SMPTE ATC 2020

Physiologically Personalized Color Management for Motion Picture Workflows

Trevor D. Canham

Universitat Pompeu Fabra, Barcelona, Spain

David L. Long

Rochester Institute of Technology, Rochester NY, USA

Mark D. Fairchild

Rochester Institute of Technology, Rochester NY, USA

Marcelo Bertalmío

Universitat Pompeu Fabra, Barcelona, Spain

Written for presentation at the

SMPTE 2020 Annual Technical Conference & Exhibition

Abstract. One of the essential mechanisms employed by the human visual system when interpreting the natural world is that of trichromatic integration of physical scene spectra by cone photoreceptors. By extension of this, different scene spectra can result in the same color sensation in an observer, a phenomenon known as metamerism. This allows imaging systems to produce realistic reproductions of scene content by the same three channel mechanism. To predict these matches, color matching functions (CMFs) are used which aim to describe the average spectral integration behavior of observers. However, the use of a single average observer CMF has been shown to result in impactful color rendering errors, as there exists significant variation in the spectral absorption characteristics of the eye within populations of color-normal observers. When this is crossed with the growing disparity between the spectral characteristics of emerging display technology it becomes evident that this inter-observer variability should be accounted for.

In lieu of more spectrally accurate imaging systems, a color management pipeline for motion picture production which considers the individual characteristics of key creative observers could remove this variability when color critical decisions are being made. For example, if at the various stages which a Director of Photography reviews imagery (on-set, dailies, color correction, VFX, etc.) their monitoring equipment is calibrated to be a metameric match to some reference considering their individual physiology, this would remove inconsistency relating to the display technology used at different stages and avoid any influence it may have on creative decisions. Additionally, it would allow an observer and display specific reference point to be established in the pipeline (for example, the colorist and primary mastering display) to which imagery could be directly related when remastering content in the future.

Asano and Fairchild present a physiologically based individual observer model, as well as a method for separating a population of observers into a limited number of observer categories. Building on this work, we present a computationally simple metameric match simulation pipeline. Using this, we perform an experiment with real display spectra and natural images to observe the variability which could occur among a population as a result of observer metamerism in a motion picture viewing scenario. The results provide further evidence that inter-observer metameric variability is a relevant problem in the context of natural images. Finally, we outline how this pipeline can be incorporated into one's color management strategy.

Keywords. *Observer metamerism, color management, personalization.*

Introduction

Through evolution, the discrimination abilities of our visual system have gradually developed according to our survival needs. This process started with simple bright versus dark discrimination in our single celled ancestors, adding spatial discrimination, color, etc. to eventually create the visual sensation that we rely on constantly, complete with its high level information management routines and constantly dynamic behavior. By virtue of its construction, this visual experience is an abbreviated probe of physical reality. For example, instead of discriminating one color from another based on their specific physical spectra, we compare the relative integrated responses of three retinal cell types (cones) which are sensitive to different portions of the spectrum. As a result of this mechanism, these three cones could produce the same sensation from stimuli that are spectrally different, a concept we refer to as metamerism.

It is because of metamerism that imaging systems can be engineered which are able to convey a certain degree of realism similarly using just three color channels. This allows engineers to employ solutions that involve the modulation of the spectral characteristics of primaries to achieve various performance goals (dynamic range and color gamut, color rendering consistency, energy efficiency, etc.) In this sense, the precise spectrum of light is percieved to be a visually non-impactful parameter, so long as it results in a matching cone response for the observer (compared to some reference representation). The problem in this strategy is that while the reference observer may percieve a match, variation in the physiology of the human eye within a population makes it such that not everyone will recieve the same sensation as a result of differing spectral interaction between scene/screen light and the eye. Furthermore, if the description of the reference observer does not match that of any real observer (ie. if it is a population average), there exists the possiblity that no viewer would see the correct color rendering.

In lieu of more spectrally accurate imaging systems, a color management pipeline for motion picture production which considers the individual characteristics of key creative observers could remove observer metamerism variability when color critical decisions are being made. For example, if for the various stages at which a Director of Photography reviews imagery (on-set, dailies, color correction, VFX, etc.) their monitoring equipment is calibrated to be a metameric match to some reference considering their individual physiology, this would remove inconsistency relating to the display technology used at different stages and avoid any influence it may have on creative decisions. Additionally, it would allow an observer and display specific reference point to be established in the pipeline (for example, the colorist and primary mastering display) which imagery could be directly related to in the future when remastering content for different formats.

A physiologically based individual observer model is introduced by Asano and Fairchild in [1], as well as a method for separating a population of observers into a limited number of observer categories [2]. Building on this work, we present a metameric match simulation pipeline that is computationaly simple and can be used practically in motion picture color management. Using this pipeline, we perform a simulation experiment with real display spectra and natural images to observe the variability which could occur among a population as a result of observer metamerism in a motion picture viewing scenario. The results provide further evidence that inter-observer metameric variability is a relevant problem in the context of natural images. Finally, we outline how this pipeline can be used in motion picture color management.

Background

In the late 1920s William David Wright [3] and John Guild [4] each performed color matching experiments where red, green, and blue lights were adjusted to match the single wavelength output of a monochromator at increments across the visible spectrum of light. The responses of observers were averaged, resulting in the CIE 1931 2-degree field of view color matching function (CMF) which has been long used in motion picture color engineering and display calibration (Figure 1a). Reliance on this average function to optimize transforms which are intended to match between displays of different spectral quality can lead to color management inconsistencies, as the physiological CMFs of individuals can vary significantly from person to person. This variation is the result of a wide range of factors. These include genetics, age, and health & lifestyle factors like whether the observer is a smoker, is overweight, has diabetes, etc. Together, these factors can result in variation in the observer's optical pathway (density variation in the lens, macula, and retinal photopigment) - altering the spectral characteristics of light absorbed by the photoreceptors [1].

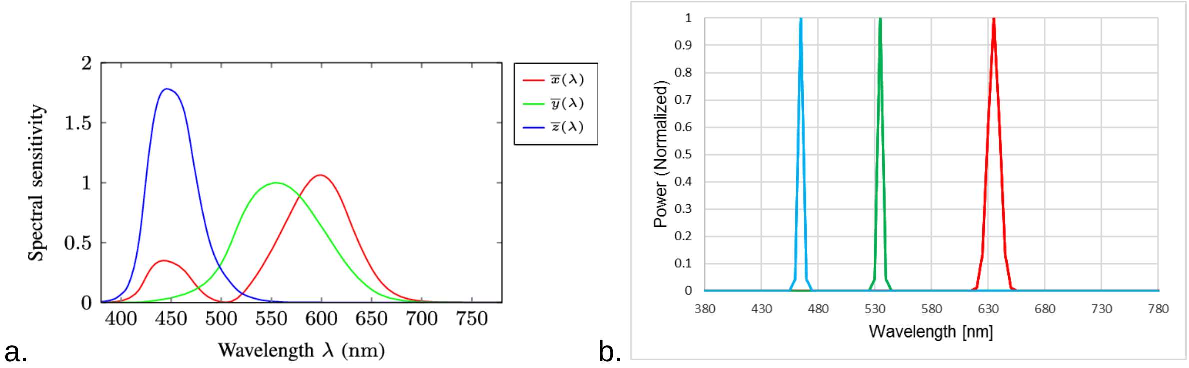

Figure 1a. CIE 1931 2-Degree field of view CMF $(\bar{x}, \bar{y}, \bar{z})$ commonly used in colorimetric measurements. b. Spectrally narrow band primary measurements for an IMAX laser projector.

While variation in these factors may only cause subtle or undetectable errors in cases where the display primaries are spectrally broadband (Xenon lamp projectors, cathode ray tube displays, film dyes, etc.), narrow-band primary displays like laser projectors (Figure 1b) can exacerbate differences [5]. Unfortunately, these narrow band spectral architectures are required to meet the high purity primaries specified in the current wide color gamut standard, Rec. 2020. In order to assess the magnitude of these variations in a practical color grading scenario, Asano et al. conducted a color matching experiment in [6] where 28 observers were asked to make matches between images displayed on an LCD monitor and on a Pico laser projector using L*a*b* controls. The results of several observers are shown in Figure 2a, compared to the match predicted using the CIE 1964 10-degree standard observer. To analyze the results, the group compared the intra to inter-observer variability and found that each observer had an average match point that was significantly different from others.

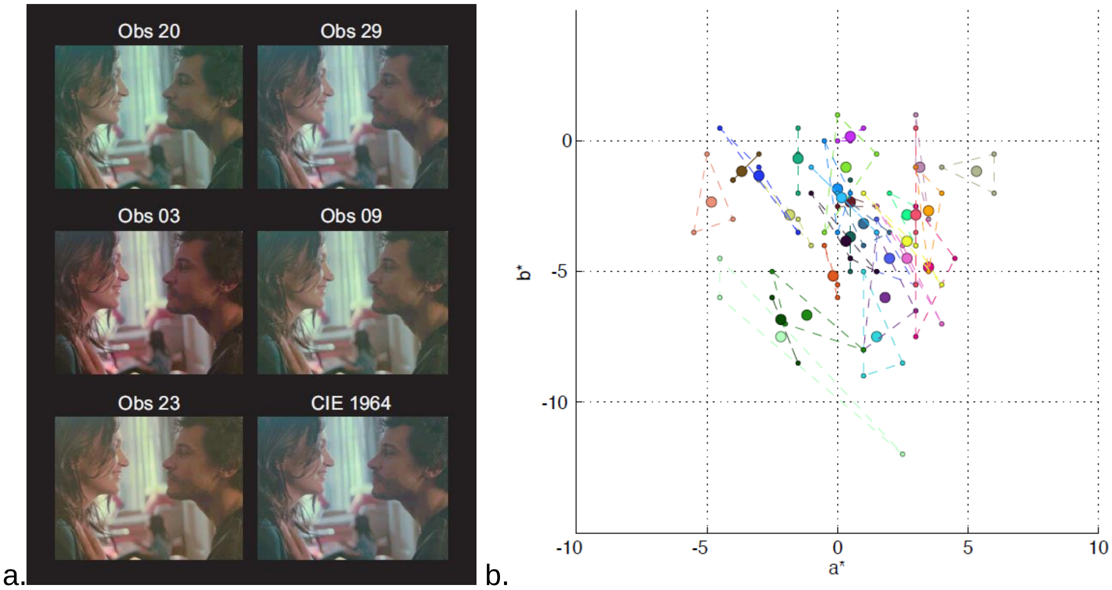

Figure 2a. Matching responses of several observers from [6], compared side by side providing a visualization of the potential variability which could result between observers matching a laser projector and LCD screen and b. average responses from [6] for each individual observer surrounded by their intra-observer error tolerances.

In order to simulate the color matching variance which could occur in a population of observers, Asano and Fairchild developed a model which considers 10 parameters (age, visual degree exposure to stimulus, lens pigment density, macular density, peak optical density for L, M, and S photopigments, and deviations in peak sensitivities for the L, M, and S cones) which can be measured from an observer's physiology [1]. Then, a series of studies were collected which reported inter-observer variability in relation to these factors, and their results were analyzed to obtain standard deviation values for each parameter. Then, a Monte-Carlo simulation generating 10,000 random color-normal observer CMFs considering these standard deviations was conducted.

Taking inspiration from the observer grouping experiment of Sarkar [7], Asano and Fairchild show that this large group of color matching functions can be simplified into a limited set of categorical observers using a k-medoids clustering algorithm [2]. This technique is similar to k-means, however its output is restricted to be a function within the set generated by the physiological model, such that the categories produced could represent actual observers rather than functions derived as an average between them. The technique of grouping observers into a small number of categories allows for the discretization of population variability, while still making a significant improvement in color matching prediction compared to the use of a single standard observer. This was tested through a series of simulations matching a white patch rendered for different display combinations. Using a large set of ground truth observers, a three-channel scaling factor was determined for each observer to make a perfect match for every display combination. Then, the match values for the test display were converted to CIE 1931 2-degree tristimulus values (XYZ) via a 3x3 matrix. The values were then converted to L*a*b* and their Delta E (DE) difference to the nearest categorical observer’s match was taken. The mean DE of the group of observers was then calculated, and the simulation was repeated with an increased number of categorical observers (adding additional categories in order of their improvement to overall error). The results are shown in Figure 3. Using a single observer category (equal to the CIE perceptual observer given an observer age of 38 and a 2-degree field size), the highest average error of 5 DE is obtained in the case of comparing a CCFL backlit LCD screen to a laser projector. It can also be seen that the improvement in terms of average error slows significantly at the 10-observer mark.

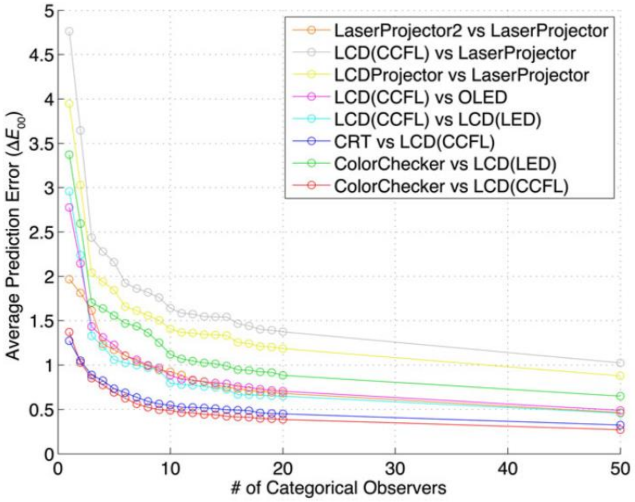

Figure 3. Average DE error between ground truth observer matches and the match generated for the nearest categorical observer, as a function of the number of categorical observers allowed.

For the categorical observers to be used in a motion picture workflow, all individuals making color critical decisions should be characterized. To accomplish this, several "observer calibration" methods have been proposed. In general, these involve method of adjustments color matching between color patches shown on spectrally dissimilar displays or are forced choice comparison experiments between different observer category renderings. Sarkar proposes a practically useful methodology in [7] where observers are presented a series of color patch pairs and are asked to report the quality of the match between the two via a three-choice scale (not acceptable, acceptable, satisfactory.) The color patch pairs are generated to each be a match for a specific observer type given the displays used in the experiment. The results are tallied via a weighted scoring scheme which puts a high penalty on unacceptable matches, a low reward for acceptable matches, and a high reward for satisfactory matches. In extension of this work, a low cost hardware observer calibrator is presented in [8]. Instead of using spectrally dissimilar displays to render color patches, this device simply uses two sets of three narrow band LEDs which emit light at wavelengths where the maximum error exists between observers.

Model

Given a categorical CMF consisting of (CMF_L, CMF_M, CMF_S) and two three-channel sets of display primary spectra (D_r, D_g, D_b), each sampled from 390-780nm at 5nm increments, we intend to find the display linear RGB scalars to send to the second display $D2$ in order to match the LMS cone excitations resulting from a given RGB triplet on the first display $D1$. The categorical CMFs are generated with the method from [1] using the 10 categorical parameters from Table 1 of [2]. No extra normalization steps are applied to this data. The calibration of the display when the primary spectra were measured is unknown, so we find scalars for each primary such that, at maximum drive value, they produce the same pseudo-cone stimulus as prescribed for the white point of Rec.709 for CMF_{ref}, which is the CIE 1931 2-degree observer.

The procedure for calibrating display D is conducted in the following way:

$$spd_D = (D_r) + (D_g) + (D_b) \tag{1}$$

$$k_D = 1/\int spd_D * CMF_{ref_Y}\, d\lambda \tag{2}$$

$$\begin{gathered} for\ i \in X,Y,Z \\ for\ j \in r,g,b \end{gathered}$$

$$i_j = k_D \int D_j * CMF_{ref_i} d\lambda \tag{3}$$

$$mat_{cal} = \begin{matrix} L_r & L_g & L_b \\ M_r & M_g & M_b \\ S_r & S_g & S_b \end{matrix} \tag{4}$$

$$cal_D = {mat_{cal}}^{-1} * XYZ_{ref_{D65}} \tag{5}$$

Then, we carry normalization factors k_D and a 3x1 calibration scalar vector $cal_D = \{cal_{D_r}, cal_{D_g}, cal_{D_b}\}$ for both displays ($D1$, $D2$) to the matching step, along with CMF_{test}, the chosen categorical observer for the match:

$$\begin{gathered} for\ i \in L,M,S \\ for\ j \in r,g,b \end{gathered}$$

$$i_j = k_{D1} \int cal_{D1_j} * D1_j * CMF_{test_i} d\lambda \tag{6}$$

$$mat_1 = \begin{matrix} L_r & L_g & L_b \\ M_r & M_g & M_b \\ S_r & S_g & S_b \end{matrix} \tag{7}$$

Now we have the forward transformation matrix mat_1, for spd_1 as seen by categorical observer CMF_{test}. Given the display linear RGB values of the image mastered on $D1$, this matrix will convert to the LMS cone signal values as observed by CMF_{test}. We now want to find the inverse transformation matrix mat_2, which we can apply to the LMS values to find the RGB values to scale spd_2 so CMF_{test} gets the same excitation while viewing $D2$. So, we repeat Equations 6 and 7 for the case of spd_2 and apply mat_1 and mat_2^{-1} to the display linear input, RGB_{lin_1}.

$$\begin{gathered} for\ i\ \in\ L,M,S \\ for\ j\ \in\ r,g,b \end{gathered}$$

$$i_j = k_{D2} \int cal_{D2_j} * D2_j * CMF_{test_i} d\lambda \tag{8}$$

$$mat_2 = \begin{matrix} L_r & L_g & L_b \\ M_r & M_g & M_b \\ S_r & S_g & S_b \end{matrix} \tag{9}$$

$$RGB_{lin_2} = mat_2^{-1} * \ mat_1 * RGB_{lin_1} \tag{10}$$

Using this procedure, the linear representation of any image mastered on $D1$ can be transformed to its linear representation for $D2$. For our purposes, we add an additional step to convert the RGB output to CIE 1931 2-degree XYZ such that we can pass our results to standard error metrics.

Starting with mat_{cal} (assuming that the CIE 1931 2-degree CMF was used as CMF_{ref}) and display D, we scale each of the coefficients by its corresponding primary calibration value $cal_{D_{rgb}}$, giving us the XYZ values for each display primary at its maximum drive value.

$$XYZ_{rgb} = \begin{matrix} X_r * cal_{D_r} & X_g * cal_{D_g} & X_b * cal_{D_b} \\ Y_r * cal_{D_r} & Y_g * cal_{D_g} & Y_b * cal_{D_b} \\ Z_r * cal_{D_r} & Z_g * cal_{D_g} & Z_b * cal_{D_b} \end{matrix} \tag{11}$$

then, convert to chromaticity values.

$$for\ i\ \in\ r,g,b \tag{12}$$

$$x_i = \frac{X_i}{X_i + Y_i + Z_i}$$

$$y_i = \frac{Y_i}{X_i + Y_i + Z_i}$$

$$z_i = \frac{Z_i}{X_i + Y_i + Z_i}$$

Finally, we derive the primary matrix PM_D.

$$C = \begin{matrix} x_r & x_g & x_b \\ y_r & y_g & y_b \\ z_r & z_g & z_b \end{matrix} \tag{13}$$

$$J = C^{-1} * \begin{matrix} \frac{x_{white}}{y_{white}} \\ 1 \\ \frac{z_{white}}{y_{white}} \end{matrix} \qquad (14)$$

$$PM_D = C * diag(J) \qquad (15)$$

Finally, by applying PM_{D2} to RGB_{lin_2}, we get the CIE 1931 2-degree tristimulus values of the observer match, XYZ_2.

$$XYZ_2 = PM_{D2} * RGB_{lin_2} \qquad (16)$$

Simulation

In order to test the pipeline presented above, we performed a series of simulations using a collection of six display spectra arranged in different combinations, a set of 41 images, and the 10 categorical observers proposed by Asano and Fairchild [2], for which the $\bar{L}\bar{M}\bar{S}$ functions are shown in Figure 4. A subset of these results (eight varied images from full test set) is shown in Table 1. Our display collection is a subset of those tested in the experiment shown in Figure 3, and includes the following: Sony BVM series CRT, Apple Cinema CCFL backlit LCD display, Samsung Galaxy S3 OLED, Dreamcolor LED backlit LCD, IMAX laser projector, PTAX200U LCD projector. In this simulation, for each image and display pair, we generate a match for every observer to be shown on the second display in the pair to match the reference image shown on the first. We then take the error between each observer match and a reference observer match (in this case we use the CIE 1931 2-degree CMF), simulating the range of error one could encounter within a population when calibrating the two displays to a single reference observer. We also performed this experiment for patches at the display maximum drive values and at white. The results are shown in Table 1 in terms of three different error metrics. The first is the mean (across all pixels in the image, all observers in the set) DE 1976. The second is the standard deviation in mean DE across the observer set. Finally, we show the observer metamerism metric of [9]. This metric is calculated as the number of pixels over a just noticeable difference threshold (here we take DE = 2), which we convert to percentage of image area.

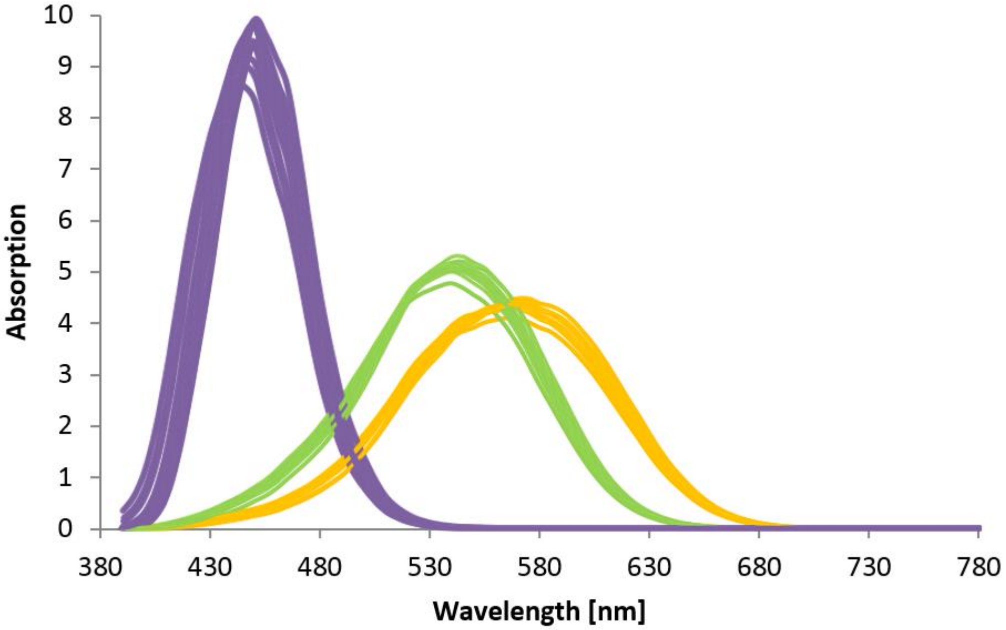

Figure 4. $\bar{L}$, $\bar{M}$, and $\bar{S}$ functions for ten categorical observers from [2]

											
Pair	red	green	blue	white	carousel2	carousel3	face9	showgirl	arrival	bark	drive
CRT v. CCFL	4.17	3.63	7.36	3.26	1.72	0.48	2.44	0.59	2.22	1.12	2.43
CCFL v. LED	3.72	4.83	8.59	3.31	1.52	0.59	2.66	1.20	2.33	1.47	2.29
CCFL v. OLED	2.76	3.73	5.19	4.67	1.67	0.67	3.66	1.23	3.04	1.69	2.94
pLCD v. IMAX	3.89	3.15	6.60	5.82	2.08	0.81	4.55	1.45	3.72	2.07	3.65
CCFL v. IMAX	4.90	4.79	8.03	7.15	2.58	1.02	5.60	1.86	4.59	2.53	4.52
CRT v. IMAX	4.27	6.98	14.66	9.12	3.99	1.29	7.04	2.04	5.90	3.06	6.20
CRT v. CCFL	0.32	1.26	5.30	1.69	1.07	0.19	1.25	0.20	1.03	0.44	1.35
CCFL v. LED	0.68	2.44	7.58	2.34	1.04	0.28	1.87	0.73	1.39	0.92	1.33
CCFL v. OLED	0.65	2.05	3.20	2.22	0.87	0.26	1.67	0.40	1.43	0.81	1.42
pLCD v. IMAX	0.91	1.37	3.57	2.88	1.02	0.37	2.23	0.67	1.87	1.08	1.78
CCFL v. IMAX	1.20	1.88	4.60	3.77	1.32	0.48	2.91	0.87	2.43	1.40	2.32
CRT v. IMAX	0.69	1.80	8.48	5.34	2.30	0.60	3.98	0.81	3.32	1.67	3.57
				CRT v. CCFL	0.32	0.05	0.48	0.04	0.38	0.13	0.41
				CCFL v. LED	0.26	0.09	0.56	0.17	0.44	0.25	0.44
				CCFL v. OLED	0.32	0.10	0.76	0.21	0.63	0.32	0.59
				pLCD v. IMAX	0.41	0.14	0.86	0.27	0.72	0.40	0.69
				CCFL v. IMAX	0.49	0.18	0.91	0.35	0.81	0.48	0.78
				CRT v. IMAX	0.60	0.22	0.93	0.36	0.78	0.52	0.73

Table 1. Metamerism simulation results for a subset of images, display pairs. Top group: mean Delta E 1976 error between observer categories and reference observer, averaged over image pixels and categories. Center group: standard deviation between mean delta E 1976 averaged over image pixels for different observer categories. Bottom group: % image area over 2 DE 1976, averaged over observer categories.

We can see from the results in Table 1 that the error values for the primary colors give a fairly good indication of the amount of error we expect to see in the image. For example, there exists a large degree of error between observers for the blue primary and for white, so images with large blue or white regions tend to have a greater degree of metamerism error. Figure 5 shows an array of images matched for a given display pair and observer. These images were generated first by transforming their display linear sRGB representations to CIE 1931 2-degree XYZ, then transforming to the first display representation (in this case, the Sony BVM series CRT) using ${PM_{D1}}^{-1}$ derived from Equations (11-15.) Then, the transform to the metameric match for a given observer category shown on the destination display (IMAX laser projector) was applied as in Equations (6-10). Finally, the images are converted back to XYZ using PM_{D2}, and later back to the sRGB representation shown. This figure should be interpreted as a visual example of the degree to which the color balance of an image may need to be shifted to make a match on the same display pair for different observers, rather than a visualization of each observer's perception. This is an important distinction as this pipeline simulates only the initial physiological stages of the visual pathway and does not consider any dynamic processing that occurs between absorption and perception which could vary greatly between observers.

Figure 5. sRGB renderings of images shifted to make a metameric match between a Sony BVM-CRT and IMAX laser projector for a set of observers . Panel a. shows the sRGB source image, and the matches for observers 3 (panel b.), 7 (panel c.), and 10 (panel d.) are shown. Test image provided by Arri Camera Systems.

Discussion

We propose the use of this model in a post-production color management pipeline in the following way. First, the displays upon which color critical observers work will need to be characterized in terms of their EOTF, as well as their primary spectra at maximum drive value. Then, a reference representation (display, observer) of the images can be selected (the colorist and their primary reference monitor, for example.) Next, all color critical observers will need to be tested for their categorical observer type. This can be done using the methods described in [7,8]. These experiments can be generalized to any display pair which is spectrally incompatible enough to produce visible metamerism errors, so to perform the test, users can select the two displays in their color management pipeline which are most likely to be problematic. The simulation procedure described above can give an indication whether metamerism error is likely to be noticeable for a given display pair and image/color patch set.

For each destination display/observer pair, the process described in equations 1-10 can be carried out, using the relevant calibration standard (observer, white point) in equations 1-5, the reference display and observer data for Equations 6 and 7 to derive mat_1, and the destination display and observer data for Equations 8 and 9 to derive mat_2. Then, taking the display linear representation of the image mastered by the colorist on their reference monitor, we apply the 3x3 transformation $mat_2^{-1} * mat_1$, apply the inverse of the EOTF, and display the result. Using this procedure, any color management inconsistency in the pipeline related to observer metamerism is minimized.

Before a recommendation for the use of this pipeline and the set of categorical observers from [2] for post-production color management can be confidently made, we plan to perform a psychophysical experiment to confirm that they consistently result in display matches for real observers. In this experiment, observers will be presented with a series of images displayed side by side on spectrally dissimilar displays and will be asked to rate the quality of the match on the following scale (1- unacceptable match, 2 - acceptable match, 3 - satisfactory match.) The image presentations will be repeated, applying transforms from the pipeline above such that they match between the two displays for each observer. Following this, the experiment will be repeated using the classification method from [7] which proposes the same procedure but employs a set of color patches instead of natural images. The results from each experiment will be tallied and compared between the natural image and synthetic stimuli (color patch) tests. If the classifications from the natural image experiment consistently match those of the synthetic stimuli experiment, we will have some confirmation that the categorical observers allow for consistent color management improvements in the context of post-production and natural images.

While the simulation results presented here indicate the metamerism error that could exist within a population, given the images and displays tested, it will be important to have psychophysical data with which to verify the use of these error metrics for natural images. To start, they rely on the imperfect perceptual uniformity of the L*a*b* space, such that equal error readings in red and blue regions, for example, might correspond to different degrees of perceptual distance to observers. Also, our use of DE 1976 averaged over image pixels could be misleading as there could be very small image regions with very high error values which would skew the metric but might not be noticed by observers. Our measurement of image area over a just noticeable difference DE could also be skewed for the opposite reason - if one image contains large regions that are just at the threshold, while another contains small but contextually significant regions that have a very significant error. It is possible that these two metrics could be combined in some way for a more accurate prediction of observer metamerism error in natural images, however this could only be verified through psychophysical experimentation. For this reason, it would be highly interesting to characterize the degree of perceptual error related to observer metamerism in still and moving images in a post-production environment.

The big question with regard to the practical use of these profiles is always how they can be leveraged for group viewing situations. If the pipeline above is applied directly to the source image, the resulting rendering would only be correct within a degree of error for certain members of the viewing group. To average between the two most prominent categories in the group would be non-trivial, as it is unclear in what representation the operation would take place, or whether or not the resulting CMF would correspond to the perception of any real observer. The experiments of Sarkar in [7] showed that for his category set, there were a few sets of 'twin' categories for which the matches were consistently similar, suggesting that rendering the image for one categorical observer may not mean that the rendering would be erroneous for all other observers. However, if we look at our pipeline, we see that the matching transformation is linear and is applied to display linear luminance, so its possible that an optical filter could be created in order to transform between a reference representation (CIE 2 degree, for example) and each categorical observer. Then, if the entire group is viewing the image through the filter corresponding to their category, everyone will perceive the correct rendering of the image.

Conclusion

A physiologically based individual observer model is introduced by Asano and Fairchild, as well as a method for separating a population of observers into a limited number of categories. Building on this work, we present a metameric match simulation pipeline that is computationaly simple and can be used practically in motion picture color management. Using this pipeline, we perform a simulation experiment with real display spectra and natural images to observe the variability which could occur among a population as a result of observer metamerism in a motion picture viewing scenario. The results provide further evidence that inter-observer metameric variability is a relevant problem in the context of natural images. Finally, we outline how this pipeline can be used in motion picture color management. For future work, a psychophysical experiment to verify the accuracy of the pipeline and the consistency of the use of categorical observers in a natural image viewing scenario will be conducted. Additional psychophysical experiments exploring the perceived degree of metamerism error in natural images would also be of interest to perform. Using this data, a more appropriate observer metamerism error metric based on image statistics, spectral differences, or a combination of the two could be derived.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 761544 (project HDR4EU) and under grant agreement number 780470 (project SAUCE), and by the Spanish government and FEDER Fund, grant ref. PGC2018-099651-B-I00 (MCIU/AEI/FEDER, UE).

References

[1] Asano Y, Fairchild MD, Blondé L (2016) "Individual Colorimetric Observer Model." PLoS ONE 11(2): e0145671. doi:10.1371/journal.pone.0145671

[2] Yuta Asano, Mark Fairchild. (2020). "Categorical observers for metamerism." Wiley Publishing. DOI: 10.1002/col.22493

[3] William David Wright (1928). "A re-determination of the trichromatic coefficients of the spectral colors." Transactions of the Optical Society. 30 (4): 141–164. doi:10.1088/1475-4878/30/4/301.

[4] John Guild (1932). "The colorimetric properties of the spectrum." Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character. 230: 149–187.

[5] David L. Long and Mark D. Fairchild (2014). "Modeling Observer Variability and Metamerism Failure in Electronic Color Displays." 22nd Color and Imaging Conference.

[6] Yuta Asano, Mark D. Fairchild, Laurent Blondé, Patrick Morvan. (2014). "Observer Variability in Color Image Matching on a LCD Monitor and a Laser Projector." 22nd Color and Imaging Conference Proceedings.

[7] Abhijit Sarkar, Laurent Blondé, Patrick Le Callet, Florent Autrusseau, Patrick Morvan, Jürgen Stauder. (2010). "Toward Reducing Observer Metamerism in Industrial Applications: Colorimetric Observer Categories and Observer Classification." 18th Color and Imaging Conference Proceedings.

[8] Patrick Morvan, Abhijit Sarkar, Jurgen Stauder, Laurent Blondé, Jonathan Kervec, Hasan Sheikh Faridul. (2011). "A Handy Calibrator for Color Vision of a Human Observer." ICME 2011, Barcelona, Spain.

[9] Hao Xie, Susan P. Farnand, Micheal J. Murdoch. (2020) "Observer metamerism in commercial displays." Journal of the Optical Society of America. DOI: 10.1364/JOSAA382228