Development of a Color Quality Scale

Wendy Davis and Yoshi Ohno
National Institute of Standards and Technology
100 Bureau Drive, Stop 8442
Gaithersburg, MD 20899-8442
Phone: 301-975-6963 (Davis) and 301-975-2321 (Ohno)
Fax: 301-840-8551
e-mail: wendy.davis@nist.gov (Davis) and ohno@nist.gov (Ohno)

A Color Quality Scale (CQS) is being developed at NIST, which evaluates several aspects of the quality of the color of objects illuminated by a light source. This metric involves several facets of color quality, including color rendering, chromatic discrimination, and observer preferences. The method for calculating the CQS is derived from modifications to the method used in the CIE's Color Rendering Index (CRI). The current CRI is based on only eight reflective samples, all of which are low to medium chromatic saturation. These colors do not adequately span the range of normal object colors. Some lights that are able to accurately render colors of low saturation perform poorly with highly saturated colors, particularly the peaked spectra of light-emitting diodes (LEDs). Instead, 15 Munsell samples were assembled that overcome these problems and are used for calculations of the CQS. Additionally, the CRI penalizes lamps for showing increases in object chromatic saturation compared to reference lights, which is actually desirable for most applications. To incorporate observer preference, we propose a new computation scheme for determining the color rendering score that differentiates between hue and saturation shifts and takes their directions into account. The uniform color space used in the CRI is outdated, so CIELAB is used in the CQS. The CRI matches the CCT of the reference to that of the test light, which can be problematic when lights are substantially bluish or reddish. Lights of extreme CCTs frequently give poor quality and smaller color gamuts (thereby decreasing color discrimination), so a system is implemented in the CQS to penalize such lights. Simple averaging of the calculated color differences, as is done in the CRI, can disguise large color shifts on a small number of samples. To ensure that large shifts of any color are adequately reflected in the score, color differences are combined with a root-mean-square (RMS). Finally, an appropriate scaling factor is chosen and the scale is made to span 0-100.

Introduction

The quality of object color under artificial lighting is an important aspect of the value of the light sources, particularly to consumers. Assessing and quantifying this dimension of lamps is complicated.

The attribute of color rendering of light sources is often interpreted as indicating object color quality. However, color rendering is actually defined as the “effect of an illuminant on the color appearance of objects by conscious or subconscious comparison with their color appearance under a reference illuminant” [1]. Color rendering only refers to color fidelity, the accurate representation of object colors compared to those same objects under a reference source, and does not include other aspects of color quality, such as chromatic discrimination and color preference. The CIE color rendering index (CRI) [2] is the only internationally-accepted metric for assessing the color rendering performance of light sources. In the calculation of the CRI, the color appearance of 14 reflective samples is simulated when illuminated by a reference source and the test source. The reference source is a Planckian radiator (if below 5000K) or a CIE Daylight source (if at or above 5000K), matched to the correlated color temperature (CCT) of the test source. After accounting for chromatic adaptation with a Von Kries correction, the difference in color appearance ΔE_i for each sample between the test and reference light sources is computed in CIE 1964 W*U*V* uniform color space. The special color rendering index (R_i) is calculated for each reflective sample by:

$$R_\mathrm{i} = 100 - 4.6\,\Delta E_\mathrm{i}$$

The general color rendering index (R_a) is simply the average of R_i for the first eight samples, all of which have low to moderate chromatic saturation:

$$R_\mathrm{a} = \frac{1}{8}\sum_{i=1}^{8} R_i$$

A perfect score of 100 represents no color differences in any of the eight samples under the test and reference sources.

The CRI has a number of problems, particularly when applied to LEDs or when used as an indicator of color quality. The uniform color space used to calculate color differences is outdated and no longer recommended for use. The red region of this color space is particularly non-uniform. Instead, the CIE currently recommends CIE 1976 L*a*b* (CIELAB) and CIE 1976 L*u*v* (CIELUV) [3] for calculating color differences. Additionally, the chromatic adaptation transform is considered inadequate. The Von Kries chromatic adaptation correction used in the CRI has been shown to perform poorer than other available models, such as the CMCCAT2000 (the Colour Measurement Committee’s chromatic adaptation transform) and the CIE CAT02 (the CIE’s chromatic adaptation transform) [4].

The CRI method specifies that the CCT of the reference source be matched to that of the test source, which assumes complete chromatic adaptation to the chromaticity of the light source. This assumption fails at extreme CCTs, however. For example, a 2000K (very reddish) blackbody source achieves R_a = 100, as does a daylight spectrum of 20,000K (very bluish). However, neither of these sources renders colors well.

None of the eight reflective samples used in the computation of R_a are highly saturated. This is problematic, especially for the peaked spectra of white LEDs. Color rendering of saturated colors can be very poor even when the R_a value is good. Further, by optimization of lamps' spectra to the CRI, R_a values can be made very high while actual color rendering is much poorer. This problem exists because too few samples are used in the calculation of R_a, and they are of too low chromatic saturation.

The eight special color rendering indices are simply averaged to obtain the general color rendering index. This makes it possible for a lamp to score quite well, even when it renders one or two colors very poorly. LEDs are at an increased risk of being affected by this problem, as their peaked spectra are more vulnerable to poor rendering in only certain areas of color space.

Finally, the very definition of color rendering is flawed for use when one is interested in the overall color quality of a light source. Color rendering is a measure of only the fidelity of object colors under the illuminant of interest and any deviations of object color appearance from under a blackbody source is considered bad. Due to this constraint, all shifts in perceived object hue and saturation result in equal decrements in CRI score. In practical application, however, increases in chromatic saturation, observed when certain sources illuminate certain surfaces, is considered desirable. Increases in saturation yield better visual clarity and enhance perceived brightness [5]. It is proposed that the absolute focus on color fidelity of the CRI is flawed and a more general metric of color quality be considered.

Color Quality Scale (CQS)

A Color Quality Scale (CQS) is being developed at NIST, which evaluates several aspects of the quality of the color of objects illuminated by a light source. The extensive description of the CRI was provided because, rather than inventing an entirely new approach to the metric, much inspiration was taken from the CRI. Borrowing from aspects of the CRI that are successful, the CQS incorporates important modifications to overcome its shortcomings and focuses on a broader definition of color quality.

The set of reflective samples tested is different from those used in the calculations of the CRI. Fifteen saturated Munsell samples are used in the CQS, with the following hue value/chroma: 7.5 P 4 / 10, 10 PB 4 / 10, 5 PB 4 / 12, 7.5 B 5 / 10, 10 BG 6 / 8, 2.5 BG 6 / 10, 2.5 G 6 / 12, 7.5 GY 7 / 10, 2.5 GY 8 / 10, 5 Y 8.5 / 12, 10 YR 7 / 12, 5 YR 7 / 12, 10 R 6 / 12, 5 R 4 / 14, and 7.5 RP 4 / 12. They were selected to have the highest chroma, span the entire hue circle in approximately even spacing, and be commercially available. Figure 1 shows these samples (bottom row) as well as the eight samples used in the calculation of R_a (top row) when illuminated by a daylight-like source (D65). This representation, as well all shown later, may be inaccurate due to the properties of the printing or display of this paper.

The uniform object color space also differs from that used in the CRI. The 1964 W*U*V* object color space is obsolete, and is very nonuniform: color differences are extremely exaggerated in the red region and suppressed in yellow and blue regions.

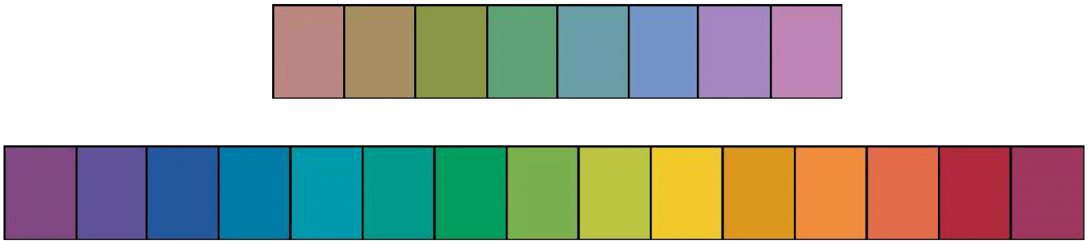

Figure 1: Representation of the eight samples used in the calculation of R_a (top row) and the 15 samples used in the calculation of the CQS (bottom row) when illuminated by D65.

So, when calculating the CQS, CIE 1976 L*a*b* ("CIELAB") is used, as it is currently recommended for use by the CIE and is considered to be reasonably uniform.

One of the major deviations that the CQS takes from the formal definition of color rendering is evident in the saturation factor. The CRI penalizes lamps for shifts in hue, chroma (chromatic saturation), and lightness, in any direction, of the reflective samples under the test source (compared to under the reference source). While a decrease in chroma always has negative effects, an increase in the chroma of objects is considered desirable in many cases. Increases in chroma yield better visual clarity and enhance perceived brightness [5]. These are positive effects and are generally preferred, though they cause deviations in color fidelity (compared to reference). In the CQS, lamps are not penalized for increasing object chroma relative to the reference source, though their scores are also not increased. The net result is that a lamp's score is only penalized for hue shifts, lightness shifts, and reductions in chroma. This is a way to take color preference (and possibly also color discrimination) into account in the CQS. For example, in Figure 2 (a) when the chroma increases under the test illuminant (with no change in hue), there is no change in score, (b) when the chroma decreases under the

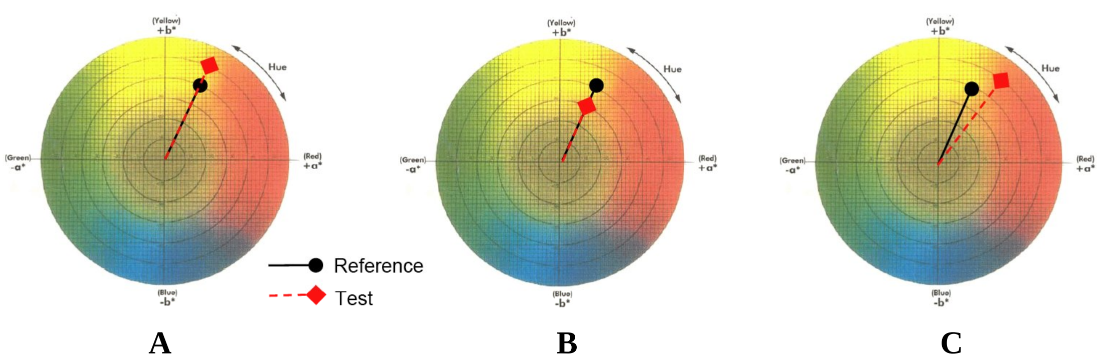

Figure 2: Effects of the saturation factor illustrated in CIELAB color space. (A) when the chroma increases under the test illuminant (with no change in hue), there is no change in score, (B) when the chroma decreases under the test illuminant, the score is decreased, and (C) when the chroma increases and the hue shifts, the score is decreased for the hue shift but not decreased for the increase in chroma.

Table 1: The gamut area of the 15 samples for reference sources of various CCTs and the CCT multiplication factor used in the CQS.		
CCT (K)	**Gamut Area**	**Multiplication Factor**
1000	2645	0.32
1500	5424	0.65
2000	6902	0.83
2500	7676	0.93
2856	7987	0.97
3000	8075	0.98
3500	8268	1.00
4000	8347	1.00
5000	8341	1.00
6000	8274	1.00
6500	8211	1.00
7000	8151	0.99
8000	8040	0.98
9000	7947	0.97
10000	7868	0.96
15000	7620	0.93
20000	7495	0.91

test illuminant, the score is decreased, and (c) when the chroma increases and the hue shifts, the score is decreased for the hue shift but not decreased for the increase in chroma.

In the CRI, the CCT of the reference source is matched to that of the test source. Therefore the CRI score is perfect (100) for reference sources of any CCT. Actual color rendering, however, is degraded at extremely low or high CCTs. This is a problem with the way the reference source is defined in the current metric, and is one of the most difficult problems to address. The perfect solution to this problem would require thorough understanding of chromatic adaptation. Such investigations have not been conducted yet, but a temporary solution has been developed. Though the CCT of the reference source is also matched to that of the test in the CQS, a multiplication factor is introduced. This CCT factor is determined based on the gamut area in CIELAB space for the 15 samples under the reference source for each CCT, as shown in Table 1. It is assumed that the color rendering performance of the reference source degrades as the gamut area decreases. The multiplication factor, as listed in the table, is the ratio of gamut area of the particular CCT with the gamut area for 6500 K. With this normalization, the multiplication factors at certain CCT ranges (e.g., 4000 K) give values slightly higher than 1, but these are truncated to 1, so that the CQS score will never be higher than 100. The exact effect of CCT on color quality is difficult to quantify, but this method offers at least a temporary solution for sources of extremely low or high CCT.

In the CRI, the color differences (ΔE) for each of the samples is averaged. This makes it possible for a lamp to score quite well, even when it renders one or two samples very poorly. This situation is even more likely with SPDs having narrowband peaks, like LEDs. To ensure that large hue shifts of any sample have notable influence on the CQS, the root-mean-square (RMS) of color shifts of each individual sample is used (rather than arithmetic mean). The RMS color differences of the 15 samples are calculated by:

$$\Delta E_{\mathrm{RMS}} = \sqrt{\frac{1}{15}\sum_{i=1}^{15} \Delta E_i^{\,2}}$$

In the CRI, the scaling factor 4.6 is used to convert color differences into color rendering indices. This factor needs to be changed for the CQS since the sample set and color space are different, in order to maintain the consistency with the CRI. For the CQS, a scaling factor of 2.81 is currently chosen so that the average score of the CQS

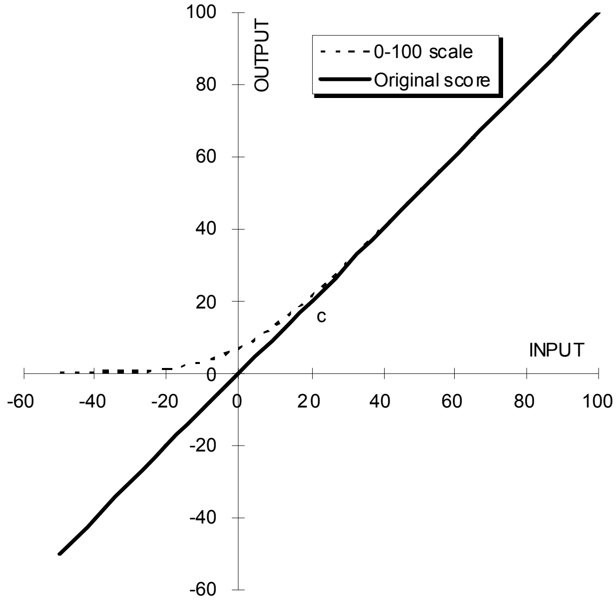

Figure 3: The 0-100 scale function (dashed) used to convert original scores (solid).

for the CIE standard fluorescent lamp spectra (F1 through F12 [6]) is equal to the average score of the current CRI R_a (=75.1) for these sources. This scaling maintains consistency of the new color quality scale with the current CRI scale for existing lamps. As the CQS formula is further modified, this scaling factor will have to be recalculated.

Negative values that the CRI reports for certain lamps are often confusing. CRI and CQS scores lower than 20 or 30 are already very poor and the linearity of the scale below these low scores is not considered important. A scale from 0 to 100 would be better understood by users. A 0-100 scale conversion has been made to avoid the confusion, and has been implemented by using the formula:

$$R_{\text{out}} = 10 * \ln\left[\exp(R_{\text{in}} / 10) + 1\right]$$

where R_{in} is the input value (which can be a negative number) and R_{out} is the output value of the conversion. This function is shown as the dashed curve in Fig. 3. Only values lower than approximately 20 are changed by this conversion, and values above 20 are scarcely affected. As such, all values within the range for usable lamps are unchanged.

The current calculations of the CQS use a Von Kries chromatic adaptation correction, which is largely considered to be outdated and incomplete. One future plan is to replace this correction formula with a more updated and accurate chromatic adaptation transform, such as the Bradford transform, CMCCAT2000, or the CIE CAT02 [4].

Though CIELAB is a substantial improvement over the 1964 W*U*V* color space, it does have its own shortcomings. Because of known nonlinearities, an adjusted method for calculating color differences has been recommended [7], called $\Delta E2000$, but it is more complicated. We plan to study the magnitude of differences in calculated scores for various known light sources with $\Delta E2000$, compared to ΔE^*_{ab}.

Computational Tests

The white LED simulation program developed at NIST is discussed in detail elsewhere [8]. A mathematical model is implemented to simulate multi-chip LEDs. The spectral power distribution of a model LED, $S_{LED}(\lambda)$, for a peak wavelength $\lambda_\circ$ and the half spectral width $\Delta\lambda_{0.5}$ is given by:

$$S_{LED}(\lambda, \lambda_0, \Delta\lambda_{0.5}) = \{ g(\lambda, \lambda_0, \Delta\lambda_{0.5}) + 2 \cdot g^5(\lambda, \lambda_0, \Delta\lambda_{0.5})\} / 3$$

$$\text{where } g(\lambda, , \lambda_0, \Delta\lambda_{0.5}) = \exp [-\{(\lambda - \lambda_\circ) / \Delta\lambda_{0.5}\}^2]$$

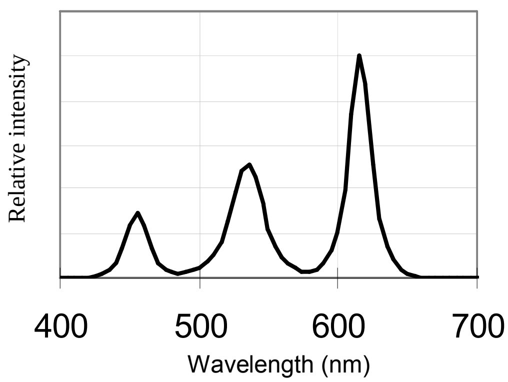

Figure 4: Spectral power distribution of modeled 3-color white LED, with peaks at 457, 534, and 616 nm.

The unit of wavelength is nm. The program can simulate three-color or four-color white LEDs, with automatic color mixing for each LED to achieve a specified chromaticity. All LED spectra shown here were produced with this simulation program; therefore, some LED spectra shown will not be representative of LEDs commercially available.

As already noted, some of the flaws of the CRI are particularly evident when applied to white LEDs. LEDs hold great potential for energy-efficient general illumination, so it is important that the CQS perform well for both traditional light sources and multi-chip LEDs.

The spectral power distribution of a three-color white LED model is shown in Fig. 4. The peaks of this spectrum are at 457, 534, and 616 nm, the CCT of the light is 3300K, and its luminous efficacy of radiation (LER) is 363 lm/W. The General Color Rendering Index (R_a) for this light source is 67—a fairly poor score. However, the CQS score for this source is 80. An examination of the colors of the 15 reflective samples under this source relative to the reference (blackbody radiation at 3300K) reveals the reason for such a large score difference between the two metrics. As shown in Fig. 5 (produced by a color simulation program [8]), several of the samples illuminated by the test source, shown in the bottom row, are more saturated than when illuminated by the reference source, in the top row. This lack of fidelity causes decreases in the R_a, but not in CQS score. As discussed earlier, increases in object chromatic saturation are generally considered a positive trait in a lamp (though excessive saturation may not be good). The visual impression of this color presentation on a calibrated computer display implies that this source may be accepted, and even preferred, by users as the CQS score indicates, though the fidelity is relatively poor. This example well-illustrates the

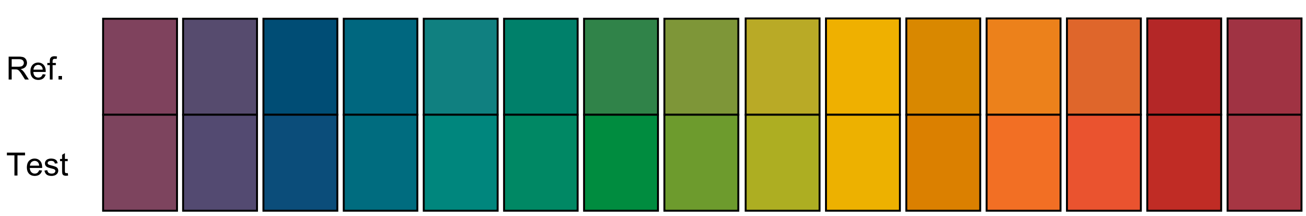

Figure 5: Simulated color appearance of 15 reflective samples when illuminated by the reference source (top row) and by the test source (bottom row).

applicability of the saturation factor and the fundamental deviation from the definition of color rendering.

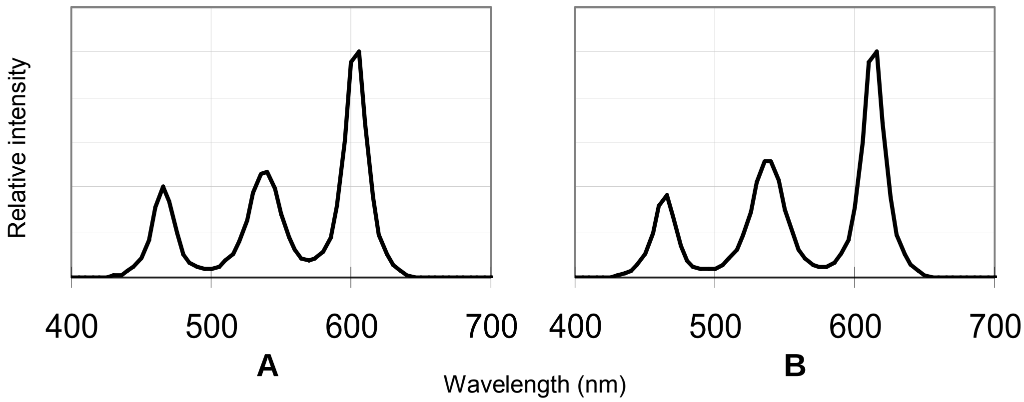

Figure 6: Spectral power distribution of modeled 3-color white LEDs with (A) peaks at 463, 538, and 603 nm and (B) peaks at 464, 538, and 613 nm.

Two additional spectra, each modeling a white source composed of RGB LEDs, are shown in Fig. 6. The peaks of the spectrum on the left, labeled A, are at 463, 538, and 603 nm, the CCT of the light is 3300K, and its LER is 406 lm/W. The spectrum shown on the right, labeled B, has peaks at 464, 538, and 613 nm, a CCT of 3300K, and a LER of 376 lm/W. The R_a for both of these sources is 80. However, there are large differences in object color between these two lights. A simulation of the appearance of each of the samples for source A is shown in Fig. 7 and for source B in Fig. 8. Source A shows fairly good color rendering of most colors but very poor color on just two or three samples (red and purple), while source B shows overall good color rendering with some samples (green and orange) of increased chromatic saturation. Source A receives a CQS score of 73, which is lower than its CRI because the RMS

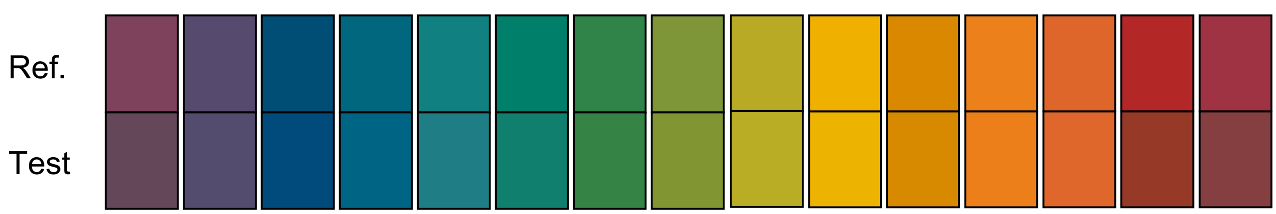

Figure 7: Simulated color appearance of 15 reflective samples when illuminated by the reference source (top row) and by test source A (bottom row).

used in the CQS enhances the effect of the large color differences of a small number of samples. Source B receives a CQS score of 85, higher than its CRI because it increases the chromatic saturation of some of the samples, which the CQS does not penalize. The CQS shows a 12 point difference between these two sources. The CQS scores are more consistent with the visual impressions of the simulated colors of the samples than the CRI scores, which fail to distinguish between the color quality of these two sources.

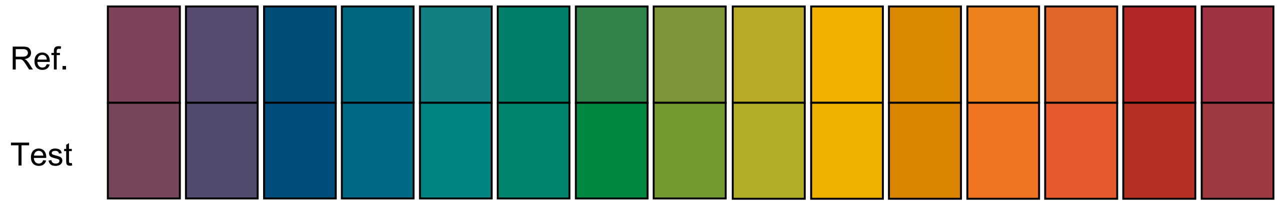

Figure 8: Simulated color appearance of 15 reflective samples when illuminated by the reference source (top row) and by test source B (bottom row).

Future Work

As previously discussed, further computational work on the new metric is planned. The chromatic adaptation transform will be updated from Von Kries to the Bradford transform, CMCCAT2000, or the CIE CAT02. Furthermore, the use of $\Delta E2000$ will be investigated as a replacement of ΔE^*_{ab} for determining color differences of samples.

A series of thorough and well-controlled vision experiments are necessary to test, improve upon, and validate the computational analyses presented here. Experiments testing observers' chromatic discrimination and absolute hue perception of illuminated objects will be complemented by subjective rankings of naturalistic scenes. Since the CQS is intended to be a metric of overall color quality, the data from several types of experiments will be used to assess and improve its performance. A new vision science laboratory is being built at NIST, with experiments testing CQS taking highest priority.

References

1. Commission Internationale de l'Eclairage. 1987. International Lighting Vocabulary, CIE 17.4: item 845-02-59.

2. Commission Internationale de l'Eclairage. 1995. Method of Measuring and Specifying Colour Rendering Properties of Light Sources, CIE 13.3.

3. Commission Internationale de l'Eclairage. 2004. Colorimetry, CIE 15.

4. Commission Internationale de l'Eclairage. 2004. A Review of Chromatic Adaptation Transforms, CIE 160.

5. K. Hashimoto and Y. Nayatani. 1994. Visual Clarity and Feeling of Contrast, *Color Res. Appl.*, 19 (No 3): 171-185.

6. Commission Internationale de l'Eclairage. 1986. Colorimetry, 2nd Edition, CIE 15.2.

7. Commission Internationale de l'Eclairage. 2001. Improvement to industrial colour-difference evaluation. CIE 142.

8. Y. Ohno. 2004. Color Rendering and Luminous Efficacy of White LED Spectra, *Proc. SPIE Fourth International Conference on Solid State lighting*, 5530: 88-98.