



# Measuring and modeling display observer metamerism

Che Shen<sup>1</sup> · Robert Wanat<sup>2</sup> · Jang Jin Yoo<sup>3</sup> · Junwoo Jang<sup>3</sup> · Mark D. Fairchild<sup>1</sup>

Accepted: 18 May 2022 / Published online: 22 June 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022, corrected publication 2022

## Abstract

Observer metamerism refers to a situation in which some observers with normal color vision will see two colors as identical while others will see them mismatch. In recent years, some issues with metameric failure have been exacerbated by displays with more saturated primaries and narrower emission spectra. Classic trichromatic colorimetry is incapable of predicting this effect because of the typical reliance on a single standard observer—a single color matching function (CMF) averaged across the population. In this paper, we present a new experiment aimed at measuring the amount of metameric failure of a single display for real observers without comparisons to other reference displays. We also propose a recommended metric of observer metamerism failure potential that uses simulated observers, represented as individual CMFs that match the distribution of color sensitivities in the general population, and compare with several other procedures.

**Keywords** Observer metamerism · Metric · Individual color matching function

## 1 Introduction

### 1.1 Observer metamerism in displays

Metamerism is a property of the human visual system (HVS) that allows colors with different spectra to visually match as long as they cause the same responses in the HVS. Such spectra are known as metamers or metameric pairs. Metamerism is caused by the fact that each of the three cone cell types (long-, medium- and short-wavelength sensitive) in the eye can be excited by photons spanning a wide range of wavelengths, with the probability of photon capture determined by the optical, chemical and physiological properties of individual components of the HVS. As long as the different spectra cause the same excitations in all cone types, they will appear the same under the same viewing conditions. Metamerism allows display devices to reproduce colors accurately without requiring identical spectral power distributions (SPDs) between the reference color and its reproduction. To determine which spectra are metamers, color matching functions (CMFs) are used that map spectral radiance of perceived light

into a trichromatic representation. The most commonly used CMFs were standardized by the International Commission on Illumination (CIE) and became known as the CIE 1931 Standard Colorimetric Observer. It represents the results of multiple color matching experiments in a two-degree bipartite field, averaged across observers.

By virtue of representing a mean response of a population, the CIE 1931 standard observer does not necessarily match the color sensitivity of any person and it is agnostic to perceptual differences between individuals. As a result, changes to color perception introduced by factors such as age, genetic mutations, and mechanical damage cannot be predicted by the model, resulting in a situation where two spectra predicted to match by the standard observer can appear as different colors to a given human observer. This metameric failure is more common when colors are reproduced using narrow-band primaries. Recently, the push within the display industry to achieve wide color gamut, such as the ITU-R Recommendation BT.2020, has resulted in the issues of individual color sensitivity differences, or observer metamerism (OM), to become more pronounced. The existence of multiple content deliverable targets, including high dynamic range (HDR) displays, standard dynamic range (SDR) displays and cinema projectors, forces content producers to evaluate and match content appearance across screens having vastly different spectral properties. This in turn causes an increasing

✉ Che Shen  
cs7607@g.rit.edu

<sup>1</sup> Rochester Institute of Technology, Rochester, USA

<sup>2</sup> LGE Zenith, Glenview, Illinois, USA

<sup>3</sup> LG Display Co., Ltd, Seoul, Korea

need for a model that can predict how likely a display is to cause significant metameric mismatches.

In 2006, CIE proposed a model [1] which outlines the prediction of fundamental cone sensitivities and related corresponding CMFs as a function of the age and field of view of an observer. Asano et al. [2] proposed a vision model for individual colorimetric observers by extending eight additional physiological parameters to the CIE 2006 model. Lens pigment density, macular pigment density, optical densities of L-, M-, and S-cone photopigments, and max shifts of L-, M-, and S-cone photopigments were all controlled by these eight factors. Most variation of individual CMFs among population can be represented by this model.

In this paper, we present an experiment aimed at measuring the degree of observer metamerism induced by display emission spectra in the general population. The experiment is designed to assess the feasibility of quantifying observer metamerism for a single display without reference to a standard display or other stimulus. We also propose three methods of estimating the same property using simulated observers, represented by CMFs proposed by Asano et al. [2], which model the distributions of likely sources of metamerism and their effect on a wide range of observers. The final result of the metrics is a representation of a statistical distribution of color differences that can be expected to occur when a color on the tested display is evaluated against a reference spectrum whose appearance it is meant to match.

## 1.2 Previous observer metamerism indices

While in color-critical applications the CMFs of individual observers can be measured and compensated for, this process is time-consuming and has to be repeated for each observer. In situations where approvals across multiple observers and devices are expected, it might be preferable to instead use displays that minimize the perceivable differences resulting from observer metamerism. Observer metamerism indices (OMIs) aim at predicting the distribution of color differences caused by observer metamerism in general population and can be used for the purpose of modeling their occurrence.

In 1989, CIE proposed the metamerism index as a method for evaluating the observer metamerism [3]. This approach, which was based on the data from Nayatani et al. [4] and Takahama et al. [5], was designed to define the CIE standard deviate observer and the metamerism index. The model was later found to underestimate the color matching variability among all observers [6]. Therefore, a more accurate representation of observer variability across large population had to be taken into consideration. Fairchild and Wyble [7] designed an OMI using a set of simulated observers based on the CIE 2006 CMFs. Using each of the generated CMFs to calculate a trichromatic representation of the emission spectra of a tested display, RGB values used to drive the tested

display were adjusted until a match was achieved between a reference spectrum and the color produced on the screen. The two color spectra, reference and test, were compared using the 1931 2° observer, and color difference metrics were used to calculate objective perceptual differences between them. Statistics of color difference dispersion were then used as the measure of observer metamerism in displays, with maximum difference and mean difference representing the worst and average scenario that can be expected in a real-world application.

Long and Fairchild [8] used a set of simulated observer CMFs, following a specification by Fedutina et al. [9] to create matches with a large number of surface color spectra. The spectra were calculated by simulating the SPDs of standardized color patches from Macbeth color charts, Munsell samples, Kodak/AMPAS test patches and metamerism test patches, illuminated by standardized light sources (CIE illuminants D65, A and F2 as well as a motion picture studio lamp). Matching colors were then reproduced on tested displays, with 1931 2° observer used as the basis for the color match calculation. The spectra of the matched colors were then evaluated using the simulated CMFs, and color differences were calculated between the reference spectrum and the matched spectrum for each such CMF. The authors presented the same statistics ( $OM_{\max}$ ) as Fairchild and Wyble [7] but also added the volume of an ellipsoid containing all of the matched spectra, represented in CIE  $L^*a^*b^*$  color space, as another metric of metameric match dispersion representative of observer metamerism. The  $OM_{\max}$  can be calculated based on Eq. 1:

$$OM_{\max} = \max(\Delta E_y, P, i) \quad (1)$$

where  $OM_x$  refers to OMI based on specific CMF sets. Color difference values ( $\Delta E_y$ ) between a reference color and test stimulus were calculated based on  $\Delta E_{ab}$ ,  $\Delta E_{94}$  or  $\Delta E_{00}$  for each patch in a patchset P for each observer i in the CMF set. Thus, this OMI represents the worst color match across all tested patches and observers.

Xie et al. [10] proposed an improved OM index ( $POM_2$ ), which was revised from  $OM_x$  following a psychophysical experiment.  $POM_2$  can be expressed as Eq. 2:

$$POM_2 = \text{Count}(\Delta E_{00} > 2) / M \quad (2)$$

where  $M$  is the number of observers. The  $POM_2$  represents the percentage of perceived mismatch larger than a detection threshold ( $\Delta E_{00}$  of 2) from observer.

### 1.3 Approach for single-display psychophysics and metric

A common problem with experiments trying to establish the amount of OM induced by a display is that any such attempt relies on the use of a reference color, usually provided on a separate display or as a hardcopy, whose appearance is matched on a tested display. However, the test spectra used as reference vary between the experimental setups and their influence affects the results of the test, reducing its generality. The chosen spectrum can be considered to act as a bias in the experiment, and therefore, the results from different experiments are not interchangeable.

Attempts have been made to design experiments without providing a reference, but the dynamic adjustment of the test content results in a floating observer adaptation state that continues to change how the test and reference patch look in contrast to each other. In this paper, an experimental design is presented where observers are tasked with selecting unique hue colors on display. Its aim is to bridge the two goals of eliminating the need for a reference but still allowing for an experimental task that is not made more difficult by a fluctuating observer adaptation.

### 1.4 Unique hues as an intrinsic reference

It is well known that there are two stages of color processing in human visual system. The first stage (trichromacy) is represented by absorption of light in long-, medium-, and short-wavelength-sensitive (L, M, S) cone receptors [11]. The next stage (opponency) is represented by combination of cone outputs in post-receptor channels [12].

Unique hues, first proposed by Hering [13], represent a concept in color science that some hues appear ‘pure,’ meaning they contain no admixtures of other hues. Unique hues are not simply linked by the opponent channels found in the early stages of vision, as Wuerger [14] pointed out that the encoding of unique hues happens at high-order color mechanisms. In addition, Webster [15] found the variations of selecting unique hue are largely independent of the sensitivity (CMFs) differences across observers which also confirmed that the interobserver variation of selecting unique hues is not only related to observer metamerism but also impacted by other visual mechanisms. The authors came up with a hypothesis that the observer variability when selecting unique hues in display are related to three aspects: observer noise, unique hue difference across observer and observer metamerism. We assume the observer noise as well as the unique hue difference across observer are the same across all tested displays, and any increase of observer variability across different display is due to observer metamerism. As such, these hues can be considered as an intrinsic reference for experiment participants.

We used that property to design an experiment where the observers can maintain a floating adaptation state without making the experimental task too difficult to complete and where there was no need to present the participants with a reference to match. At the same time, we expected that individual differences in color perception would still influence results sufficiently to serve as a metric of metameric stability of a display—OMI.

## 2 Unique hues experiment

A method of adjustment psychophysical experiment was conducted in which the observer task was to select unique hue colors on a tested display. In addition, observers were also required to choose four unique hue colors among Farnsworth-Munsell 100 (FM-100) color plates under a simulated D65 lighting condition.

### 2.1 Display characterization

The stimuli for the unique hues experiment were shown on four displays. The displays used in the experiment were selected because they use different technologies with a disparity in their emission spectra causing varying degrees of OM. Display1 was an OLED TV, Display2 used Mini-LEDs, Display3 was an LCD display, and Display4 was a QLED display. After an hour warm-up time, the MATLAB graphical user interface (GUI) was used to display square test color patches on all monitors. The patches were showing maximum red, green, and blue values and gray patches sampled in RGB space from 0 to 95 with an increment of 5 and from 105 to 255 with an increment of 10. Konica Minolta CS2000 spectroradiometer was used to measure all the test colors in a dark environment. Figure 1 demonstrates the spectral power distributions of white points of the individual displays. 3D look-up tables (LUTs) were used to characterize the displays [16]. The results of the characterization were then verified

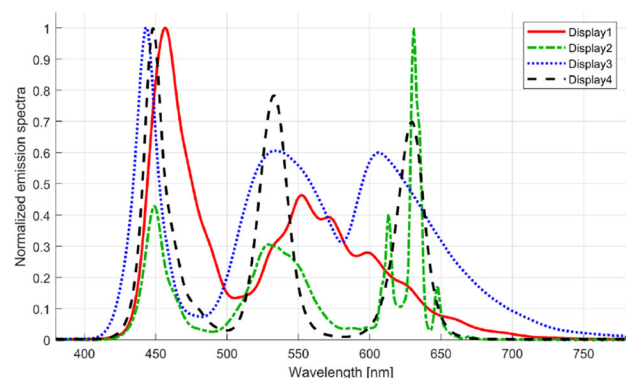


Fig. 1 SPDs of the four displays' white point

**Table 1** Characterization accuracy (CIEE2000) of the LUT model for each display

Statistic	Display1	Display2	Display3	Display4
Mean	0.54	0.35	2.05	8.61
Minimum	0.14	0.04	0.38	0.28
Maximum	1.04	1.21	4.92	38.70

using a set of test patches. Table 1 shows the minimum, mean, and maximum CIE  $\Delta E_{00}$  prediction accuracy measured on the four displays for 216 random verification colors.

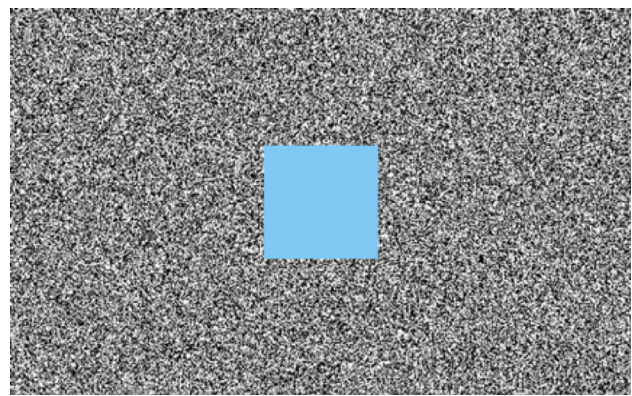
The characterization results all met the accuracy level of our subsequent psychophysical testing, with the exception of the Display4, which is expected of a consumer grade display. However, this LUT was still used when rendering colors in the experiment. To ensure the accuracy of the experimental results, unique hues chosen on Display4 by the experiment participants were measured with a spectroradiometer during the experiment rather than relying on the characterization results.

## 2.2 Unique hues as an intrinsic reference

The experiment was carried out in a dim environment with D65 ambient light source positioned behind the displays. This was determined to help stabilize adaptation and improve observer precision in pilot experiments. The experiment user interface (UI), which was designed in MATLAB, is shown in Fig. 2. Each stimulus was displayed in the center of the screen, overlaid over a luminance noise background. The observers were seated in front of the display at a distance such that each color stimulus covered  $2^\circ$  of visual angle.

In the first part of the experiment, observers were asked to select unique hues on the display. Using the arrow keys on the keyboard, viewers adjusted the hue angle of a randomized color patch, with a fixed lightness and chroma value, until the desired unique hue was achieved. Table 2 lists the chroma ( $C^*$  from CIELAB  $L^*C^*h$ ) and lightness ( $L^*$  from CIELAB) values of the color patches shown on each display. Two chroma ( $C^*$ ) values, high and low, were created for Display1 in order to verify the relationship between chroma and unique hue angle. Each observer made 36 selections—4 different hue hues each with 9 repetitions.

In the second part of the experiment, observers selected unique hue colors from a set of physical samples. The observers selected four unique hues from Farnsworth Munsell 100 (FM-100) hue test for color vision deficiencies. The color plates were viewed under simulated D65 lighting in a Macbeth Spectralight booth, with each observer making 36 selections—4 unique hue colors each with 9 repetitions. The reflectance spectra of FM-100 samples are shown in Fig. 3. This phase of the experiment was included to provide results

**Fig. 2** The setup of the psychophysical experiment. Color stimuli were shown on the display against a noise background. Observers were sitting in front of the display and using keyboard to change the color. The test patches spanned visual angle of  $2^\circ$ **Table 2** Chroma and lightness value of five different display experiment sessions

	Red ( $C^*/L^*$ )	Green ( $C^*/L^*$ )	Yellow ( $C^*/L^*$ )	Blue ( $C^*/L^*$ )
Display1 (HC)	90/53	60/80	80/75	40/74
Display1 (LC)	25/53	25/80	25/75	15/74
Display2	90/53	60/80	80/75	40/74
Display3	80/58	60/80	60/86	40/67
Display4	75/50	30/85	75/80	30/85

$C^*$  and  $L^*$  are CIELAB  $L^*C^*h$  chroma and lightness, respectively, relative to the actual display white points (nominally D65). *HC* High chroma, *LC* Low chroma

with broadband spectra that are more relevant to colors that can be shown on modern display stimuli.

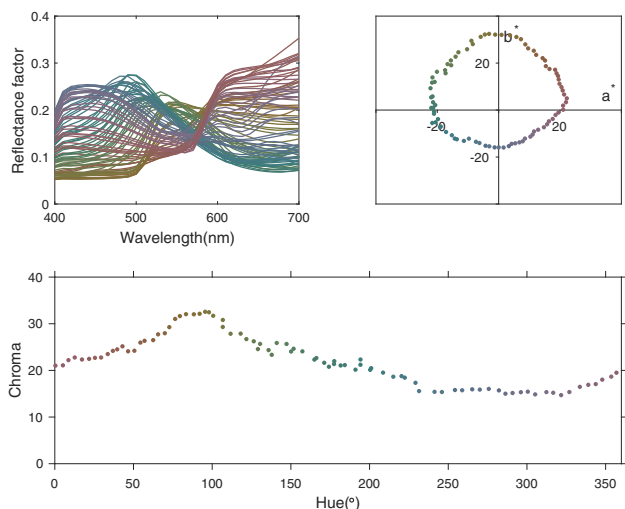
## 2.3 Observers

Ten observers (four females and six males) participated in the experiment. One observer was unable to complete the psychophysical experiment on Display4. All observers were evaluated to have normal color vision. Four of the ten observers were color science experts, and the others were naïve observers.

## 3 Experimental results

The results of the experiment were analyzed to assess whether the experimental procedure is sufficiently robust to measure the amount of observer metamerism expected in a display.





**Fig. 3** Reflectance spectra of FM-100 samples (top left) and their CIELAB  $L^*a^*b^*$  coordinates projected onto  $a^*b^*$  chromaticity plane (top right). The bottom plot shows the chroma value of these samples

### 3.1 Data

The violin plots of unique hue selections across displays are shown in Fig. 4. Each subplot in Fig. 4 represents a unique hue. Our initial goal was to find a clear relationship between display type and observer variation. For example, observer variation on Display4 is larger than Display3 because QLED display with a narrow emission spectrum is expected to show a higher rate of observer metamerism than a broad emission spectrum LCD. However, in order to use individual differences in unique hue perception as a metric of metameric stability of a display, a clear relationship has to be established between measured observer variation and different display technologies. We conducted additional analysis to establish how much of the result variation could be explained by the observer noise.

### 3.2 Intra and inter observer variation

Two types of intra-observer variation were considered. One is the intra-observer variation per display (Intra-PD), which represents the observer noise in unique hue selection for a single display, or the noise from the observer selecting the unique hues inconsistently. The other is intra-observer variation across displays (Intra-AD), representing the noise averaged across all tested displays. There are also two types of interobserver variation: interobserver variation per display (Inter-PD) which represents the observer difference in unique hue perception on a specific display caused by individual differences in perception and interobserver variation across displays (Inter-AD), representing average differences across all displays.

Based on our hypothesis, any increase of Intra-AD (compared to Intra-PD) and Inter-AD (compared to Inter-PD) is due to metamerism. Therefore, We could use the difference either between Intra-AD and Intra-PD or Inter-AD and Inter-PD to serve as a OMI to evaluate single display. However, our observers’ data are not consistent with the ideal situation. Both intra- and interobserver variations are calculated based on Eqs. 3–6.

$$\text{Intra-PD} = \frac{\sum_{i=1}^{N_{\text{obs}}} \sigma(h_{i,k})}{N_{\text{obs}}} \tag{3}$$

$$\text{Inter-PD} = \sigma[\overline{h_{1,k}}, \dots, \overline{h_{N_{\text{obs}},k}}] \tag{4}$$

$$\text{Intra-AD} = \frac{\sum_{i=1}^{N_{\text{obs}}} \sigma(h_{i,\text{all}})}{N_{\text{obs}}} \tag{5}$$

$$\text{Inter-AD} = \sigma[\overline{h_{1,k}}, \dots, \overline{h_{N_{\text{obs}},\text{all}}}] \tag{6}$$

The symbol of the standard deviation of variable (hue angle) is “ $\sigma$ ”;  $h_{i,k}$  is hue angle chosen by observer  $i$  on display  $k$ ,  $N_{\text{obs}} = 9$ ,  $h_{i, \text{all}}$  is the hue angle of observer  $i$  averaged across all displays.

Intra- and interobserver variation of our experiment data is shown in Fig. 5. Each subplot represents each unique hue, respectively. As can be seen in Fig. 5, in most cases Intra-PD is larger than Intra-AD and Inter-PD is larger than Inter-AD. In other words, the differences from the same observer conducting the same experiment multiple times are higher than the differences measured between the different observers.

We also conducted a one-side t-test for each unique hue to test the null hypothesis that the Intra-AD minus Intra-PD (or Inter-AD minus Inter-PD) is greater than or equal to zero. The alternative hypothesis is Intra-AD minus Intra-PD (or Inter-AD minus Inter-PD) less than zero. The null hypothesis has been rejected for all test samples at a 5% significance level.

### 3.3 Conclusions from the experiment

This study has identified that the within-individual variation in setting unique hue in this experiment was larger than the variation between observers, which suggests that the proposed approach cannot be used to measure the added variation due to observer metamerism. To use this technique, we need the individual variation in the setting to be lower than the variation between observers. Because the opposite is true, it is impossible the separate the experimental noise from the

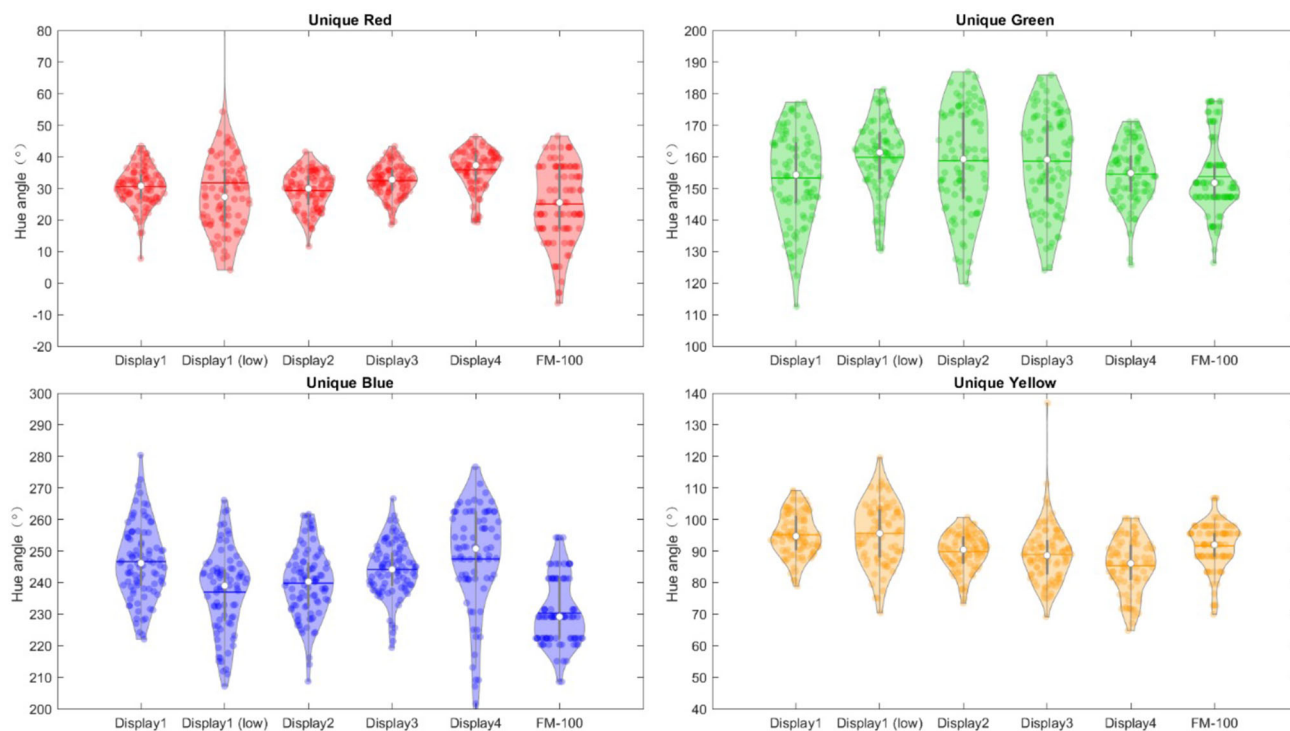


Fig. 4 Violin plots of the hue angle from CIE  $L^*C^*h$  for each unique hue selection

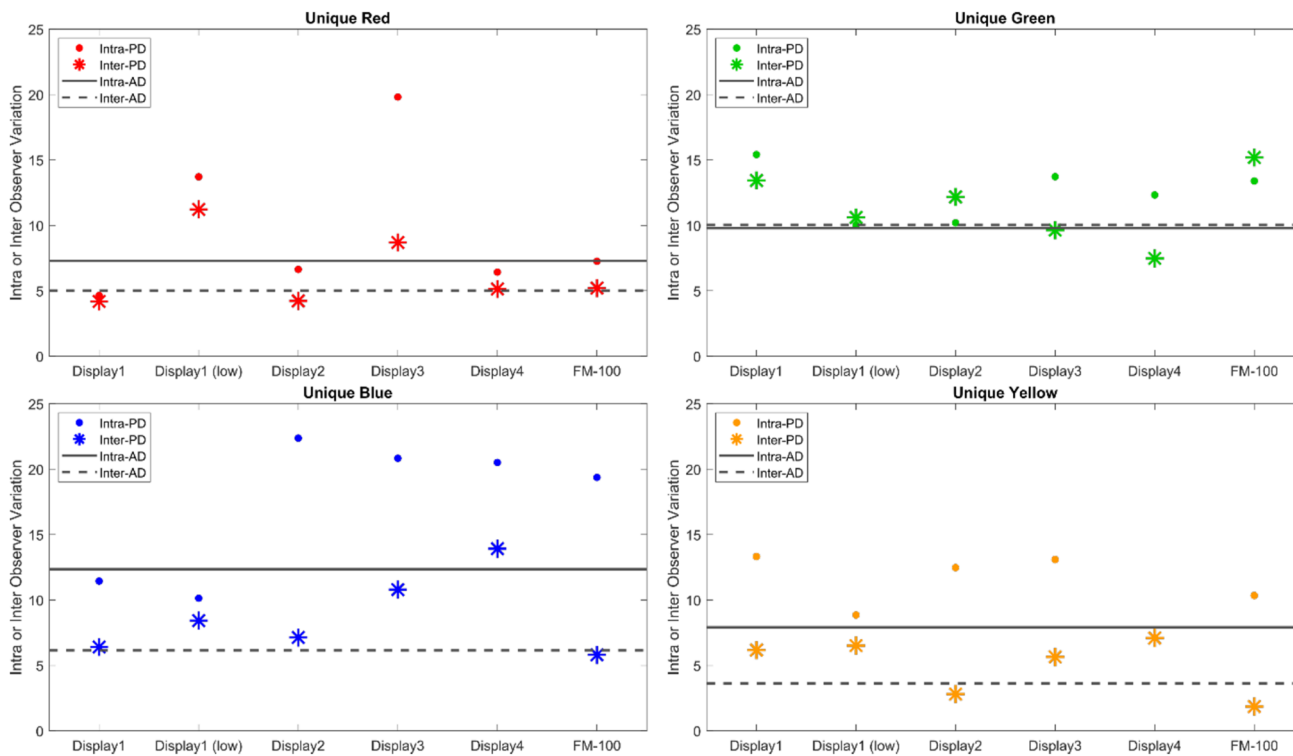
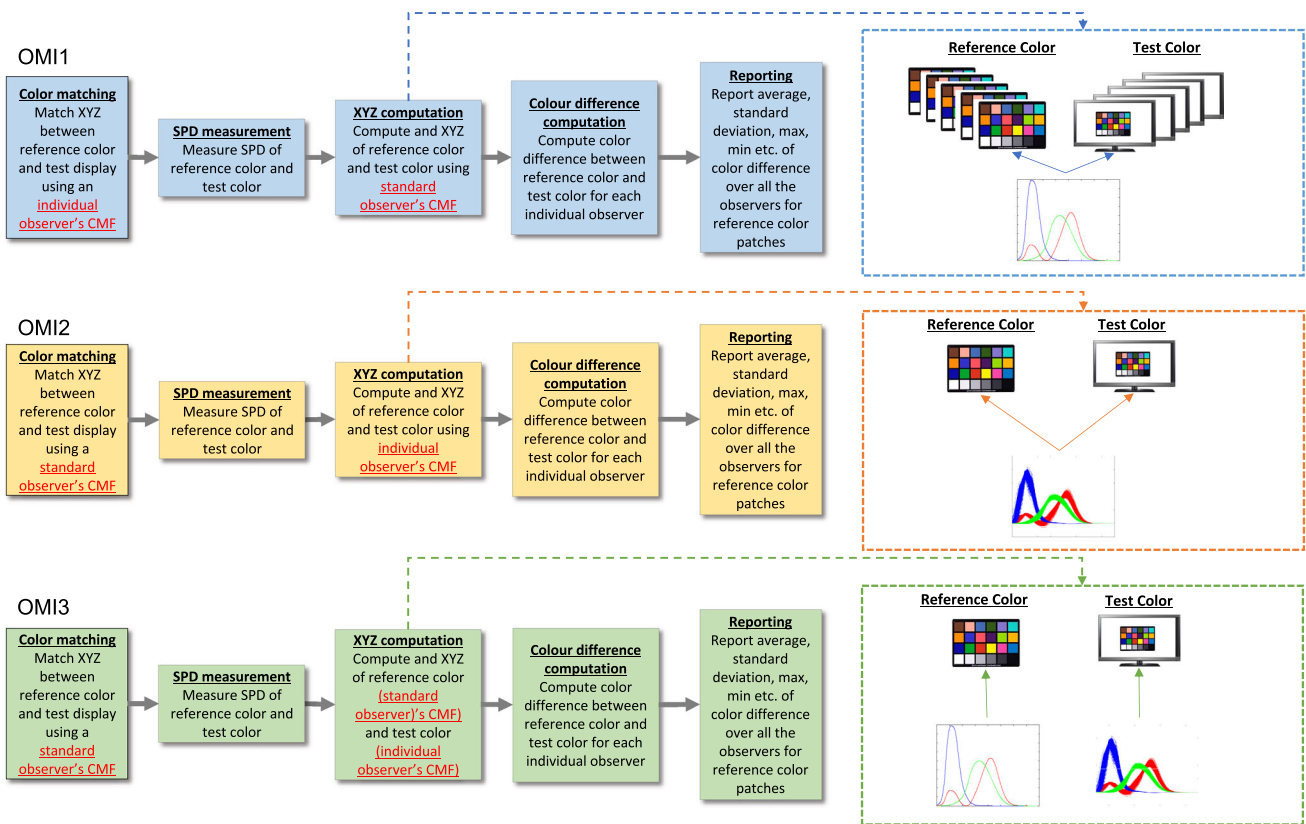


Fig. 5 Results of the variation analysis of the data from the psychophysical experiments. Results are shown per-display (PD, stars and dots) and averaged across displays (AD, continuous and dashed lines). In

almost all cases, intra-observer variations are higher than the interobserver variations, indicating that the experimental noise is higher than the individual perception differences measured by our method



**Fig. 6** Flowchart diagram of three tested OMI methods. While the individual steps remain the same, the differences (marked in red and underlined in the diagrams) lie in the CMFs used in the color matching and XYZ computation steps

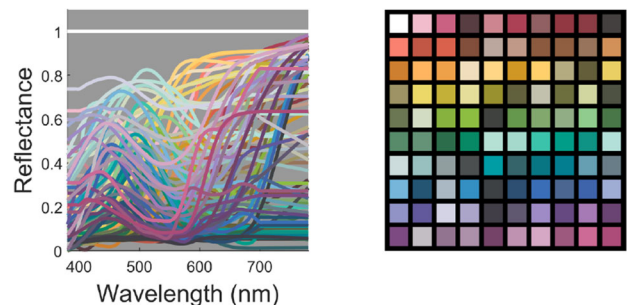
individual observer differences. Based on our results, detecting observer metamerism in chromatic colors when displays are viewed individually with no reference stimuli remains a difficult task.

### 4 Observer metamerism indices

In this section, we describe several proposed OMIs that are intended to serve as descriptors of the propensity of a display to cause observer metamerism.

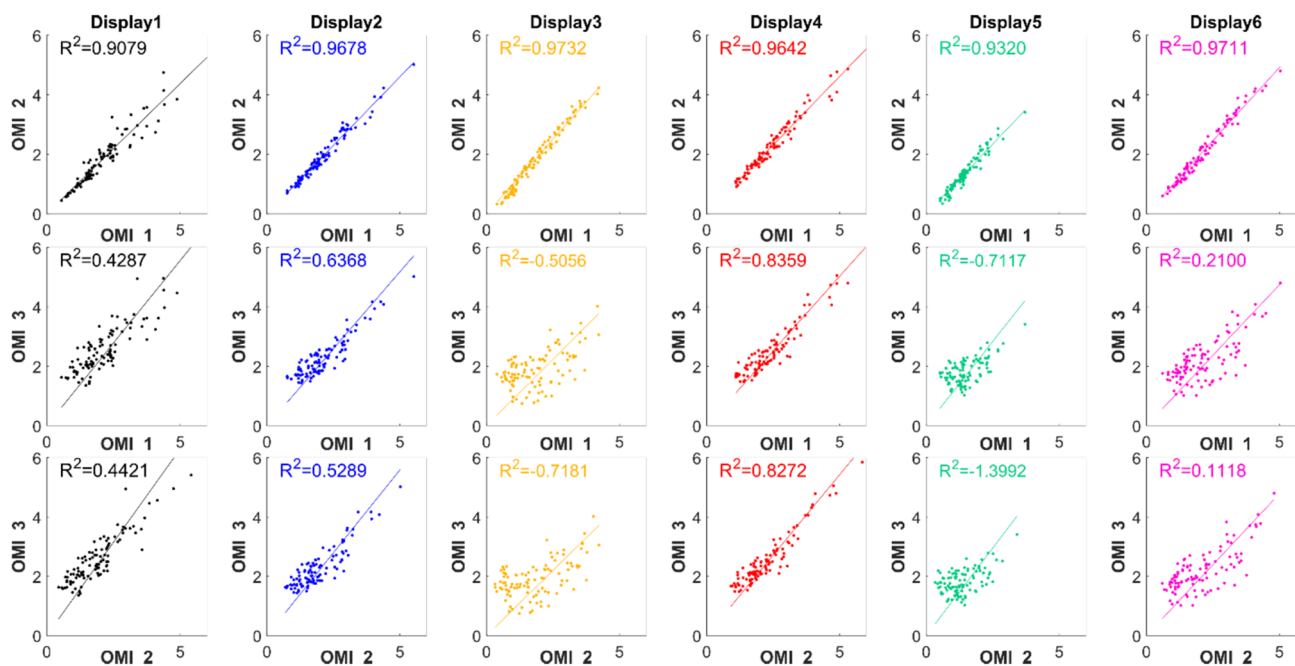
#### 4.1 Proposed methods

In this paper, three methods of evaluating OMI of a display are described, as shown in Fig. 6. Two of them, OMI1 and OMI2, are similar to two metrics proposed in the past by Park and Murdoch [17], except using spectra of real objects as reference metamers. The 100 reference spectra of real objects (shown in Fig 7) were selected from David et al. [18]. In addition to the four displays that were used in the experiment, two other displays were also included (Display5: LCD and Display6: QLED display).



**Fig. 7** Reflectance spectra of 100 real objects from David et al. [18]

In OMI1, 1000 individual observer CMFs from Asano model [2] were chosen to create a set of metamers, matching all 100 reference spectra for each simulated observer. CIE 1931 2° standard observer was used to calculate the trichromatic representation of each color and color difference metrics were calculated between the reference spectrum and the matched spectrum. All color differences were calculated for one standard observer in OMI1, which is most commonly used method in OMI research. This procedure simulates a situation where each of the 1000 individual observers adjusts the RGB drives on the tested display to achieve a perceptual match to the reference spectrum and then a standard light



**Fig. 8** Relationship between OMIs of six tested displays. Note the  $R$ -squared correlation value can be negative as a result of setting a specific intercept. In our model, the intercept is set at (0,0) because of the

measurement device (LMD) calculates the 1931 XYZ representation of the colors, used to calculate color difference metrics for a standard observer. However, OMI1 is complex and time-consuming because the perceived match has to be calculated for each individual observer.

In OMI2, CIE 1931  $2^\circ$  standard observer CMF is used to generate the set of metamers, requiring a set 100 matches per display. Individual observer CMFs from Asano [2] are then used to calculate color differences. This procedure simulates the situation where the display is first calibrated to match the test spectra using a standard LMD and then the observers were asked to judge to amount of perceived mismatch between the reference spectrum and the color produced by tested display. OMI2 is simpler because individual color matching process is omitted; instead, only one match is needed per display per spectrum. This method results in non-standard color difference calculation but previous research [17] and our simulation results (see Figures 10,11) both show the highly linear relationship between OMI1 and OMI2.

In OMI3, CIE 1931  $2^\circ$  standard observer CMF was also used to create the set of metamers but the color difference values were calculated between tristimulus values from reference color using standard observer CMF and tristimulus values from test color using individual observer CMFs. This method also results in nonstandard color difference value but it is even less computationally expensive than OMI2 so it was included in the comparison.

assumption that if there is no observer metamerism, both OMI results should be 0. The negative  $R$ -squared still represents worst fitting results

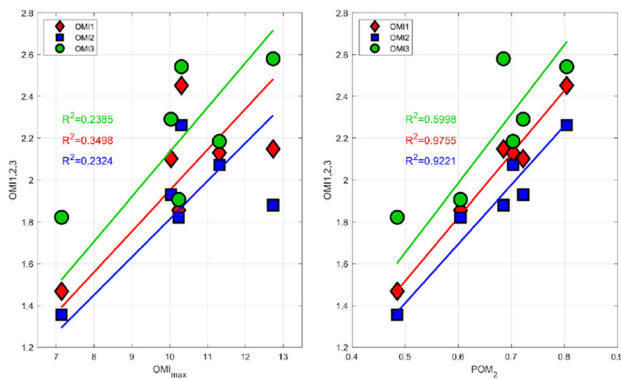
## 4.2 Justification of colorimetry of proposed method

Figure 8 shows the relationship between OMI1, OMI2 and OMI3. Each color dot in Fig. 8 represents the average color difference value (CIE  $\Delta E_{00}$ ) between one reference color patch and its metamers. As mentioned earlier, OMI1 is highly correlated with OMI2. However, OMI3 is not correlated well with either OMI1 and OMI3 (especially in Display5) compared with the relationship between OMI1 and OMI2.

## 4.3 Comparison of proposed method with other OMI suggestions

To assess the performance of the tested OMIs, two different previous OMIs ( $OM_{max}$  and  $POM_2$ ) were selected for comparison. These previous OMIs can only represent the OM effect between two real displays, one test display and one reference display. These metrics were still used by using real world object spectra from David et al. [18] as a reference display emission spectrum. The assumption is that very low observer metamerism effect can be expected in real world objects because their reflectance spectra are smooth. If that assumption is correct then both  $OM_{max}$  and  $POM_2$  between the smooth spectrum and its reproduction on the display are entirely attributable to the observer metamerism caused by the tested display. The evaluated methods were calculated based on mean value of color difference metrics from all





**Fig. 9** Plots showing the correlation the three tested OMIs with previous OMIs (left:  $OMI_{max}$ , right:  $POM_2$ ) across 6 displays. Each point (circle, square or diamond) represents an OMI value of a single display

color matching patches, a pair for each of 1000 observers and 100 colors. As confirmed by early research [10],  $POM_2$  performs better than  $OMI_{max}$  because  $OMI_{max}$  cannot represent the average performance of display. Figure 9 illustrates the correlation between previous OMIs and the OMIs tested in this paper. It can be seen that the evaluated OMIs are not correlated well with  $OMI_{max}$  while the  $POM_2$  is highly correlated with OMI1 and OMI2.

#### 4.4 Recommendations

OMI1 is the theoretically correct and best performing metric to evaluate the OM effect. In practice, OMI1 is time-consuming and complicated. OMI2 generates nonstandard color difference values but it is less computationally expensive and statistically close to OMI1. OMI3 is not highly correlated with  $POM_2$  which has been shown to align well with psychophysical data. Therefore, OMI2 is recommended as a general-purpose OMI.

First, there is significant interobserver variation in the unique hue setting. This is not a new or unexpected result, but our experiment results confirm that. In this experiment, the Intra-PD was larger than Intra-AD and Inter-PD was larger than Inter-AD. Thus, we were unable to apply this method to assess the additional variance caused by individual variations in CMFs. As mentioned before, to use our proposed method, the observer noise should be lower than interobserver variation. Based on our results, we cannot use this method to evaluate the OM because the variation due to observer metamerism is overwhelmed by observer noise.

This paper also described three methods to modeling OM. OMI2, with metamers generated by standard observer CMF and color difference value calculated from individual observers CMF, is recommended by authors. OMI2 also correlates well with previous OMI ( $POM_2$ ).

## 5 Conclusions

Measuring and modeling observer metamerism effect for a display is significantly important in display industry. This paper examined whether unique hue method can be used to measure OM in displays. We found that the methodology was too noisy to produce a clear answer but some analysis results is also useful.

## Declarations

**Conflict of interests** All authors declare that they have no conflicts of interest

## References

1. CIE: Fundamental Chromaticity Diagram with Physiological Axes-Part I. CIE Pub. 170-1:2006. CIE, Vienna (2006)
2. Asano, Y., Fairchild, M.D., Blondé, L.: Individual colorimetric observer model. *PloS one* **11**(2), e0145671 (2016)
3. CIE: Special Metamerism Index: Change in Observer, vol. 80. CIE Pub., Vienna (1989)
4. Nayatani, Y., Hashimoto, K., Takahama, K., Sobagaki, H.: Comparison of methods for assessing observer metamerism. *Color Res. Appl.* **10**(3), 147–155 (1985)
5. Takahama, K., Sobagaki, H., Nayatani, Y.: Prediction of observer variation in estimating colorimetric values. *Color Res. Appl.* **10**(2), 106–117 (1985)
6. Berns, R.S.: Billmeyer and Saltzman's Principles of Color Technology. Wiley, New York (2019)
7. Fairchild, M.D., Wyble, D.R.: Mean observer metamerism and the selection of display primaries. In: Color and Imaging Conference, vol. 2007, no. 1. Society for Imaging Science and Technology, pp. 151–156 (2007)
8. Long, D.L., Fairchild, M.D.: Modeling observer variability and metamerism failure in electronic color displays. In: Color and Imaging Conference, vol. 2014, no. 2014). Society for Imaging Science and Technology, pp. 14–27 (2014)
9. Fedutina, M., Sarkar, A., Urban, P., Morvan, P.: (How) Do observer categories based on color matching functions affect the perception of small color differences?. In: Color and Imaging Conference, vol. 2011, no. 1. Society for Imaging Science and Technology, pp. 2–7 (2011)
10. Xie, H., Farnand, S.P., Murdoch, M.J.: Observer metamerism in commercial displays. *JOSA A* **37**(4), A61–A69 (2020)
11. Von Helmholtz, H.: Treatise on Physiological Optics, vol. 3. Courier Corporation, Chelmsford (2013)
12. Fairchild, M.D.: Unique hues and principal hues. *Color Res. Appl.* **43**(6), 804–809 (2018)
13. Hering, E.: Outlines of a Theory of the Light Sense. Harvard University Press, Cambridge (1964)
14. Wuerger, S.M., Atkinson, P., Cropper, S.: The cone inputs to the unique-hue mechanisms. *Vis. Res.* **45**(25–26), 3210–3223 (2005)
15. Webster, M.A., Miyahara, E., Malkoc, G., Raker, V.E.: Variations in normal color vision. II. Unique hues. *JOSA A* **17**(9), 1545–1555 (2000)
16. Day, E.A., Taplin, L., Berns, R.S.: Colorimetric characterization of a computer-controlled liquid crystal display. *Color Res. Appl.* **29**(5), 365–373 (2004)

17. Park, Y., Murdoch, M.J.: Efficiently evaluating the effect of color gamut and spectral bandwidth on observer metamerism in high dynamic range displays. *J. Soc. Inform. Disp.* **29**(9), 704–722 (2021)
18. David, A., Fini, P.T., Houser, K.W., Ohno, Y., Royer, M.P., Smet, K.A., Whitehead, L.: Development of the IES method for evaluating the color rendition of light sources. *Opt. Express* **23**(12), 15888–15906 (2015)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Che Shen** received his BS in Gem and Material Engineering from Hebei GEO University and his MS in Gemology from the China University of Geosciences. He is currently a Ph.D. student in the Munsell Color Science Laboratory at Rochester Institute of Technology, majoring in Color Science.



**Robert Wanat** is a staff researcher at LGE Zenith. His work revolves around creating models of the human visual system that simulate visual perception and how it's affected by changing viewing conditions. He graduated with a Ph.D. from Bangor University in Bangor, Wales, and an M.Sc. from the West Pomeranian University of Technology in Szczecin, Poland.



**Jang Jin Yoo** is currently a research fellow at LG Display. He received his BS and MS in chemistry and physics at Sogang University, Korea, in 1993 and 1996. He also studied color science at Leeds University, UK, for his Ph.D. He has more than 25 years of research experience in display technologies and display picture quality. Specialty areas include picture quality evaluation, enhancement, and international standardization.



**Junwoo Jang** is a senior research engineer at the image quality development department of LG Display. He received his BSc and MSc in electronic engineering from the Chung-Ang University of Seoul. He has been working on image quality enhancement and display metrology.



**Mark D. Fairchild** is a Professor at Rochester Institute of Technology and Director of the Program of Color Science and Munsell Color Science Laboratory. He received his B.S./M.S. degrees in Imaging Science from Rochester Institute of Technology and his M.A./Ph.D. degrees in Vision Science from the University of Rochester. He is a Fellow of IS&T and OSA.