COLOR DISCRIMINATION MAY BE HUE AGNOSTIC: A PILOT STUDY

Tony Esposito, PhD

Philips Lighting Research NA, Cambridge, MA USA
Tony.esposito@philips.com

Abstract

Past studies have hinted that average color rendering measures, with a particular focus on average gamut area measures, cannot predict the color discrimination ability of highly structured spectra. A recent study by Esposito and Houser solidified this finding, and showed that a new measure, $R_d$—an error score based on the light-source induced hue transpositions of the FM-100 hue test—could reliably predict error scores. The current study optimized spectra using $R_d$ (and accompanying sub-indices) to determine if several individual transpositions, each with a different nominal hue, would result in different color discrimination ability. The results show that four separate stimuli, each of which transposes one cap at different hue angles, do not produce statistically different mean error scores. Assuming the FM-100 hue test is indicative of color discrimination ability and knowing that cap transpositions are a strong predictor of FM-100 error scores, the results suggest that color discrimination ability may not differ based on hue.

Keywords: colour rendering, colour discrimination, transposition, juxtaposition, Farnsworth-Munsell 100 hue test, light emitting diodes, FM-100 adjusted error score

1 Introduction

1.1 Background

In 1972, WA Thornton introduced the concept of color discrimination as an important aspect of color quality and defined it as “the extent to which the illumination allows the observer to discriminate among a large variety of object colors simultaneously viewed” [Thorton 1972]. To quantify this ability, Thornton proposed the Color Discrimination Index (CDI), computed as the area enclosed by the eight test color samples of the CIE CRI calculation [CIE 1995] in the CIE 1960 UCS. Thornton proposed using this gamut area as a way to predict color discrimination and suggested that as gamut area increases, so does color discrimination ability. Several studies show that this is not necessarily true for highly structured spectra (shark peak and valleys) [Mahler and others 2009, Royer and others 2012, Wei and Houser 2012].

The latter two studies [Royer and others 2012, Wei and Houser 2012] demonstrated the inability of available average indices, particularly average gamut indices, to predict the error scores of various Farnsworth-Munsell 100 hue tests (FM-100). A recent study by Esposito and Houser [2017] explored the color discrimination ability of 24 LED light spectra with strategically varied average fidelity, average gamut, and gamut shapes. Results showed that average gamut indices—CDI [Thorton 1972], FM-100 Gamut Area (FMG) [Boyce and Simons 1977], FMG$\text{CIECAM02}$, Gamut Area Index (GAI) [Rea and Freyssinier-Nova 2008], Color Quality Scale (CQS) $Q_6$ [Davis and Ohno 2010], and IES TM-30-15 $R_g$ [IES 2015]—all fail to reliably predict color discrimination (as measured with an adjusted FM-100 Total Error Score, $TES_{adj}$). Pairing each gamut index with a fidelity index increased model prediction, but models were still notably weak. The authors conclude that no measure of gamut is highly correlated with the results, and they propose a new measure of color discrimination ($R_d$) with strong predictive ability of error scores ($R^2 = 0.860$).

1.2 An explanation of $TES_{adj}$ and $R_d$

The FM-100 test is a hue discrimination test consisting of 85 colored caps, of gradually changing hue, presented in four separate test trays. The score for any individual cap “is the sum of the difference between the number of that cap and the numbers of the caps adjacent to it” minus 2 [Farnsworth 1957]. The standard Total Error Score (TES) is the sum of the error scores for each of the four test trays, and only considers the order of caps as arranged by a participant.
The adjusted FM-100 Total Error Score ($TES_{adj}$) is a modified Farnworth error score which considers the interaction of the light source Spectral Power Distribution (SPD) and test chip spectral reflectance distribution (SRD). $TES_{adj}$ reconciles the discrepancy between a light source-induced transposition and a transposition arranged by a participant performing the test. It is based on the assumption that a participant should not be attributed an error for correctly responding to a transposition caused by the light source. For example, if a light source causes the order of caps $29$-$31$-$30$-$32$, and a participant arranges these caps accordingly in their physical test tray, they would not be attributed an error (as opposed to an error score of 4 which would be assigned using the standard scoring procedure) (Figure 1). $TES_{adj}$ is the sum of the error score associated with each of the four trays of the FM-100 test:

$$TES_{adj} = \sum_{i=1}^{4} iES_{adj} = AES_{adj} + BES_{adj} + CES_{adj} + DES_{adj}$$

where

- $AES_{adj}$ is the adjusted error score for tray A;
- $BES_{adj}$ is the adjusted error score for tray B;
- $CES_{adj}$ is the adjusted error score for tray C;
- $DES_{adj}$ is the adjusted error score for tray D.

Tray A is nominally red to red-orange, B is yellow to yellow-green, C is green to green-blue, and D is indigo to indigo-magenta (Figure 2).

The Total Light Source Error Score ($R_d$) is a Farnsworth-type error score that is applied directly to the light source, and quantifies the number of cap transpositions that it causes. A hue transposition is computed in the $a'b'$ plane of the CIECAM02-UCS using the CIE $10^\circ$ standard observer. For example, if a light source causes the configuration $29$-$31$-$30$-$32$ (as in Figure 1), the light source is directly attributed an error score of four (4). An error score of zero (0) means the source causes no transpositions; one transposition is an error of 4, two transpositions an error of 8, three transpositions an error of 12, and so on. $R_d$ is the sum of the light source error scores for each of the 4 trays of the FM-100 test:

$$R_d = \sum_{i=1}^{4} R_{d,i} = R_{d,A} + R_{d,B} + R_{d,C} + R_{d,D}$$

where

- $R_{d,A}$ is the light source error score for tray A;
- $R_{d,B}$ is the light source error score for tray B;
- $R_{d,C}$ is the light source error score for tray C;
- $R_{d,D}$ is the light source error score for tray D.

1.2 Goals and hypotheses

An $R_d$ value of 4 indicates that a source spectra causes one transposition, but does not indicate in which hue the transposition occurs. Thus, the primary goal of this experiment was to determine if several individual cap transpositions, located at different hue angles, would cause significantly different error scores. The a priori hypothesis was that hue angle (or hue location) of the transposition would not cause significantly different mean error scores. Said another way, all single cap transpositions, regardless of their hue, will produce similar mean error scores.
2 Methodology

2.1 Lighting Conditions

Spectra were designed using an 11-channel spectrally tuneable LED source (Thouslite LEDcube). The goal of spectral optimization was to achieve spectra which met the following 8 conditions: \( R_{d,A} = 0, 4; R_{d,B} = 0, 4; R_{d,C} = 0, 4; R_{d,D} = 0, 4 \). One spectra did not transpose any caps \( (R_{d,A} = R_{d,B} = R_{d,C} = R_{d,D} = 0) \). One spectra transposed many caps, but exactly one in tray A \( (R_{d,A} = 4) \); one spectra transposed exactly one cap, located in tray B \( (R_{d,B} = 4) \); a final spectra transposed many caps, but exactly one transposition in each of trays C and D \( (R_{d,C} = R_{d,D} = 4) \). These four \( (4) \) spectra were used to deliver the eight desired stimuli (Figure 3, Table 1, APPENDIX A).

All spectra were designed to be a metameric match to Blackbody radiation at a Correlated Color Temperature (CCT) of 3500 K and produce an illuminance (E) of 600 lux at the base of the booth to permit direct comparison to results from Esposito and Houser [2017]. Actual measurements showed the following variations: \( E \pm 10 \) lux, \( CCT \pm 20 \) K, and all chromaticities within a 1-step MacAdam ellipse of a 3500 K Blackbody radiator.

2.2 Apparatus

All experiments were performed in a 71 x 71 x 45 cm viewing booth (Figure 4). The LEDcube rested in a 30 x 30 cm cutout in the top of the booth. The booth was painted with a high reflectance, approximately spectrally neutral white paint. Spectra were measured with a Jeti Specbos 1211 spectroradiometer.

The room containing the apparatus was dark during all experimental trials. A small reading light, separated from the experimental apparatus with blackout curtains, was used by the researcher for recording data and was never visible to participants.
2.3 Participants

Fifteen (15) people participated in this experiment. Three people were identified as having some form of color-blindness, as tested with the 14-plate Ishihara test, and their data were removed from analysis. The current analysis used the responses from the remaining 12 participants. All but three participants were male, and all but one participant identified as Asian/Pacific Islander. The average age was 27 years. Ten of the 12 participants (83.3%) spoke English as a second language.

2.4 Variables/Experimental design

The independent variable in this experiment is $R_d$, which has 4 levels ($R_{d,A} = 4$, $R_{d,B} = 4$, $R_{d,C} = 4$, and $R_{d,D} = 4$). Additionally, a spectrum which transposed no caps ($R_{d,A} = R_{d,B} = R_{d,C} = R_{d,D} = 0$) was included as a baseline comparison. The dependent (or response) variable is the appropriate, corresponding adjusted error score ($AES_{adj}$, $BES_{adj}$, $CES_{adj}$, or $DES_{adj}$). For example, when the experimental stimuli considers tray A, $R_{d,A}$ captures the number of transposition caused by the light source in that tray—which takes on a value of 0 or 4 by design in this experiment—and $AES_{adj}$ captures the error score of the participant who arranged the caps in the physical tray illuminated by that light source.

Spectra were administered in random order, subject to the constraint that a spectra with an $R_{d,j}$ of 0 always appeared first. Additionally, randomization was restricted such that the same tray did not appear sequentially, to allow the researcher time to score the results from the previous trial.

2.5 Procedure

Upon entering the lab, the participant completed the informed consent form, a demographics survey, and an Ishihara test for color deficiency [Ishihara 2016]. The participant proceeded to the viewing booth, overhead lights were turned off, and background information and instructions were read to the participant. The narrative took at least two minutes, which provided time for substantial chromatic adaptation [Fairchild and Reniff 1995]. The apparatus was pre-loaded with the appropriate spectra and contained the first FM-100 tray. If there were no questions, participants proceeded with the first tray. When finished, the participant was asked to roll to the side of the booth so that the tray could be collected, and the next tray administered. The participant was instructed to return to the booth and close their eyes while the experimental spectra was changed. Upon opening their eyes, participants began working on the next tray. This process was repeated for the remaining six spectra. Participants were required to spend at least two minutes with each tray—which provided intermediate adaptation to compensate for the closing of the eyes—though they could spend as much additional time as they needed. Experimental trials took approximately 30 minutes per participant.

![Figure 3](image-url)  
**Figure 3** – Absolute SPDs for the experimental spectra. SPD 1 transposed no caps, and was used to deliver four stimuli ($R_{d,A} = 0$, $R_{d,B} = 0$, $R_{d,C} = 0$, and $R_{d,D} = 0$). SPD 2 delivers a transposition in tray A ($R_{d,A} = 4$), SPD 3 delivers a transposition in Tray B ($R_{d,B} = 4$), and SPD 4 delivers a transposition in tray C and D ($R_{d,C} = R_{d,D} = 4$).
Table 1 – Characteristics of the four experimental spectra. The color discrimination values that are bolded, italicized, and underlined (e.g. $R_{d,A} = 0$ for SPD 1), represent the conditions (stimuli) that were used in this experiment. For example, SPD 1 was used four times (for each $R_{d,i} = 0$), and SPD 3 was used once to test $R_{d,B} = 4$.

<table>
<thead>
<tr>
<th>SPD ID</th>
<th>CCT [K]</th>
<th>$D_{uv}$</th>
<th>$R_{d,A}$</th>
<th>$R_{d,B}$</th>
<th>$R_{d,C}$</th>
<th>$R_{d,D}$</th>
<th>$R_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3503</td>
<td>0.0000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3498</td>
<td>0.0002</td>
<td>4</td>
<td>12</td>
<td>20</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>3487</td>
<td>0.0000</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>3502</td>
<td>0.0003</td>
<td>0</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>28</td>
</tr>
</tbody>
</table>

ID: Identification number for each experimental light source; CCT: Correlated Color Temperature measured in Kelvin; $D_{uv}$: Delta uv; $R_{d,A}/R_{d,B}/R_{d,C}/R_{d,D}$: light source error score for Tray A, Tray B, Tray C, and Tray D, respectively; $R_d$: total light source error score.

Figure 4 – The experimental apparatus containing the Thouslite LEDcube and the viewing booth painted with high reflectance, approximately spectrally neutral paint. The Jeti Specbos 1211 spectroradiometer, with an irradiance measurement attachment, is shown at the approximate measurement location used for photometric and colorimetric calibration.

3 Results

A one-way Analysis of Variance (ANOVA) shows that the four transpositions, strategically located in each of the FM-100 hue test trays, did not produce statistically different mean adjusted error scores ($p = 0.541$) (Figure 5). A Tukey mean comparison shows that the mean responses cluster into the same group. This provides initial support for the hypothesis that transpositions at different nominal hues will not produce significantly different mean error scores.

Figure 5 – mean adjusted error scores ($iES_{ad}$) for each of the 4 experimental spectra. Values are an average of 12 participant responses, and error bars show the 95% confidence interval. The four mean adjusted error scores were not statistically different ($p = 0.541$). The letter in the name of the scene (i.e. “A4”) indicates the concerned tray; the number indicates the $R_{d,i}$ value (i.e. $R_{d,A} = 4$).
A one-way ANOVA shows that all eight experimental stimuli did not produce statistically different mean adjusted error scores (Figure 6). That is, four spectra which caused one transposition ($R_{a,i} = 4$), did not produce statistically different mean adjusted error scores from each other, or from spectra which caused no transpositions ($R_{a,i} = 0$). A Tukey mean comparison shows that all means cluster into the same group.

![Figure 6 – mean adjusted error scores ($iES_{adj}$) for each of the 4 experimental conditions, plus the 4 baseline conditions. Values are an average of 12 participant responses, and error bars show the 95% confidence interval. The mean adjusted error scores were not statistically different ($p = 0.259$).](image)

4 Discussion

4.1 Data variability

Overall, there was a large degree of variability between participants for a single stimulus. One particular participant was a substantial outlier; their scores were double the next highest scores. Removing this participant’s data from the analysis, however, does not change this study’s conclusions.

In an early test trial of the experiment, the research assistant ran the current author as a test subject (data not included in the current analysis). The current author had the lowest average score across all participants. The FM-100 test is an objective test, and there is no reason to expect that knowledge of the test would distort results. Additionally, past research suggests there is no learning effect on error scores [Boyce and Simons 1977]. This may suggest that a portion of the variability in the resulting error scores may be due to a lack of care and attention. That is, participants who are more motivated to perform well, will actually perform better.

Requiring that a participant spend at least two minutes with a tray is partially an attempt to reduce carelessness. The intra-participant variability may suggest the need for pre-experiment screening measures to include, as best as possible, highly motivated individuals. A focused pool of participants may reduce variability and better isolate the underlying effect of the SPD.

4.2 $R_{a,i} = 0$ vs. $R_{a,i} = 4$

On average, the expectation is that an increasing number of transpositions will result in increased confusion, reduced performance, and higher mean adjusted error score [Esposito and Houser 2017]. Though there was no statistical difference between means, in all but one case the mean adjusted error score was higher for stimuli that transposed one cap, as compared to that which transposed no cap (Figure 6). For example, mean $AES_{adj}$ is higher for the stimulus which transposed one cap in tray A ($R_{a,A} = 4$), as compared to the stimulus which transposed no caps in tray A ($R_{a,A} = 0$). This is true for all tray-specific stimuli comparisons except tray B, in which case a hue transposition resulted in a lower mean error score. The error scores, however, fall within an expected range (Figure 7).

5 Conclusion

Results of the current study show that source-induced cap transpositions in each of the four trays of the FM-100 hue test do not produce statistically different mean adjusted error scores. If we can assume the FM-100 hue test is indicative of color discrimination ability, and cap transpositions are predictive of the error scores of the FM-100 hue test, the results of the current study suggest that color discrimination ability may not depend upon hue. Mean error scores correspond well to past research, which increases confidence in this methodology.
A follow up investigation should be performed with a larger sample size (to increase the power of the statistical test) and with a motivated set of participants (to reduce variability).

**Figure 7** – Comparison of experimental mean adjusted error scores for tray B ($BES_{adj}$) to the experimental data of Esposito and Houser [2017]. An increase from 0 to 1 transpositions in tray B decreases mean error scores, though both means fall within the expected range of variability for spectra which cause 0 and 1 transposition(s), respectively.

**Acknowledgments**

The author would like to thank Tanushree Kakani for her help with data collection.

**References**


Esposito T, Houser KW. 2017. A new measure of colour discrimination for LEDs and other light sources. Lighting Research and technology. Accepted for publication.


Wei M, Houser KW. 2012. Colour discrimination of seniors with and without cataract surgery under illumination from two fluorescent lamp types. CIE x037:359-368
Figure 8 – Order of the caps of the FM-100 hue caps when illuminated by the four experimental spectra. Labels indicate the cap number; transposed caps are bolded and underlined. (Top left) SPD 1 which transposes no caps and has $R_{d,A} = R_{d,B} = R_{d,C} = R_{d,D} = 0$. (Top right) SPD 2 transposes many caps overall, but creates exactly one transposition (cap 3 and 4) in tray A ($R_{d,A} = 4$). (Bottom left) SPD 3 creates only one transposition overall (cap 28 and 29), which is located in tray B ($R_{d,B} = 4$). (Bottom right) SPD 4 transposes many caps overall, but creates exactly one transposition in tray C (caps 46 and 47) and one in tray D (caps 78 and 79) ($R_{d,C} = R_{d,D} = 4$). For reference, dashed lines show the $a'b'$ plane partitioned into 16 hue angle bins, and the solid line shows the gamut area of the FM-100 hue test chips illuminated by CIE Standard Illuminant C.