

Variability In Visual Small Color Difference Assessment: What It Means for Color Difference Formula Performance

*David Hinks, Renzo Shamey, Lina Cardenas, Seung Geol Lee,
Rolf Kuehni, Warren Jasper
North Carolina State University
Raleigh, NC 27695*

Introduction

The majority of color-using industries require effective color difference prediction that accurately represents average perceptual assessments of the magnitude of color difference between two objects. Color difference formulas today are a critical tool for effective electronic color communication for color management in a product supply chain. Existing formulas are based on several different sets of perceptual data that have been established under various experimental conditions, using samples representing a diverse range of substrates and different groups of observers. In the textile industry, for example, the CMC (2:1) color difference formula is used as an international standard [1-2]. In 2001, the International Commission on Illumination (CIE) recommended the CIEDE2000 formula, replacing the lower performing CIE94 [3]. Luo et al. reported accuracy of prediction for several formulas against average data from a large aggregate visual data set using four separate experiments. The aggregate data set was established as a fundamental part of the experiment that resulted in establishment of CIEDE2000. Using the PF/3 performance method, a value of 67.4 for the CIEDE2000 formula vs. 62.1 for CMC (2:1) was reported [4]. This paper reviews attempts to date to field test the performance of the latest CIE formula, proposes root cause variables to be investigated, and reports on a new set of visual color difference experiments currently underway.

Field Testing of CIEDE2000

While the latest CIE formula is the highest performing equation when measured against the aggregate dataset, several independent experiments, each incorporating separate color difference pairs and visual observers, have not shown CIEDE2000 to perform statistically significantly better than CMC (2:1). In textiles, four unrelated field tests of CIEDE2000 vs. CMC (2:1) in three different countries resulted in a similar level of accuracy for the two formulas [5-8]. In the paints and coatings industry, initial work (e.g., by the Detroit Color Council) has also not shown conclusively that the new formula outperforms previous existing formulas [9]. Independent validation of any model is, of course, an essential first step to establishing confidence for its use in research fields or commercial applications. While the shortcomings of color difference equations in current industry standards are well known, industry-wide adoption of a major new protocol often takes many years to complete, and at considerable cost.

To date, no data in the literature appears to provide a definitive answer to the disparity between the theoretical performance and the actual performance of CIEDE2000.

However, the performance of a model is only as good as the noise in the data upon which it was developed. We contend that the disparity is likely due to one or more of the following factors:

- Significant variability in experimental methodology used in each of the field tests,
- Significant variability in the aggregate visual data set upon which CIEDE2000 was developed, including observer panels and judgment methods, substrates, and viewing conditions,
- Insufficient fit of the formulas to mean observer data,
- Errors in mathematical fitting of visual data set to specific geometries.

At recent meetings of the AATCC Color Measurement and Test Methods Committee (RA36), the following two questions were discussed: 1) *In practice*, does CIEDE2000 perform better than CMC (2:1) for critical areas of color space? 2) What performance improvement justifies adoption of the new formula? We also contend that a third question must be addressed: What is the maximum performance possible for a color difference formula?

According to the independent field testing conducted to date, the answer to question 1 is no [5-8], even when color centers were used for which standard formulas are known to perform poorly, namely a navy, a high chroma blue and a light and medium gray [7]. The root causes of the lack of agreement between theory and field-testing of CIEDE2000 must be ascertained before questions two and three can be addressed.

An Internationally Replicated Visual Color Difference Experiment

As part of a study funded by the U.S. Department of Commerce through the National Textile Center, we have developed the following working hypothesis: The disparity between CIEDE2000 performance and independent field testing is mainly due to large observer variability, and minimizing the uncertainty in the method of practical color difference assessment is necessary for establishment of a globally optimized color difference formula. To test the hypothesis, a set of experiments has been devised to determine the optimum psychophysical experimental method in terms of intra- and inter-observer variability. There are, of course, numerous variables that affect observer variability, and a series of sub-experiments are first required to determine the psychophysical method resulting in least variability. Once the optimized experimental conditions (for textile samples) have been established highly controlled replication experiments will be performed under identical conditions in different regions of the world (US, Europe and Asia).

One variable to be addressed is: Do naïve and expert observers differ in terms of intra- and inter-observer variability in perceptual color difference assessments? We recently reported that the intra-variability of assessments among naïve observers was not statistically significant between the first experimental run and after three repetitions of the experiment when using a conventional gray scale method for rating color differences of a set of textile samples [10]. Hence, no significant training of observers was detected, and comparison of data obtained from expert assessors for the same experiment showed

using naïve observers for color difference optimization was valid for the protocol employed.

Sub-experiments are also underway to determine the effect on variability of using 1) an anchor gray pair; 2) using large 2” by 2” uniform grayscale and sample pair comparisons; 3) a perceptually linear gray scale (not a geometric scale which is currently used in standard protocols), all using an identical observer panel. The final replication experiment will employ color centers throughout color space, but with a separate study in the blue region. We anticipate completing the experimental design and commencing the full replication study in summer 2007 (for more information see www.ntcresearch.org/current/FY2004/C04-NS11.pdf).

References

1. AATCC Test Method 173-1998, CMC: Calculation of small color differences for acceptability. AATCC Technical Manual, pp. 311-315, 2005.
2. ISO International Standard 105-J03:1995.
3. CIE Technical Report: Improvement to industrial colour-difference evaluation. CIE Publication No. 142-2001. Vienna: Central Bureau of the CIE, 2001.
4. Luo MR, Cui G, Rigg B. The Development of CIE 2000 Colour Difference Formula: CIEDE2000. *Color Res Appl* 2001; 26:340-350.
5. Aspland JR, Shanbhag P. Comparison of Color Difference Equations for Textiles: CMC (2:1) and CIEDE2000. *AATCC Review* 2004; 6: 26-30.
6. Gay J, Hirschler R. Field Trials for CIEDE2000 – Correlation of Visual and Instrumental Pass/Fail Decisions in Industry. In: 25th Session of the CIE. San Diego; 2003, ISBN 3 901 906 21, June 25-July 2.
7. Noor K, Hinks D, Laidlaw A, Treadaway G, Harold R. Comparison of the Performance of CIEDE2000 and DE_{CMC}. In: Proceedings of the AATCC International Conference and Exhibition. Greenville 2003; 140-147.
8. Gibert GM, Daga JM, Gilabert EJ, Valldeperas J. Evaluation of Color Difference Formulae. *Coloration Technol.* 2005; 121(3): 147-151.
9. Personal communication.
10. L. Cardenas, D. Hinks, R. Shamey, R. Kuehni, W. Jasper and M. Gunay, *Comparison of Naïve and Expert Observers in the Assessment of Small Color Differences between Textile Samples*, Imaging Science and Technology's 3rd European Conference in Color in Graphics, Imaging and Vision, Book of Papers, Leeds University, 2006.

Author Biography

David Hinks received his BS and PhD in Colour Chemistry from the University of Leeds, U.K. He is Associate Professor and Program Director of the Polymer and Color Chemistry Program at North Carolina State University. He is Academic Editor of *Coloration Technology*, and is a past Chair and Secretary of AATCC's Color Measurement and Test Methods Committee.