Toward an improved color rendering metric

Wendy Davis and Yoshi Ohno
Optical Technology Division, National Institute of Standards and Technology, 100 Bureau Drive,
MS 8442, Gaithersburg, MD 20877-8442

ABSTRACT

Several aspects of the Color Rendering Index (CRI) are flawed, limiting its usefulness in assessing the color rendering capabilities of LEDs for general illumination. At NIST, we are developing recommendations to modify the CRI that would overcome these problems. 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. This is particularly prominent with light sources with peaked spectral distributions as realized by solid-state lighting. We have assembled 15 Munsell samples that overcome these problems and have performed analysis to show the improvement. Additionally, the CRI penalizes lamps for showing increases in object chromatic saturation compared to reference lights, which is actually desirable for most applications. We suggest 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 The uniform color space used in the CRI is outdated and a replacement will be recommended. The CRI matches the CCT of the reference to that of the test light. This can be problematic when lights are substantially bluish or reddish. Lights of extreme CCTs are frequently poor color renderers, though they can score very high on the current CRI. An improved chromatic adaptation correction calculation would eliminate the need to match CCT and an updated correction is being considered.

Keywords: colorimetry, color rendering, spectrum, light source, CRI, color appearance, white LED, solid-state lighting

1. INTRODUCTION

1.1 Color rendering index (CRI)

Color rendering is 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] 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 two 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:

$$R_i = 100 - 4.6\,\Delta E_i \tag{1}$$

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:

$$R_a = \frac{1}{8}\sum_{i=1}^{8} R_i \tag{2}$$

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

Fifth International Conference on Solid State Lighting, edited by Ian T. Ferguson,
John C. Carrano, Tsunemasa Taguchi, Ian E. Ashdown, Proc. of SPIE Vol. 5941
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.615388

1.2 Problems with the CRI

The CRI has a number of problems, particularly when used to assess the color rendering of LEDs. 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] 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 (Colour Measurement Committee's chromatic adaptation transform) and the CIE CAT02 (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 is because the eight color samples used in the calculation of R_a are all of medium saturation and the number of samples is too few.

The eight special color rendering indices are combined by a simple averaging 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 many applications. 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.

2. NEW METRIC

2.1 Color Quality Scale (CQS)

Rather than a complete re-invention, the fundamental method of the CRI is maintained and modified. However, an important change is proposed, involving a deviation away from definition of color rendering. Rather than assess only color fidelity, this improved metric is intended to assess overall color quality of light sources and is aptly named *Color Quality Scale (CQS)* to avoid confusion with the CRI.

2.2 New reflective sample set

One of the most serious problems with the CRI is that color rendering of saturated colors can be very poor even when the R_a value is good. For the CQS, the eight samples used in the calculation of R_a have been replaced with 15 samples of high chromatic saturation spanning the entire hue circle. The chromaticity a^*, b^* of these samples under illumination by D65 are shown in Figure 1. All are currently available Munsell samples of 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.

2.3 Updated uniform color space

The 1964 W*U*V* object color space used in the CRI is obsolete, and is very nonuniform: color differences are extremely exaggerated in the red region and suppressed in yellow and blue regions. CIE 1976 L*a*b* ("CIELAB") is currently recommended by CIE and is widely used in many applications (CIELUV is also under current recommendation, but it is not as uniform as CIELAB and not as widely used). In the CQS, the W*U*V* color space has been replaced by CIELAB for determining perceived color differences between samples under test and reference sources. A comparison of the spacing of the 15 new samples in each of the color spaces is shown in Figure 1.

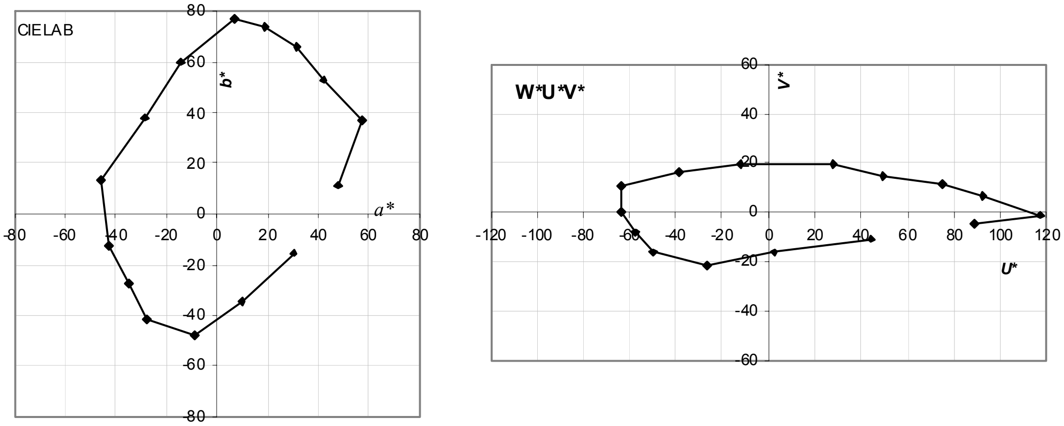

Figure 1: The 15 new reflective samples used in the CQS plotted in CIE 1976 L*a*b* (left) and CIE 1964 W*U*V* (right) color spaces.

2.4 Saturation factor

The CRI penalizes lamps for hue, chromatic saturation, and lightness shifts (in any direction) of the reflective samples between reference and test sources. However, an increase in object saturation (chroma) under the test source is considered desirable. 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

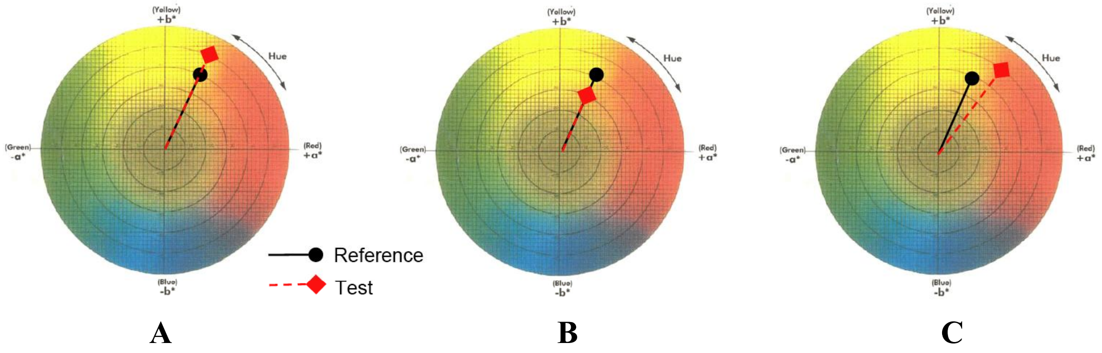

Figure 2: Effects of the saturation factor illustrated in CIELAB color space. (A) when the chroma increases under the test illuminant (with no change in hue), there is no change in score, (B) when the chroma decreases under the test illuminant, the score is decreased, and (C) when the chroma increases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

chroma. This is a way to take color preference (and possibly also color discrimination) into account in the CQS. For example, as illustrated 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 test illuminant, the score is decreased; (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 CTs 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

2.5 CCT factor

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 matched to that of the test also 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 Table 1, is the relative gamut area normalized at 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 at extremely low or high CCT.

2.6 Use of RMS

In the CRI, the color difference (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. 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 difference of the 15 samples is calculated by:

$$\Delta E_{\mathrm{RMS}} = \sqrt{\frac{1}{15}\sum_{i=1}^{15}\Delta E_i^2} \tag{3}$$

2.7 New scaling factor

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 new scaling factor is determined so that the average score of the CQS for the CIE standard fluorescent lamp spectra (F1 through F12[6]) is equal to the average score of the CRI (R_a=75.1) for these sources. This scaling maintains consistency of the new color quality scale with the CRI scale for existing lamps. As the CQS formula is further modified, this scaling factor (currently 3.01) will have to be recalculated.

2.8 Conversion to a 0-100 scale

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:

$$R_{\text{out}} = 10 * \ln\left[\exp(R_{\text{in}} / 10) + 1\right] \qquad (4)$$

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 Figure 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.

2.9 Other components

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]

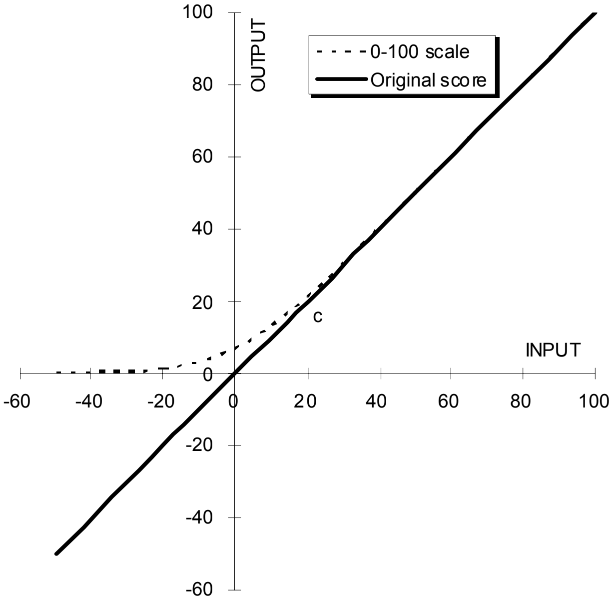

Figure 3: The 0-100 scale function (dashed) used to convert original scores (solid).

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 ΔE2000, but it is more complicated. We plan to study the magnitude of differences in calculated scores for various known light sources with ΔE2000, compared to ΔE^*_{ab}.

3. COMPUTATIONAL TESTING OF NEW METRIC

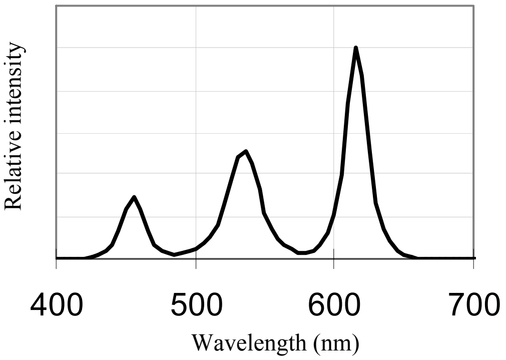

Figure 4: Spectral power distribution of modeled 3-color white LED, with peaks at 457, 534, and 616 nm.

3.1 White LED simulation program

The white LED simulation 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, $S_{\text{LED}}(\lambda)$, for a peak wavelength λ_0 and the half spectral width $\Delta\lambda_{0.5}$ is given by:

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

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

The unity 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.

3.2 Comparison of CRI and CQS for LED models

As noted in Section 1, some of the flaws of the CRI are particularly evident when applied to white LEDs. The spectral power distribution of a three-color white LED model is shown in Figure 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 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 CIELAB in Figure 5, several of the samples illuminated by the test source, shown in black, are more saturated than when illuminated by the reference source, as shown in grey. This lack of fidelity causes decreases in the R_a, but not in CQS score. As mentioned earlier, increases in object chromatic saturation is generally considered a positive trait in a lamp (though excessive saturation may not be good). Simulations of the color appearance of the 15 reflective samples under the test and reference sources are shown in Figure 6 (this representation will be limited by the production quality of the printing and does not account for chromatic adaptation). Visual impression of this color presentation on a calibrated computer monitor implies that this source may be preferably accepted as the CQS score indicates, though the fidelity is relatively poor. This example well-illustrates the applicability of the saturation factor and the fundamental deviation from the definition of color rendering.

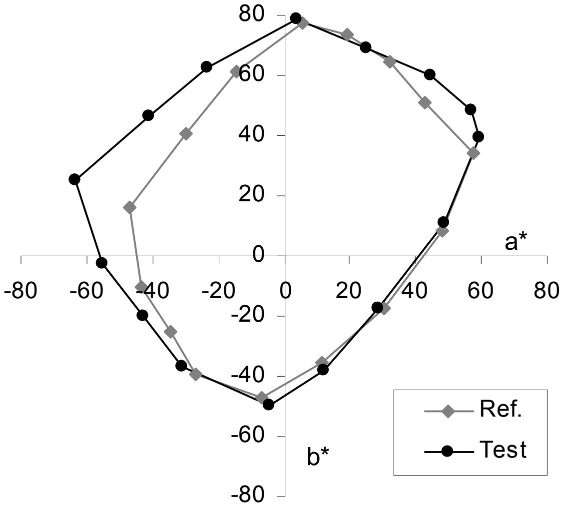

Figure 5: CIELAB coordinates of the 15 reflective samples when illuminated by the test source (black) and the reference source (grey).

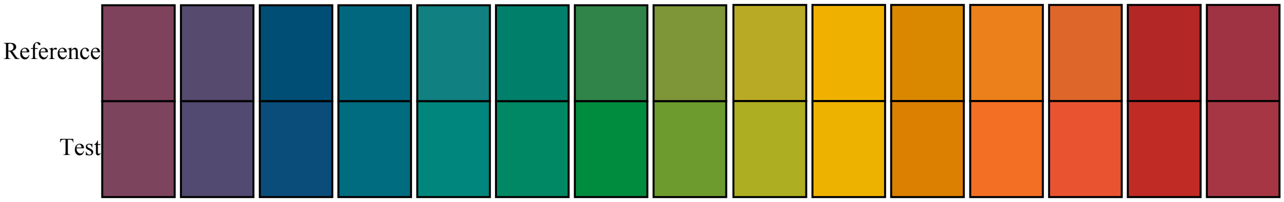

Figure 6: Simulated color appearance of 15 reflective samples when illuminated by the reference source (top rows) and by the test source (bottom rows).

Two additional spectra, each modeling a white source composed of RGB LEDs, are shown in Figure 7. 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, the 15 samples used in the CQS in CIELAB color space under both sources, as shown in Figure 8, reveals large differences between the two lights.

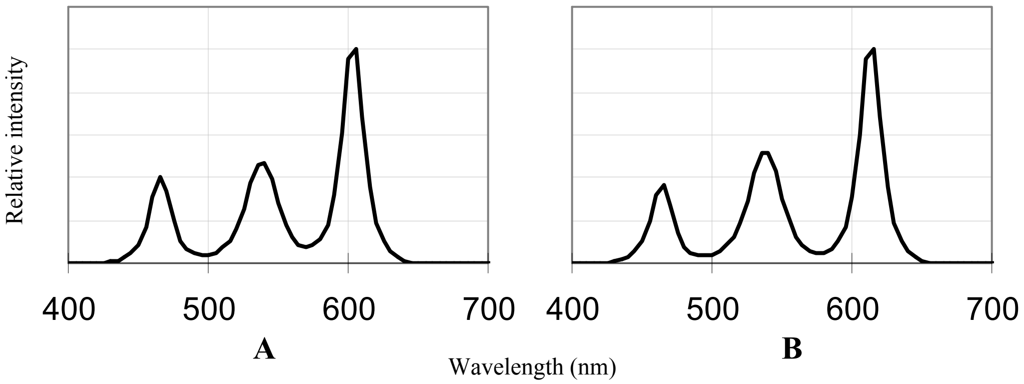

Figure 7: Spectral power distribution of modeled 3-color white LEDs with (A) peaks at 463, 538, and 603 nm and (B) peaks at 464, 538, and 613 nm.

Source A shows large hue and saturation shifts in the red region, while source B shows more, much smaller shifts. Additionally, source B increases the apparent saturation of several samples in the green region. A simulation of the appearance of each of the samples for source A is shown in Figure 9 and for source B in Figure 10. Source A receives a CQS score of 73, which was lower than its CRI because the RMS used in the CQS enhanced the effect of the large color differences of reddish 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 fails to distinguish between the color rendering properties of these two sources.

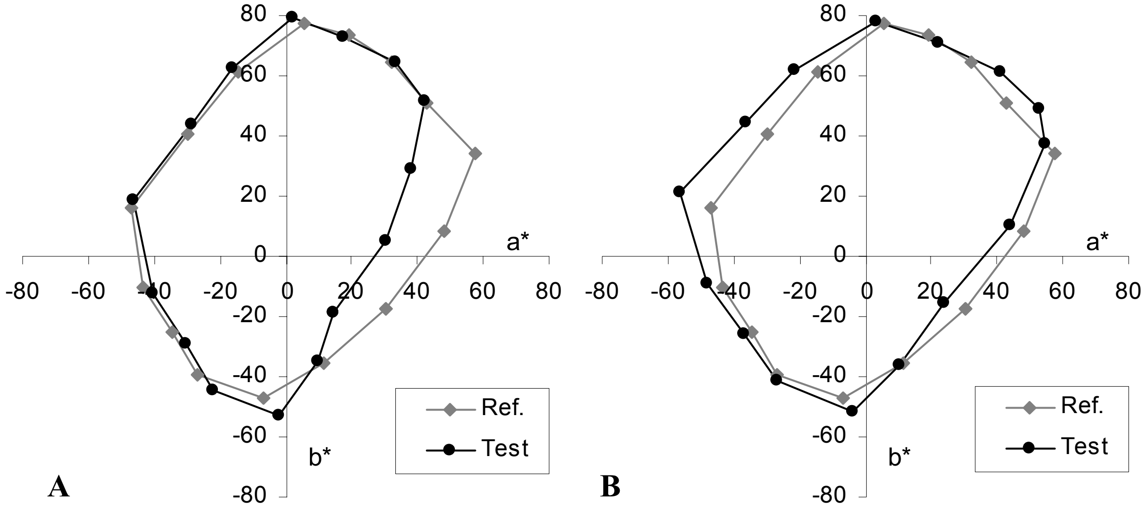

Figure 8: CIELAB coordinates of the 15 reflective samples when illuminated by the test source (black) and the reference source (grey) for spectrum A and spectrum B.

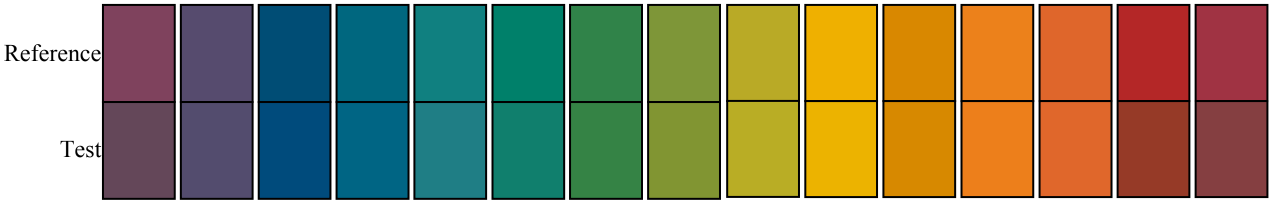

Figure 9: Simulated color appearance of 15 reflective samples when illuminated by the reference source (top row) and by test source A (bottom row).

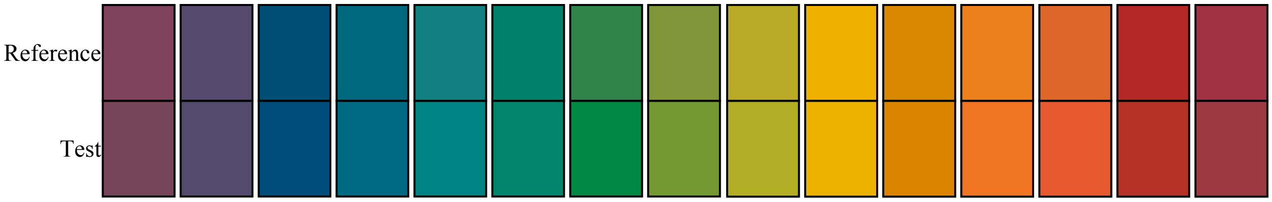

Figure 10: Simulated color appearance of 15 reflective samples when illuminated by the reference source (top row) and by test source B (bottom row).

4. FUTURE WORK

4.1 Computational

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.

4.2 Vision experiments

A series of thorough and well-controlled vision experiments are necessary to test 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.

5. REFERENCES

1. Commission Internationale de l'Eclairage. *International Lighting Vocabulary,* item 845-02-59, CIE 17.4-1987.
2. Commission Internationale de l'Eclairage. *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE 13.3-1995.
3. Commission Internationale de l'Eclairage. *Colorimetry*, CIE 15:2004.
4. Commission Internationale de l'Eclairage. *A Review of Chromatic Adaptation Transforms*, CIE 160:2004.
5. K. Hashimoto and Y. Nayatani, "Visual Clarity and Feeling of Contrast," Color Res. Appl. **19-3**, 171-185 (1994).
6. Commission Internationale de l'Eclairage. *Colorimetry*, 2nd Edition, CIE 15.2-1986.
7. Commission Internationale de l'Eclairage. *Improvement to industrial colour-difference evaluation*. CIE 142-2001.
8. Y. Ohno, "Color Rendering and Luminous Efficacy of White LED Spectra," *Proc. SPIE Fourth International Conference on Solid State lighting*, 5530, 88-98, Denver, CO, 2004.