(How) Do Observer Categories Based on Color Matching Functions Affect the Perception of Small Color Differences?

***Maria Fedutina*[1]*, Abhijit Sarkar*[2,3]*, Philipp Urban*[1]*, and Patrick Morvan*[2]**
[1] ***Institute of Printing Science and Technology, Technische Universität Darmstadt, Magdalenenstr. 2, 64289 Darmstadt, Germany***
[2] ***Technicolor Research, Rennes, France***
[3] ***LUNAM Université, Polytech'Nantes, IRCCyN UMR CNRS 6597 (Institut de Recherche en Communications et Cybernétique de Nantes), Rue Christian Pauc, BP50609, 44306, Nantes, France***
Corresponding author sarkar@abhijitsarkar.com

Abstract

In this paper we investigate the impact of colorimetric observer categories on the prediction of the average suprathreshold color difference perception. The observer categories were obtained from an observer classification experiment, while the color difference data were obtained from an experiment involving a liquid crystal display (LCD) with fluorescent backlight. The same observer panel with normal color vision participated in both experiments. Results obtained from the observer classification experiment were consistent with the average observer threshold for color difference judgment. This analysis demonstrates that the observer categories, determined based on individual differences in cone spectral sensitivities (and thus color matching functions), have an influence on the prediction of average suprathreshold color difference perception for a given observer population.

Introduction

Do color matching functions (CMFs) influence our perception of small suprathreshold color differences? If yes, color difference perception would be affected by the spectral power distribution of the color stimulus that is completely discounted for predicting color differences today. Relying on previous studies on observer metamerism, i.e. colors that match for one observer and deviate for another, this would be especially apparent for narrowband stimuli [1]. Such stimuli become more and more important due to the increasing use of energy saving lamps, e.g. fluorescent lamps or light emitting diodes (LED) particularly utilized in modern displays.

Some reports in literature yield to our hypothesis that the perception of small color differences is influenced by individual CMFs. Kuehni stated that there are highly reliable observers that deviate significantly from the average in their perception of small color differences [2]. He called such observers *"extreme observers"*. Furthermore, he evaluated a color-difference experiment conducted by Mangine [3] and found that interobserver variability is highest at threshold differences and declines if the distance becomes larger [2]. Threshold discrimination (also refered as just noticeable distance (JND)) is a measure of uncertainty and variability. Typically it is determined by color matching, which is highly affected by individual CMFs. If the color difference becomes larger (suprathreshold) we assume that the influence of the CMF on the perceived color difference decreases continuously, i.e. there is some impact of CMFs on perceived small suprathreshold color differences in addition to higher order processes. Proving this assumption is difficult, because it requires a color difference experiment where the spectral stimuli are recorded and CMFs of the observer panel are known. To our knowledge only colorimetric values are provided in previous color difference experiments. Furthermore, the observer panel is typically not available any more to evaluate individual CMFs.

Therefore, we conducted two experiments in this work: 1. A color matching experiment to classify observers into different CMF categories according to Sakar et al. [4] and 2. a color difference experiment with the same observer panel allowing us to correlate individual color difference judgments with CMF categories.

We chose a liquid crystal display (LCD) with fluorescent backlight for our color difference experiments because we assume that a possible impact of color matching functions is highest for narrowband stimuli resulting from the display's primaries [1, 5].

In the following we describe both experiments as well as our approach to visualize the relationship of CMF categories with the corresponding color-difference judgments.

Color Difference Experiment

The first part of our investigation was a color-difference experiment on a liquid crystal display (LCD). We used an *EIZO ColorEdge 301W* LCD with fluorescent backlight for this experiment. Using a hardware calibration the luminance was set to 120 cd/m^2 and the gamma to 2.2. The LCD was calibrated to the D65 white point. The calibrations were done with the X-Rite i1 spectrophotometer. For accurately displaying the stimuli we considered the warm-up time, colorimetric stability, calibration performance and spatial display uniformity.

The colorimetric characterization of the display was performed based on a method similar to that of Day et al. [6] using test colors measured by a Konica Minolta CS1000A spectroradiometer. This device was also used to measure all stimuli shown to observers. The difference between measured colors and colors used for the subsequent evaluation was in average $\Delta E_{00} = 0.31$.

We used the *method of constant stimuli* [7] to determine color differences around five CIE color centers (CIE Gray, CIE Red, CIE Yellow, CIE Green, and CIE Blue [8]). These color differences are perceived equally to the color difference in the anchor pair consisting of two neutral gray tones with a color difference of $\Delta V = \Delta E^*_{ab} = \Delta L^* = 2.2$. 14 directions around each color center were investigated (Figure 1). Along each direction five test colors were chosen resulting in $14 \times 5 = 70$ color comparisons for each color center and a total of 350 comparison for the whole experiment.

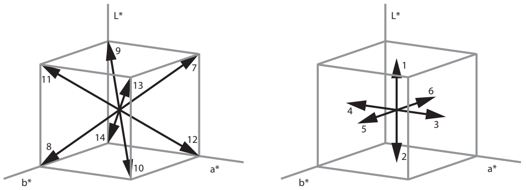

Figure 1. *Investigated directions around the color centers [9]*

The arrangement of patches is shown in Figure 2. The software to run the experiment was developed in MATLAB utilizing the Psychophysics Toolbox Version 3 [10, 11]. The observer was sitting in front of the LCD having his head placed in a chin rest 70 cm away from the monitor. The display was the only light source in the black-painted room. To let the observer adapt to the monitor's white point a white border ((L*, a*, b*) = $(100,0,0)$) was displayed. The color pairs were placed on a neutral gray background ((L*, a*, b*) = $(50,0,0)$). Each test pair was composed of a color center and one of the test colors. The patches cover approx. 10° of the visual field. The positions of the displayed test and anchor pairs were switched randomly as well as the color positions within the pairs. To reduce the adverse effect of afterimages a random monochromatic noise was shown after each observer's choice for 1.5 seconds before the next patches appear. Observers were asked to choose the color pair (anchor or test pair) with the largest perceived color difference. A detailed description of the experiment as well as further evaluations can be found in Ref. [12].

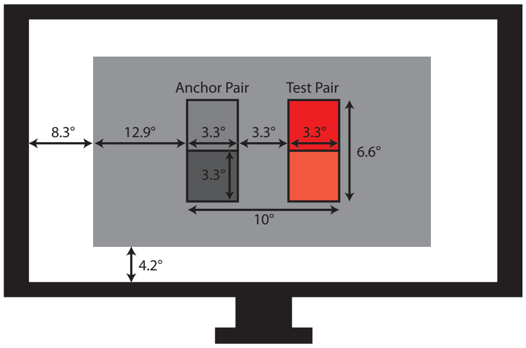

Figure 2. *Experimental setup on LCD. ° - degree of visual field*

Probit analysis is typically used to determine the T50 distance for each color center and direction. The T50 distance is the color difference from the color center that is judged by 50% of the observers to be smaller than the perceived color difference of the anchor pair and to be larger by the remaining 50% of the observers.

Since we want to determine an individual color difference threshold for each observer, probit analysis cannot be used. Therefore, we directly computed individual thresholds based on the binary choices of each single observer as shown in figure 3. Please note that this threshold is biased by quantization error due to the small number of binary choices.

Observer Classification Experiment

Observer Categories: A method for deriving seven distinct colorimetric observer categories was proposed at the 18th Color and Imaging Conference [4]. The method comprised of two steps. In the first step, five representative L, M and S cone fundamentals (a total of 125 combinations) were derived through a cluster analysis on the combined set of 47-observer data from 1959 Stiles-Burch study, and 61 color matching functions derived from the CIE 2006 model corresponding to 20-80 age parameter range. Squared Euclidean distance measure (in cone fundamental space) was used in this analysis, and thus was fundamental in nature. In the second step, a reduced set of seven representative observers were derived through an iterative algorithm, using several predefined criteria on perceptual color differences ΔE_{00} with respect to actual color matching functions of the 47 Stiles-Burch observers, computed for the 240 Colorchecker samples viewed under D65 illumination. Thus the goal was to come up with a minimum set of observer models that would satisfy all predefined color difference criteria for each of Stiles-Burch observers. The derivation of the reduced set of seven observer models is more applied in nature in comparison to the model cone fundamentals derived in the first step. However, the spectral power distributions of the Colorchecker samples under D65 are broadband in nature, and so are unlikely to manifest significant observer variability. For deriving the reduced set of observer models, a better dataset can be obtained from a color system with narrow-band primaries. It was decided that the new Observer Calibrator prototype, described in the next section, was the most appropriate device for this purpose, since it is capable of producing highly metameric color signals. Accordingly, the 240 stimuli used in the second step were replaced by 5832 estimated spectral power distributions obtained by using the LED primaries in the right half of the bipartite field of the prototype. These colors are characterized by high observer variability. As shown in Figure 4, these color samples cover a wide color gamut formed by the prototype primaries. Rest of the method to derive the reduced set of observer models was the same as before [4].

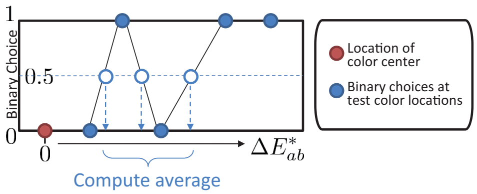

Figure 3. *Calculation of the individual threshold for a single observer. If there are multiple crossings of the 0.5 line the corresponding ΔE^*_{ab} distances are averaged for thresholding.*

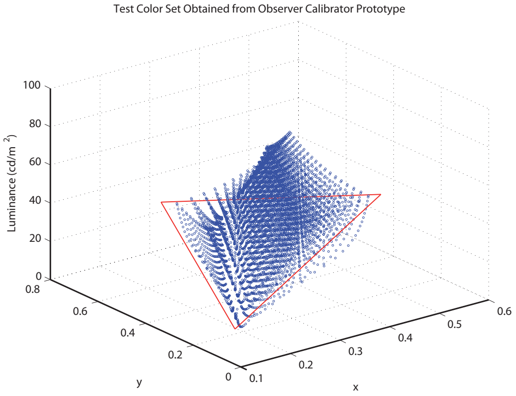

Figure 4. *Test color set obtained from Observer Calibrator Prototype*

Using the new set of test colors, a total of eight colorimetric observer categories were obtained. In these categories, there are four unique x- functions, three y- functions and four z- functions, with more variability in the x- functions than in others. These updated categories were used in the observer classification experiment described in this paper.

Observer Calibrator Prototype: A portable, LED-based instrument prototype has recently been developed, and was demonstrated at the 18th Color and Imaging Conference. This prototype replicates the observer classification experimental setup using two displays, one broadband and the other narrow-band, as described in [4].

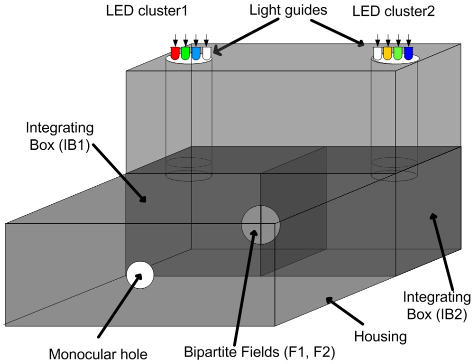

Figure 5. *Configuration of the Observer Calibrator Prototype*

The prototype configuration is shown in Figure 5. The illumination system in the prototype is composed of two clusters of four LEDs, two adjacent integrating boxed (IB1, IB2) for light mixing, an LED driver and a computer. Out of the four LEDs in each half-field, one is a white LED and is used only for generating an adaptation field. The colors can be viewed monocularly in the 10° bipartite field (F1, F2). The LEDs for the two fields were carefully selected such that the peak wavelengths of the LEDs in one field (F2) fall in the region of high variability in the observer categories, while those of the LEDs in the other field (F1) coincide with region of low variability in the observer categories. Thus, when observers look at different versions of color matches in the prototype, the left half of the bipartite field stays relatively constant, while the right half tends to change. Figure 6 shows the spectral power distributions of the observer calibrator primaries in two fields, and the color gamuts obtained from them.

Observer Classification Method: Nine matching colors are produced in each half of the 10° bipartite field corresponding to nine observer models, namely the CIE 10° standard observer and the eight reduced set of observer categories. The test software allows these nine versions of color matches to be presented in a random order in each trial, allowing the observer to browse through them with the help of a user control. His or her task is then to follow a multi-step method and classify these nine versions of color matches into superior, average or inferior categories. Based on several such trials (for different base colors), the category that most often produces the best match is identified, and is the category assigned to the given observer. In our experiment, there were eight trials for each observer. More details of the experimental method can be found in [4].

27 observers (10 female and 17 male with average age of 34.5 years) with normal color vision according to the Ishihara and Farnsworth D-15 tests participated in both experiments.

Results

In analyzing the data from the color difference experiment, an individual threshold was calculated for each observer. The color difference between the color center and the color, indicated by this threshold, is perceived by the current observer similar to the color difference of the anchor pair. Based on the individual thresholds an average observer was calculated for the observer panel. Figure 7 shows the mean deviation of the observers from average observer with respect to colorimetric categories. Note that here average thresholds and individual thresholds are calculated in the CIELAB color space. As shown in the diagram, two observers belonged to Cat. 1, one observer to category 3, eight observers to category 4, six observers to category 5 and 6 each, three observers to category 8 and one observer to category 9. No observer belonged to categories 2 and 7. Thus categories 4, 5 and 6 were most popular. The two observers belonging to category 1 are closer to the standard observer than others in this observer population.

Separate analysis was conducted for investigating the correlation between average observer color difference thresholds and observer categories. This analysis involved the use of CMFs for various categories, thus CIEXYZ color space was preferred over CIELAB since the conversion of CIEXYZ to CIELAB is valid only for 2° or 10° standard observer. On the other hand, CIEXYZ coordinate system is purely computational, and is not restricted to any specific CMF. The xyY chromaticity diagram is defined by the specific monochromatic primaries used in obtaining the original color matching functions. Since all our categories are essentially based on Stiles-Burch 10° CMFs, chromaticity coordinates obtained by using individual categories can be compared and even plotted on the same diagram. However, it is important to note that the distances in CIEXYZ color space are not perceptual, and the scale is not uniform in different areas of color space. At this point, a perceptual space for these categories does not exist, and so there is no appropriate perceptual metric available to us.

From the spectral power distributions (SPDs) of all test stimuli and the color matching functions (CMF) for each of the nine categories (category 1 being CIE 10° standard observer), CIEXYZ and category-specific XYZ (henceforth CatXYZ) values were computed. Since all observer thresholds were originally computed in CIELAB using the CIE 10° standard observer, these had to be converted to category-specific XYZ. For each category, a transformation matrix was computed in a least square sense from CIEXYZ and CatXYZ data of all color stimuli obtained before. Observers' average color difference threshold data were then converted from CIELAB to CIEXYZ, which were then converted to CatXYZ by multiplying with the transformation matrices. Eq (1) explains these two steps:

$$\begin{aligned} \mathrm{M_{Cat}} &= \mathrm{XYZ}^{-1}_{\mathrm{Stimuli,Cat}} \cdot \mathrm{XYZ}_{\mathrm{Stimuli,Std}} \\ \mathrm{LAB}_{\mathrm{ObsAv,Std}} \longrightarrow \mathrm{XYZ}_{\mathrm{ObsAv,Std}} &\xrightarrow{\mathrm{M_{Cat}}} \mathrm{XYZ}_{\mathrm{ObsAv,Cat}} \end{aligned} \tag{1}$$

These computations allow us to plot observer data organized by categories. In this analysis, root-mean-square (RMS) distances between XYZ coordinates of observer thresholds and color centers have been considered, with the hypothesis that around a given color center, small color differences in a given direction can be assumed to be Euclidean. Figure 8 shows the RMS distances between the test colors along various directions and a given color center. The central line

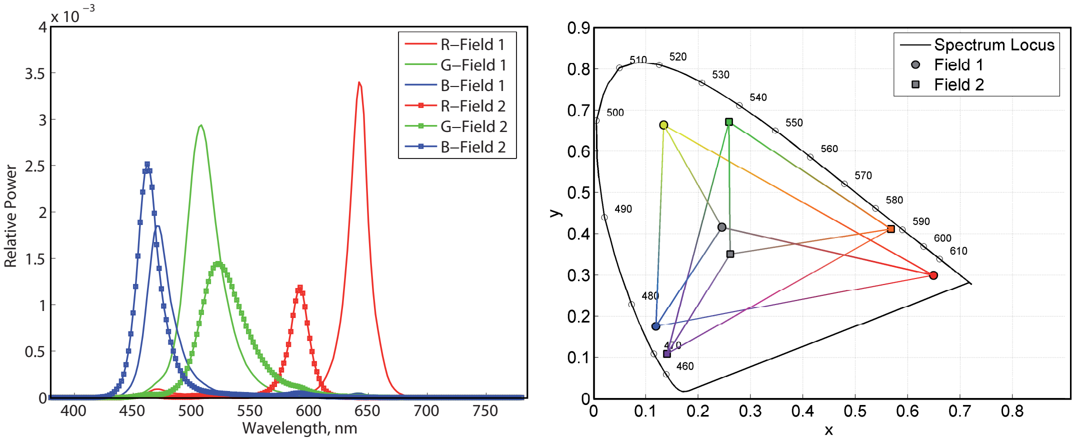

Figure 6. *Spectral Power Distributions of Observer Calibrator Primaries (left) and the gamuts of two fields (right); Field 1 has low and Field 2 high observer variability*

is the color center. All distances are measured from this color center and represented on two sides of the line. The fourteen directions (refer to Figure 1) are organized in pairs along the ordinate, with each direction having seven colored lines corresponding to various categories. Note that the first line is for CIE standard observer. The rest are for categories 3, 4, 5, 6, 8 and 9. Each colored line joins the five test stimuli in a given direction for any given category, shown as black dots. As mentioned before, the RMS distances in this figure are not perceptual. But conveniently, comparing the lengths of these lines gives an idea of the relative distance scales in various directions and categories for a given color center, thus allowing us to compare the RMS distances of observer thresholds.

The empty circles represent the RMS distances between color difference thresholds averaged over all observers and the color center. The average over all observers were obtained by multiplying the within-category averages by the respective number of observers belonging to the categories, and then summing and dividing by the total number of observers. Such scaling took into account the fact that the categories were not equally populated. For the first colored line in each direction, they represent CIEXYZ RMS distances (Cat 1), while for the rest they represent the RMS distances in respective CatXYZ spaces. The filled circles represent similar RMS distances where color difference thresholds are computed for observers grouped by their assigned categories. So the filled circles in the first line are for observers belonging to Cat 1, so on and so forth. Finally, the blue star on the first line are for color difference thresholds computed only for observers belonging to the three dominant categories, namely 4, 5 and 6. As before, weightings based on number of observers were applied.

In a perceptual color space optimized for each category the average color difference thresholds (filled circles) would ideally form a vertical line, which is not the case here. The transformations between CIEXYZ and CatXYZ are approximate. For these reasons, threshold points for some categories do not always fall on the colored lines, implying the RMS distances in CatXYZ space can in some cases exceed the distance of farthest test stimulus. The distances between the global average observer thresholds mapped to various categories (empty circles) and average thresholds within categories (filled circles) indicate which category is further from the averaged observer data. For example, typically categories 3, 8 and 9 have the largest distances. In the observer classification experiment, categories 8 and 9 rejected color matches corresponding to the CIE standard observer with high certainty, for all seven test colors. This bolsters the inference that these categories are indeed quite different from the standard observer. In this experiment, all observers belonging to these two categories were in the highest age-group, but other experiments (yet to be reported) have indicated that some young observers can also belong to these categories. In case of category 3, distances from the color center are generally less than that in case of other categories, which indicates the observer (only one) had better color discrimination than average observers in other categories. However, as per Figure 7, this observer had the highest deviation from the mean color difference threshold, which is also consistent with Figure 8.

In many cases, RMS distances between global average thresholds and average thresholds within category 1 (CIE standard observer) are larger than those in case of categories 4, 5 and 6, which indicates observers belonging to category 1 are relatively further away from the average observer data. This indicates that the perception of such individuals can still be strongly distinct from the statistical mean of a certain observer population. On the other hand, color difference thresholds averaged for observers in categories 4, 5 and 6 (blue stars), are in general significantly closer to the global average (empty circles). Over 70% observers belonged to these three categories.

The distances between global average thresholds and average thresholds within categories are typically larger for categories 3, 8 and 9, compared to other categories. This implies that there exists possibility to improve average color difference prediction for observers belonging to these categories by using color matching functions that are more appropriate than the standard observer. These observers stand to gain the most by a practical implementation of the concept of observer classification. For example, we can consider the absolute difference between global average thresholds and average thresholds within categories for the blue color center. These differences are relatively significant for categories 3, 8 and 9 along several directions. Examples are the distances along directions 5, 7 and 8 for category 3, along directions 1, 2, 7 and 8 for category 9, and to a lesser extent, along directions 2 and 5 for category 8. These relatively large distances are an indication that observers in these categories will tend to have high disagreement in color difference judgment in blues with the rest of the population.

It is also interesting to note that in Figure 7, the two category 1 observers are in the middle of the graph, indicating their deviation from the average is closer to the mean of all observers. Categories 4, 5 and 6 observers are spread above and below these two observers, while those belonging to categories 8 and 9 are located widely apart from category 1. Data corresponding to category 3 observer is markedly different from all others. This is consistent with earlier observation that observer data corresponding to categories 3, 8 and 9 can be well distinguished from those of other categories.

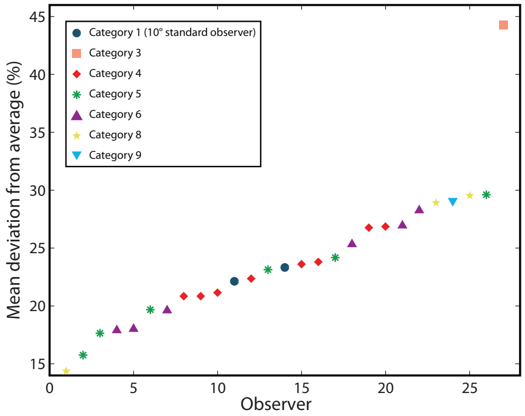

Figure 7. *Mean deviation from average observer*

Conclusions

Two main inferences emerge from the correlation analysis of observer classification data and color difference judgments. Firstly, color difference thresholds for categories that are very different from the CIE standard observer, as indicated by the observer classification results, have large differences from the averages of observer thresholds. Secondly, average thresholds for observers belonging to dominant categories are generally very close to the observer thresholds averaged over the whole population. The consistency between observer categories and color difference data lead us to conclude that colorimetric observer categories, derived from classical color matching data, have an influence on average suprathreshold color difference perception for a given observer population. The extent of this influence needs further investigation, requiring additional visual data and appropriate metrics. But this preliminary study opens up an important issue for future discussion: is it possible to customize color difference equations for individual observer categories, and even derive more uniform color spaces for these categories? More importantly, can we use our knowledge of observer categories to derive a better representative average observer from a limited amount of visual data? Can this help us to establish perceptually more uniform color spaces, and simpler color difference equations, than what is possible today with a single standard observer?

References

[1] M. D. Fairchild and D. R. Wyble. Mean Observer Metamerism and the Selection of Display Primaries. In *IS&T/SID, 15th Color Imaging Conference*, pages 151–156, Albuquerque, New Mexico, 2007.

[2] R. G. Kuehni. Variability in estimation of suprathreshold small color differences. *Color Research and Application*, 34(5):367–374, 2009.

[3] H. N. Mangine. Variability in experimental color matching conditions: Effects of observers, daylight simulators, and color inconstancy, 2005. PhD. thesis. Ohio State University, 2005.

[4] A.Sarkar, L. Blondé, P. Le Callet, F. Autrusseau, J. Stauder, and P. Morvan. Toward Reducing Observer Metamerism in Industrial Applications: Colorimetric Observer Categories and Observer Classification. In *Color and Imaging Conference 2010*, 2010.

[5] A. Sarkar, L. Blondé, P. Le Callet, F. Autrusseau, P. Morvan, and J. Stauder. A color matching experiment using two displays: design considerations and pilot test results. In *CGIV*, pages 414–422, Joensuu, Finland, 2010.

[6] Ellen A. Day, Lawrence Taplin, and Roy S. Berns. Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display. *COLOR research and application*, 29, 2004.

[7] Ethan D. Montag and David C. Wilber. A Comparison of Constant Stimuli and Gray-Scale Methods of Color Difference Scaling. *COLOR research and application*, 28, 2003.

[8] CIE Publication No. 101. Parametric Effects in Colour-difference Evaluation. Technical report, Central Bureau of the CIE, Vienna, Austria, 1993.

[9] R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. R. Balonon-Rosen. Visual determination of suprathreshold color-difference tolerances using probit analysis. *Color Research and Application*, 16(5):297–316, 1991.

[10] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10, 1997.

[11] Psychtoolbox. Available from: http://psychtoolbox.org.

[12] P. Urban, M. Fedutina, and I. Lissner. Analyzing small suprathreshold differences of LCD-generated colors. *Journal of the Optical Society of America A*, 28(7):1500–1512, 2011.

Author Biographies

Maria Fedutina received her Printing Science diploma from Moscow State University of Printing Arts (Russia) in 2007 and her MS degree in Paper Science and Technology from the Technische Universität Darmstadt (Germany) in 2010. She currently is a research assistant and doctoral candidate at the Institute of Printing Science and Technology of the Technische Universität Darmstadt (Germany).

Abhijit Sarkar received his bachelor's degree in electrical engineering from India, and two MS degrees specializing in lighting and color science, from the Pennsylvania State University, USA and the Rochester Institute of Technology, USA, respectively. He is currently a PhD student at Technicolor Research, Rennes, France, and is affiliated to the Ecole Polytechnique de l'université de Nantes, France. His research interests include digital color imaging, color vision and perception.

Philipp Urban has been head of an Emmy-Noether research group at the Technische Universität Darmstadt (Germany) since 2009. His research focuses on color science and spectral imaging. From 2006-2008 he was a visiting scientist at the RIT Munsell Color Science Laboratory. He holds a MS in mathematics from the University of Hamburg and a PhD from the Hamburg University of Technology (Germany).

Patrick Morvan received his Electronic Engineering diploma from Polytech' Nantes (France) in 1989. Since then he has worked in Thomson/Technicolor Research Labs in Rennes. His work is now focused on color management topics for the motion picture industry.

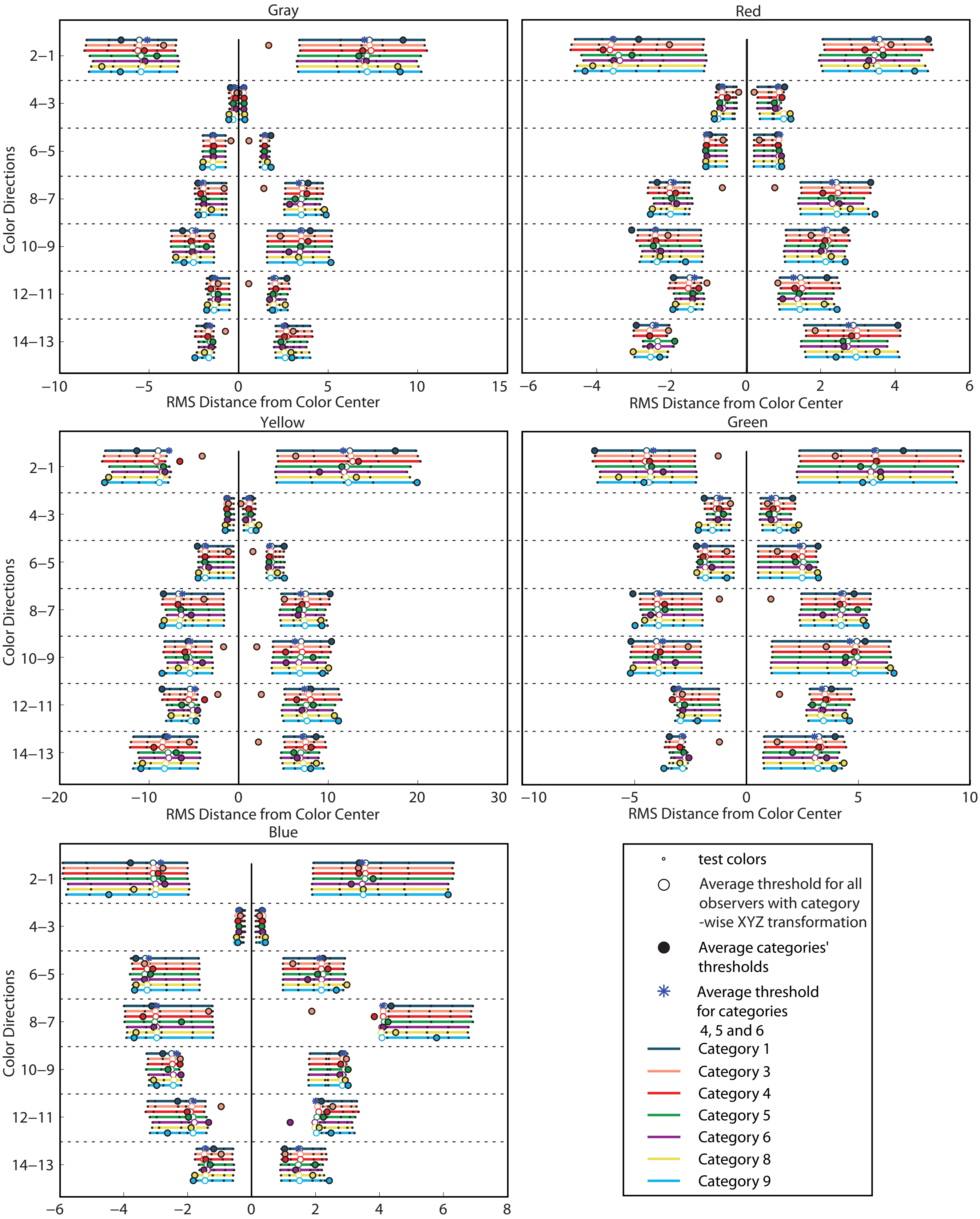

Figure 8. *RMS Distances in CIE XYZ*