An observer-metamerism sensitivity index for electronic displays

Steven Le Moan
Tejas Madan Tanksale
Roman Byshko
Philipp Urban

Abstract **—** The effect of observer metamerism induced by electronic displays depends to a large extent on their primary spectra (red, green, and blue in the most common case). In particular, for narrow-band primary spectra whose peak wavelength lies in the range of high variability of the observer's color-matching function, some observers can experience very large differences between actual surface colors (e.g. in a light booth) and displayed colors if the monitor is optimized for the International Commission on Illumination (CIE) 1931 standard observer. However, because narrow-band light-emitting diodes lead to larger color gamuts, more and more monitors with very narrow band primaries are coming onto the market without manufacturers taking into account the associated problem of observer variations. Being able to measure these variations accurately and efficiently is therefore an important objective. In this paper, we propose a new approach to predict the extent of observer metamerism for a particular multiprimary display. Unlike existing dedicated models, ours does not depend on a reference illuminant and a set of reflectance spectra and is computationally more efficient.

Keywords — *color display, observer metamerism.*

DOI # 10.1002/jsid.605

1 Introduction

Electronic displays create colors by combining a small number of color primaries, typically red, green and blue (RGB). In order to obtain large color gamuts, these primaries are often designed so that they are as narrow banded as possible (i.e. ideally monochromatic). The recent International Telecommunication Union - Radiocommunication Sector (ITU-R) Rec. 2020 standard for ultra-high-definition television recommends that RGB primary are coincident with the spectrum locus. However, such primaries may have the adverse effect of increasing the variability of color perception across a range of observers.[1,2] In particular, for narrow-band primary spectra whose peak wavelengths lie in the range of high variability of the observer's color-matching functions, some observers can experience noticeable differences between actual colors (e.g. in a light booth, for softproofing) and monitor colors if the monitor is optimized for the CIE 1931 standard observer.

Being able to measure this effect is particularly important not only to help users choose the right display for color-critical applications but also to help manufacturers design new display technologies.[3–5] Oicherman *et al*.[6] studied the effect of observer metamerism (OM) on color matching of display and surface colors and suggested that traditional color-difference formulae such as CIEDE2000 can model it well. Aiming at developing a single index that allows to compare several displays, Long and Fairchild[7] proposed several indices to determine the sensitivity of a monitor to observer variability. Each metric uses a set of reference reflectance spectra and a certain illuminant type (e.g. CIE D65) to measure the variations of color perception across a group of representative observers (given their individual color-matching functions) in a nearly perceptually uniform color space like CIELAB. For one particular reflectance, the colors perceived by the group of observers span a volume in CIELAB that can be coarsely represented by an ellipsoid, the average or maximum volume of which then serves as a measure of the monitor's proneness to inducing metameric errors.

The main problem with such an approach is that it depends on the reference reflectance spectra and illuminant. In particular, the choice of the latter is difficult for softproofing applications because light booths differ in their use of different light sources. To simulate a CIE D50 light type, light-emitting diode clusters, filtered halogen lamps or fluorescent lamps are used in light booths, all of which lead to almost the same tristimulus values for the CIE 1931 standard observer but differ significantly in the spectral power distribution of the emitted light. Consequently, an index that depends on a particular type of light is not optimal in practice because of the variability of the possible light types in applications like softproofing.

Received 06/25/17; accepted 09/10/17.
This work was partially supported by the AIF IGF grant 18178N of the German Federal Ministry of Economics and Technology.
Steven Le Moan is with the School of Engineering and Advanced Technology, Massey University, New Zealand; e-mail: steven.lemoan@gmail.com.
Tejas Madan Tanksale and Philipp Urban are with Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany.
Philipp Urban is also with the Norwegian Colour and Visual Computing Laboratory, NTNU, Gjøvik, Norway.
Roman Byshko is with Fogra Graphic Technology Research Association, Munich, Germany.

It would be desirable to have an index that is invariant to the type of light and thus would be valid for virtually any real softproofing conditions. Finally, it is the monitor that should be assessed and not the variability in the softproofing conditions. In other words, the metric should depend only on the display primaries. Long and Fairchild's OM indices ($OM_{x,\,var}$ and $OM_{x,\,varmax}$)[7] are defined as follows:

$$OM_{x,var} = \overline{Vol\left(\Delta(L^*,a^*,b^*)_{\mathbf{P}}\right)} \tag{1}$$

and

$$OM_{x,varmax} = \max\left[Vol\left(\Delta(L^*,a^*,b^*)_{\mathbf{P}}\right)\right], \tag{2}$$

where $Vol(\Delta(L^*,a^*,b^*)_{\mathbf{P}})$ is the mean CIELAB ellipsoid volume constructed from colour matching function-based error vectors in L^*, a^*, and b^* from each patch in patchset **P**, under one-reference illuminant. Both OM indices are particularly dependent on that reference illuminant, as shown in Table 1. The latter reports correlation values, each of them calculated between two sets of calculated index values, where each set corresponds to a different index and each value in the sets corresponds to a different display. It can be seen that OM index values based on different reference illuminants do not fully correlate with each other. Even though the correlations are fairly high, they are significantly different from 1. In other words, errors will be made when ranking displays with the OM indices if the exact reference illuminant is unknown. This means that, in a softproofing application, one needs to know the exact spectral power distribution of the illuminant used in the light booth for the metric to be valid. Note that the results of these indices for different illuminants can be combined (e.g. averaged) but at the expense of computational efficiency (see Section 3). Furthermore, they always depend on the reflectances used as reference.

In this paper, we propose a new index that, unlike the OM indices, does not depend on a reference illuminant and a set of reflectance spectra and is computationally more efficient.

TABLE 1 — Pearson and Spearman correlation between the OM indices[7] scores for three different illuminants (CIE D50, CIE A, and a Fluorescent D50 simulator) computed for the Standard Object Color Spectra (SOCS) database[8] and 1269 matte Munsell color chips,[9] as well as for each display of our benchmark (see Section 3).

		SOCS	Munsell
$OM^{D50}_{x,var}/OM^{A}_{x,var}$	PLCC	0.911	0.899
	SROCC	0.813	0.886
$OM^{D50}_{x,varmax}/OM^{A}_{x,varmax}$	PLCC	0.924	0.901
	SROCC	0.726	0.659
$OM^{D50}_{x,var}/OM^{Fluo}_{x,var}$	PLCC	0.915	0.900
	SROCC	0.815	0.602
$OM^{D50}_{x,varmax}/OM^{Fluo}_{x,varmax}$	PLCC	0.854	0.877
	SROCC	0.644	0.706

SOCS, Standard Object Color Spectra; OM, observer metamerism; PLCC, Pearson Linear Correlation Coefficient; SROCC, Spearman Rank Order Correlation Coefficient.

It is based on the so-called Vora value[10] and measures the resemblance between the spectral power distribution of the display primaries and the color-matching functions of a given set of observers.

2 Proposed index

2.1 Related work in defining a figure of merit for camera filters

The proposed index is conceptually derived from the idea to compute the degree of deviation of spectral camera sensitivities from the Luther-Ives condition[11,12] in camera technology: the so-called Vora value.[10] The idea behind the Vora value is to compare the vector spaces spanned by the color-matching function of the standard observer and the sensitivities of a camera. The principal angles between the vector spaces are measured and transformed into a number between 0 and 1. If the angles are all 0, a linear transformation exists between camera sensitivities and spectral value curves of the standard normal observer (Luther condition is fulfilled). The larger the angles, the smaller the Vora value is.

If a camera fulfills the Luther-Ives condition, its spectral sensitivities can be transformed by a linear mapping into the color-matching functions of the standard observer. Because cameras almost always do not meet the Luther-Ives condition in practice, a transformation of the camera RGB values in CIE XYZ always engenders an error. The index can be used to classify camera systems according to the degree of error independently of the original used[13] and thus give a good guideline value over the expected color-correction error.

2.2 A new observer-metamerism sensitivity index

In the case of monitors, we compare the main angles of the vector spaces spanned by the monitor stimuli multiplied by the color-matching functions of the CIE 1931 standard observer as well as those of an individual observer. Note that the standard observer and linear camera models should be considered as conceptually similar.

Let $\mathbf{C}_s = \begin{pmatrix} \mathbf{C}_s^{\bar{x}} \\ \mathbf{C}_s^{\bar{y}} \\ \mathbf{C}_s^{\bar{z}} \end{pmatrix}$ and $\mathbf{C}_i = \begin{pmatrix} \mathbf{C}_i^{\bar{x}} \\ \mathbf{C}_i^{\bar{y}} \\ \mathbf{C}_i^{\bar{z}} \end{pmatrix}$ be $3 \times N$ matrices

whose rows are the color matching of, respectively, the CIE 1931 standard observer and the individual observer (where N is the number of equidistant samples covering the visible wavelength range), and let the $M \times N$ dimensional matrix **P** contain the M primary spectra of the monitor to be evaluated as line vectors ($M = 3$ for an RGB monitor). The observer's individual index is then calculated as follows:

1. Calculate two new $M\times 3N$ dimensional matrices:

$$\mathbf{C}'_s = \begin{pmatrix} \mathbf{C}_s^{\bar{x}}\otimes\mathbf{P}_1 & \mathbf{C}_s^{\bar{y}}\otimes\mathbf{P}_1 & \mathbf{C}_s^{\bar{z}}\otimes\mathbf{P}_1 \\ \mathbf{C}_s^{\bar{x}}\otimes\mathbf{P}_2 & \mathbf{C}_s^{\bar{y}}\otimes\mathbf{P}_2 & \mathbf{C}_s^{\bar{z}}\otimes\mathbf{P}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{C}_s^{\bar{x}}\otimes\mathbf{P}_M & \mathbf{C}_s^{\bar{y}}\otimes\mathbf{P}_M & \mathbf{C}_s^{\bar{z}}\otimes\mathbf{P}_M \end{pmatrix} \quad (3)$$

and

$$\mathbf{C}'_i = \begin{pmatrix} \mathbf{C}_i^{\bar{x}}\otimes\mathbf{P}_1 & \mathbf{C}_i^{\bar{y}}\otimes\mathbf{P}_1 & \mathbf{C}_i^{\bar{z}}\otimes\mathbf{P}_1 \\ \mathbf{C}_i^{\bar{x}}\otimes\mathbf{P}_2 & \mathbf{C}_i^{\bar{y}}\otimes\mathbf{P}_2 & \mathbf{C}_i^{\bar{z}}\otimes\mathbf{P}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{C}_i^{\bar{x}}\otimes\mathbf{P}_M & \mathbf{C}_i^{\bar{y}}\otimes\mathbf{P}_M & \mathbf{C}_i^{\bar{z}}\otimes\mathbf{P}_M \end{pmatrix}, \quad (4)$$

where $\otimes$ is the element-wise multiplication, $\mathbf{C}_s^{\bar{x}}$, $\mathbf{C}_s^{\bar{y}}$, and $\mathbf{C}_s^{\bar{z}}$ are the CIE 1931 standard color-matching functions corresponding respectively to the X, Y, and Z components (and similarly for $\mathbf{C}_i$), and $\mathbf{P}_j$ is the spectral emission curve corresponding to the display's jth primary.

2. Calculate their singular value decompositions:

$$\mathbf{U}_s\mathbf{S}_s\mathbf{V}_s^T = \mathbf{C}_s'^T$$

and

$$\mathbf{U}_i\mathbf{S}_i\mathbf{V}_i^T = \mathbf{C}_i'^T.$$

3. Determine n_1 and n_2 the number of nonzeros singular values in $\mathbf{S}_s$ and $\mathbf{S}_i$ that are above a threshold τ (because of numerical errors of the singular value decomposition, we used a small threshold $\tau = 10^{-3}$ in the succeeding text below which the singular values are treated as zero) and calculate $n = \max(n_1, n_2)$ (Note that n_1 and n_2 are the same if the primaries are not linearly dependent on each other). In $\mathbf{U}_s$ and U_i, extract the n columns with largest singular values to form $\mathbf{U}_s'^T$ and $\mathbf{U}_i'^T$.
4. Compute the singular value decomposition of their product $\mathbf{USV}^T = \mathbf{U}_i'^T\mathbf{U}'_s$.
5. The observer-individual index is then calculated as the sum of the singular values in $\mathbf{S}$:

$$\overline{\Theta}(\mathbf{C}_s, \mathbf{C}_i, \mathbf{P}) = \frac{1}{\mathrm{num}(s)} \sum_{i=1}^{\mathrm{num}(s)} s_i^2, \quad (5)$$

where s_i is the ith singular value in $\mathbf{S}$ and num(s) corresponds to the number of nonzero singular values in $\mathbf{S}$.

The value computed in Eq. (5) is one if the colors generated by the monitor's primaries for the individual observer $\mathbf{C}_i$ are not different from those of the standard observer. The larger the differences, the smaller the metric score is. Because differences between spectral curves of individual observers and the standard observer are relatively small, at least as compared with the differences in the spectral sensitivities of cameras to the spectral values of the standard observer, the dynamic range of the metric's score is relatively small and close to 1 for all individual observers. Therefore, we propose to rescale the score as an error permillage (0: best score, 1000: worst score):

$$\Theta(\mathbf{C}_s, \mathbf{C}_i, \mathbf{P}) = 1000\big(1 - \overline{\Theta}(\mathbf{C}_s, \mathbf{C}_i, \mathbf{P})\big). \quad (6)$$

While the theoretical range of values for our index is indeed from 0 to 1000, only a portion of this range will be useful in practice. This is due to the fact that a score of 1000 would mean that there is absolutely no correlation between the set of display primaries and the observer's color-matching functions, which virtually never happens in practice.

In order to define a general index for a group of observers, a representative set of individual color-normal observers (noted $\boldsymbol{x} = \{\mathbf{C}_i^1, \ldots, \mathbf{C}_i^K\}$) is necessary, and we propose to use the following two statistics of the metric across this set as final indices:

$$\Theta_{x,mean}(\mathbf{C}_s, \mathbf{P}) = \frac{1}{K}\sum_{j=1}^{K} \Theta\Big(\mathbf{C}_s, \mathbf{C}_i^j, \mathbf{P}\Big)$$

and

$$\Theta_{x,max}(\mathbf{C}_s, \mathbf{P}) = \max\Big\{\Theta\Big(\mathbf{C}_s, \mathbf{C}_i^j, \mathbf{P}\Big) \,|\, j\in[1\ldots K]\Big\}.$$

In the next section, we analyze the performances of $\Theta_{\boldsymbol{x},\,mean}$ and $\Theta_{\boldsymbol{x},\,max}$.

3 Results

3.1 Benchmark

In order to demonstrate the efficiency and reliability of our indices, we used 76 sets of individual color-matching functions measured for color-normal observers[14,*] (see Figure 1) and two sets of reference natural reflectance spectra: the SOCS database (53,490 reflectance spectra, 31 bands)[8] as well as the reflectance spectra of 1269 matt Munsell color chips.[9]

In terms of displays, we used 20 sets of primaries in total:

- Ten sets were measured from actual RGB monitors. Note that three measurements were made on each monitor in order to account for measurement errors and inhomogeneities of the display. However, we did not observe any significant influence on the indices. In other words, our indices as well as Long and Fairchild's are robust to measurement inaccuracies and local inhomogeneities.

*Note that we also calculated these results for a set of 100-simulated individual color-matching functions, following the method described in Fairchild and Heckaman[15] and obtained very similar results that support the same conclusions.

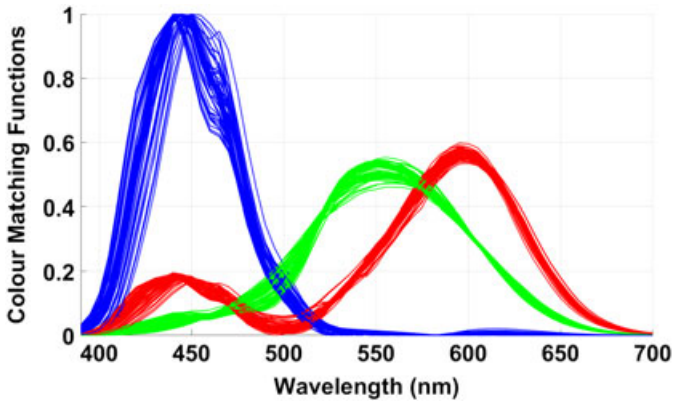


FIGURE 1 — Seventy-six individual color-matching functions.

- One set represents a simulated ITU-R Rec. 2020-compatible display with monochromatic primaries at 630, 532, and 467 nm.
- One set represents a simulated chromaticity-gamut-optimized eight-primary laser projector.[7]
- Eight sets were generated artificially so that they contain three primaries, each of them being a Gaussian curve with random standard deviation and random mean (where the latter is of course constrained to be in the range of visible wavelengths).

This means that half of the benchmark consists of measured data and the other half is simulated.

We then compared $\Theta_{x,\,mean}$ and $\Theta_{\mathbf{x},\,max}$ with $OM_{x,\,var}$ and $OM_{x,\,varmax}$ in terms of their correlation to the reference indices $OM^{all}_{x,var}$ and $OM^{all}_{x,varmax}$, that is, the "global" $OM_{x,\,var}$ and $OM_{x,\,\mathrm{varmax}}$, pooled across a representative set of 40 illuminants (10 daylights, 10 incandescent lights, 10 fluorescent lights, and 10 light-emitting diode lights, selected randomly from the National Gallery's set[16] as well as from the University of Eastern Finland's daylights set[17] and the CIE standard illuminants). The OM^{all}_{x} indices are considered hereas reference because they account for many illuminants and reflectances and work in a "brute force" manner. They are however particularly expensive computation wise (see Table 4).

Note that all colorimetric values and volumes were calculated based on the CIE 1931 observer color-matching functions and the ΔE^*_{ab} color difference. Note that there exist more advanced color-difference metrics such as CIEDE2000 or the hue linearized ΔE_{00HL}.[18] However, Long and Fairchild's original indices are based on ΔE^*_{ab}, so we report only results with the latter for the sake of consistency and fair comparison. We also implemented the OM indices with ΔE_{00HL} and used it again for the comparison with our index, but we found no significant difference in the results.

Tables 2 and 3 show the Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficient (SROCC) between the metric scores on the benchmarked displays and, respectively, reference mean indices and reference max indices. A correlation coefficient of 1 shows a perfect match. In order to assess whether two PLCC

TABLE 2 — Pearson and Spearman correlations between the indices values and reference mean index $OM^{all}_{x,var}$.

		SOCS	Munsell
$\Theta_{x,\,mean}$	PLCC	0.609	**0.623**
	SROCC	**0.842**	**0.795**
$OM^{D50}_{x,var}$	PLCC	**0.959**	**0.839**
	SROCC	**0.908**	**0.889**

Values in bold are not significantly different from the maximum corresponding correlation in the same column. SOCS, Standard Object Color Spectra; OM, observer metamerism; PLCC, Pearson Linear Correlation Coefficient; SROCC, Spearman Rank Order Correlation Coefficient.

TABLE 3 — Pearson and Spearman correlations between the indices scores and reference max index $OM^{all}_{x,varmax}$.

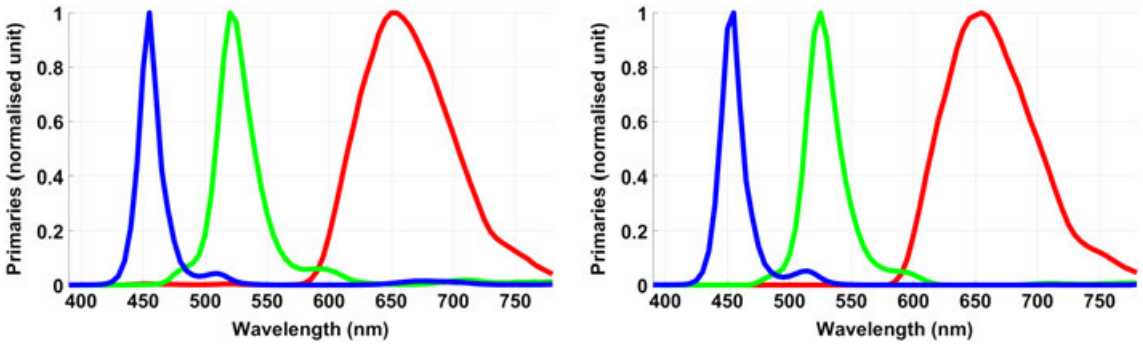
		SOCS	Munsell
$\Theta_{x,\,max}$	PLCC	0.595	0.470
	SROCC	**0.699**	**0.722**
$OM^{D50}_{x,varmax}$	PLCC	**0.906**	**0.910**
	SROCC	**0.820**	**0.850**

Values in bold are not significantly different from the maximum corresponding correlation in the same column. SOCS, Standard Object Color Spectra; OM, observer metamerism; PLCC, Pearson Linear Correlation Coefficient; SROCC, Spearman Rank Order Correlation Coefficient.

are significantly different from each other, we used Steiger's Z test[19] with a p value of 0.05. Note that seemingly large discrepancies between PLCC and SROCC are nothing unusual, because PLCC takes into account the distance between values, and some outliers can strongly influence the correlation. SROCC takes only the rank into account without considering the magnitude of the value. Table 1 shows that some displays have extremely different OM values for two considered illuminant (e.g. D50 vs. Fluo) even though the OM values show a reasonable high-rank-order correlation. Figures 2 and 3 show respectively the two best and two worst sets of primaries according to $OM^{\mathrm{all}}_{x,\,\mathrm{var}}$.

3.2 Discussion

From Tables 2 and 3, we notice that our indices achieve no significantly different performances in terms of SROCC than $OM^{D50}_{x,var}$ and $OM^{D50}_{x,varmax}$ when it comes to predicting $OM^{all}_{x,var}$ on both reflectance datasets. However, in terms of linear correlation, our indices are outperformed on the SOCS dataset for $\Theta_{x,\,mean}$ and both datasets in the case of $\Theta_{x,\,max}$. This is in accordance with an observation made by Vora *et al.*[10] that "[…] the measure should not be used for fine tuning a set of filters [in our case: a set of primaries] but can give a good indication of performance for larger differences in measure when knowledge of the signal statistics is not available." Nevertheless, our indices offer particularly well balanced performances considering that they do not use any reference illuminant or reflectance data and that they are substantially more computationally efficient than any of the OM indices, as demonstrated by the computation times reported in Table 4.



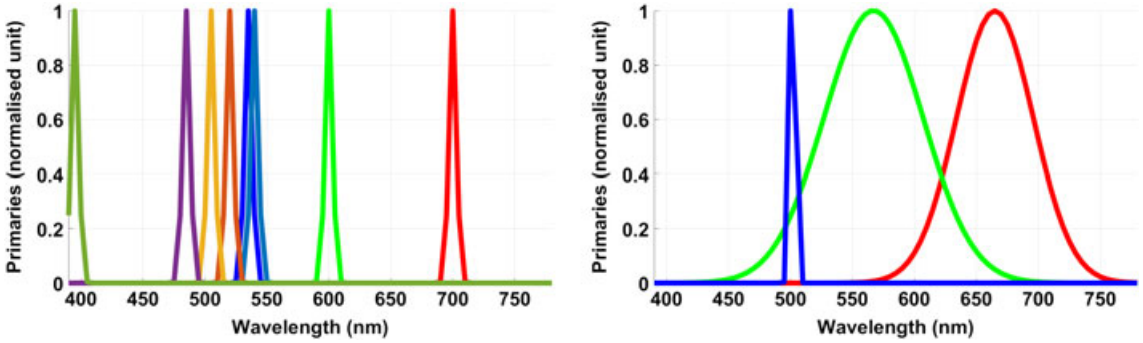
$\Theta_{x,mean} = 6.47$ $\quad$ $\Theta_{x,mean} = 6.64$

$\Theta_{x,max} = 16.67$ $\quad$ $\Theta_{x,max} = 17.86$

$\mathrm{OM}^{all}_{x,var} = 5.79k$ $\quad$ $\mathrm{OM}^{all}_{x,var} = 5.88k$

$\mathrm{OM}^{all}_{x,varmax} = 28.03k$ $\quad$ $\mathrm{OM}^{all}_{x,varmax} = 27.36k$

FIGURE 2 — Two best sets of display primaries in terms of $\mathrm{OM}^{\mathrm{all}}_{x,\mathrm{var}}$. The spectral power distributions were normalized to a 5-nm-sampling step and divided by their peak values for visualization. While the two sets look very similar, they were measured on two different models. The reported OM values were calculated using the SOCS database. Note that *k* stands for ×1000.



$\Theta_{x,mean} = 9.03$ $\quad$ $\Theta_{x,mean} = 12.16$

$\Theta_{x,max} = 31.65$ $\quad$ $\Theta_{x,max} = 47.64$

$\mathrm{OM}^{all}_{x,var} = 7.47k$ $\quad$ $\mathrm{OM}^{all}_{x,var} = 23.77k$

$\mathrm{OM}^{all}_{x,varmax} = 41.90k$ $\quad$ $\mathrm{OM}^{all}_{x,varmax} = 390.66k$

FIGURE 3 — Two worst sets of display primaries in terms of $\mathrm{OM}^{\mathbf{all}}_{\mathbf{x},\mathrm{var}}$. The spectral power distributions were normalized to a 5-nm-sampling step and divided by their peak values for visualization. The set on the left corresponds to the simulated chromaticity-gamut-optimized eight-primary laser projector.[7] The reported OM values were calculated using the SOCS database. Note that *k* stands for ×1000.

These times were measured in Matlab, based on optimized and parallelized code (eight workers). Note that the times measured to compute $\mathrm{OM}_{x,\mathrm{var}}$ and $\mathrm{OM}_{x,\mathrm{varmax}}$ were identical (as well as for $\mathrm{OM}^{\mathrm{all}}_{x,\mathrm{var}}$ and $\mathrm{OM}^{\mathrm{all}}_{x,\mathrm{varmax}}$). Based on these results, we suggest that $\Theta_{x,\mathrm{mean}}$ and $\Theta_{\boldsymbol{x},\mathrm{max}}$ are particularly interesting for display manufacturers for designing their primaries. They can for instance maximize the size of the gamut and simultaneously minimize the observer-metamerism sensitivity index, without having to choose one particular illuminant or set of reflectances as reference.

TABLE 4 — Average computation times (in seconds).

	SOCS	Munsell
Θ_x	**0.08**	
$\mathrm{OM}_{x,\mathrm{var}}$	2.07	0.45
$\mathrm{OM}^{\mathrm{all}}_{x,\mathrm{var}}$	81.70	17.79

SOCS, Standard Object Color Spectra; OM, observer metamerism.

Note that the simulated ITU-R Rec. 2020-compatible display ranks high in our benchmark (4 out of 20), which does not mean that it induces no observer variability but merely that most of the other displays are worse in this regard. In order to demonstrate that the ITU-R Rec. 2020 standard display is only suboptimal in terms of OM, we ran a brute-force optimization on the central wavelengths of its primaries (initially

at 630, 532, and 467 nm), with $\Theta_{x,\,\mathrm{mean}}$ as objective function. With this procedure, we found that moving the narrow-banded primaries central wavelengths to 615, 547, and 460 nm reduces $\Theta_{x,\,\mathrm{mean}}$ from 6.88 to 3.53. The corresponding $\mathrm{OM}^{\mathrm{all}}_{x,\,\mathrm{var}}$ score dropped from 5.95*k* to 5.61*k* (with the SOCS database as reference). Therefore, we can reduce observer variability of ITU-R Rec. 2020-compatible displays with primaries centered at slightly different wavelengths. This would only slightly affect the total size of the display's color gamut as we only moved the primaries along the spectrum locus.

When it comes to measuring the significance of the (difference between) index values, we believe that there is no direct and sound way to do so. First, significance depends heavily on the target application and particularly on the kind of reflectance spectra used in that application. We propose a general index, which accounts for possible metameric mismatches in the whole range of visible colors and consequently some displays may yield very similar scores although they respectively induce OM in very different hue ranges. Secondly, in this context, significance needs to carry some sort of perceptual meaning. While there exists a set of guidelines to assess the significance of color differences (e.g. ΔE^*_{ab}), based on the notion of just noticeable difference, our index does not carry the same perceptual meaning. Our experiments showed that Θ correlates to mean and max color differences only to an extent, which makes it difficult to translate the threshold of noticeability in terms of values produced by our indices. Note that the same can be said about the OM indices. Differences of volumes in a color space are not trivial to interpret, regardless of whether the space is perceptually uniform or not (what would then be a just noticeable difference in terms of e.g. $\mathrm{OM}^{\mathbf{D50}}_{x,var}$?). Furthermore, the same can be said about most image-quality indices such as the SSIM index, yet they have been used in a number of applications with great success. The proposed index can be used to sort and optimize displays in terms of induced metamerism, without reference data and with minimal computational effort.

One alternative to help assessing the significance of $\Theta_{x,\,mean}$ or $\Theta_{x,\,max}$ in a coarse fashion would be to generate a 1D look-up table transforming our max-index values to $\mathrm{OM}^{all}_{x,varmax}$ employing a large number of (artificially created) display primaries. The radius of a sphere with the same volume as $\mathrm{OM}^{all}_{x,varmax}$ would then be measured, resulting in a ΔE-like metric, the significance of which we can assess.

4 Conclusions

The recent ITU-R Rec. 2020 standard for ultra-high-definition television recommends that RGB primary are coincident with the spectrum locus. However, this approach has the adverse effect of increasing the variability of color perception across a range of observers[1,2] if the primaries are not selected carefully. In particular, for narrow-band primary spectra whose peak wavelengths lie in the range of high variability of the observer spectral curves, some observers can experience noticeable differences between actual surface colors (e.g. in a light booth, for softproofing) and monitor colors if the monitor is optimized for the CIE 1931 standard observer. Being able to measure this effect is particularly important to help users choose the right display for color-critical applications but also to help manufacturers design new display technologies. In this paper, we proposed a new approach to predict the extent of OM for a particular multiprimary display, based on the principal angles between the vector subspaces spanned by the display primaries multiplied with the CIE 1931 standard observer and individual observers. Compared with existing indices, results demonstrate that ours are significantly faster and depend only on the display primaries as well as a set of representative individual color-matching functions.

References

1 M.D. Fairchild and D. R. Wyble, "Mean observer metamerism and the selection of display primaries," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2007, vol. 2007, pp. 151–156.

2 A. Sarkar *et al*., "Toward reducing observer metamerism in industrial applications: colorimetric observer categories and observer classification," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2010, vol. 2010, pp. 307–313.

3 R. Ramanath, "Minimizing observer metamerism in display systems," *Color. Res. Appl.*, **34**, No. 5, 391–398 (2009).

4 A. Sarkar, Identification and assignment of colorimetric observer categories and their applications in color and vision sciences, Ph.D. thesis, Université de Nantes, 2011.

5 H. Xie, *et al*., "Pareto optimal primary designs for color displays," *Electronic Imaging*, **2017**, No. 18, 84–90 (2017).

6 B. Oicherman, *et al*., "Effect of observer metamerism on colour matching of display and surface colours," *Color. Res. Appl*, **33**, No. 5, 346–359 (2008).

7 D. L Long and Mark D Fairchild, "Modeling observer variability and metamerism failure in electronic color displays," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2014, vol. 2014, pp. 14–27.

8 Technical Committee ISO/TC 130, Graphic technology, "ISO/TR 16066: standard object colour spectra database for colour reproduction evaluation (SOCS)," Tech. Rep., 2003.

9 J. Hiltunen, Munsell colors matt (spectrophotometer measured): https://www2.uef.fi/fi/spectral/munsell-%20colors-matt-spectrofotometer-measured, Accessed on May. 20th 2017.

10 P. L. Vora and H. Joel Trussell, "Measure of goodness of a set of color-scanning filters," *JOSA A*, **10**, No. 7, 1499–1508 (1993).

11 R. Luther, "Aus dem gebiet der farbreizmetrik," *Z. Tech. Phys.*, **8**, 540–558 (1927).

12 H. E. Ives, "The transformation of color-mixture equations from one system to another," *J. Frankl. Inst.*, **16**, 673–701 (1915).

13 P. Urban and R. R. Grigat, "Metamer density estimated color correction," *SIViP*, **3**, No. 2, 171–182 (2009).

14 Kraushaar, A, "Einflussgrößen auf die hochqualitative monitordarstellung hinsichtlich einer standardisierung von softproofsystemen," Tech. Rep., FOGRA, 2016.

15 M. D. Fairchild and R. L. Heckaman, "Measuring observer metamerism: the nimeroff approach," *Color. Res. Appl.*, **41**, No. 2, 115–124 (2016).

16 "Spectral power distribution curves, the national gallery: http://research.ng-london.org.uk/scientific/spd/," Accessed on June. 1st 2017.

17 "Daylight spectra, university of eastern Finland: http://cs.joensuu.fi/spectral/databases/download/daylight.htm," Accessed on June. 1st 2017.

18 I. Lissner and P. Urban, "Toward a unified color space for perception-based image processing," *Transactions on Image Processing*, **21**, No. 3, 1153–1168 (2012).

19 J. H. Steiger, "Tests for comparing elements of a correlation matrix," *Psychol. Bull.*, **87**, No. 2, 245 (1980).

Steven Le Moan received his PhD degree in Image Processing from the University of Burgundy, Dijon, France, in 2012. He was then enrolled for 2 years as an experienced researcher at the Technische Universität Darmstadt, Darmstadt, Germany, followed by 2 years at Gjøvik University College, Norway. He is currently a lecturer at the School of Engineering and Advanced Technology, Massey University, Palmerston North, New Zealand. His research is mainly focused on multispectral imagery, image quality, and visual perception.

Tejas Madan Tanksale received his MS degree in Information and Communication Engineering from the Technische Universität Darmstadt, Darmstadt, Germany, in 2015. Since then, he has been working on Appearance and Color Management techniques for 3D printers at the Competence Center 3D Printing Technology at the Fraunhofer Institute for Computer Graphics Research, Darmstadt, Darmstadt, Germany. His research interests include optics, mathematical modelling, and image recognition.

Roman Byshko received his master's degree in Applied Mathematics from Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. He later worked at the University of Konstanz, Konstanz, Germany, as Research Assistant. Since 2016, he is working at Fogra Forschungsinstitut für Medientechnologien e.V., Aschheim, Germany. His research interests include pattern recognition, image processing, and color science.

Philipp Urban received the MS degree in mathematics from the University of Hamburg, Hamburg, Germany, and the PhD degree from the Hamburg University of Technology, Hamburg, Germany, in 1999 and 2005, respectively. From 2006 to 2008, he was a Visiting Scientist with the Munsell Color Science Laboratory, Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA, and headed afterward the Color Research Group at the Institute of Printing Science and Technology, Technische Universität Darmstadt, Darmstadt, Germany. Since 2013, he has been the Head of the Competence Center 3D Printing Technology with the Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany. His research interests include spectral imaging, image quality, and material appearance reproduction.