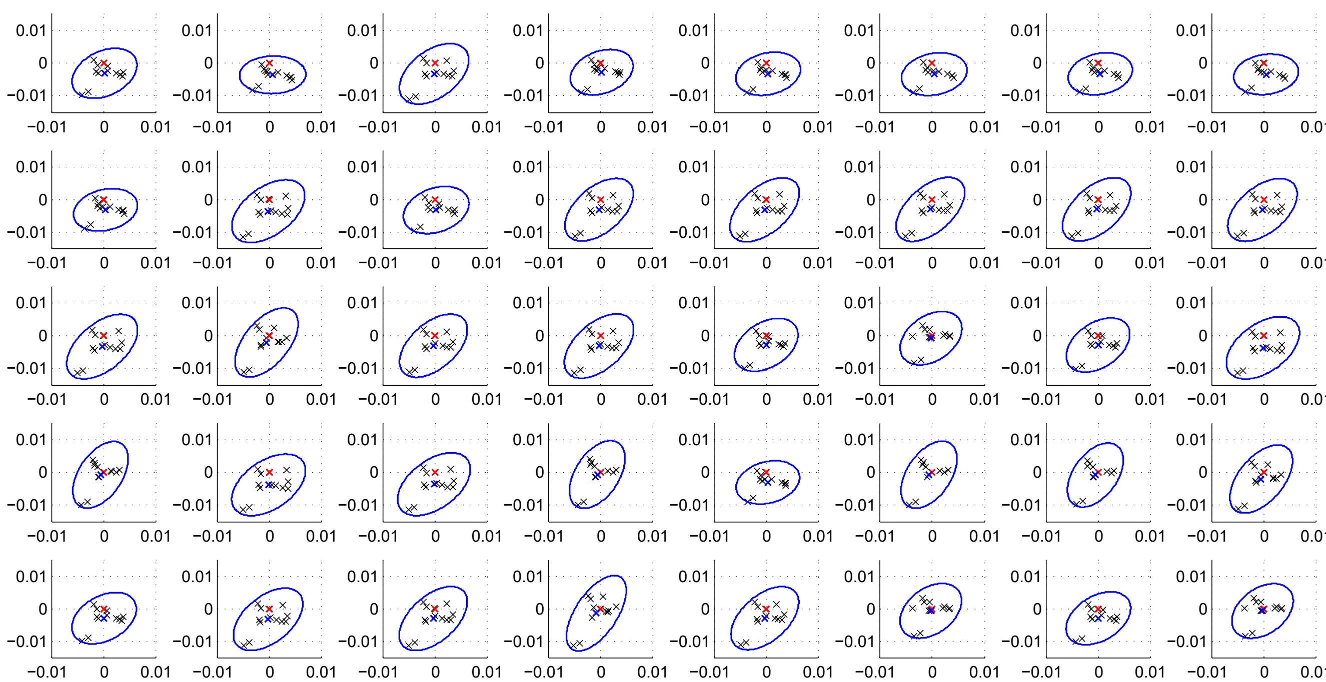


Fig. 9: Assessing intra-display by computing corresponding chromaticity offsets for 40 NEC PA242W LED-backlit LCD displays. Plot legend and axes consistent with that of Figure 7.

non-CRT display technology through 2° field of view via an opponent-based CIELAB-like controller mapped to a traditional keyboard. The selection of a CIELAB-like controller is based on the intuitively understood nature of the color space with minimal explanation. Figure 10 illustrates the relationship between the arrow keys and the corresponding color adjustment direction. The colored square is printed and affixed to the keyboard as a reference for the observer.

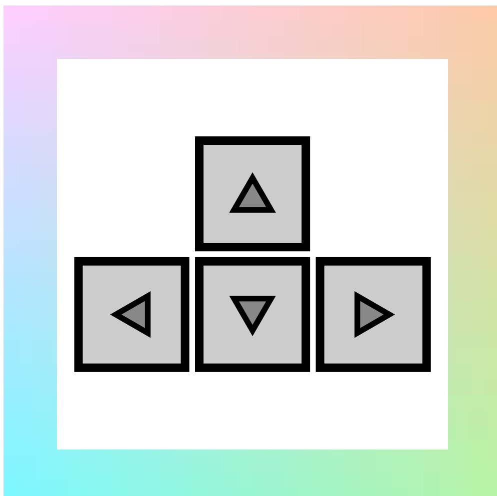


Fig. 10: Corresponding color directions of the CIELAB-based color adjustment controller mapped to the arrow keys of a traditional keyboard.

Display drive code values are sent to be rendered with respect to display RGB primary amounts, not CIELAB values. As a result, observer chromaticity adjustments in the CIELAB-based space are first transformed to XYZ tristimulus values (with respect to the spectrum of the D65 standard illuminant as a white point) followed by a transformation to RGB drive code values. A single button press of an arrow key does not change the drive signal by more than a single 8-bit code value.

The displays are masked with black foam core board such that an observer sitting at an appropriate viewing distance will assess a visual match via a window which subtends two 2° field of views, one for the reference CRT display stimulus and the other for the new display stimulus, as shown in Figure 11.

The choice to select a 2° field of view is consistent with the calibration tools and methodologies used by professionals and creative individuals who perform colorimetric measurements with respect to the CIE 1931 2° standard observer.

Two display models with different illuminant technologies are chosen for the psychophysical display matching experiment: an NEC PA242W LED-backlit LCD reference monitor and a Panasonic BT-LH1700WP CCFL-backlit LCD field monitor. The NEC display is chosen given its popular use in color correction. It also effectively represents the emerging class of LED-backlit displays available on the market today. The Panasonic display is chosen for its consistent use on production sets; this highlights the industry-placed importance of capturing the correct image on set.

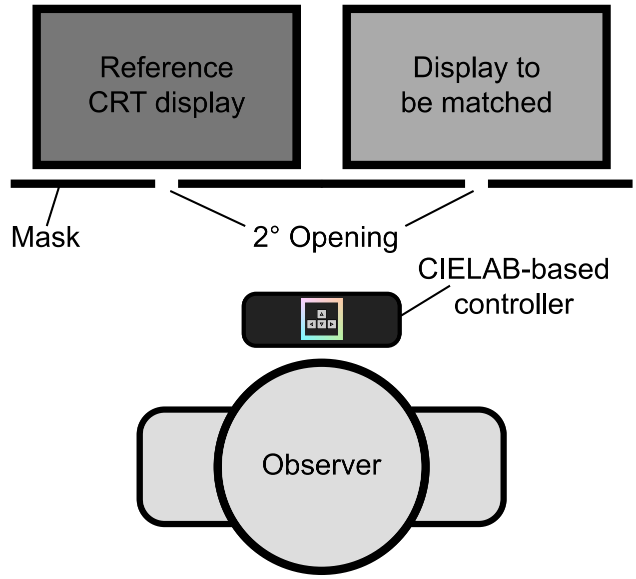


Fig. 11: Psychophysical display matching experiment setup.

B. Chromaticity Drift

Given the physical properties of phosphors in a CRT display, it is possible for fluctuations in voltage as a function of time to render a potentially observable shift in chromaticity.

To characterize the chromaticity drift of the reference CRT display, the SPD of D65 white on the display is measured with the PR-655 at one minute intervals for 146 minutes. The xy-chromaticity coordinates are calculated at each minute using Equations 1 through 5 with respect to the CIE 1931 2° standard observer. The chromaticity drift as a function of time is illustrated in Figure 12. The impact of the chromaticity drift with respect to the empirical assessments is discussed in Section VI.

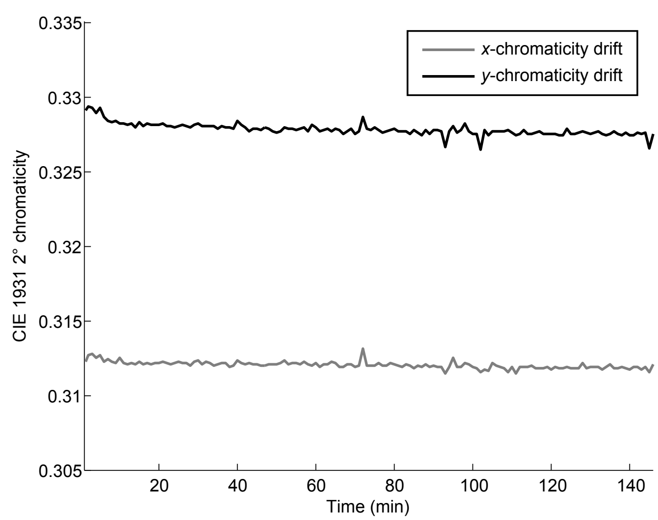


Fig. 12: Measured CRT phosphor chromaticity drift, 146 minutes.

C. Running the Experiment

For each display, 34 unique observers perform the experiment. These observers are given the option to perform the experiment on either or both displays, but never the same display twice. The vast majority of observers are between the ages of 19 and 22.

At the beginning of each day of experimentation, the reference CRT display is given 20 minutes to warm up. Prior to the start of each iteration of the experiment, the chromaticity of the CRT display is set to D65 white (0.3127, 0.3290) and its luminance is set to 100 cd/m^2. The luminance of the display to be matched is also set to 100 cd/m^2. The exact luminance of this display is subject to change insignificantly with respect to the use of the CIELAB-based chromaticity adjustment tool by the observer.

Beginning the experiment, the lights are turned off in the room and the SPD of the reference CRT display is measured with the PR-655. This records any inherent chromaticity drift or fluctuation of the phosphors of the CRT between experiment iterations. The measurement is read five times to reduce measurement error. Calculation of the chromaticity offset is done with respect to the xy-chromaticity of the reference CRT display as measured at the start of the experiment for each observer as opposed to the fixed chromaticity standard of (0.3127, 0.3290). These measured differences in corresponding chromaticity from a slightly imperfect D65 white calibration between displays are translated prior to assessment so that all offsets are reported with respect to (0.3127, 0.3290).

On the display to be matched, the observer is presented with a random stimulus within ± 5 a^*b^* from (0, 0) with respect to the mismatched display. He/she is then asked to adjust the chromaticity of the display to be matched using the CIELAB-based controller (Figure 10) until he/she believes a satisfactory visual match is achieved between the CRT reference display and the display to be matched. Once the observer affirms that a satisfactory match has been made, a spectral reading of the white stimulus on the display to be matched is measured with the PR-655. The observer completes the visual matching process a total of four times to reduce intra-observer variability error.

VI. Results

For effective comparison between simulated results and empirical assessment, a mean SPD for both the NEC display and Panasonic display is calculated given the SPD of 40 like-modeled NEC displays and 5 like-modeled Panasonic displays. These average SPDs or these displays are illustrated in Figures 14a and 14b. This allows conclusions to be drawn about the application of a single empirically assessed chromaticity offset with identical model displays. The gamuts of the display models used in the experiment compared to the gamut of the reference CRT display are illustrated in Figure 13.

The simulated chromaticity offsets for the mean SPDs computed for the NEC and Panasonic displays are illustrated in Figures 15a and 15b. Comparatively, the empirically assessed chromaticity offsets for the NEC and Panasonic displays are illustrated in 15c and 15d. The confidence ellipses alone calculated from both simulation result offsets and empirical result offsets are plotted in Figures 15e and 15f.

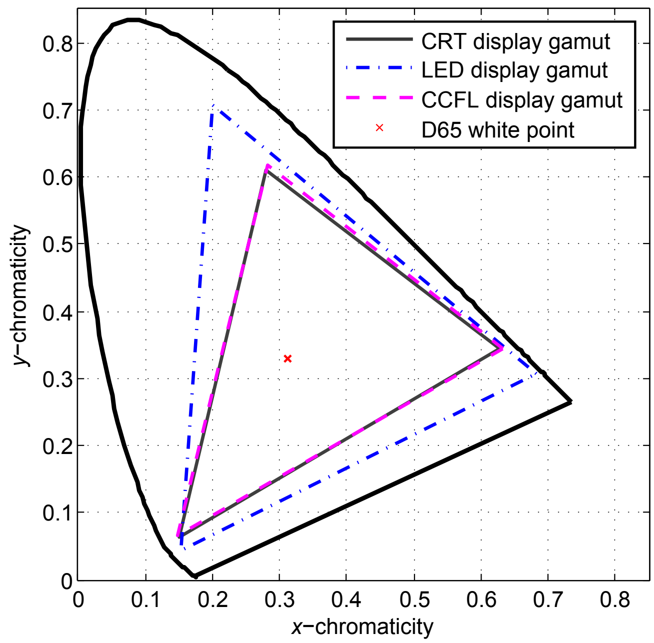
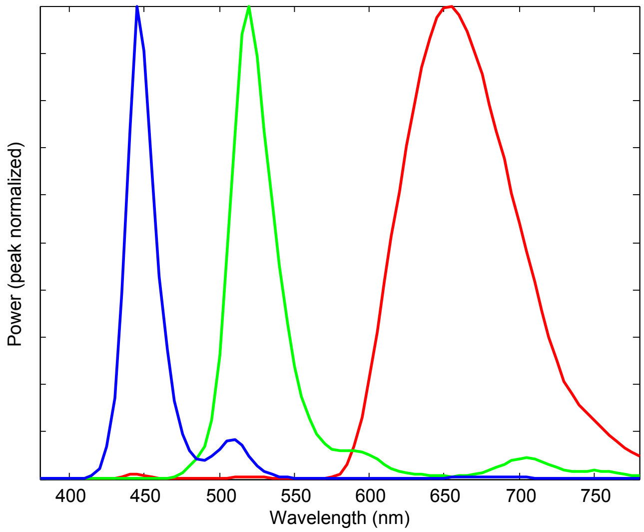


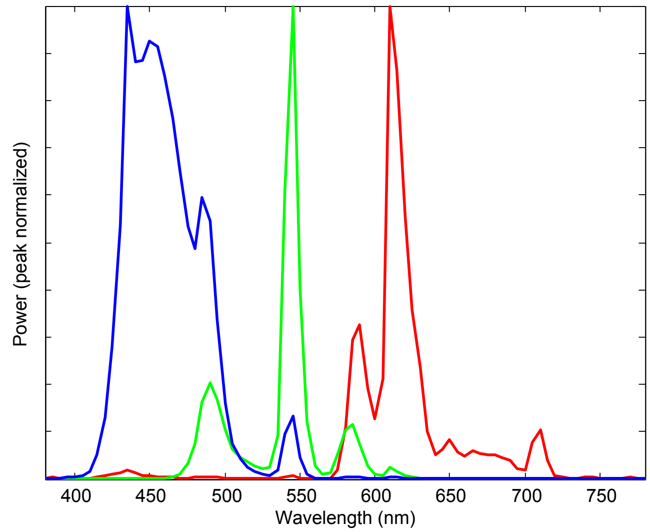
Fig. 13: Gamut boundaries computed from the SPDs of the reference CRT (Figure 1), LED (Figure 14a) and CCFL (Figure 14b) displays used for the simulated corresponding chromaticity calculation comparison.

Immediately observable is a more prominent variance in corresponding chromaticity estimates determined using an empirical assessment for both displays. It can be expected that an empirical assessment will be less precise than a simulation given the changing nature of observer color sensitivity as an observer transitions from photopic to scotopic vision, the discrete 8-bit limitations of the color adjustment tool and uncertainties in observable matches. Some observers reported being bothered by temporal flickering when viewing the reference CRT display stimulus and believed that this may have impaired their decision to accurately assess a visual match. Measurement error may also exist as a result of stimulus-rendered reflected light in the testing area and measurement device noise.

For a sample of 32 observers, the average elapsed time of the experiment was 14 minutes and 40 seconds, and the maximum elapsed time was 22 minutes and 50 seconds. Included in this experimentation time is 4 minutes and 12 seconds for which the PR-655 made five measurements of the SPD of the reference CRT display and five measurements of each of the four visual matches made by observers. Given the maximum measured time utilized by an observer to complete the experiment (rounded to 23 minutes), the drift in chromaticity varied with a standard deviation of $\pm$(0.0002, 0.0004) about the mean calibrated chromaticity in this interval. Using a single standard deviation in either direction, the corresponding change in xy-chromaticity amounts to a ΔE_{94} color difference of 0.4593 over the course of 23 minutes. While this may impact assessments of intra-observer variability during a single experiment iteration, the ΔE_{94} color difference is less than 1; errors as a result of chromaticity shifts due to drift alone are not compensated for.



(a) SPD of the simulated LED display derived from the average of 40 peak-normalized, individually measured, fully driven primary sets of NEC PA242W LED-backlit LCD displays, peak normalized.



(b) SPD of the simulated CCFL display derived from the average of 5 peak-normalized, individually measured, fully driven primary sets of Panasonic BT-LH1700WP field monitors, peak normalized.

Fig. 14: Computed SPDs of display technologies used to compare the xy-chromaticity offset simulation with the empirical assessment. Plot colors correspond to display primary color designation.

For the LED display, the simulated assessment and empirical assessment predict mean chromaticity offsets consistently with one another. While there is greater variability amongst real observers than simulated observers, the color difference between the mean simulated offset and mean observer offset varies by a negligible ΔE_{94} of 0.4029. For both assessments, this mean offset is similar in direction and magnitude from the original chromaticity aim of D65 white; the simulated assessment and empirical assessment are precise with one another. The consistency in the means of both assessment methods indicate that the simulation proposed in Section IV-A may be sufficient for determining a chromaticity offset for LED displays spectrally similar to those used in this assessment. Tabulated statistical results for the population of observers in the simulated assessment and empirical assessments for the LED-backlit display are listed in Tables V and VI, respectively.

TABLE V: Statistics for the simulated corresponding chromaticity offset, LED-backlit display.

	x-offset	y-offset
Mean	-0.0001	-0.0027
Min.	-0.0044	-0.0102
Max.	0.0035	0.0018
Std. dev.	0.0025	0.0035

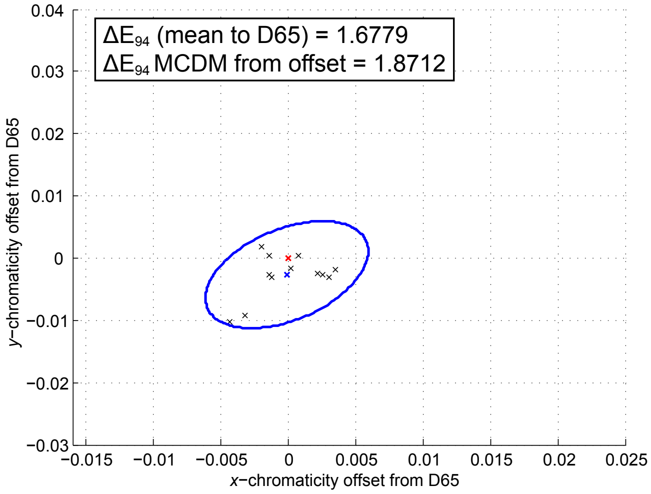
The simulated and empirical assessments of mean chromaticity offsets for the CCFL-backlit field monitor, on the other hand, bear only subtle resemblances with one another; their mean chromaticity shift is transposed towards a blueish hue with varying magnitudes. The variability amongst real observers of the CCFL-backlit display is more compact than the variability amongst real observers of the LED-backlit display. This may be a result of the wider bandwidth and increased overlap between the SPDs of the red, green and blue primaries of the CCFL-backlit display.

TABLE VI: Statistics for the empirically assessed corresponding chromaticity offset, LED-backlit display.

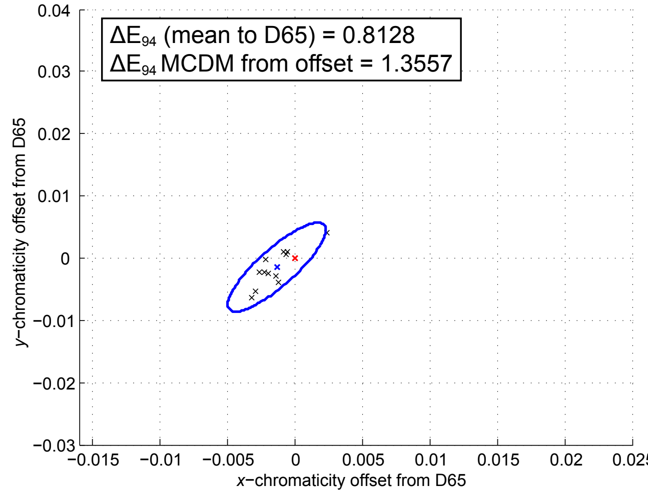
	x-offset	y-offset
Mean	0.0002	-0.0032
Min.	-0.0121	-0.0206
Max.	0.0238	0.0329
Std. dev.	0.0063	0.0086

The simulated assessment of the CCFL-backlit display predicts a mean chromaticity offset from D65 ΔE_{94} of 0.8128, which is less than 1. This indicates that chromaticity calibration for a CCFL-backlit display assessed by the 1931 2° standard observer may more likely render metameric matches for a number of observers. The empirical assessment, however, suggests otherwise. The means of the two assessments are colorimetrically separated by a ΔE_{94} of 3.8793. A simulated white point xy-chromaticity offset for a CCFL-backlit display may not satisfactorily render a metamerically matched white point for a diverse population of observers. Tabulated statistical results for the population of observers in the simulated assessment and empirical assessments for the CCFL-backlit display are listed in Tables VII and VIII, respectively.

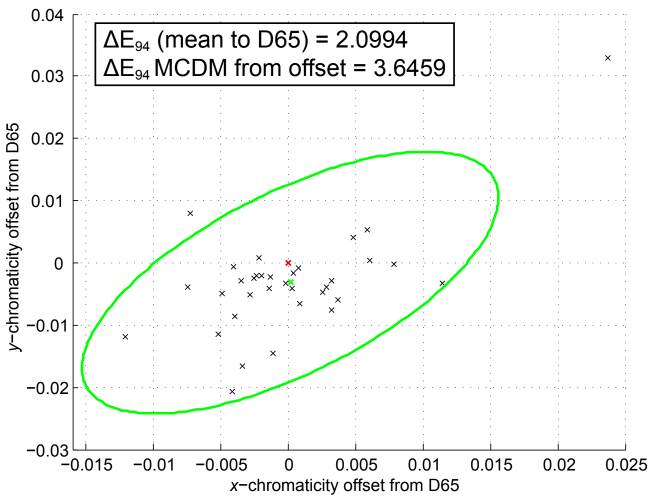
Importantly, the mean chromaticity offsets as determined by simulation or empirical assessment are within the 95% confidence ellipses of one another. This indicates that it is possible for some observers to be satisfied with either the simulated or empirically assessed mean corresponding chromaticity offset if it is to be used for calibrating the white



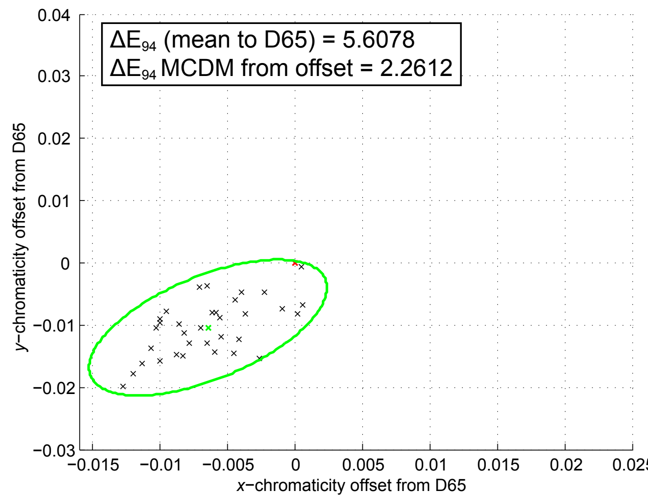
(a) Simulated corresponding chromaticity offset, LED display.



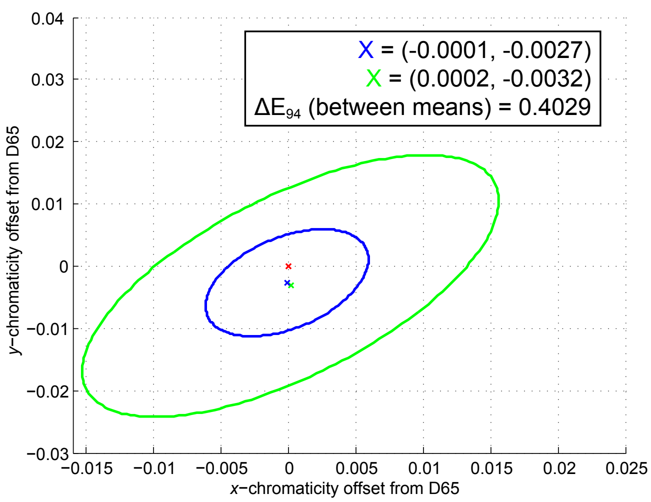
(b) Simulated corresponding chromaticity offset, CCFL display.



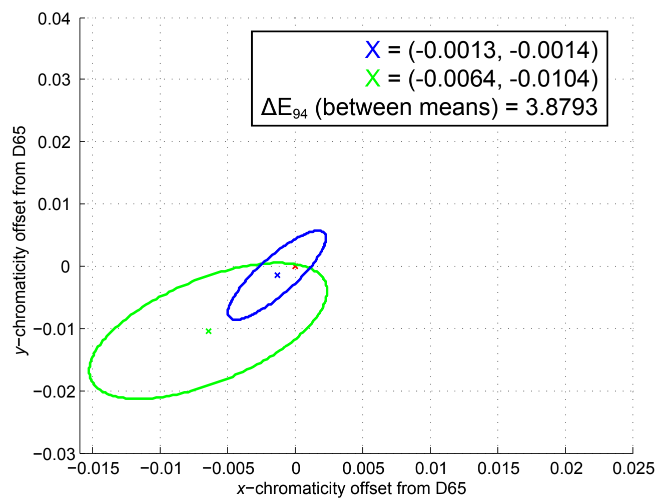
(c) Empirically assessed corresponding chromaticity offset, LED display.



(d) Empirically assessed corresponding chromaticity offset, CCFL display.



(e) Simulated and empirically assessed tolerance ellipses from corresponding chromaticity offsets from D65 white, LED display.



(f) Simulated and empirically assessed tolerance ellipses from corresponding chromaticity offsets from D65 white, CCFL display.

Fig. 15: Simulated and empirically assessed corresponding chromaticity offsets from D65 white with 95% tolerance ellipses. The origin, marked with a red ×, represents D65 white chromaticity. Each black × represents the corresponding chromaticity offset for a given observer model, and the blue or green × highlight the mean corresponding chromaticity offsets, simulated or empirically assessed, respectively, for a given display. 95% confidence ellipses encircle individual observer model offsets in either blue or green for simulated or empirical assessments, respectively.

TABLE VII: Statistics for the simulated corresponding chromaticity offset, CCFL-backlit display.

	x-offset	y-offset
Mean	-0.0013	-0.0014
Min.	-0.0032	-0.0064
Max.	0.0024	0.0042
Std. dev.	0.0015	0.0029

.

TABLE VIII: Statistics for the empirically assessed corresponding chromaticity offset, CCFL-backlit display.

	x-offset	y-offset
Mean.	-0.0064	-0.0104
Min.	-0.0127	-0.0198
Max.	0.0006	-0.0006
Std. dev.	0.0036	0.0045

.

point for both LED-backlit and CCFL-backlit displays. Since the CCFL-backlit display means are dissimilar, however, an observer may be equally unsatisfied with one or the other.

The empirical assessment of the NEC was completed by 17 observers who identified as male and 17 observers who identified as female. After removing 3 outliers: 2 males and 1 female from their respective smaller populations, the mean ΔE_{94} offset between the two classes of observers is 0.3875. The outliers were selected visually from corresponding chromaticity offset plots as significantly variant from their respective populations. This minimal color difference suggests that color perception amongst normal trichromatic observers does not change significantly as a function of gender.

VII. Conclusions

The results of both the simulation and empirical assessment reveal noticeable variances in observer color perception which is consistent with present research and publications investigating observer metamerism and new, more effective models to quantify and represent observer variability.

Given the results of this experiment and the results of the white paper published by Sony, it can be concluded that a simulated or empirically assessed offset can be used to modify traditional white point calibration procedures for displays with LED and OLED illuminants, while the methods proposed in this assessment may be ineffective for CCFL-backlit display technologies. The corresponding chromaticity offset proposed by Sony is not a mean corresponding chromaticity offset for their population of observers. The recommendation to use traditional colorimeters to calibrate Sony OLED displays to a white point chromaticity offset (0.3067, 0.3180) is based on a visual match assessed by the Judd-modified color matching function. While this chromaticity offset is not the mean chromaticity offset for the population of Sony observers, the Judd-modified color matching function predicts a corresponding chromaticity closer to the mean than (0.3127, 0.3290). The tolerance ellipse in Figure 8c encompasses the proposed (-0.006, -0.011) chromaticity offset.

The use of an xy-chromaticity as a standard has been practiced repeatedly for industry display calibration. It is important to acknowledge the origins of this color appearance which were derived from the daylight spectrum of Standard Illuminant D65 rendered on a CRT display and measured with the CIE 1931 2° standard observer. Industry professionals have become accustomed to relying on a simplistic representation of colorimetry which resulted in calibration issues when multiple observers viewed colorimetrically matched and calibrated stimuli on a variety of displays.

It is important to reiterate that, given the nature of narrow-band three-primary displays, assessed xy-chromaticity offset will not render exact color matches between observers. A single observer model is not "wrong" in its prediction of chromaticity, rather, it may be classified as satisfactory or unsatisfactory for any one real observer amongst a population of diverse observers. In other words, a single observer model is not necessarily sufficient for describing the color perception attributes of a diverse population of observers.

Given the need for a calibration standard for narrow-band three-primary television displays commercially available on the market today and the cost of a spectroradiometer, it is recommended for a manufacturer to use a spectroradiometer to perform a single measurement of the fully-driven SPDs of each of the primaries of their LED or OLED displays when new primary illuminant technologies are used. The manufacturer should then run the simulation with the observer models listed in Section IV-A and compute a corresponding chromaticity offset with respect to the CIE 1931 2° standard observer to report to users of their displays.

For clarification, the simulation does not determine a best color matching function to view a display, rather, it simply provides the appropriate xy-chromaticity offset to be used by post production facilities utilizing colorimeters that report chromaticity with respect to the CIE 1931 2° standard observer.

Given the consistency in the results for LED and OLED displays, it is unnecessary and time consuming to repeat an empirical assessment for all commercially available displays with these illuminants.

Observer variability is still present in the mean corresponding chromaticity offset predictions. ΔE_{94} mean color difference from the mean offset prediction for the LED-backlit and CCFL-backlit displays (Figures 15a and 15b) in simulation vary colorimetrically with magnitudes of 1.8712 and 1.3557, respectively. Since these MCDM values are slightly greater than 1, simulated corresponding chromaticity offsets suggest observer variabilities that may result in metameric failures for a population of observer models. Extending to empirical assessments, these MCDM values are larger in magnitude with ΔE_{94} values of 3.6459 and 2.2612 for chromaticity offset predictions for the LED-backlit (Figure 15c) and CCFL-backlit (Figure 15d) displays, respectively. As a function of variations in human perception, real observer variability is subject to greater differences in color perception than the simulation suggests.

Importantly, however, the original xy-chromaticity coordinates of D65 white with respect to the CIE 1931 2° standard observer are always within the empirically assessed and simulation calculated 95% tolerance ellipses; the chromaticity coordinates of D65 white are also considered to be a statistically relevant part of the population.

Commercially available displays today are manufactured with larger screen areas than their predecessors, and this trend in increased screen size is growing. Further iterations of this assessment could involve an empirical evaluation where observers determine color matches utilizing stimuli which each subtend a 10° field of view to be compared to solely 10° observer models in simulation. Greater variability amongst observer visual matches with a wider field of view could be expected given increased diversity in observer perception present in peripheral vision, particularly with respect to short wavelength visible energy.

Following present trends in the evolution of display illuminant technologies, it can be anticipated that a common goal for commercially available displays is to render the greatest gamut of color possible. One way to achieve such a gamut would involve the use additional display primaries. Recent work by Long and Fairchild suggest the use of optimized ideal multispectral primaries for an increased likelihood of metameric image reproduction [22]. This would enable displays to render more spectrally relevant colors rendering satisfactory color reproduction.

Acknowledgment

The author would like to personally acknowledge his undergraduate and senior project advisor, David Long, for continued support and flexibility; critical, thought-provoking discussions; and encouragement throughout the his entire undergraduate career and for the entire duration of his project.

Additionally, he would like to thank Bill Miller, along with the SMPTE 10E standards committee for the insightful opportunity to be the first student fully involved and immersed with a SMPTE standards committee. He extends further thanks to Bill for proofreading the final draft of this report.

Continuing, the author would like to thank Alex Forsythe, Scott Dyer and Joe Di Gennaro of the Academy of Motion Picture Arts and Sciences for their support, discussion and initial execution of this experiment. He would also like to thank Bill Landers of the Rochester Institute of Technology (RIT) School of Film and Animation (SoFA) for assistance with additional equipment loans through the SoFA cage to complete the empirical assessment.

Lastly, he would like to thank the 51 observers (52 total, including himself) who volunteered their time to participate in either or both iterations of the empirical assessment portion of this project. He hopes that all those who participated enjoyed the baked goods that followed.

References

[1] Wolfe, Jeremy, Keith Kluender, and Dennis Levi. "The Perception of Color." In *Sensation & Perception*, 108-109. 2nd ed. Sunderland: Sinauer Associates, 2009.

[2] "NTSC Signal Specifications." *Proceedings of the IRE* (1954): 17-19.

[3] "ISO 10526:1999/CIE S005-1998." International Commission on Illumination, 1998. Web. http://www.cie.co.at/publ/abst/s005.html.

[4] "SMPTE C Color Monitor Colorimetry." The Society of Motion Picture and Television Engineers RP 145-2004 (2004).

[5] "Recommendation ITU-R BT.601-7." *International Telecommunication Union*, 2011.

[6] "Recommendation ITU-R BT.709-5." *International Telecommunication Union*, 2002.

[7] Smith, T., & Guild, J. (1932). The C.I.E. Colorimetric Standards and Their Use. *Transactions of the Optical Society*, 73-134.

[8] Westland, S., Ripamonti, C., & Cheung, V. (2013). *Computational Colour Science using MATLAB* (2nd ed.). Wiley Series in Imaging Science and Technology.

[9] Stiles, W.S., and J.M. Burch. "N.P.L. Colour-matching Investigation: Final Report." *Optica Acta: International Journal of Optics* 6, no. 1 (1959): 1-26.

[10] Wyszecki, Gunther, and W. S. Stiles. "Visual Equivalence and Visual Matching." In *Color Science Concepts and Methods, Quantitative Data and Formulae*, 330-333. 2nd ed. New York: John Wiley & Sons, 2000.

[11] Fairchild, Mark. "Human Color Vision," "Colorimetry." *Color Appearance Models*. 3rd ed. Hoboken: John Wiley & Sons, 2013. 19-21, 79-80.

[12] Bruce Lindbloom, *Delta E (CIE 1994)*, Accessed 2 November 2014. Web. http://www.brucelindbloom.com/Eqn_DeltaE_CIE94.html

[13] Alfvin, Richard, and Mark Fairchild. "Observer Variability in Metameric Color Matches Using Color Reproduction Media." *Color Research & Application* 22.3 (1997): 174-88.

[14] "Fundamental Chromaticity Diagram With Physiological Axes - Part 1." *International Commission on Illumination Technical Report*, 2005.

[15] Sarkar, A., Blond, L., Le Callet, P., Autrusseau, F., Morvan, P., & Stauder, J. (2010). Toward Reducing Observer Metamerism in Industrial Applications: Colorimetric Observer Categories and Observer Classification. *Color Imaging Conference Final Program and Proceedings, 18*, 307-313.

[16] Fedutina, M., Sarkar, A., Urban, P., & Morvan, P. (2011). (How) Do Observer Categories Based on Color Matching Functions Affect the Perception of Small Color Differences? *Color Imaging Conference Final Program and Proceedings, 19*.

[17] Fairchild, Mark, and Rodney Heckaman. "Metameric Observers: A Monte Carlo Approach."

[18] Sony Corporation (2013 Feb 15). Colour Matching Between OLED and CRT White Paper. V1.0.

[19] Long, David, and Mark Fairchild. "Modeling Observer Variability And Metamerism Failure in Electronic Color Displays." *Journal of Imaging Science and Technology* 58.3 (2014).

[20] "Colour Matching Functions." Web. 15 July 2014. http://cvrl.ioo.ucl.ac.uk/cmfs.htm.

[21] "Psychtoolbox-3." Web. 17 Sept. 2014. http://psychtoolbox.org/.

[22] Long, David, and Mark Fairchild. "Optimizing Spectral Color Reproduction in Multiprimary Digital Projection."

Matthew Donato is working towards the completion of a B.S. degree in Motion Picture Science at the Rochester Institute of Technology and is a recipient of the SMPTE 2014 The Louis F. Wolf Jr. Memorial Scholarship. In the summer of 2014, Matthew held an imaging science internship at the Academy of Motion Picture Arts and Sciences. Prior to this experience, he had the opportunity to work as a news editor for NBCUniversal.