Apparatus and method for measuring individual color-matching functions

LUVIN MUNISH RAGOO,[1,2] **ID IVAR FARUP,**[1,3] **ID**
AND JAN HENRIK WOLD[1,4] **ID**

[1]*Norwegian University of Science and Technology, Norway*
[2]*luvin.m.ragoo@ntnu.no*
[3]*ivar.farup@ntnu.no*
[4]*jan.h.wold@ntnu.no*

Abstract: This article presents a compact visual colorimeter for the purpose of measuring color-matching functions (CMFs) of individual observers through psychophysical experiments. Constructed from 3D-printed parts, optical elements, and LED-based light engines, the colorimeter facilitates the juxtaposition of two fields to create a bipartite field. The system underwent characterization to evaluate factors that may impact color-matching experiments, such as LED-light stability, spatial homogeneity of the bipartite field, and potential stray-light leakage. The study aimed to assess the accuracy and performance of the system in measuring individual observer CMFs. Results indicate that the system is stable enough to measure both intra- and inter-observer variations in CMFs.

1. Introduction

In color science, color-matching functions (CMFs) constitute the experimental basis for specifying color through measurement. CMFs are three functions that, for a sequence of monochromatic test-color stimuli originating from spectral lights evenly distributed across the visible spectrum, determine the amounts of each of three specified reference color stimuli yielding respective color matches in a color-matching experiment [1]. Precise determination of CMFs is vital for numerous applications, such as colorimetry, color reproduction, color management, and the perception of material appearance. The Commission Internationale de l'Éclairage (CIE) has undertaken the task of collecting and standardizing CMFs, pooling observers' color-matching data from several authors to derive a set of standardized functions approximating the mean color-matching response of the human population with normal color vision [2].

Despite these efforts, several studies have revealed notable disparities between predicted color matches based on standard CMFs and actual experimental matches made by color-normal observers [3–6]. These discrepancies underscore the need for characterizing individual differences in observers' CMFs. Several authors have developed apparatus and methods to classify observers into categories [3,6–10]. However, most of these methods rely heavily on estimations, and direct measurements of individual observers' CMFs is often avoided, as it is a laborious endeavor that requires special hardware [11,12].

In this article, we extend upon previous work [13] and present an improved visual colorimeter design and method for the purpose of measuring individual observers' CMFs with a good degree of confidence.

2. Background

2.1. Standard observer color-matching functions

The CIE 1931 2° standard colorimetric observer, derived from the works of Wright [14] and Guild [15] and the data of the CIE 1924 photopic luminous efficiency function [16,17], $V(\lambda)$,

stands as the oldest and likely most utilized CMF set. However, criticisms arise due to the $V(\lambda)$ function's insensitivity at shorter wavelengths, contributing to predictive errors in the CIE 1931 CMFs, especially in the blue/violet end of the visible spectrum [5,18,19]. Replacing the incorporated $V(\lambda)$ by a corrected luminous efficiency function, Judd later proposed an improved set of 2° CMFs, which were further refined by Vos, resulting in the Judd–Vos modified 2° CMFs [20]. However, these CMFs were never standardized by the CIE.

In 1955, Stiles and Burch undertook a color-matching investigation involving 10 observers for the visual angle of 2°, resulting in the publication of the Stiles and Burch 1955 2° CMFs [21]. Unlike the Wright–Guild functions and the Judd–Vos modification, the Stiles and Burch CMFs are directly measured, thus avoiding the reconstruction errors present in the former two. Subsequently, in 1959, Stiles and Burch published a comprehensive set of 10° CMFs for 49 observers, which remains the most meticulously measured CMF dataset to date [22]. This dataset, together with color-matching data from Speranskaya [23], later served as the foundation for the development of the CIE 1964 10° CMFs [24,25]. Furthermore, the Stockman and Sharpe cone fundamentals [26], later adopted as characteristic of the CIE 2006 physiological observer (CIEPO06) [27], were also based on the seminal Stiles and Burch 1959 dataset.

2.2. Observer variability

For practical purposes, standardized CMFs aim to provide a mean specification of color for the general population [28]. However, they perform poorly in predicting color matches for individual observers. As a result, two fields of light that are calculated to be in color match for a standard observer, such as the CIE 1931 2° standard colorimetric observer, may appear as fields of distinguishable colors to an individual [29]. Such predictive failures arise from the fact that there are significant perceptual differences between individuals, which are not accounted for by the standard observer models.

Individual differences in vision can be attributed to many physiological factors. Lens optical density, for example, varies significantly with age, accounting for about 47 % of its variations [30]. According to Pokorny *et al.* [31], the optical density of the lens grows by 38 % between the ages of 20 years and 60 years. Furthermore, diseases like diabetes and cataracts have also been shown to have an impact on the lens [32]. Macular pigment optical density is another important factor that influences visual variability, particularly in the 400 nm–525 nm region of the visible spectrum [6]. Macular pigment concentration decreases beyond the central region of the retina, resulting in different spectral sensitivity measures between the foveal and peripheral retinal areas [33,34].

Furthermore, the photopigment chemical compositions also vary among the population, affecting cone absorption properties [35]. Genotypic variations in the L- and M-cone photopigments can cause significant shifts in their spectral sensitivities. For example, the different genetic variants of the L-cone photopigment, such as L(S180) and L(A180), differ significantly in the wavelength of the peak spectral absorbance, by about 2 nm–7 nm [36]. Aside from variations in the photopigments, their concentrations as well as the respective lengths of the cone outer segments in which they are contained may also vary among individuals, leading to differences in cone photopigment optical densities (cf. Beer's law [37]).

Additionally, each individual's unique retinal mosaic, characterized by a distinct arrangement of cone types, adds another layer of complexity to observer variability in cone sensitivities [38].

An understanding of these sources is thus critical for interpreting color vision data correctly.

2.3. Factors affecting color-matching experiments

Central to the color-matching experiments are Grassmann's laws [39], which provide essential principles governing color matching. Including the properties of symmetry, transitivity, proportionality and additivity, these laws outline the linear relationship between color matches, which

in turn enable the calculation of heterochromatic matches from the spectral power distributions (SPDs) of the relevant color stimuli and the CMFs particular to the observer [40,41]. (Here, and in the following, the SPD of a light or corresponding color stimulus is understood to be the spectral radiance of the field from which the light emanates.)

Perhaps the most critical factor affecting color-matching experiments is the participation of rods (rod intrusion), especially under low light conditions. Trezona discussed extensively on the effect of rod participation on color matching [42]. When rods are active, their responses can interact with cone signals, leading to complex perceptual effects that may not adhere to simple additive mixing [43,44]. This may explain the failure of additivity in color-matching experiments reported by several authors [45–48]. However, Oleari *et al.* performed foveal color-matching experiments and confirmed that Grassmann's laws hold to a good approximation for color fields of visual angles less than 2° [49]. In the corresponding retinal region, the rod density is at a minimum [50] and rod participation thus effectively mitigated. Other methods to reduce rods' involvement are to use rod-saturating techniques, such as a bright surround [18,51].

Given that rod density, macular pigment concentration, and optical density of the cone photopigments are all linked with retinal topology, it becomes obvious that field size is a critical factor to consider when designing color-matching experiments. Li *et al.* conducted a series of achromatic matching experiments in bipartite fields with visual angles of 2°, 4°, 6°, 8°, and 10°, using spectrally narrow-band reference lights [52]. The objective was to evaluate the predictive accuracy of the CMFs in the CIE 1931 2° and CIE 1964 10° standard colorimetric systems and of the field size-dependent cone fundamentals derived from the CIEPO06 model. They found that the CIEPO06 2° cone fundamentals performed best in predicting color matches in a 2° field. They also observed the largest inter-observer variability in 2° matches.

The experimental method used may also impact color-matching experiments. There are generally two types of color-matching methods that have been employed in various color-matching studies: Maxwell's method and the "maximum saturation" method. Maxwell's method involves matching a mixture of a monochromatic test light and any two of three reference lights to a fixed white light. In the "maximum saturation" method, aptly named by Crawford [53] and Lozano and Palmer [54], a mixture of two reference lights is matched to a monochromatic test light that, in certain spectral regions, must be slightly desaturated by a third reference light for a color match to be obtained. If the laws of additivity and proportionality were to strictly hold in trichromatic color matching, both of these methods should yield the same CMFs [41]. In his experiments, Thornton found that Maxwell-type color matches deviated more from the chromaticities calculated using the CIE 1964 standard observer CMFs than "maximum saturation"-type matches did [7,8]. However, Maxwell matching has the advantage that matches are performed in a field of constant luminance and chromaticity, whereby possible non-linearities introduced by changes in luminance and chromaticity can be avoided. The suggestion is that Maxwellian color matching tend to yield better predictions for matches of colors close to the reference white, whereas for matches involving highly saturated colors the "maximum saturation" method might be more suitable [41].

Another factor affecting color-matching experiments is the choice of reference lights. Thornton identified specific "prime-color" wavelengths of monochromatic red–green–blue (RGB) reference lights that are deemed most effective in additive mixtures of light [7], while Li *et al.* found that sets of near-spectral reference lights of peak wavelengths 636 nm, 521 nm, and 447 nm result in the most stable matching performance [28].

Finally, Boynton *et al.* discovered that introducing a small gap between chromatic fields enhances the ability to discriminate between them, suggesting that spatial separation between the two halves of the bipartite field can improve visual discrimination by reducing interference and enhancing the distinctiveness of the color [55].

2.4. Previous work

The Stiles and Burch datasets (1955 and 1959) are extensive enough to give us an appreciation of how the general population matches color. However, the individual observer CMFs therein have no precise measures of individual matching uncertainty. Therefore, it is difficult to attribute the differences in individual observer CMFs to actual inter-observer variability or observer matching uncertainty.

Efforts to characterize inter-observer variability have been conducted by Asano *et al.* [3,4]. Their approach focuses on maximizing inter-observer variability to highlight the differences among observers. Their color-matching experiments, though simplified – limited to five matches per observer, repeated three times – yielded results reported as mean color difference from the mean (MCDM), which effectively measures overall differences between observers. Although MCDM is useful for quantifying the extent of variability, it does not reveal how individual observers' CMFs differ, nor does it indicate whether region-specific differences in these CMFs are statistically significant.

Viénot's study [56] is closely related to our research on individual CMFs. In her work, "Relations between inter- and intra-individual variability of color-matching functions," Viénot examined CMFs for ten observers, highlighting significant dispersion in the short-wavelength region due to challenges in color matching caused by the Maxwell's spot and luminance variations. She found strong intra-individual correlations across all wavelengths, indicating consistency within observers, while minimal inter-individual correlations in the green/yellow part of the visible spectrum pointed to distinct individual differences.

Recently, Shi et al. [57,58] introduced a new LED-based multi-primary trichromator, which they used to conduct a series of color-matching experiments aimed at estimating individual differences in color matches and cone spectral sensitivities among 51 young adults. In these experiments, observers adjusted various light triplets in a Maxwell-type matching procedure to a fixed reference white, allowing the researchers to infer individual cone sensitivities. Rather than directly measuring CMFs, these studies estimated cone spectral sensitivities by analysing the matches made by the observers. The results indicated small but consistent deviations from the CIE 2006 recommendations, highlighting the variability in individual cone sensitivities. It is important to note that these works infer cone fundamentals from the matching data rather than directly measuring the CMFs themselves. Ultimately, the true test of an observer's inferred CMFs lies in their ability to accurately predict color matches for that individual.

An earlier version of the visual colorimeter was presented by Ragoo and Farup [13]. This initial iteration exhibited several critical issues, including insufficient light intensity in the 400 nm –440 nm region, poor spatial homogeneity of the bipartite field, and light leakage between hemispheres. Additionally, the light emanating from the bipartite field diverged outward, resulting in significant stray light that distracted the observers during the experiment while also reducing the amount of light reaching the observers' eyes.

3. Color-matching apparatus

In this section, a thorough description of the color-matching apparatus is provided. The purpose of this detailed outline is to facilitate potential re-investigations and continuations of the experiments presented in this pilot study.

3.1. Design

The visual colorimeter is constructed from several 3D-printed parts, optical elements, and LED light sources. A schematic of the colorimeter is shown in Fig. 1. The 3D-printed parts include the integrating chambers, the lens-mounting sockets, a mirror-mounting socket, and a central enclosure that all the other parts are assembled into. The integrating chambers are constructed

from two mating parts that, when assembled, create fully spherical integrating spheres within. This allowed for the internal surface of integrating spheres to be painted with several layers of a white Barium Sulphate coating to create a diffuse and near-Lambertian surface. The internal integrating spheres have a diameter of 36 mm and an exit aperture of 8 mm. Two types of integrating spheres have been designed: one with a single input hole for the RGB LED light engine on the reference side (i.e., the side where the reference lights that are not used for desaturation of the test light, are mixed by intensity adjustments), and another with two input holes, one for the light yielding the test-color stimulus and another for the RGB LED light engine generating the desaturation component on the test side. On the outside of the exit aperture of each integrating chamber, an optical iris diaphragm is mounted. This makes the exit aperture of the integrating chambers adjustable between 1 mm and 8 mm.
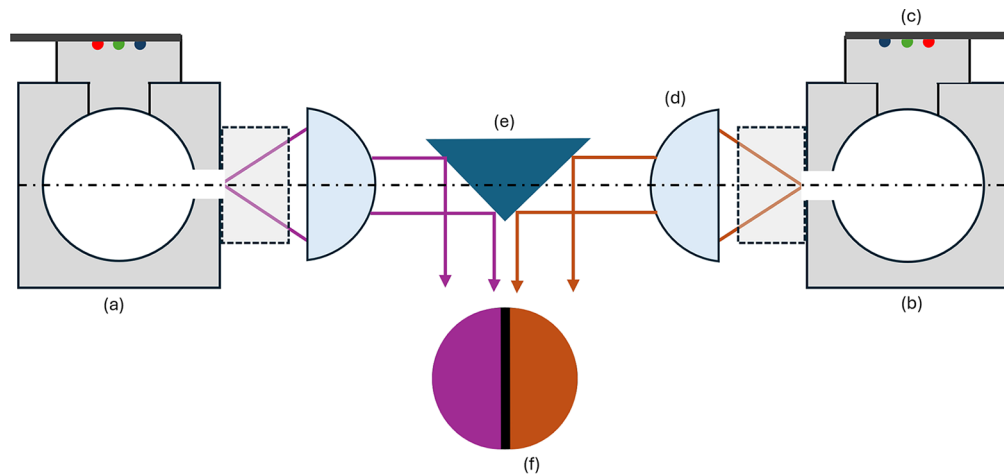

Fig. 1. Schematic diagram of the visual colorimeter. (a) Integrating chamber for the test field, which has two inputs, the near-spectral test lights (not shown on the diagram) and the RGB LED light engine. (b) The integrating chamber for the reference field, which has only one input, for the RGB LED light engine. (c) RGB LED light engine. (d) Aspheric condenser lens. (e) Knife-edge right-angle prism mirror. (f) Resulting bipartite field with adjustable gap width.

Lens enclosures were also 3D-printed to securely house condensing lenses of 45 mm diameter with a focal length of 32 mm, shown in Fig. 2. The enclosures were designed in such a way that the lens within are centered around the exit aperture of the integrating chambers. The latter has threaded sockets, which allow the lens enclosures to be screwed on securely. The lenses serve to make the light rays emanating from the integrating chambers collimated. The lens enclosures are meant to fit firmly into circular sockets on each side of a central enclosure, within which a knife-edge right-angle mirror is placed. The collimated light from the lens is incident on the right-angle mirror at a 45° angle and is reflected at 90° toward the observer. When viewed from the front, the central enclosure has a circular exit hole of 2 cm diameter through which the near-parallel light from each side of the mirror escapes toward the observer, effectively creating a bipartite field as seen in Fig. 1. The knife-edge right-angle mirror, being placed on a sliding socket (visible in Fig. 3), can move horizontally normal to the central axis of the lenses. This mechanism makes the position at which the mirror intercepts with the parallel beam of light from the lenses, adjustable. Thus, the width of the central separation between the left and right fields also becomes adjustable, whereby the mirror's position can be adjusted to create two almost perfectly juxtaposed fields, suitable for a minimally-distinct-border type of experiment [59], or to create a small dark gap between the two fields. In this study, the separation between the two

fields was set to approximately 5 arcmin. The motivation for having a gap originates from studies indicating that, for color stimuli with dominant wavelengths falling within certain wavelength intervals, a small separation between fields can enhance color discrimination [55,60].
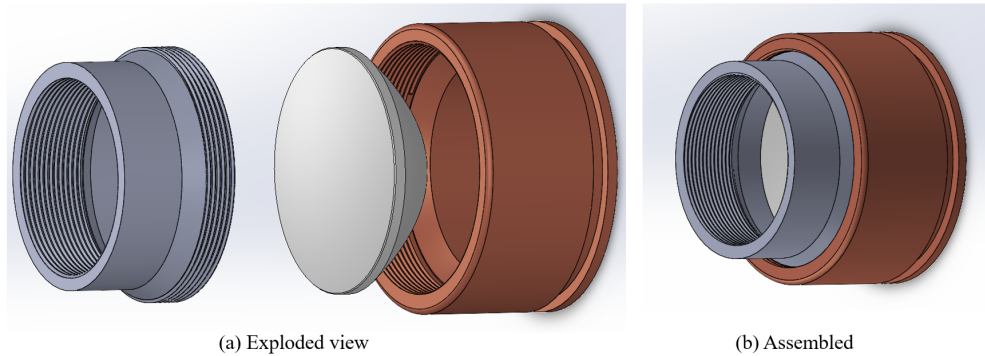

(a) Exploded view   (b) Assembled

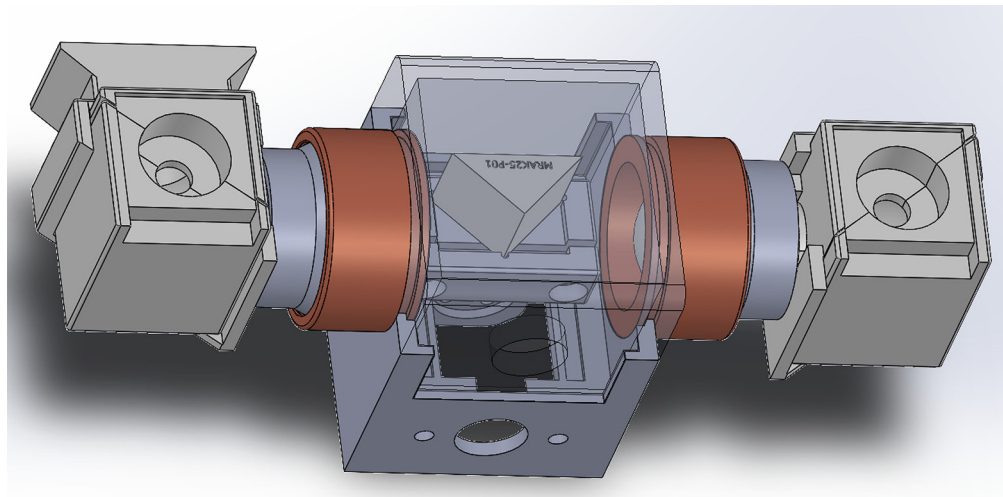**Fig. 2.** 3D drawings showing the lens within its mounting socket.

Fig. 3. 3D drawing of assembled visual colorimeter. The mirror enclosure body is made transparent in this drawing to show the right-angle mirror within.

The 3D drawing of the assembled visual colorimeter is shown in Fig. 3. It is worth noting that in the figure, the 3D-printed parts are shown in different colors to highlight individual parts. In reality, however, they were all printed in the same material and were all black in color.

In the experimental setup, the test lights were near-spectral lights, with peak wavelengths ranging from 400 nm to 720 nm at 10 nm intervals, generated from a monochromator (Bentham TMc300). A Xenon lamp (Bentham Il7 Xenon source) is used as input to the monochromator. The output slit of the monochromator was adjusted such that the lights had a full width at half maximum (FWHM) of about 9 nm. To control for the viewing field, the bipartite field, which had a diameter of 2 cm, was placed at a distance of approximately 80 cm from the observer, corresponding to a visual angle of roughly 1.4°.

The RGB LED light engine is a metal-core printed circuit board onto which four LEDs are soldered in close proximity: two red, one green, and one blue. The peak wavelengths for

the red, green, and blue LEDs are 634 nm, 534 nm, and 452 nm, respectively. These LEDs yield the reference lights for the color-matching experiments, and they were selected based on commercially available high-power LEDs for which the peak wavelengths of the emitted lights are near those of the Stiles and Burch 1955 RGB reference color stimuli.

Dimming strategies when operating LEDs have been shown to have an impact on spectral shifts experienced by such lighting devices. Notably, amplitude modulation tends to produce significantly larger spectral shift than pulse width modulation (PWM) dimming [61]. Thus, PWM was chosen as the dimming method for the LEDs in the light engines. Since the LED lights had much higher luminous flux than the spectral lights from the monochromator, their forward current was limited to only 35 mA, which is quite low compared to their typical rated current of 700 mA. This resulted in their light output being within a range that was perceptually comparable to that of the spectral lights. Additionally, running the LEDs at much lower currents had the added benefit of keeping them cool during their operation, thus limiting potential spectral shifts due to temperature. The light engines were powered by a TLC59711 PWM LED Driver from Adafruit, which can control 12 channels of 16-bit PWM output independently. The PWM frequency is around 150 Hz, which is well above the fusion frequency of the average human observer [62]. Only six of the LED driver channels were used, three for the reference-field LED light engine and three for the engine providing the LED lights used for desaturating the test field. The driver boards were controlled using an Arduino Mega microcontroller. A look-up table was used to map the 65 536 linear levels of the LED driver to 1024 gamma-corrected levels with a correction value $\gamma = 2$. This was done to ensure that the adjustments of the brightness of the LED light-emitting fields were smooth and perceptually uniform.

The control interface has three rotary encoders, each of which controls one of the channels in the RGB LED light engine. Rotating the encoder clockwise increases the intensity of its corresponding channel on the reference side. If a light channel is active on the reference side, rotating its corresponding encoder counter-clockwise will reduce the light intensity; otherwise, it increases the intensity of the desaturating light on the test side. This means that the same light cannot be active on the reference side and the test side simultaneously. Additionally, a reset button and a toggle switching between coarse and fine adjustments are integrated into the system.

3.2. System characterization

3.2.1. Test-color stimuli

The near-spectral lights (hereafter referred to as test lights) yielding the test-color stimuli were selected at regular intervals of 10 nm from 400 nm to 720 nm. Their measured SPDs are plotted in Fig. 4. In later sections, these SPDs, normalized to unit power, are denoted as $\mathbf{S}_{\lambda_i}$, where λ_i is the peak wavelength of the respective test light.

The luminances of the test lights are shown in Fig. 5. Some of the test lights are probably below the photopic threshold, although it is worth noting that the latter varies between individuals and is loosely defined in the literature. In this study, we consider the photopic threshold to be defined as the luminance level where rod saturation begins according to Stockman *et al.* [63], at roughly 3 cd/m^2. The horizontal red dashed line in Fig. 5 represents that photopic threshold.

3.2.2. RGB reference color stimuli

It is well documented that LED devices are subject to spectral shifts both during operation and over lifetime [64,65]. If, in a color-matching experiment, the reference lights are LED lights (hereafter referred to as RGB LED lights), it is crucial to characterize these spectral shifts. If the shifts are much smaller than the average observer matching uncertainty, their impact on the resulting CMF measurements can be considered to be insignificant.

To characterize spectral shifts of the RGB LED lights during their operation, the SPD of each light was measured from 5 % to 100 % intensity at every 5 % intensity interval. The SPDs are
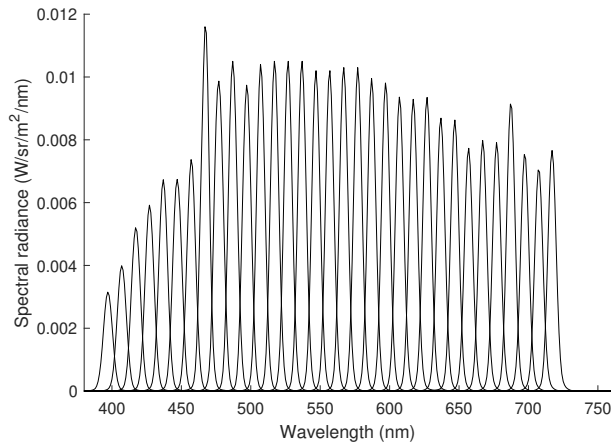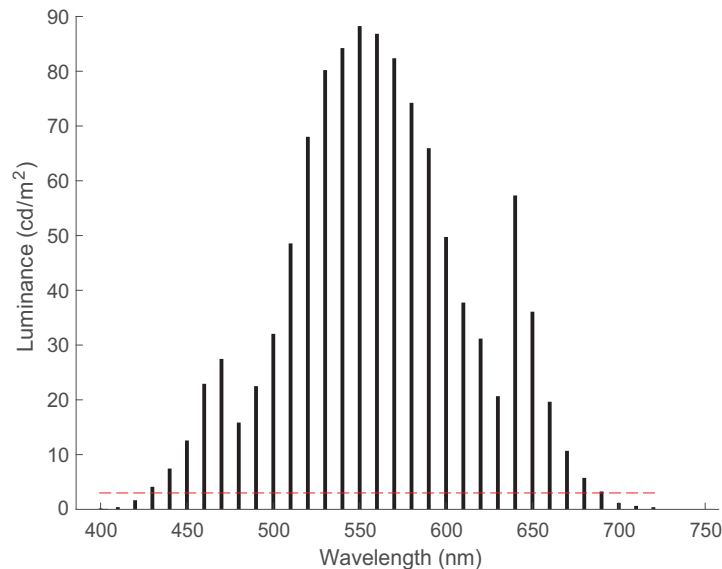
Fig. 4. SPDs of the test lights.

Fig. 5. Luminances of the test lights by their peak wavelengths (the red dashed line represents the photopic threshold of 3 cd/m^2.)

shown in Fig. 6. To make any shifts in the spectral shape of the distribution more apparent, these SPDs were then normalized to yield corresponding relative spectral radiances with peak values all equal to 1, as shown in Fig. 7. As apparent from the graphs, the light from the green LED is subject to larger spectral shifts than the other two.

For each LED-type (red, green, and blue), the u′v′-chromaticity coordinates of the color stimulus yielded by the LED light at each intensity level were computed from the measured SPDs shown in Fig. 6. The mean $\Delta_{u'v'}$ was then calculated between the chromaticity computed from the 5 %-intensity SPD, which was used as reference, and the chromaticities computed from the SPDs at the other intensity levels, as shown in Table 1.

To mitigate the impact of LED spectral shifts, the different relative SPDs in Fig. 7 were averaged for each of the three LED light channels. As a result, a single mean relative SPD for each LED light channel was obtained, which represents the mean shape of the distribution
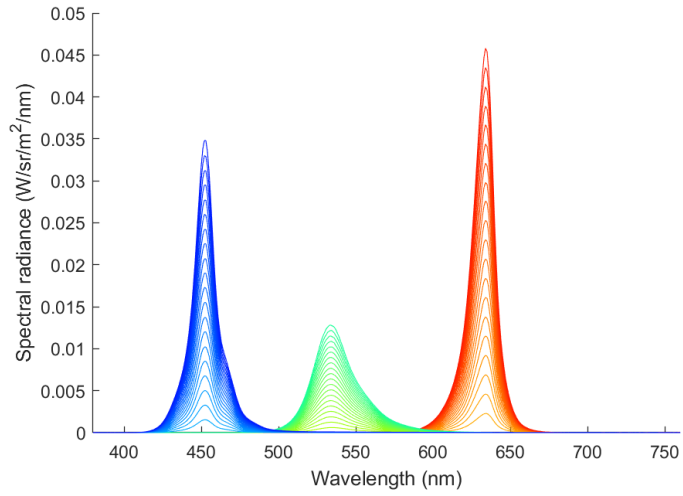
Fig. 6. SPDs of the RGB LED lights at every 5 % intensity interval.
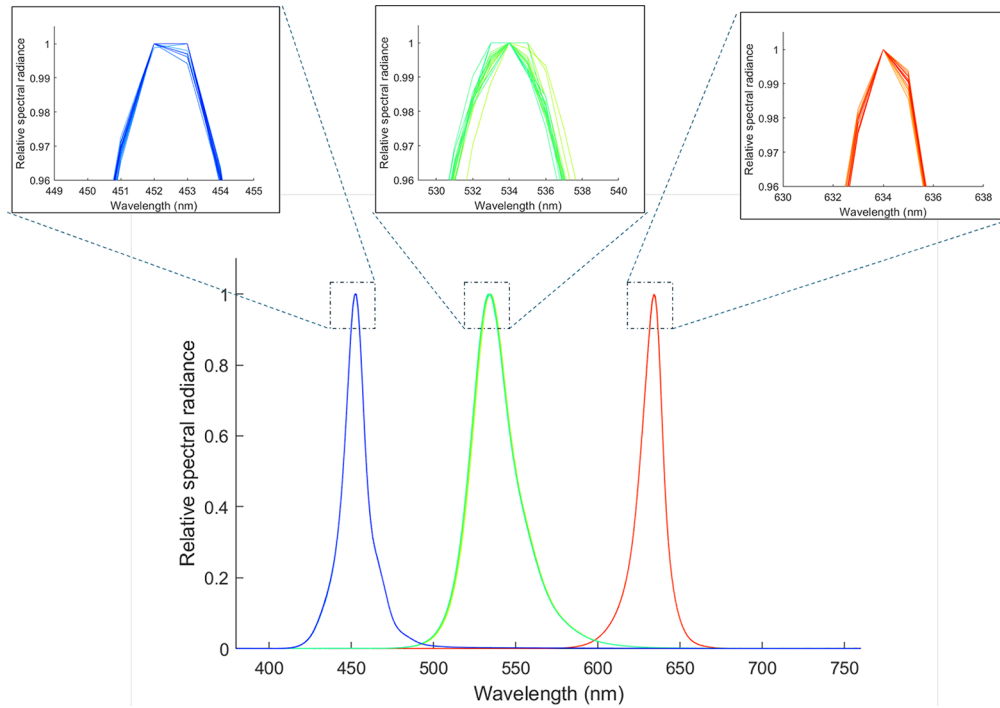
Fig. 7. Relative SPDs of the RGB LED lights (normalization to peak value of 1).

Table 1. Mean $\Delta_{u'v'}$ for the color stimuli yielded by the RGB LED lights, between the chromaticity computed from the measured SPD of the LED light at 5 % intensity and the chromaticities computed from the SPDs (of the lights from the same LED) as measured at the other intensity levels (i.e., from 10 % to 100 % in steps of 5 %)

LED	Red	Green	Blue
Mean $\Delta_{u'v'}$	1.5×10^{-4}	1.9×10^{-3}	5.6×10^{-4}

throughout its operation (when adjusting from 0 % to 100 % intensity). These mean relative SPDs were then scaled such that their integrals (i.e., the respective sums of the values at the sampling wavelengths of the spectroradiometer) were all equal to 1, as shown in Fig. 8. Scaled in this way, they serve as the column vectors **R**, **G**, and **B** defining the reference color stimuli used in the colorimetric computations described in later sections.
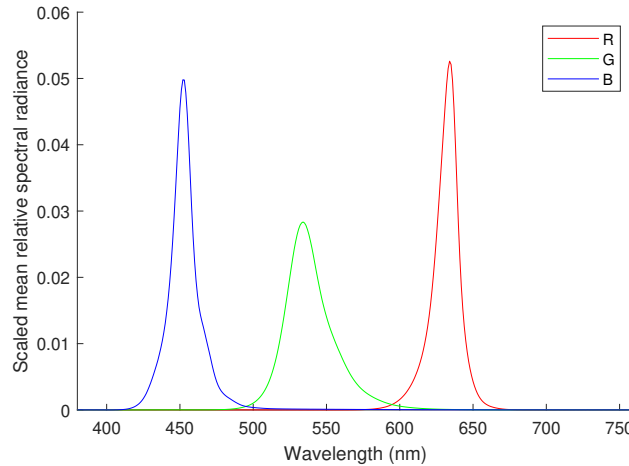

Fig. 8. Scaled mean relative SPDs of the RGB LED lights, the scaling making their integrals (i.e., sums of their values at the sampling wavelengths of the spectroradiometer) all equal to 1 (The SPDs define the reference color stimuli, **R**, **G**, and **B**.)

3.2.3. Homogeneity of the bipartite field

To characterize the spatial homogeneity of the bipartite field, 17 non-overlapping points that spanned the entire region of each half of the bipartite field were measured with the spectroradiometer. Spatial homogeneity is affected by how well the LED lights are mixed. If only one type of LED light was active, the effect of mixing on spatial homogeneity would not be apparent. Thus, for the test field, a near-spectral light of peak wavelength 490 nm was set, which required all three RGB LED lights to be active for a match to be obtained. The average signal for those measurements was computed. Then, the root-mean-square error (RMSE) values between the average signal and each measured signal were calculated.

For the test field, the average RMSE was 5.9×10^{-5}, which when compared to the peak value of the average signal, corresponds to 0.4 %. The maximum RMSE was 1.3×10^{-4}, corresponding to 0.9 % of the average signal's peak value. In the reference field, the average and maximum RMSE were 1.5×10^{-5} and 2.6×10^{-5}, equivalent to 0.3 % and 0.5 % of the average signal's peak value, respectively.

3.2.4. Light leakage between the halves of the bipartite field

It was observed that the two halves of the bipartite field were not entirely isolated from each other internally, resulting in some light leakage between them. For instance, when a near-spectral light of peak wavelength 540 nm was applied to the test side, a minor component leaked into the reference side of the bipartite field, as illustrated in Fig. 9.

To characterize this internal stray light leakage, various near-spectral lights were applied to the test field of the apparatus, with no light inputs to the reference field. The SPD of the emitted light was measured for both fields, and the radiometric values at the peak wavelength were compared. In all cases, the leakage was found to be less than 0.6 %. Although this leakage is
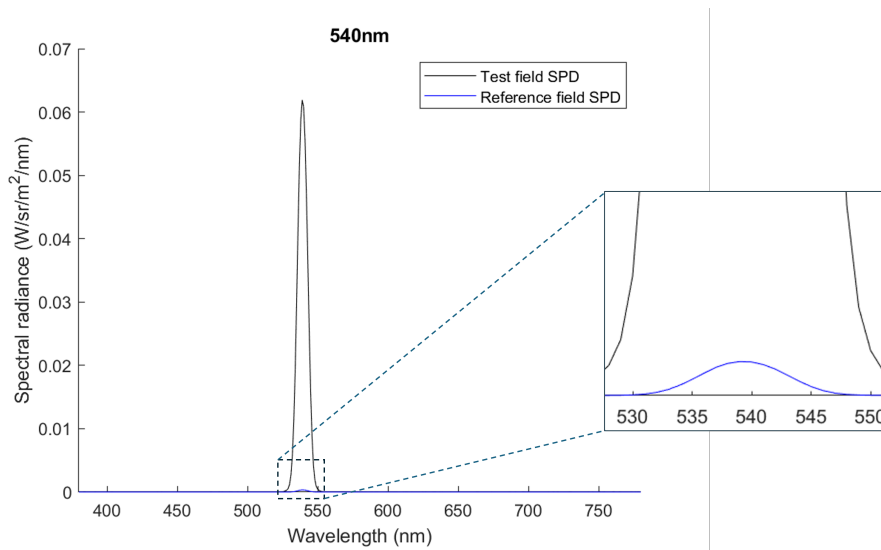
Fig. 9. Light leakage from test field to reference field

potentially negligible, it, along with overall ambient stray light, can be addressed by subtracting the measured SPD of the light of the reference field from that of the test field after each match. The resulting difference between the SPDs of the lights emanating from the two halves of the bipartite field is denoted as $\mathbf{b}_{\lambda_i}$ in subsequent sections.

4. Pilot study

4.1. Experimental method

The experiment was performed in a dark room, where the bipartite field is the only surface from which light is emanating. There was no surround field, since color discrimination has been shown to be best when observers are adapted to the color stimulus itself [66].

In Fig. 10, the experimental setup's schematic is shown. The observer's eyes would be roughly positioned around 80 cm from the bipartite field, corresponding to a visual field of 1.4°. Given the latter being less than 2°, small changes of the visual field due to head movements would not significantly impact color matching. As a result, no head restraints were used. However, observers were instructed to loosely maintain the same viewing position during the experiment.

Observers were briefed on the study's objective and given verbal instructions on how to operate the controls of the setup. They were given time to adapt to the viewing conditions and familiarize themselves with the controls by attempting to obtain matches for some selected lights.

The experiment began with a near-spectral test light of peak wavelength λ_i set on the test side of the bipartite field, while the RGB LED lights, which in the experimental context are referred to as reference lights, were off. The observer would then scan the bipartite field freely with both eyes and attempt to match the color of the reference field to that of the test field. In this, they were instructed to adjust the three reference lights such that the differences in hue, brightness, and saturation between the two fields were minimized. They were informed that for some test lights, it could be necessary to desaturate the test field with one of the reference lights to obtain a match. Observers were also advised to aim for an acceptable match using the coarse control before switching to fine adjustments, and to ignore potential differences due to Maxwell's spot in the centre of the bipartite field [53]. After a match was reached for the near-spectral test
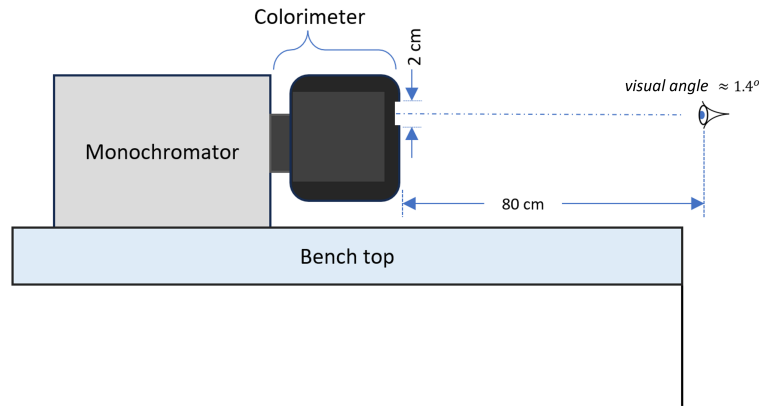
Fig. 10. Experimental setup schematic

light of peak wavelength λ_i, the SPDs of the light emanating from the test field, $\mathbf{T}_{\lambda_i}$, and from the reference field, $\mathbf{M}_{\lambda_i}$, were measured using a spectroradiometer (Konica Minolta CS-2000). The observer would then set the intensities of the three reference lights back to zero, before an additional measurement was made for the test light, $\mathbf{S}_{\lambda_i}$, all by itself. $\mathbf{T}_{\lambda_i}$, $\mathbf{M}_{\lambda_i}$, and $\mathbf{S}_{\lambda_i}$ are 401×1 column vectors representing the respective spectral radiances over the interval 380 nm–780 nm at a resolution of 1 nm.

This procedure was repeated for all the test lights within the 400 nm–720 nm range of the peak wavelengths, at every 10-nm interval.

Three color-normal observers participated in the pilot study, male age 51 (I), male age 62 (J) and female age 35 (T). The experiment was done at least three times for each observer, with some additional matches made with test lights from those spectral regions where the observer had exceptionally high matching variability. Due to the time-consuming nature of the experiments (around 15–20 hours per observer), matches were made over several sessions spanning several days. Table 2 shows the number of matches each observer made per test light.

4.2. Data processing

4.2.1. Normalization and scaling

For each test light (specified by its peak wavelength, λ_i), three measurements were made, the SPD of the light from the test field, $\mathbf{T}_{\lambda_i}$, the SPD of the light from the reference field, $\mathbf{M}_{\lambda_i}$, and the SPD of the test light alone, $\mathbf{S}_{\lambda_i}$, as described in the previous sections. Since the test lights (of peak wavelengths from 400 nm to 720 nm) varied in intensity, the measured spectral radiances of the corresponding test fields (with any desaturating lights turned off) needed to be normalized: First, the radiance, L_{e,λ_i}, was computed by summing the elements, $s_{\lambda_i,k}$, of $\mathbf{S}_{\lambda_i}$, as shown in Eq. (1). Then, by multiplying by the reciprocal of L_{e,λ_i}, a normalized vector $\bar{\mathbf{S}}_{\lambda_i}$ was obtained whose elements sum to 1, as shown in Eq. (2) and Eq. (3).

$$L_{e,\lambda_i} = \sum_k s_{\lambda_i,k} \tag{1}$$

$$f_{\lambda_i} = 1/L_{e,\lambda_i} \tag{2}$$

$$\bar{\mathbf{S}}_{\lambda_i} = f_{\lambda_i} \mathbf{S}_{\lambda_i} \tag{3}$$

The vector $\bar{\mathbf{S}}_{\lambda_i}$ thus derived can be interpreted as representing the SPD of the color stimulus yielded by the test light of peak wavelength λ_i as given per unit radiance of the test field, when all reference lights are off.

Table 2. Number of color matches made per test light for each observer (each light specified by its peak wavelength, λ_i)

λ_i (nm)	Obs-I	Obs-J	Obs-T	λ_i (nm)	Obs-I	Obs-J	Obs-T
400	3	3	3	570	3	4	3
410	3	3	3	580	4	4	3
420	3	3	3	590	3	4	3
430	3	3	3	600	3	3	3
440	4	3	3	610	5	3	5
450	3	4	3	620	9	3	3
460	4	5	5	630	3	3	3
470	5	7	3	640	3	3	4
480	5	3	4	650	3	3	3
490	3	3	3	660	3	3	3
500	3	3	3	670	3	3	3
510	3	3	3	680	3	3	4
520	5	3	3	690	3	3	3
530	4	3	3	700	3	3	3
540	3	4	3	710	3	3	3
550	5	3	3	720	3	3	3
560	4	3	3				

The corresponding SPDs of the light from the test field, $\mathbf{T}_{\lambda_i}$, and the light from the reference field, $\mathbf{M}_{\lambda_i}$, were scaled accordingly, using the scaling factor computed in Eq. (2), to yield the scaled SPDs, $\bar{\mathbf{T}}_{\lambda_i}$ and $\bar{\mathbf{M}}_{\lambda_i}$, of the corresponding color stimuli. It is worth noting that the elements of vectors $\bar{\mathbf{S}}_{\lambda_i}$, $\bar{\mathbf{T}}_{\lambda_i}$, and $\bar{\mathbf{M}}_{\lambda_i}$ have unit 1 (i.e., are "unitless").

4.2.2. Computing tristimulus values

As noted earlier, a small amount of light is leaked from the reference field into the test field and vice versa, due to stray light in the central mirror enclosure. Assuming Grassmann's laws hold true, this can be handled if we deal with the difference between $\bar{\mathbf{T}}_{\lambda_i}$ and $\bar{\mathbf{M}}_{\lambda_i}$,

$$\mathbf{b}_{\lambda_i} = \bar{\mathbf{T}}_{\lambda_i} - \bar{\mathbf{M}}_{\lambda_i} \tag{4}$$

where the 401×1 column vector $\mathbf{b}_{\lambda_i}$ can be interpreted as representing the "unitless" SPD of a virtual "black color stimulus", characterized by its tristimulus values being all equal to zero. (The black color stimuli together constitute a set of black metamers.) Since this difference spectrum is given by the three reference color stimuli and the test-color stimulus, we can form the matrix $\mathbf{A}_{\lambda_i}$ as

$$\mathbf{A}_{\lambda_i} = \begin{bmatrix} \mathbf{R} & \mathbf{G} & \mathbf{B} & \bar{\mathbf{S}}_{\lambda_i} \end{bmatrix} \tag{5}$$

where $\mathbf{R}$, $\mathbf{G}$, $\mathbf{B}$, and $\bar{\mathbf{S}}_{\lambda_i}$ are 401×1 column vectors representing the "unitless" SPDs of the reference color stimuli (cf. Fig. 8) and the normalized test-color stimulus of peak wavelength λ_i, respectively.

Introducing $\mathbf{x}_{\lambda_i}$ as a 4×1 column vector where the first three rows are the coefficients of the reference color stimuli at color match and the final row is the corresponding coefficient of the

test-color stimulus, this can be found as the least-squares solution to

$$\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i} \cong \mathbf{b}_{\lambda_i} \tag{6}$$

which is

$$\mathbf{x}_{\lambda_i} = (\mathbf{A}_{\lambda_i}^T\mathbf{A}_{\lambda_i})^{-1}\mathbf{A}_{\lambda_i}^T\mathbf{b}_{\lambda_i} \tag{7}$$

The elements of matrix $\mathbf{x}_{\lambda_i}$ are then normalized such that the coefficient of the test-color stimulus, $\bar{\mathbf{S}}_{\lambda_i}$, is -1. This ensures that the coefficients referring to the reference lights active on the reference side of the bipartite field are always positive, while the coefficients referring to those active as desaturation component on the test side are negative.

Repeating these steps for all test lights applied in the experiments, the tristimulus values of the corresponding (near-spectral) test-color stimuli can be derived. Ultimately, the CMFs of the individual observer are obtained by expressing the resulting tristimulus values as three functions of wavelength defined on the set $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$, where n is the number of test lights included in the color-matching experiments.

4.2.3. Assessing estimation accuracy and matching uncertainty with PSNR

After estimating the coefficient vector, $\mathbf{x}_{\lambda_i}$, that minimizes the difference between the SPDs of the color stimuli originated from respective halves of the bipartite field through least-squares regression, the fidelity of the estimation can be evaluated. The peak signal-to-noise ratio (PSNR) metric is widely utilized in signal processing tasks to gauge the fidelity of an estimated signal compared to a reference signal. In this context, the reference signal is represented by the vector $\mathbf{b}_{\lambda_i}$, which equals the difference between the vectors $\bar{\mathbf{T}}_{\lambda_i}$ and $\bar{\mathbf{M}}_{\lambda_i}$ representing the "unitless" SPDs of the color stimuli originated from the test field and the reference field, respectively, as shown in Eq. (4). The estimated signal, derived from the computed coefficients and represented by $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ in Eq. (6), reflects the model's prediction of the difference between the two SPDs. Thus, the PSNR can be computed between the difference vector $\mathbf{b}_{\lambda_i}$, obtained directly from measured light spectra, and the product of matrix $\mathbf{A}_{\lambda_i}$ and the computed coefficient vector, $\mathbf{x}_{\lambda_i}$, using Eq. (8), where the numerator of the fraction corresponds to the maximum value of vector $\mathbf{b}_{\lambda_i}$, while the denominator is the root-mean-square error of $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ with respect to $\mathbf{b}_{\lambda_i}$.

$$\text{PSNR}_{\mathbf{Ax},\lambda_i} = 20\log_{10}\left(\frac{\max(\mathbf{b}_{\lambda_i})}{\text{RMSE}(\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}, \mathbf{b}_{\lambda_i})}\right) \tag{8}$$

The main factors that would affect PSNR in $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ versus $\mathbf{b}_{\lambda_i}$, are the first three columns of the matrix $\mathbf{A}_{\lambda_i}$, that is, the vectors $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$ representing the reference color stimuli. It is worth noting that these vectors were derived in the calibration step and not through measures after every match, unlike $\mathbf{S}_{\lambda_i}$. Moreover, they were derived from the average SPDs of the corresponding LED lights during their operation, and thus do not reflect the exact SPDs of these when a particular match was reached. As mentioned in Section 3.2, there are minor shifts in the RGB LED lights' respective SPDs during their operation. To assess whether those minor shifts would affect a color match, it is therefore important to quantify the matching uncertainty of the observer for a specific test-color stimulus. If the matching uncertainties are larger than the shifts in SPDs of the LED lights, the latter can be considered to be insignificant. This analysis can also be done in terms of PSNR.

In the experiments, several matches were made for each test light. Since each match corresponds to a slightly different $\mathbf{b}_{\lambda_i}$, a representative difference vector $\bar{\mathbf{b}}_{\lambda_i}$ was obtained by averaging. The latter then served as a reference signal against which the PSNR for each measurement-determined
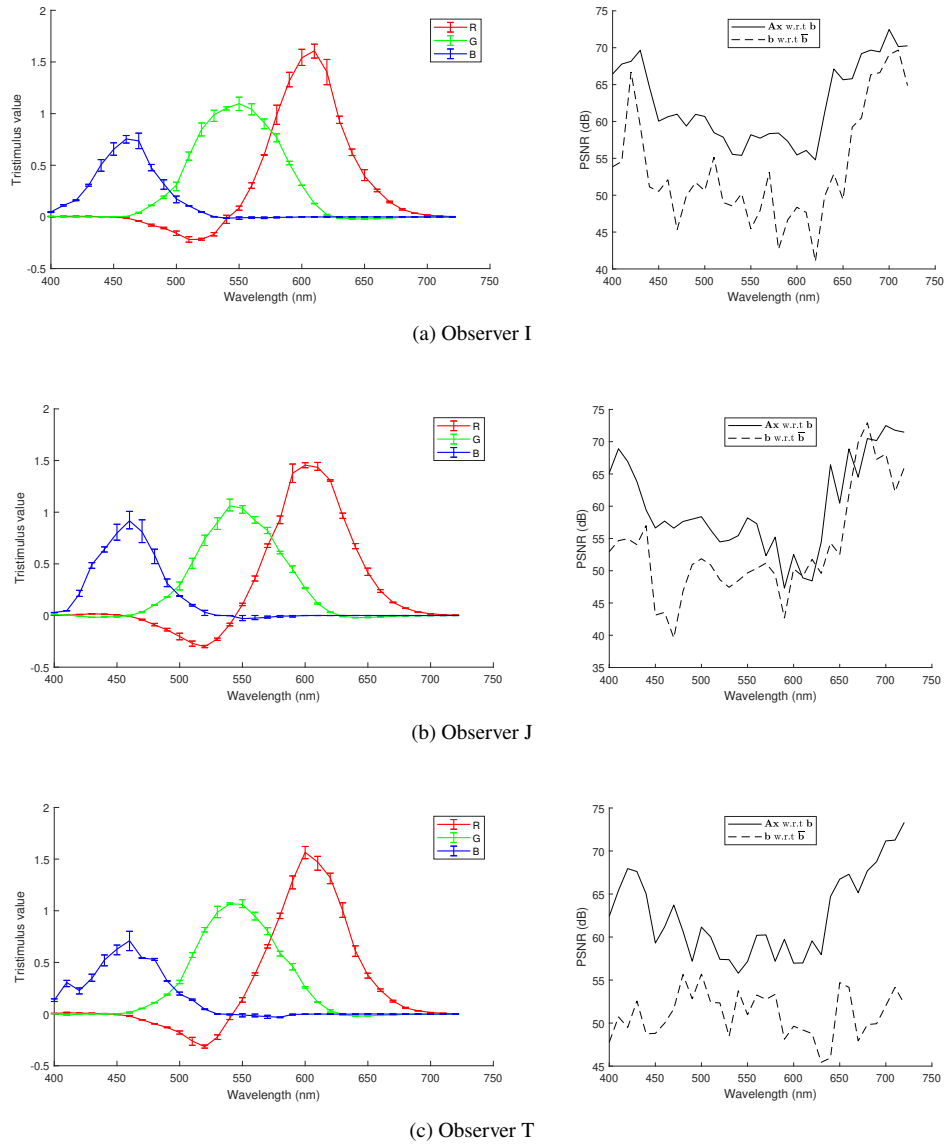
(a) Observer I

(b) Observer J

(c) Observer T

Fig. 11. Computed CMFs for each observer, determined relative to unit-energy normalized reference color stimuli, **R**, **G**, and **B**, with peak wavelengths of 634 nm, 534 nm, and 452 nm, respectively (right). Corresponding estimation error compared with matching uncertainty, expressed in terms of PSNR (left) (data points joined by straight lines)

$\mathbf{b}_{\lambda_i}$ was calculated using Eq. (9):

$$\mathrm{PSNR}_{\mathbf{b},\lambda_i} = 20 \log_{10} \left(\frac{\max(\overline{\mathbf{b}}_{\lambda_i})}{\mathrm{RMSE}(\mathbf{b}_{\lambda_i}, \overline{\mathbf{b}}_{\lambda_i}))} \right) \qquad (9)$$

Several computations of $\mathrm{PSNR}_{\mathbf{b},\lambda_i}$ were made for each λ_i, based on the number of difference vectors $\mathbf{b}_{\lambda_i}$ determined. Considering that the difference vectors, $\mathbf{b}_{\lambda_i}$, refer to the color matches an observer makes for a given test light, an appreciation of the matching uncertainty per wavelength was depicted by averaging the calculated values of $\mathrm{PSNR}_{\mathbf{b},\lambda_i}$ for each test-color stimulus and

plotting the mean values against the respective peak wavelengths, λ_i. Similarly, the several computations of $\text{PSNR}_{\mathbf{Ax},\lambda_i}$ for each test-color stimulus were averaged and plotted against λ_i. The resulting plots for each observer are shown in Fig. 11.

4.2.4. Estimating uncertainty in mean tristimulus values

For each observer, several matches (at least three) were made for each test light. While these repeated matches may give us an intuition of the observer's own matching uncertainty, they are insufficient in estimating a measure of the latter. Traditionally, a match would have to be repeated many times before any statistically significant measure of uncertainty can be computed. This is laborious and incredibly time-consuming. To circumvent this challenge, the set of tristimulus values for each test-color stimulus was resampled using bootstrapping statistics [67,68]. By resampling the observed data, we were able to derive estimates of the mean tristimulus values along with their 95 % confidence intervals, at each wavelength λ_i. In particular, the MATLAB function *bootci* was used to compute the 95 % confidence interval of the mean tristimulus values from 1000 bootstrap samples. These confidence-interval estimations can then serve as a measure of uncertainty in the mean tristimulus values obtained.

5. Results

5.1. Color-matching functions

The complete CMFs as determined relative to unit-energy normalized reference color stimuli, $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$, with peak wavelengths of 634 nm, 534 nm, and 452 nm, respectively (cf. Fig. 8), are plotted for each observer in Fig. 11. The uncertainties in the computed means of the tristimulus values are expressed as error bars representing the 95 % confidence intervals. In the same figure, the average estimation error of the tristimulus values and the average observer matching uncertainty per test-color stimulus are also compared. A higher PSNR value suggests a higher degree of fidelity in the least-squares estimate of $\mathbf{x}_{\lambda_i}$ in $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ with respect to $\mathbf{b}_{\lambda_i}$. While a higher value in the PSNR of $\mathbf{b}_{\lambda_i}$ with respect to $\bar{\mathbf{b}}_{\lambda_i}$, suggests a low matching uncertainty.

To get an appreciation of how the individual CMFs of each observer are different from one another, they are plotted in the same diagram, shown in Fig. 12. By plotting the mean tristimulus values alongside their respective confidence intervals as error bars, a comparative analysis of the observers' CMFs becomes possible. Notably, regions where the error bars do not overlap indicate statistically significant differences in CMFs among the individuals. This observation suggests distinct variations in CMFs among the observers at those wavelengths.

5.2. Chromaticity coordinates

Given that there are several estimates of the tristimulus values per test-color stimulus for each observer, their corresponding mean rg-chromaticity coordinates can also be calculated. Using the same bootstrapping procedure as outlined in Section 4.2.4, the set of rg-chromaticity coordinates were therefore resampled to obtain estimates of the mean rg-chromaticity coordinates and corresponding 95 % confidence intervals for each test-color stimulus included. The resulting mean rg-chromaticity coordinates for each observer, along with the respective error bars representing the 95 % confidence intervals in the r and g coordinates, are plotted in Fig. 13. The underlying reference color stimuli are the same as for the CMFs in Fig. 11.
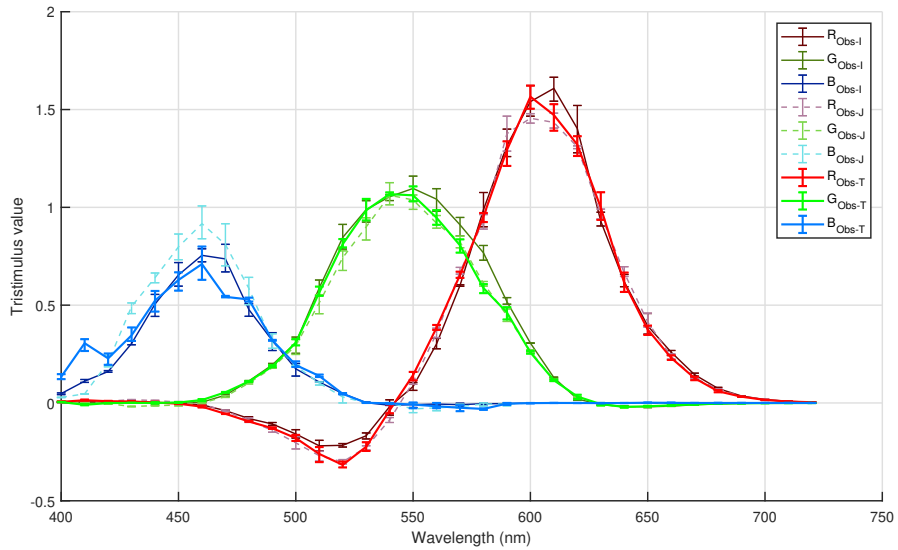
Fig. 12. Comparison of the CMFs of three observers for inter-observer differences (data points joined by straight lines)
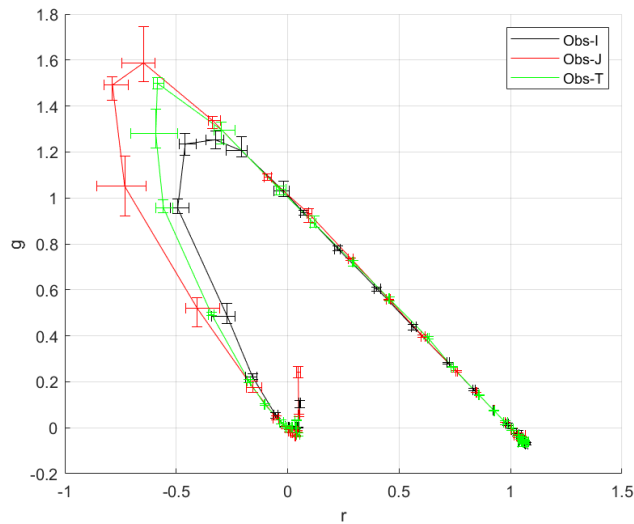

Fig. 13. rg-chromaticity diagrams with error bars of observer I, J and T (data points joined by straight lines). The diagrams refer to unit-energy normalized reference color stimuli, **R**, **G**, and **B**, with peak wavelengths of 634 nm, 534 nm, and 452 nm, respectively.

6. Discussion

6.1. Apparatus

The apparatus used in this study is markedly different from its predecessor [13], and features several improvements. In the improved visual colorimeter, a knife-edge right-angled first-surface mirror replaces the 3D-printed central wedge structure from the previous design. This modification not only sharpens the bipartite edge, but also reflects most of the incident light toward the observer, significantly reducing beam divergence. As a result, stray light emanating from the bipartite field was significantly reduced. One consequence of this, was that the possible viewing angle was much narrower. The bipartite field was only visible when viewed directly in front, and was invisible when seen from the side.

The integrating hemispheres from the previous setup have been replaced with fully spherical integrating chambers, improving the light mixing properties of the system. The addition of condensing lenses ensured that light exiting the integrating chambers was near-parallel and homogeneous. Since the spatial homogeneity of the previous colorimeter's bipartite field was not measured, no quantitative comparison can be made. However, observers who participated in experiments involving both setups reported that the bipartite field of the improved colorimeter was noticeably more homogeneous than the predecessor and the matching experience was better overall.

Another important improvement over the old design, was the increase in luminance at the edges of the visible spectrum, specially in the 400 nm–450 nm region. Compared with its predecessor, the colorimeter's luminance in this region has increased threefold. Despite those improvements, for test-color stimuli with peak wavelengths within the intervals 400 nm–420 nm and 700 nm–720 nm, the corresponding test fields are still not bright enough and remain non-photopic. It is worth noting that a color match may still be achievable, although not a colorimetric one, if the field's luminance is below the photopic threshold. Moreover, it should be emphasized that the photopic threshold varies between individuals and is somewhat loosely defined in the literature.

Previous color-matching experiments often employed a Maxwellian view system to focus as much light as possible through the centre of the pupil and lens, onto the retina. In so doing, retinal illumination could be significantly improved [69]. However, this type of viewing system limits the experiment to be done with one eye only and is far from the natural viewing condition of a regular person. In daily life, an observer would use both eyes, and light would fall on a larger portion of the lens, not just the centre. In particular, the spectral transmittance of the central section of the eye lens may not be representative of the actual transmittance of the lens as a whole. For these reasons, and to keep the experimental setup relatively simple, we opted not to have a Maxwellian view system.

6.2. Colorimeter performance

Analysis of the PSNR plots in Fig. 11 offers valuable insights into the performance of the colorimeter in measuring the CMFs of individual observers. Across observers I, J, and T, variations in the PSNR curves highlight differences in the stability and accuracy of the coefficient-estimation process. Consistently lower PSNR values of $\mathbf{b}_{\lambda_i}$ versus $\bar{\mathbf{b}}_{\lambda_i}$ compared to the PSNR of $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ versus $\mathbf{b}_{\lambda_i}$ indicate that observer matching uncertainty predominantly influences the variability observed in the measurement-based difference vector $\mathbf{b}_{\lambda_i}$. However, subtle deviations are observed for observer J, where the PSNR of $\mathbf{b}_{\lambda_i}$ versus $\bar{\mathbf{b}}_{\lambda_i}$ occasionally surpasses the PSNR in $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ versus $\mathbf{b}_{\lambda_i}$. After averaging the PSNR curves for all observers, it becomes evident that the colorimeter demonstrates stability in measuring intra-observer variability, as shown in Fig. 14. Notably, the small system variations of the colorimeter caused by spectral shifts of the

LED lights are likely negligible, as the mean PSNR of $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ versus $\mathbf{b}_{\lambda_i}$ is consistently higher than the mean PSNR of $\mathbf{b}_{\lambda_i}$ versus $\overline{\mathbf{b}}_{\lambda_i}$.
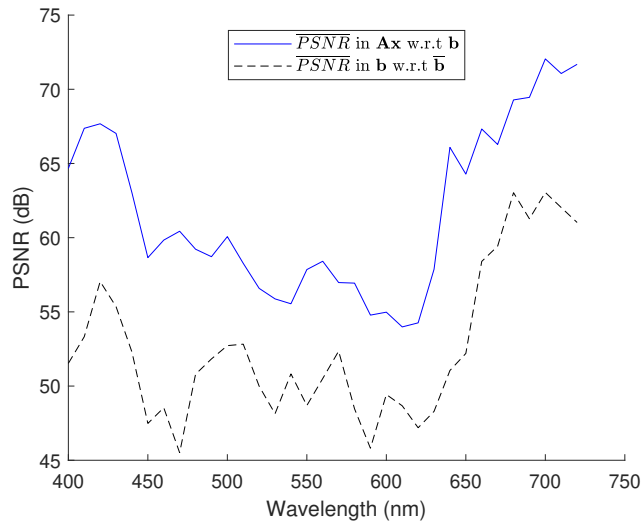

Fig. 14. Mean PSNR for all observers (data points joined by straight lines)

In addition to observer-specific variability, distinct trends observed in PSNR across wavelength regions provide further insights into the influence that spectral shifts of the LED lights have on the accuracy of coefficient estimation. The higher PSNR of $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ versus $\mathbf{b}_{\lambda_i}$ observed in the short- and long-wavelength regions suggests the stability of the blue and the red LED. Conversely, the drop in PSNR in the medium-wavelength region corresponds to the greater spectral shifts observed for the green LED. These observations underscore the link between the spectral shifts of the LED lights and the accuracy of coefficients estimation, highlighting the importance of accounting for such shifts in assessing colorimeter performance.

When analysing the mean PSNR curve of $\mathbf{b}_{\lambda_i}$ versus $\overline{\mathbf{b}}_{\lambda_i}$ in Fig. 14, it appears that its shape somewhat follows a similar pattern as that of the mean PSNR curve of $\mathbf{A}_{\lambda_i}\mathbf{x}_{\lambda_i}$ versus $\mathbf{b}_{\lambda_i}$. PSNR tends to be high at the edges of the visible spectrum, particularly in the long-wavelength region, while it is noticeably lower in the middle of the visible spectrum. Our hypothesis to explain this trend is that in the central region of the visible spectrum, where typically all three reference lights are activated to achieve a color match, there is greater uncertainty introduced in the matching process. In contrast, toward the edges of the spectrum, typically only two of the three reference lights are required to reach a match, resulting in reduced uncertainty in the matching process. Observers also report that matching is easier toward the edges of the spectrum than in the middle, reinforcing our hypothesis.

6.3. Intra- versus inter-observer variability

Observers exhibit significant variation in their approach to color-matching, with some displaying meticulousness in striving for absolute similarity between fields, while others are more expedient in reaching approximate matches. Most observers reach the vicinity of a color match effortlessly. However, the time they spend fine-tuning differs significantly among them. Repeated color-match settings reveal that the matching uncertainty for test-color stimuli of certain peak wavelengths can be considerable. This would counter-intuitively mean that the observer reaches a match faster and would be satisfied with a wider range of settings. Conversely, challenging color matches tend

to entail smaller uncertainty, reflecting a narrower range of satisfactory settings due to heightened observer discernment.

In previous studies [22,56], relative standard deviation has often been employed to assess variability in color-matching data. However, we contend that it is not a suitable metric for evaluating variability because it yields infinite values for uncertainty where the CMFs cross zero. This arises from the implicit assumption that uncertainty in tristimulus values should be zero when the coefficients themselves are zero, which does not accurately reflect the physical reality of color-matching experiments. In our experiments, it is common for observers to register a small component of a reference light either above or below zero when they intend for that light to be exactly at zero. Therefore, we utilize the PSNR metric to assess matching uncertainty. PSNR provides an indication of uncertainty relative to the peak value of the signal, rather than the signal's current level. It is worth noting that if relative standard deviation was calculated using the peak signal's value rather than the value at each wavelength, it would give the same information as the PSNR metric.

From the individual PSNR plots, in Fig. 11, it appears that the wavelength where individual observers tend to have the highest matching uncertainty, seems to vary. From the mean PSNR plots in Fig. 14, test-color stimuli at 470 nm and 590 nm stand out as having the largest intra-observer matching uncertainty. When it comes to assessing inter-observer variability, Fig. 12 is useful. Regions where there are clear differences in the mean tristimulus values between the different observers, can be clearly identified; inter-observer variability is only statistically significant in regions where error bars do not overlap. Notably, the largest inter-observer differences can be observed in the short-wavelength region of the visible spectrum, especially below 480 nm in the CMF referring to the "blue" reference color stimulus. Significant differences can be observed in the CMF referring to the "green" reference color stimulus as well, notably in the 560 nm–590 nm region. For the CMF referring to the "red" reference color stimulus, the tristimulus values are significantly different at the wavelengths of 520 nm and 610 nm. Above 620 nm, no significant inter-observer differences can be observed.

Viénot [56] observed significant inter-observer variations in the blue and green regions of the visible spectrum, with the most substantial variations occurring in the blue region. This is consistent with our findings.

6.4. Validity of the measured color-matching functions

By definition, a set of CMFs and the underlying set of reference color stimuli have to be biorthogonal with each other, i.e., $\mathbf{C}^T \begin{pmatrix} \mathbf{R} & \mathbf{G} & \mathbf{B} \end{pmatrix} = \mathbf{I}$, where $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$ represent the SPDs of the reference color stimuli, $\mathbf{C}$ is the $N \times 3$ matrix whose columns represent the corresponding CMFs, and $\mathbf{I}$ is the 3×3 identity matrix. In a matching context, this implies that if a mixture of specific quantities of the reference lights is used as test light, the observer should, in theory, match it (within the bounds of matching uncertainty) with a superposition of the same quantities of those reference lights.

Previous authors often transform CMFs to some arbitrarily chosen set of reference color stimuli before presenting them. This forces the biorthogonality condition to be true. In particular, the Stiles and Burch data were transformed in this way, generating non-experimental tristimulus values that are (1,0,0), (0,1,0) and (0,0,1) at the wavelengths of the monochromatic reference color stimuli. However, for our three observers, the individual means of the directly measured CMFs – represented here by the columns of respective 401×3 matrices $\overline{\mathbf{C}}$ – do not satisfy the

biorthogonality criterion:

$$\overline{\mathbf{C}}_{\text{Obs-I}}^{T}\begin{pmatrix}\mathbf{R} & \mathbf{G} & \mathbf{B}\end{pmatrix} = \begin{pmatrix} 0.9201 & 0.0276 & -0.0057 \\ 0.0077 & 0.9551 & 0.0199 \\ 0.0002 & 0.0092 & 0.6316 \end{pmatrix} \tag{10}$$

$$\overline{\mathbf{C}}_{\text{Obs-J}}^{T}\begin{pmatrix}\mathbf{R} & \mathbf{G} & \mathbf{B}\end{pmatrix} = \begin{pmatrix} 0.9274 & -0.0090 & -0.0014 \\ 0.0068 & 0.8913 & 0.0086 \\ 0.0001 & 0.0043 & 0.7700 \end{pmatrix} \tag{11}$$

$$\overline{\mathbf{C}}_{\text{Obs-T}}^{T}\begin{pmatrix}\mathbf{R} & \mathbf{G} & \mathbf{B}\end{pmatrix} = \begin{pmatrix} 0.9285 & 0.0091 & -0.0109 \\ 0.0078 & 0.9304 & 0.0232 \\ -0.0002 & 0.0107 & 0.5996 \end{pmatrix} \tag{12}$$

Each row of the rightmost matrices represents the tristimulus values of each reference color stimulus for the given observer. If the biorthogonality condition were satisfied, the expected results of Eqs. (10)–(12) would be 3×3 identity matrices. The off-diagonal elements of the resulting matrices are not exactly zero, although quite close. This is expected given the matching uncertainty of the observers. However, the values of the diagonal elements are significantly lower than the ideal value of one, especially so for the diagonal element referring to the "blue" reference color stimulus.

The consistent deviation of the tristimulus values for the "blue" reference color stimulus from their expected values of (0,0,1) across all three observers may indicate the presence of systematic errors in the measurements. To investigate this further, we have, for each observer, determined the set of CMFs – represented as the columns of a 401×3 matrix $\mathbf{C}$ – that are as close as possible to the mean of the individual's directly measured CMFs, while satisfying the biorthogonality constraint. This was accomplished by solving the following quadratic programming problem for $\mathbf{C}$:

$$\min\left(||\mathbf{C} - \overline{\mathbf{C}}||^2 + \alpha||\mathbf{T}\mathbf{C}||^2\right) \quad \text{s.t.} \quad \mathbf{C}^T\begin{pmatrix}\mathbf{R} & \mathbf{G} & \mathbf{B}\end{pmatrix} = \mathbf{I} \tag{13}$$

where $\mathbf{T}$ is a Tikhonov regularization matrix.

In Fig. 15, the regularized CMFs that satisfy the biorthogonality criterion (i.e., the functions represented by the vectors $\mathbf{C}_R$, $\mathbf{C}_G$, and $\mathbf{C}_B$ formed from the columns of matrix $\mathbf{C}$) are plotted for each observer alongside the mean of their respective directly measured CMFs (represented by the vectors $\overline{\mathbf{C}}_R$, $\overline{\mathbf{C}}_G$, and $\overline{\mathbf{C}}_B$ formed from the columns of matrix $\overline{\mathbf{C}}$).

From the figure, the regularized CMFs referring to the "red" reference color stimulus appear, for all three observers, to lie within (or close to) most of the corresponding mean CMFs' error bars, indicating an agreement between the measured and regularized data for this reference color stimulus. Regarding the regularized CMFs referring to the "green" reference color stimulus, the situation is more nuanced. For observers I and T, the respective graphs predominantly fall within the error bars of the corresponding mean CMFs, deviating slightly at some data points. For observer J, however, the graph shows a more significant deviation, which is also apparent in Eq. (11), where the tristimulus value for the "green" reference color stimulus are further away from (0,1,0) than they are in the case of the other two observers. Conversely, for the CMFs referring to the "blue" reference color stimulus, the regularized versions deviate significantly from the corresponding measured mean CMFs for all three observers, particularly in the 450 nm–460 nm region. The magnitude of the deviations is clearly correlated to how far the tristimulus values for the "blue" reference color stimulus are from their theoretical values
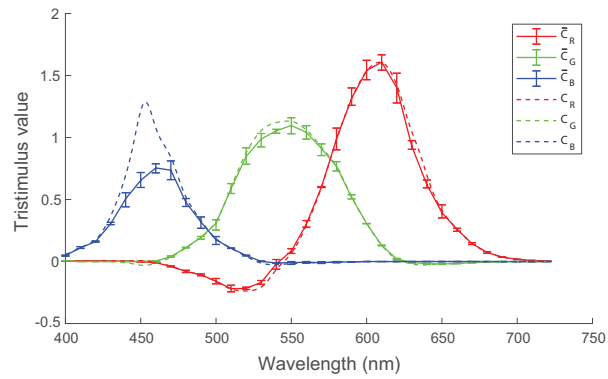
(a) Observer I
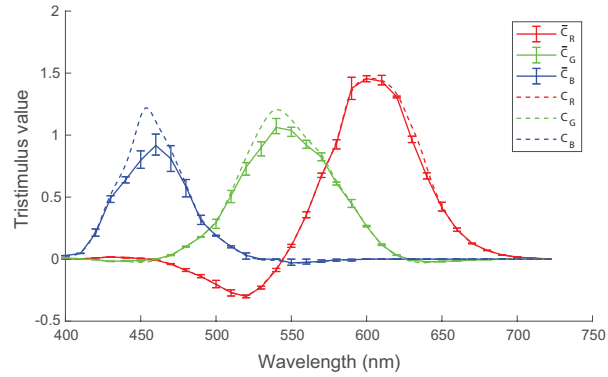
(b) Observer J
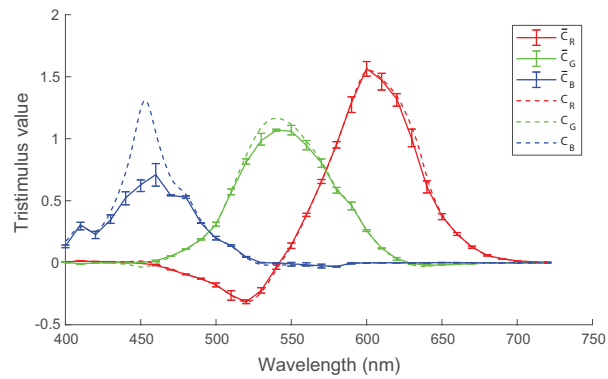
(c) Observer T

Fig. 15. Computed mean CMFs, $\overline{\mathbf{C}}_R$, $\overline{\mathbf{C}}_G$, and $\overline{\mathbf{C}}_B$, for each observer compared with regularized CMFs, $\mathbf{C}_R$, $\mathbf{C}_G$, and $\mathbf{C}_B$, that would satisfy the biorthogonality criterion (data points joined by straight lines)

of (0,0,1) (cf. Eqs. (10)–(12)). This may suggest a more systematic issue in the measurement process and/or particular challenges in accurately matching blue aperture colors.

Since the reference lights in the color-matching experiments of this study are heterochromatic (albeit relatively narrow-band) LED lights, it is necessary to add a desaturating component in the test field, even when the peak wavelength of the near-spectral test light is close to the peak wavelength of one of the reference lights superimposed in the reference field. If the desaturation is insufficient and the reference light similar to the test light contributes more to brightness than to saturation, the light emanating from the reference field will thus have a smaller than expected component of this reference light. This underscores the greater influence of brightness over other color attributes, such as saturation, which appears to be a recurring trend in our color-matching experiments using the "maximum saturation" method. We believe this could explain the lower settings of the blue reference light for near-spectral test lights with peak wavelengths near that of the blue reference light itself.

To test our hypothesis, additional measurements were conducted for the test lights with peak wavelengths λ_i of 450 nm, 460 nm, and 470 nm. In these new experiments, the observers were specifically instructed to over-desaturate the test field using both the red and the green reference lights before starting the matching process. In Table 3, the tristimulus values obtained through this new procedure are compared with those from the previous experiments.

Table 3. Tristimulus values for near-spectral test stimuli with peak wavelengths λ_i near the wavelength of maximum intensity for the blue reference light, as obtained for each observer using the original matching procedure (Old tristimulus values) and the new matching procedure (New tristimulus values)

Observer	λ_i	Old tristimulus values (R, G, B)			New tristimulus values (R, G, B)		
	450 nm	−0.002	0.002	0.6546	0.0106	−0.0232	0.9259
Obs-I	460 nm	−0.0094	0.0037	0.7547	−0.0046	0.0011	0.9888
	470 nm	−0.0393	0.04	0.7364	−0.036	0.042	0.866
	450 nm	0.005	−0.0109	0.7979	0.0238	−0.0308	1.3324
Obs-J	460 nm	−0.0071	−0.0003	0.9166	0.0011	−0.0088	1.2454
	470 nm	−0.0397	0.0325	0.8101	−0.042	0.0426	0.9419
	450 nm	−0.0038	0.0022	0.6288	0.0204	−0.032	1.2716
Obs-T	460 nm	−0.019	0.0152	0.7101	0.001	−0.0212	1.3869
	470 nm	−0.0554	0.0545	0.5425	−0.0517	0.048	0.9015

As apparent from the old tristimulus values in Table 3, for a test stimulus with a peak wavelength of 450 nm, i.e., near the intensity peak for the blue reference light, observers I and T did not sufficiently desaturate the test field with the green reference light, resulting in that side of the bipartite field appearing less bright. Since the blue reference light contains components also of longer wavelengths, which contribute more to brightness than the near-spectral test light with intensity peak at 450 nm, a smaller amount of the blue reference light was required to achieve a match in brightness. Observer J, however, desaturated more strongly with the green reference light, leading to a higher intensity for the blue reference light at match with that same test light, but still not sufficient to approach the intensity setting expected from theory. A similar pattern is observed for the near-spectral test light with maximum intensity at 460 nm. In the new measurements, when observers were instructed to start the color matching by adding a desaturating component, they generally registered higher settings for the blue reference light and commented that it was *easier* to achieve a match. This supports the hypothesis that brightness takes precedence over saturation in color matching, though further investigation is needed.

However, for a test stimulus at 470 nm, further away from the peak wavelength of the blue reference light, there is no clear relationship between desaturation with the green reference light
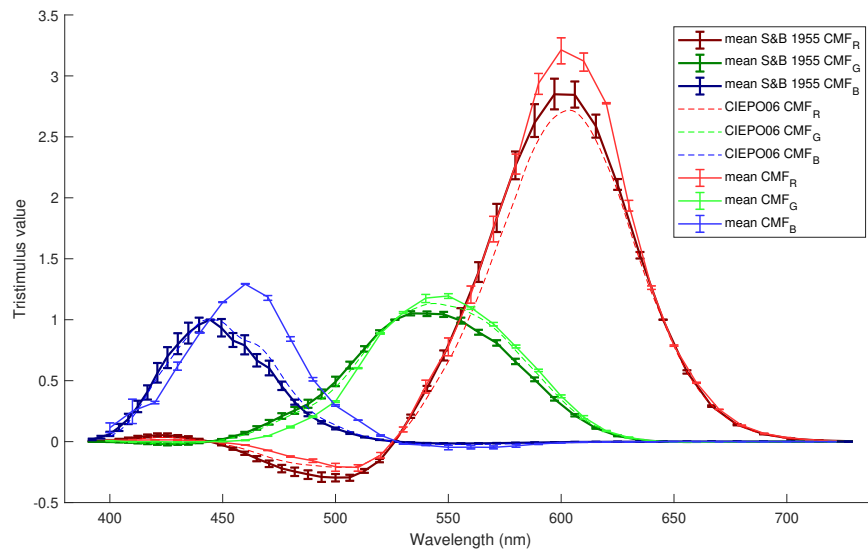
Fig. 16. Mean CMFs of the three observers in this study compared with the Stiles and Burch 1955 mean 2° CMFs and the CIEPO06 model's prediction for an observer of age 49 and a visual angle of 1.4° (data points joined by straight lines)

and increased intensity of the blue. In fact, in that wavelength region, observers start to desaturate with the red reference light, and their desaturation settings do not vary significantly in the new measurements compared to the original ones.

While this pilot study reveals interesting patterns in color matching, future research must ensure that the biorthogonality criterion is met. Particular attention should be given to confirming that observers are matching across all color attributes, including brightness, saturation, and hue.

6.5. Comparison with Stiles and Burch's 1955 CMFs

In this study, an attempt at replicating the seminal color-matching experiment conducted by Stiles and Burch in 1955 was made, albeit with only three observers. From the previous section, it was found that the measured CMFs in this study do not satisfy the biorthogonality criterion. From further analysis, there are grounds to believe that there are some systematic errors in the measured CMFs, especially in the blue region of the visible spectrum.

Nonetheless, it is worth comparing the measured CMFs, albeit with possible systematic errors, with the Stiles and Burch 1955 mean 2° color-matching data [70], to see how much they deviate from them. To facilitate such comparison, the CMFs of our three observers were averaged and the resulting mean CMFs were then transformed to the same basis of monochromatic reference color stimuli as that of the Stiles and Burch 1955 study, notably with wavelengths of 444 nm, 526 nm, and 645 nm for the "blue", the "green", and the "red" stimulus, respectively. This adjustment enabled a direct comparison with the mean observer established by Stiles and Burch. Subsequently, the data were scaled accordingly for accurate plotting.

Considering that the average age of the observers in this study is around 49 years, which is 19 years older than the age of the Stiles and Burch 1955 mean observer, comparison with the CIEPO06 model's prediction for an observer of age 49 and a visual angle of 1.4° might provide some additional insights. The mean CMFs of the three observers were therefore plotted alongside the Stiles and Burch 1955 mean 2° CMFs and the CIEPO06 model's predicted CMFs in Fig. 16.

The error bars in the Stiles and Burch 1955 mean CMFs represent the 95 % confidence interval of the mean and were generated by bootstrapping from the individual CMFs data of the ten
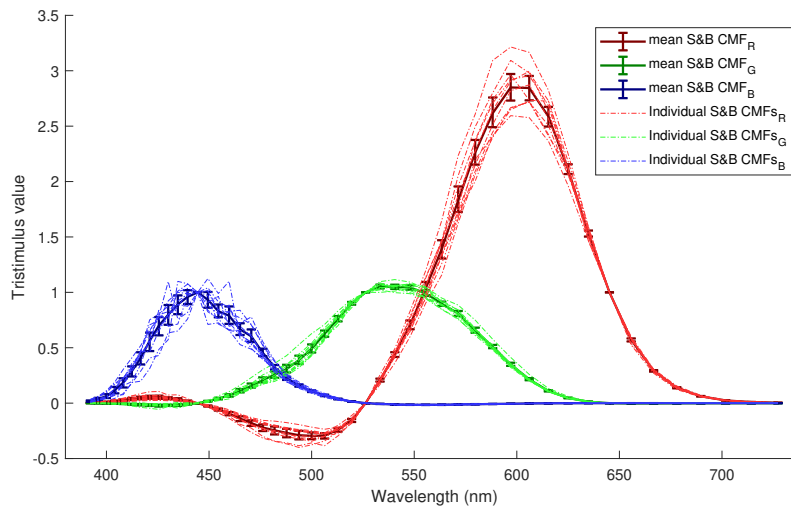
Fig. 17. Stiles and Burch 1955 mean 2° CMFs plotted with error bars representing 95 % confidence intervals. The dashed lines are the CMFs of individual observers in the original Stiles and Burch 2° pilot dataset. (data points joined by straight lines)

observers who participated in the study, shown in Fig. 17. In contrast, the error bars in our mean observer CMFs were generated by bootstrapping from the set of all tristimulus values available at each test-color stimulus for all three observers. Since our CMF dataset includes repeated measurements from the three observers, the error bars are typically smaller than those of the Stiles and Burch 1955 mean observer CMFs, which have only one data point for each of the 10 observers.

When comparing the mean CMFs to those of Stiles and Burch, we observed a significant shift in the function referring to the "blue" reference color stimulus and a lesser shift in the one referring to the "green". The CIEPO06 model's predicted CMFs also show a shift in the corresponding two functions toward the red end of the spectrum, although to a much lesser extent than our mean CMFs suggest. It is important to highlight that the CIEPO06 model is derived from the Stiles and Burch datasets (both 2° and 10°), so it is unsurprising that it remains closely aligned with the mean 2° CMFs from their work.

Several factors may contribute to these observations. Given the higher average age of the observers in this study and the fact that ageing often leads to increased "yellowing" of the lens, this could partially account for the observed shifts. Additionally, it is important to note that in the wavelength region below 450 nm, our mean CMFs exhibit a lesser degree of desaturation with the "green" reference color stimulus. As a result, lower settings for the "blue" reference color stimulus were recorded for all three observers. This may explain why the tristimulus values associated with the "blue" reference color stimulus are lower than those reported by Stiles and Burch in this wavelength range, leading to an apparent shift toward the red end of the spectrum.

As shown in Fig. 17, when examining the individual CMFs from Stiles and Burch (1955), at least one set appears to somewhat resemble our mean CMFs. However, since we do not have access to the raw color-matching data from Stiles and Burch, it is not possible to determine if their data meet the biorthogonality requirement. The available data have already been transformed to monochromatic primaries, which would artificially impose this condition.

7.   Conclusion and future works

This study demonstrates the development and implementation of a visual colorimeter and method for measuring individual CMFs through psychophysical experiments. Designed with 3D-printed components, a few optical elements, and LED-based light engines, this apparatus successfully facilitated precise and stable measurements of CMFs for individual observers.

The system is stable and precise enough to effectively characterize intra-observer variability. From the matching experiments conducted, regions with significant inter-observer differences in CMFs were identified. When comparing the mean observer CMFs from this study with the Stiles and Burch 1955 2° mean CMFs, we find that the shapes of the functions generally agree, though significant shifts were observed in the CMFs referring to the "blue" and, to a lesser extent, the "green" reference color stimuli. These shifts may be partly due to age-related "yellowing" of the lens, which is expected given the mean age difference between our observers and those in the Stiles and Burch study. The systematic errors identified in Section 6.4 may be another reason for the shifts observed.

The colorimeter presented in this study is significantly improved from its predecessor [13]. Critical refinements were made, particularly enhancing the homogeneity and overall luminance of the bipartite field, especially in the short-wavelength region. These improvements have mitigated previous limitations and resulted in a more consistent and reliable measurement process. Despite these advancements, certain limitations persist, particularly regarding the stimuli from the violet end (400 nm–420 nm) and the far-red end (700 nm–720 nm) of the visible spectrum, which remain non-photopic. Future work could address this issue through further refinements to the apparatus.

On the matter of validity of the measured CMFs, it is evident that the biorthogonality criterion was not satisfied. It was observed that, at certain wavelengths, brightness may exert a greater influence than saturation when observers attempt to achieve a color match. As a result, observers were not desaturating enough for certain test stimuli. In future experiments, precautions should be taken to ensure that observers are matching across all color attributes and that the biorthogonality criterion is adequately satisfied. Although stepping away from "raw" measurements, regularization to find the optimal functions that would satisfy this condition while adhering to the bounds of matching uncertainty may be a worthy addition as a post-processing step. Ultimately, the true test for any set of individual CMFs, is its ability to predict color matches for that individual observer.

It is important to acknowledge that the sample size in this study remains limited, comprising color-matching data from only three observers. Expanding the dataset to include a more diverse and representative demographic would offer more profound insights into the variability of human color vision. Such a comprehensive database could also serve as a foundation for developing methods to estimate CMFs from a reduced set of measurements.

Nonetheless, the insights gained from this research underscore the importance of considering individual differences in color vision, paving the way for more personalized and accurate color-management solutions.

Disclosures.   The authors declare no conflicts of interest.

Data availability.   Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. N. Ohta and A. R. Robertson, *Colorimetry: Fundamentals and Applications* (John Wiley & Sons, 2005). Chap. 2.5, pp. 57–61.
2. J. Schanda, "CIE Colorimetry," in *Colorimetry: Understanding the CIE System*, J. Schanda, ed. (John Wiley & Sons, 2007), pp. 25–78.
3. Y. Asano, M. D. Fairchild, L. Blondé, *et al.*, "Color matching experiment for highlighting interobserver variability," Color Res. Appl. **41**(5), 530–539 (2016).
4. Y. Asano, "Individual colorimetric observers for personalized color imaging," Ph.D. thesis, Rochester Institute of Technology (2015).
5. P. Csuti and J. Schanda, "Colour matching experiments with RGB-LEDs," Color Res. Appl. **33**(2), 108–112 (2008).
6. A. Sarkar, "Identification and assignment of colorimetric observer categories and their applications in color and vision sciences," Ph.D. thesis, Université de Nantes (2011).
7. W. A. Thornton, "Toward a more accurate and extensible colorimetry. Part I. Introduction. The visual colorimeter-spectroradiometer. Experimental results," Color Res. Appl. **17**(2), 79–122 (1992).
8. W. A. Thornton, "Toward a more accurate and extensible colorimetry. Part II. Discussion," Color Res. Appl. **17**(3), 162–186 (1992).
9. R. L. Alfvin and M. D. Fairchild, "Observer variability in metameric color matches using color reproduction media," Color Res. Appl. **22**(3), 174–188 (1997).
10. P. Morvan, A. Sarkar, J. Stauder, *et al.*, "A handy calibrator for color vision of a human observer," in *2011 IEEE International Conference on Multimedia and Expo* (Institute of Electrical and Electronics Engineers, 2011), pp. 1470–1473.
11. K. R. Lee, A. J. Richardson, E. Walowit, *et al.*, "Predicting color matches from luminance matches," J. Opt. Soc. Am. A **37**(4), A35–A43 (2020).
12. E. Walowit and L. Tahoe, "Determination of individual-observer color matching functions for use in color management systems," Electron. Imaging **31**(14), 96-1–96-5 (2019).
13. L. M. Ragoo and I. Farup, "A simple and cost effective colorimeter for characterising observer variability in colour matching experiments," in *London Imaging Meeting*, vol. 4 (Society for Imaging Science and Technology, 2023), pp. 50–54.
14. W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colors," Trans. Opt. Soc. **30**(4), 141–164 (1929).
15. J. Guild, "The colorimetric properties of the spectrum," Phil. Trans. R. Soc. Lond. A **230**(681-693), 149–187 (1931).
16. Commission Internationale de l'Éclairage, "Principales Décisions de la Commission Internationale de l'Éclairage, 6e Session, 1924," in *Recueil des travaux et compte rendu des séances, Sixième session*, Commission Internationale de l'Éclairage, ed. (Bureau Central de la Commission, The National Physical Laboratory Teddington, Cambridge University, Cambridge, 1924), pp. 67–70.
17. "Photometry – The CIE system of physical photometry," ISO 23539:2005(E) / CIE S 010/E:2004.
18. A. Stockman and L. T. Sharpe, "Cone spectral sensitivities and color matching," in *Color vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, eds. (Cambridge University, 1999), pp. 53–87.
19. P. Csuti and J. Schanda, "A better description of metameric experience of LED clusters," Light and Engineering **18**(1), 44–50 (2010).
20. J. J. Vos, "Colorimetric and photometric properties of a 2° fundamental observer," Color Res. Appl. **3**(3), 125–128 (1978).
21. W. S. Stiles and J. M. Burch, "Interim report to the Commission Internationale de l'Éclairage, Zürich, 1955, on the National Physical Laboratory's investigation of colour-matching (1955)," Opt. Acta **2**(4), 168–181 (1955).
22. W. S. Stiles and J. M. Burch, "N. P. L. colour-matching investigation: Final report (1958)," Opt. Acta **6**(1), 1–26 (1959).
23. N. I. Speranskaya, "Determination of spectrum color coordinates for twenty-seven normal observers," Opt. Spectrosc. **7**, 424–428 (1959).
24. CIE, "Official Recommendations, Committee E-1.3.1 – Colorimetry," in *Compte rendu, Quinzième session; Vienne; Juin 1963*, Commission Internationale de l'Éclairage, ed., Publication CIE 11-64, vol. A (Bureau Central de la Commission, 1964), p. 35.
25. D. B. Judd, "Judd's method for calculating the tristimulus values of the CIE 10° observer," in *Proceedings of the CIE Symposium '93 on Advanced Colorimetry*, J. Schanda, *et al.*, eds., Publication CIE x007-1993 (CIE Central Bureau, 1993), pp. 107–114.
26. A. Stockman and L. T. Sharpe, "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype," Vision Res. **40**(13), 1711–1737 (2000).
27. CIE, "Fundamental chromaticity diagram with physiological axes – Part 1," Tech. Rep. CIE 170-1:2006 (CIE Central Bureau, 2006). ISBN: 978-3-901906-46-6.
28. J. Li, P. Hanselaer, and K. A. G. Smet, "Impact of color-matching primaries on observer matching: Part II – Observer variability," Leukos **18**(2), 127–144 (2022).
29. C. Shen, R. Wanat, J. J. Yoo, *et al.*, "Measuring and modeling display observer metamerism," Vis. Comput. **38**(9-10), 3301–3310 (2022).

30. J. M. Artigas, A. Felipe, A. Navea, *et al.*, "Spectral transmission of the human crystalline lens in adult and elderly persons: color and total transmission of visible light," Invest. Ophthalmol. Vis. Sci. **53**(7), 4076–4084 (2012).
31. J. Pokorny, V. C. Smith, and M. Lutze, "Aging of the human lens," Appl. Opt. **26**(8), 1437–1440 (1987).
32. M. Lutze and G. H. Bresnick, "Lenses of diabetic patients 'yellow' at an accelerated rate similar to older normals," Investigative Ophthalmology and Visual Science **32**(1), 194–199 (1991).
33. V. C. Smith, J. Pokorny, and S. J. Starr, "Variability of color mixture data—I. Interobserver variability in the unit coordinates," Vision Res. **16**(10), 1087–1094 (1976).
34. 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 Res. **16**(10), 1095–1098 (1976).
35. P. B. M. Thomas, M. A. Formankiewicz, and J. D. Mollon, "The effect of photopigment optical density on the color vision of the anomalous trichromat," Vision Res. **51**(20), 2224–2233 (2011).
36. A. Stockman, L. T. Sharpe, S. Merbs, *et al.*, "Spectral sensitivities of human cone visual pigments determined *in vivo* and *in vitro*," in *Methods in Enzymology*, vol. 316 (Elsevier, 2000), pp. 626–650.
37. Beer, "Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten," Ann. Phys. Chem. **162**(5), 78–88 (1852).
38. R. S. Berns, *Billmeyer and Saltzman's Principles of Color Technology* (John Wiley & Sons, 2019).
39. H. Grassmann, "Zur Theorie der Farbenmischung," Ann. Phys. Chem. **165**(5), 69–84 (1853). [English version: "On the theory of compound colours", Philos. Mag. Ser. 4 **7**, 254–264 (1854)].
40. D. H. Krantz, "Color measurement and color theory: I. Representation theorem for Grassmann structures," J. Math. Psychol. **12**(3), 283–303 (1975).
41. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (John Wiley & Sons, 1982).
42. P. Trezona, "Rod participation in the 'blue' mechanism and its effect on colour matching," Vision Res. **10**(4), 317–332 (1970).
43. S. L. Buck, "Influence of rod signals on hue perception: evidence from successive scotopic contrast," Vision Res. **37**(10), 1295–1301 (1997).
44. S. L. Buck, R. Knight, G. Fowler, *et al.*, "Rod influence on hue-scaling functions," Vision Res. **38**(21), 3259–3263 (1998).
45. W. S. Stiles and G. Wyszecki, "Field trials of color-mixture functions," J. Opt. Soc. Am. **52**(1), 58–75 (1962).
46. R. D. Lozano and D. A. Palmer, "The additivity of large-field colour matching functions," Vision Res. **7**(11-12), 929–939 (1967).
47. Q. Zaidi, "Adaptation and color matching," Vision Res. **26**(12), 1925–1938 (1986).
48. W. A. Thornton, "Toward a more accurate and extensible colorimetry. Part III. Discussion (continued)," Color Res. Appl. **17**(4), 240–262 (1992).
49. 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 Res. Appl. **33**(4), 271–281 (2008).
50. C. A. Curcio, K. R. Sloan, R. E. Kalina, *et al.*, "Human photoreceptor topography," J. Comp. Neurol. **292**(4), 497–523 (1990).
51. A. Stockman, L. T. Sharpe, and C. Fach, "The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches," Vision Res. **39**(17), 2901–2927 (1999).
52. J. Li, P. Hanselaer, and K. A. G. Smet, "Impact of matching field size on color matching (functions) accuracy," Color Res. Appl. **48**(1), 88–102 (2023).
53. B. H. Crawford, "Colour matching and adaptation," Vision Res. **5**(1-3), 71–78 (1965).
54. R. Lozano and D. Palmer, "Large-field color matching and adaptation," J. Opt. Soc. Am. **58**(12), 1653–1656 (1968).
55. R. M. Boynton, M. M. Hayhoe, and D. I. A. MacLeod, "The gap effect: chromatic and achromatic visual discrimination as affected by field separation," Opt. Acta **24**(2), 159–177 (1977).
56. F. Viénot, "Relations between inter- and intra-individual variability of color-matching functions. Experimental results," J. Opt. Soc. Am. **70**(12), 1476–1483 (1980).
57. K. Shi, M. R. Luo, A. T. Rider, *et al.*, "A multi-primary trichromator to derive individual color matching functions and cone spectral sensitivities," Color Res. Appl. **49**(5), 449–464 (2024).
58. K. Shi, M. R. Luo, A. T. Rider, *et al.*, "Individual differences in color matches and cone spectral sensitivities in 51 young adults," Opt. Express **32**(13), 23597–23616 (2024).
59. R. M. Boynton, "Implications of the minimally distinct border," J. Opt. Soc. Am. **63**(9), 1037–1043 (1973).
60. R. T. Eskew Jr., "The gap effect revisited: Slow changes in chromatic sensitivity as affected by luminance and chromatic borders," Vision Res. **29**(6), 717–729 (1989).
61. S. Bęczkowski and S. Munk-Nielsen, "LED spectral and power characteristics under hybrid PWM/AM dimming strategy," in *2010 IEEE Energy Conversion Congress and Exposition* (Institute of Electrical and Electronics Engineers, 2010), pp. 731–735.
62. CIE, "Visual aspects of time-modulated lighting systems," Tech. Rep. CIE 249:2022 (CIE Central Bureau, 2022). ISBN: 978-3-902842-68-8.
63. A. Stockman and L. T. Sharpe, "Into the twilight zone: the complexities of mesopic vision and luminous efficiency," Ophthal. Phys. Opt. **26**(3), 225–239 (2006).

64. M. Dyble, N. Narendran, A. Bierman, *et al.*, "Impact of dimming white LEDs: chromaticity shifts due to different dimming methods," Proc. SPIE **5941**, 291–299 (2005).

65. J. L. Davis, K. C. Mills, G. Bobashev, *et al.*, "Understanding chromaticity shifts in LED devices through analytical models," Microelectron. Reliab. **84**, 149–156 (2018).

66. J. Krauskopf and G. Karl, "Color discrimination and adaptation," Vision Res. **32**(11), 2165–2175 (1992).

67. A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application* (Cambridge University, 1997).

68. T. J. DiCiccio and B. Efron, "Bootstrap confidence intervals," Statistical Sci. **11**(3), 189–228 (1996).

69. H. Leibowitz, "The use and calibration of the 'Maxwellian view' in visual instrumentation," Am. J. Psychol. **67**(3), 530–532 (1954).

70. P. W. Trezona, "Individual observer data for the 1955 Stiles–Burch 2° pilot investigation," J. Opt. Soc. Am. A **4**(4), 769–782 (1987).