Assessing Large Color Differences Using 3 Step Color Series

Rafael Huertas, María J. Rivas, Manuel Melgosa; Department of Optics, Faculty of Sciences, University of Granada, Spain

Abstract

We report on the results of a visual experiment using forty-nine 3 step color series supplied by the Federal Institute for Materials Research and Testing (BAM, Berlin), which were assessed 3 times under a D65 source by a panel of 20 observers with non-defective color vision. This experiment tries to contribute to current CIE TC 1-63 work, analyzing the performance of 7 advanced color-difference formulas (CIELAB, CIE94, CIEDE2000, DIN99d, CAM02-LCD, CAM02-UCS, and OSA-GP) for color pairs with large color differences in a very wide range: 10-95 CIELAB units. After earlier results, boxplot analyses detected that some answers from specific observers (9.2% of the whole results) were outliers. These answers were removed, and a total of 2669 visual assessments were managed. In overall the CIELAB formula provided the best predictions of our visual results (53.5 STRESS units), followed by the CAM02-LCD formula, the average inter-observer accuracy in our experiment being 43.6 STRESS units. These STRESS values are greater than those usually found in experiments involving small-moderate color differences, like the ones conducting to the proposal of the CIEDE2000 formula.

Introduction

The magnitude of a color-difference has been largely considered a relevant parametric factor affecting the visual evaluation of color differences [1]. Thus, the two last CIE-recommended color-difference formulas, CIE94 [2] and CIEDE2000 [3], were proposed for color-differences in the range 0-5 CIELAB units. However, larger color-differences are also interesting, particularly in color reproduction and other industrial applications. Currently, CIE TC 1-63 is analyzing the validity of the CIEDE2000 formula predicting visual assessments of color differences with different magnitudes. The goal of this paper is to report on the results of an experiment using 3 step color series, which involves color differences in the range 10-95 CIELAB units, analyzing the predictions of the visual results by different advanced color-difference formulas: CIELAB, CIE94, CIEDE2000, DIN99d [4], CAM02-LCD [5], CAM02-UCS [5], and OSA-GP [6].

Methods

The color samples employed in our experiment come from a set of 3 and 5 step color series which were carefully printed at the Federal Institute for Materials Research and Testing (BAM) in Berlin, and are now being used by different laboratories involved in CIE TC 1-63 activities. Each one of the 3 step color series is composed by 3 adjacent samples (Figure 1) with a size of 1.5 x 1.5 cm, which are positioned in the horizontal direction on a neutral background (L^*=49.85, a^*=-3.17, b^*=0.95) with dimensions 7.5 x 13.0 cm. The left and right color samples in our 3 step color series, which had fixed scores of 0 and 10 respectively, were black, white, or one of the primary or secondary colors (according to ISO/IEC 15775:1999) with maximum or intermediate chroma.

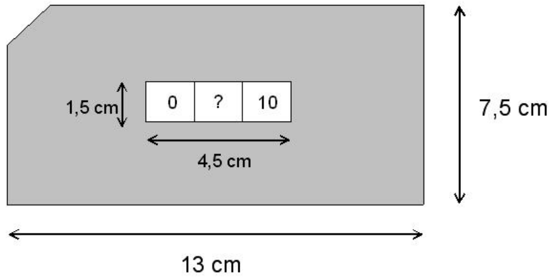

Figure 1. *Scheme of one of the forty-nine 3 step color series employed in our experiment. Visual assessments assigned a score to the central sample (indicated with a question mark), after comparison of the central-left and central-right color differences.*

Figure 2 shows the forty-nine 3 step color series employed in our experiment, ordered in 8 groups according to the similar quality of the left and/or right color samples in each group. For example, the left sample in all 3 step color series in group 1 is white, while in group 2 it is black, etc. Note that the 3 step color series in group 4 had very different hues. It is also noticeable that color differences between the left and right samples are lower in groups 5 to 8 than in the remaining ones. Each group had six 3 step color series, except group 2. The 3 step color series were assessed by a panel of 20 non-defective observers (12 males and 8 females, staff and graduate students in our Department). All the assessments were made in a portable VeriVide cabinet with a D65 source. The illuminance at the plane of the samples was 1200 lx, and the distance sample-observer 50 cm, approximately. Instrumental color measurements of samples were performed with a PR-704 spectroradiometer positioned at the same position than the observers. Each one of the forty-nine 3 step color series was assessed at a time, without time limit, in random order. Each observer replicated 3 times the whole experiment in different days. The observer's task was to compare the color difference between the central-left and central-right samples in each one of the 3 step color series, scoring the central sample between 0 (left) and 10 (right). For example, a visual score of 5 means that these two color differences were considered visually identical by the observer.

Both the visual and computed data considered in our next analyses were the ratio of the color differences between central-left and central-right samples. This transformation is necessary because the color differences between left and right samples in our 3 step color series change from group to group, and also inside some groups. Thus, initial visual scores of 1, 2, 3, 4, 5, 6, 7, 8, and 9 were transformed to 1/9, 2/8, 3/7, 4/6, 5/5, 6/4, 7/3, 8/2 and 9/1, respectively. Analogously, instrumental color differences between pairs of samples were transformed to ΔE_R ratios using the next equation:

$$\Delta E_R = \frac{\Delta E(Central - Left)}{\Delta E(Central - Right)} \quad (1)$$

where ΔE is the color difference computed from any of the color-difference formulas mentioned before.

From initial 2940 visual assessments (49 series x 20 observers x 3 replications), boxplot analyses detected some moderate and severe outliers, mainly coming from 4 specific observers, which were removed. A total of 2669 assessments have been considered in our next analyses.

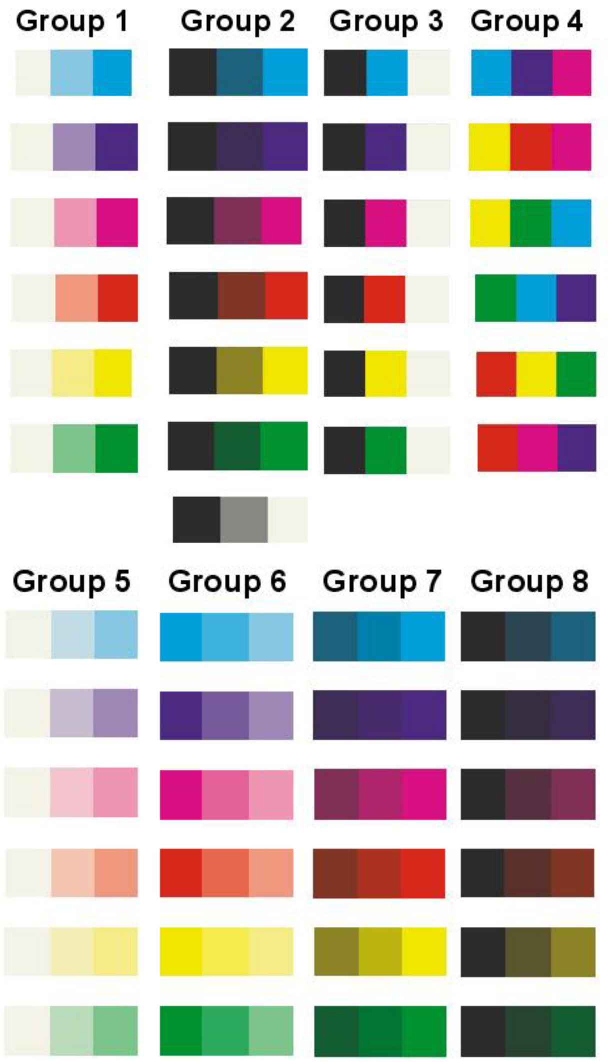

***Figure 2.** Approximate color reproduction of the forty-nine 3 step color series employed in our experiment, distributed in 8 groups according to common qualities of the left and/or right color samples.*

The STRESS index [7] was used to measure both the performance of different color-difference formulas with respect to the visual results, and the inter-observer variability. Zero STRESS means perfect agreement with visual results or null variability, respectively. The STRESS index is given as a percentage in the range 0-100, and is defined by the next equations:

$$STRESS = 100\left(\frac{\sum(\Delta V_i - F\,\Delta E_i)^2}{\sum \Delta V_i^2}\right)^{1/2} \quad (2)$$

$$F = \frac{\sum \Delta E_i \Delta V_i}{\sum \Delta E_i^2} \quad (3)$$

where ΔVi and ΔEi are visual and computed ratios of color differences, respectively, for the i=1…N 3 step color series. The statistical significance of the difference between any two color-difference formulas has been also tested here using STRESS values [7], assuming a 95% confidence level.

Results and discussion

Inter-observer accuracy (average of the visual differences between each observer and the mean) was 43.6 STRESS units in our experiment. This value is higher than usual ones found in experiments involving small to moderate color differences. This fact may be considered a consequence of the difficulty of the current experiment: in particular, the performance of the experiment was not easy for the 3 step color series in group 4 even for our most experienced observers.

Figure 3 shows STRESS values found for the different color difference formulas considered in this work for the whole experiment. The lowest STRESS value (best performance) was 53.5 and corresponds to the CIELAB formula. The CAM02-LCD color difference formula, with a STRESS value of 55.5, is not significantly different than CIELAB. The remaining color difference formulas have similar slightly higher STRESS values, and are not significantly different amongst them. It can be noted that the STRESS values found for these advanced color difference formulas predicting other experimental results of small-moderate color differences (0-5 CIELAB units, approximately) are in the range 20-35 STRESS units [8,9]. This means that predictions of results in current experiment are considerably worse than for small-moderate color differences. This is not a surprising result, because all these advanced color-difference formulas were not developed from experimental datasets with very large color differences like the ones involved in our current experiment.

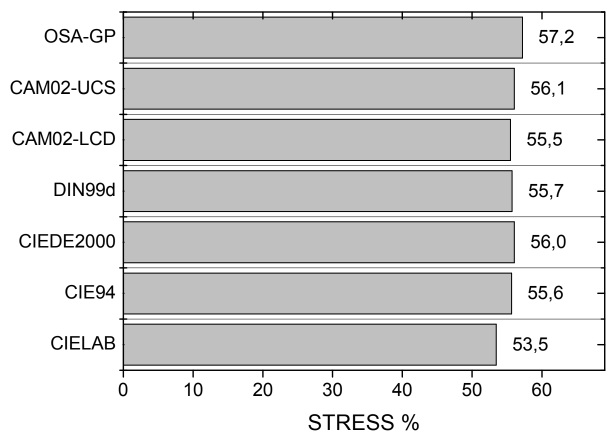

***Figure 3.** Percentage STRESS values for 7 color difference formulas with respect to the 2669 visual assessments in the current experiment.*

It can be added that the good results achieved by CIELAB in our current experiment have been confirmed by results found at the LCAM laboratory of the University of Liberec (Czech Republic) for 17 normal observers who performed 5 replications and used the same 3 step color series [10]. Results from the University of Leeds [11,12] indicate that CIELAB improves CIE94 and CIEDE2000 for color-differences in the range 20-40 CIELAB units, in agreement with results found at BAM (Berlin) [13], and the performance of CIELAB slightly decreases for color differences in the range 40-110 CIELAB units. Researchers from the University of Leeds have reported that the CIEDE2000 color-difference formula performs best for a whole range of color differences including small (<0.5 CIEDE2000 units), medium (around 3 CIEDE2000 units) and large (larger than 25 CIELAB units) color differences, but for large color differences CIELAB improves CIEDE2000 [12].

Figure 4 shows STRESS values for the 8 groups of 3 step color series considered in our experiment. In agreement with results shown in Figure 3, the CIELAB formula provides the best results in almost all groups, and results achieved by the remaining color-difference formula are very close (perhaps with the exception of groups 2, 3, and particularly group 4). Considerably high STRESS values (bad performance) were found for all formulas for the 3 step color series in group 4 where there were strong hue changes and very large color differences which made observer's task considerably difficult. The smallest STRESS values were found for groups 5, 6 and 7 where average color differences are smaller. In particular, some formulas in group 6 achieved STRESS values lower than 40, which approach to the 20-35 STRESS units usually found for datasets involving small-moderate color differences [8,9].

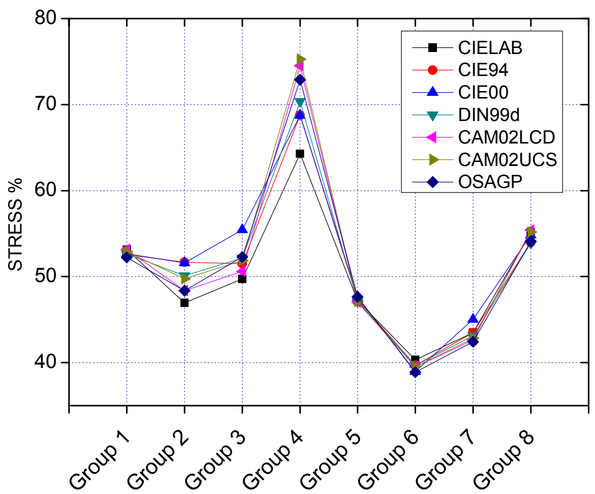

Figure 4. Percentage STRESS values for the 7 color-difference formulas considering the 3 step color series in each of the 8 groups (see Figure 2).

Our first results without removing outliers observers' answers showed that for color differences in the range 40-80 CIELAB units (average of central-left and central-right samples in the 3 step color series) all formulas performed considerably worse than for smaller color differences, and the CIELAB formula achieved the best results (lowest STRESS). These results are in agreement with findings from other laboratories [11,12]. The average size and standard deviation of the CIELAB color differences in the 3 step color series (average of central-left and central-right samples) for the 8 groups are shown in Table 1. It can be noted that largest average color differences are in group 3 and do not correspond with the highest STRESS values, which are achieved for samples in group 4 (see Figure 4). It can also be noted that for the dark samples in group 8 the STRESS is considerable higher than for samples in groups 5, 6 and 7, while the average CIELAB color difference is similar in all these four groups. In summary, the results of our experiment do not provide a clear relationship between the size of the color differences and the performance of the formulas (STRESS values).

Table 1. CIELAB color differences for samples in the 3 step color series in each of the 8 groups (see Figure 2)

Series	Average ΔE^*_{ab}	St. Dev. ΔE^*_{ab}
Group 1	38.2	3.5
Group 2	35.6	9.2
Group 3	73.9	11.3
Group 4	68.5	9.4
Group 5	20.4	2.0
Group 6	19.0	6.2
Group 7	18.4	1.5
Group 8	17.5	5.3

After removing outliers, Figure 5 shows again that there is not a clear relationship between STRESS values and average size of assessed color differences, considering the whole range of 10-95 CIELAB units available in our experiment.

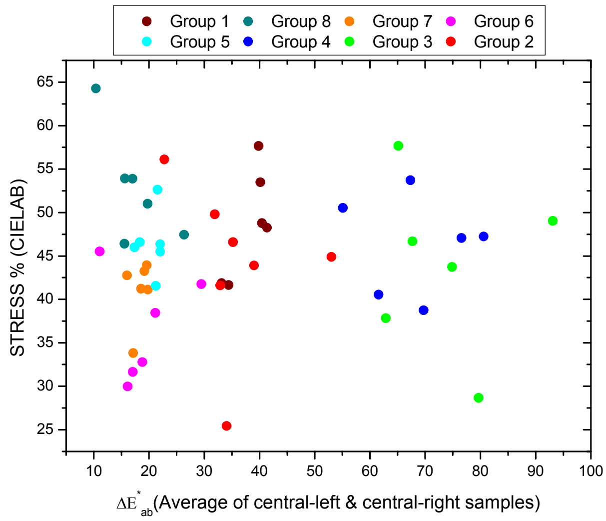

Figure 5. Percentage STRESS values for the CIELAB formula with respect to the average size of the color differences. Results for all 3 step color series in each group are distinguished.

Figure 5 also illustrates that there are important differences amongst results for the 8 groups. In agreement with previous literature [11,12] and earlier rough results of our current experiment (i.e results without removing outliers), we feel that in the range 40-80 CIELAB units all color difference formulas perform worse than for smaller color differences, but this question requires more detailed studies in the future.

Conclusions

1. In the experiment carried out with 3 step color series the CIELAB color-difference formula provides the best predictions. At 95% confidence level the CIELAB formula is statistically significant better than any of the other formulas, except CAM02-LCD. The remaining formulas analyzed (CIE94, CIEDE2000, DIN99d, CAM02-UCS and OSA-GP) perform identically in this experiment. This good performance of CIELAB for very large color differences is consistent with previous literature. The good performance of the CAM02 formulas is a promising result for future color difference research.

2. Both the average inter-observer accuracy (43.6 STRESS units) and the best predictions achieved in this experiment (53.5 STRESS units for the CIELAB color-difference formula) are considerably higher than in previous experiments using small to moderate color differences (let's say color differences in the range 0-5 CIELAB units), like the ones conducting to the last CIE recommended formulas CIE94 and CIEDE2000. This is not a surprising result because these formulas were not developed from experimental datasets with very large color differences, like the ones in the current experiment.

3. Our experiment does not provide clear information on changes in the performance of color-difference formulas with the size of the color differences. It seems that all color-difference formulas become worse for color differences above 40 CIELAB units, approximately. More research should be useful to clarify this point.

4. The next valuable comment [14] on our current research shows the special characteristics of the experiment described in this paper: "This experiment somewhat resembles the original OSA-UCS experiment … I don´t think it is possible (already Judd has shown some data in this respect) that the small difference metric can apply to very large differences. There you get into categorical differences rather than differences within a category as in small difference data. The categorical nature of large hue differences brings in new cognitive aspects that large lightness and chroma differences (except in yellow) usually do not".

Acknowledgments

To Dr. Klaus Richter (BAM), CIE TC 1-63 chairman, who provided the color samples employed in this experiment and useful advice. We also thank you very much valuable discussions and ideas from many colleagues, in particular Dr. Pedro A. García (University of Granada, Spain), Dr. Michal Vik and Eng. Martina Viková (Technical University of Liberec, Czech Republic), Dr. János Schanda (University of Pannonia, Hungary), Dr. M. Ronnier Luo and Dr. G. Cui (University of Leeds, UK), and Mr. Rolf G. Kuehni (North Carolina University, USA).

We also thank you very much all the volunteer observers who kindly accepted to participate in our experiment.

This research was supported by Research Project FIS2007-64266, Ministry for Education and Science (Spain), with FEDER support.

References

[1] A. R. Robertson, "CIE guidelines for coordinated research on colour-difference evaluation," Color Res. Appl. 3, 149-151 (1978).

[2] CIE Publication No. 116, "Industrial colour-difference evaluation". CIE Central Bureau, Vienna, 1995.

[3] CIE Publication No. 142, "Improvement to industrial colour-difference evaluation". CIE Central Bureau, Vienna, 2001.

[4] G. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference formula," Color Res. Appl. 27, 282-290 (2002).

[5] M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. 31, 320-330 (2006).

[6] R. Huertas, M. Melgosa, and C. Oleari, "Performance of a color-difference formula based on OSA-UCS space using small-medium color differences," J. Opt. Soc. Am. A 23, 2077-2084 (2006).

[7] P. A. García, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," J. Opt. Soc. Am. A 24, 1823-1829 (2007).

[8] M. Melgosa, R. Huertas, R.M. García-Monlleó, M.J. Rivas, L. Robledo, R. Roa, M. García, J.A. García, E. Hita, J.M. Roa, M. Sánchez-Marañón, A. González, and P.A. García, "Relaciones entre diferencias de color percibidas y calculadas" in "Algunas cuestiones de ciencia" Book in Tribute of Prof. Manuel Quintanilla, Universidad de Zaragoza (España), ISBN 978-84-7733-921-2.

[9] M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulae using the Standardized Residual Sum of Squares index," J. Opt. Soc. Am. A. Accepted with mandatory revision on May 5, 2008.

[10] M. Vik, K. Richter, M. Viková, V. Juricková, and P. Kittelmann, "Validity test of the range of CIEDE2000 using BAM color visual uniform samples". Private communication at CIE TC1-63 working meeting hosted by Dr. K. Richter in Berlin on December 4, 2007.

[11] Private communication. "Report to TC1-63 validity of the range of CIEDE2000". Leeds University. 29 September 2007.

[12] H. Wang, G. Cui, and M. R. Luo, "Evaluation of colour difference formulae for different colour difference magnitudes". Department of Colour Science, University of Leeds, UK. Private communication.

[13] P. Kittelmann, "Visual assessment of large colour differences using 3 and 5 step color series," Ph.D. Thesis. Technical University of Berlin, 2005.

[14] R. G. Kuehni. Private communication. November 24, 2007.

Authors Biography

Rafael Huertas earned a Ph.D. in Physics in 2004 at the University of Granada (Spain). Currently he is Assistant Professor at the Department of Optics of the same university. His research interests include basic and applied colorimetry: color-difference formulas, color spaces, color in images, and food color.

María José Rivas obtained a M.S. in optometry from the University Complutense of Madrid (Spain). Since 1990 she has been in the Department of Optics at the University of Granada. She has conducted a number of training courses in optical technology and optometry. Her research interests are focused in color applications related to optometry and ophthalmology.

Manuel Melgosa, Ph.D. in Physics in 1989, is Associate Professor at the Department of Optics of the University of Granada and President-elect of the Color Committee of Spain (Optical Society of Spain). He is the head of the basic and applied colorimetry laboratory of this university: http://www.ugr.es/~basapplcolor/