Effects of Inter-Observer Variation on Color Rendering Metrics

Michael J. Murdoch & Mark D. Fairchild
Munsell Color Science Lab, Rochester Institute of Technology, Rochester, NY, USA

Abstract

Metrics for describing the color rendering characteristics of light sources are based on CIE standard observers. However, the range of natural variation in color sensitivity over people with normal color vision means any individual may see something different than the standard observer. Modeling results quantify the effects of these inter-observer differences on color rendering metrics CIE 13.3 (CRI) and IES TM-30-15. Inter-observer differences are smallest at high color fidelity values and generally larger for light source spectra with steep transitions and narrow peaks.

Background

Advancements in lighting technology, coupled with better understanding of the human visual system and population distributions, mean we have both the opportunity and need to describe color in our built environments in more nuanced ways.

Inter-Observer Variability

Colorimetry has functioned successfully for nearly a century using sets of color matching functions designed to represent the visual sensitivity of an average human observer. As with any statistical average, the mean functions do not necessarily represent any individual observer. Within any given population of people with normal color vision (i.e., excluding color vision deficiencies), there is a natural range in spectral sensitivities caused by variations in cone spectral absorptivity, ocular media density, and other anatomical and physiological parameters. Some of these variations are age dependent, others are genetic, and some have other causes such as diet or environment.
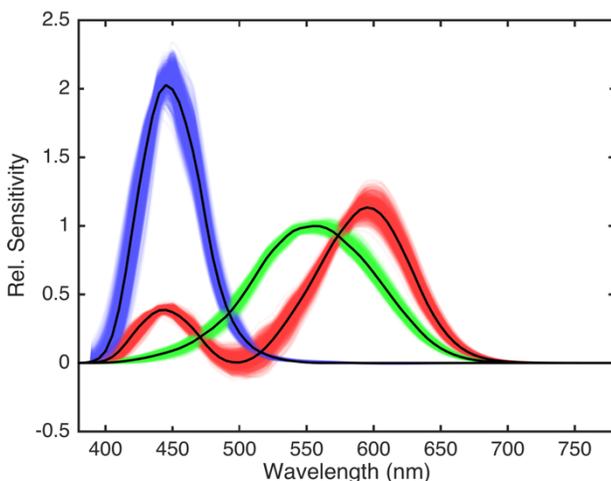
Figure 1: A population of 1000 color matching functions derived from the individual colorimetric observer model (colored lines) for a visual field of 10 degrees and an age distribution based on United States census data. Also shown is the CIE 1964 10-degree standard observer (black).

Recently, Asano et al. created an individual colorimetric observer model that takes into account the natural distributions of anatomical and physiological features such as physical densities of ocular components and wavelength and density shifts of cone photopigments based on age, size of visual field, and random population variability [1]. Using Monte Carlo simulations, their model is able to create populations of individual color matching functions (CMFs) with any given set of parameters (e.g. age, field size, genetic markers, etc.). These populations of observers have been verified using population statistics and on an individual basis by predicting observed experimental color matches. Such populations of observers, represented by collections of individual color matching functions, are used to analyze color rendition metrics for lighting to see if individual differences might impact consumer satisfaction with selected light sources.

Color Rendering of Light Sources

The spectral characteristics of an illuminant have an obvious and direct effect on the reflected spectral power distribution of objects being illuminated. Experience tells us that broad-spectrum light sources such as daylight appear to render object colors in a natural way. This can be contrasted with narrow-band light sources such as sodium vapor lamps, which render object colors very poorly by not giving our eyes enough spectral content to allow good color discrimination. In lighting, efficacy is often at odds with color rendering, motivating the development of metrics that can be used to describe color rendering characteristics and allow specifiers and consumers to make informed decisions.

There have been a variety of color rendering metrics proposed, especially since the development of fluorescent light sources. The legacy industry standard is the CIE 13.3 color rendering index (CRI), which provides a summary number (Ra) describing the average accuracy in color saturation over a small set of standard colors, relative to a reference light source of the same correlated color temperature (CCT) [2]. CRI Ra was created in part to quantify the disappointment people found in the color rendering of fluorescent lighting. In some cases, norms have followed, for example requiring a certain minimum CRI (e.g. CRI Ra 80 for offices in Europe), and manufacturers responded by producing products tuned to meet the Ra values the market demanded.

Analogously to how fluorescent lighting spurred CRI, LED lighting, with the variety of spectral power distributions the technology can provide, has led to the development of newer metrics. The Illuminating Engineering Society (IES) has published a new standard for North America, commonly known as TM-30-15 (TM30, herein) [3]. TM30 builds on CRI, adds additional metrics beyond color accuracy, and attempts to address a variety of criticisms of the legacy metric. IES provides an Excel spreadsheet that performs the computations described in the TM30 document. The application of TM30 to a test light source includes these steps:

- Determine the CCT of the test source using a standard 2-degree observer.
- Define a reference source of the same CCT that is Planckian if below 4500K, using the CIE daylight model if above 5500K, and a linear mix in between.
- Compute the coordinates of 99 spectral color evaluation samples (CES) in CIECAM02 UCS using a standard 10-degree observer for both test and reference light sources.
- Compute the color difference as a Euclidean distance in CIECAM02 UCS, and report a fidelity metric Rf based on the mean color difference.
- Compute a polygon in the opponent-color (a,b) plane of CIECAM02 UCS based on binned hues of the 99 CES for both test and reference sources, and report a gamut Rg based on the ratio between the areas of the test and reference polygons.
- Provide bar charts of hue-binned fidelity and relative saturation along with a hue-circle distortion graphic.
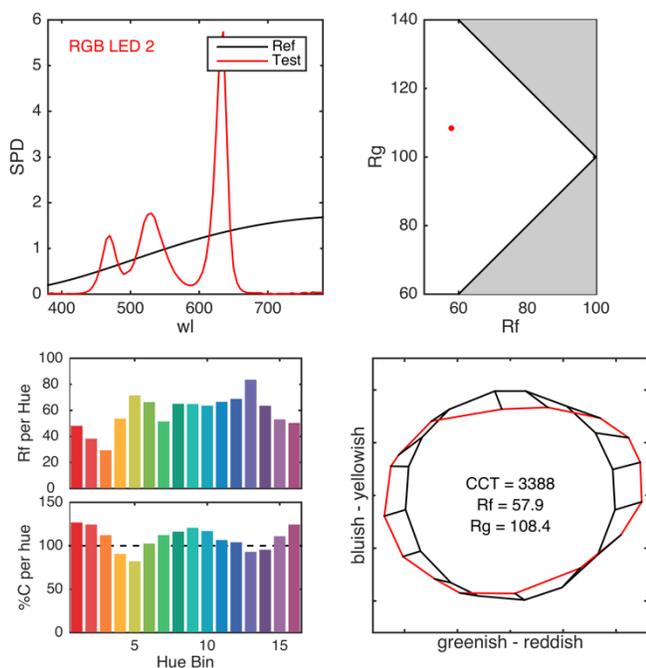
Figure 2: Example TM30-style graphics for an RGB LED light source. UL shows the SPDs of test (red) and reference (black) sources; UR displays gamut metric Rg versus fidelity metric Rf; LL bar charts show Rf and relative chroma per hue bin; and LR color vector graph shows color shifts of the normalized circular reference (black) to the test source (red).

Part of TM30's message is that a single metric or even a short list of numbers is not enough to adequately describe the color rendering characteristics of light sources, hence its heavy use of plots and graphics. Examples of these for a single light source are shown in Figure 2. Importantly, TM30 is not a quality or preference metric, but of course trends in Rf, Rg, and the enhancement and diminishment of color saturation in different parts of the hue circle have been shown to correlate with preference, depending on context [4][5]. If the variability among observers moves or blurs the points

of these metrics, that could lead to disappointment or increased cost. With that in mind, we modeled individual observer dependencies.

Modeling

A Matlab implementation of the TM30 metrics was written, based on the IES' TM30 document and related Excel spreadsheet. This re-implementation was necessary to allow the flexibility to vary the color matching functions (and thus diverge from IES' recommendation). The 10-degree CMFs used in the computation of CIECAM02 UCS values for both test and reference sources were exposed so that the individual colorimetric observers could be used. Similarly, the computation for CRI was modified to accept individual colorimetric observers in place of the 2-degree standard. Note that both CRI and TM30 use CIE 1931 2-degree CMFs to calculate CCT and define the reference source, and this was unchanged in the present model. Thus, the simulation is of real-world populations of observers assessing the comparison of test and reference illuminants (selected in the standard way), and potentially seeing differences thanks to their individual sensitivities.

In the present work, we simulated two different populations. A first population of 1,000 observers was computed using an age distribution taken from the most recent (2010) U.S. census, which is a wide distribution with a mean age of 39.9 years and standard deviation of 17.2. This sampling is as described by Asano et al. [1]. Both 2- and 10-degree sensitivities were simulated for the population. The populations' 10-degree CMFs are all shown in Figure 1 along with the CIE 1964 standard observer.

Second, because age is one of many drivers of individual differences, and one of the few that may be practically addressed in the market, another observer population was generated similarly. This second population includes 500 observers of age 25 years and 500 observers of age 65 years, and their 10-degree CMFs can be seen in Figure 3. With the younger observers plotted in transparent red and the older in transparent blue, biases are clearest at short wavelengths, while in other areas they blend together as purple.
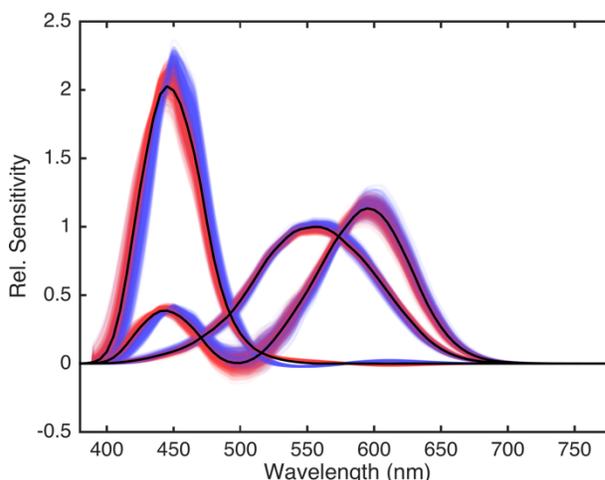
Figure 3: A population of color matching functions derived from the individual colorimetric observer model for two age subgroups of 500 observers each: 25 years (red) and 65 years (blue). Also shown is the CIE 1964 10-degree standard observer (black).

A variety of real light sources was selected to cover new and old technologies and differing spectral characteristics, including both broadband and mixed narrowband SPDs. These eight are listed in Table 1 and shown in Figure 4.

Table 1: List of selected light sources used in the model, showing their correlated color temperature (CCT), CIE CRI (Ra), and IES TM-30-15 Rf and Rg values.

	Source Type	CCT	Ra	Rf	Rg
A	Tungsten	2812	100	100	100
B	CIE D5500K	5500	100	100	100
C	Blue-pump LED	2732	97	95	103
D	RGB LED	3304	80	74	89
E	Blue-pump LED 2	3815	65	60	97
F	RGB LED 2	3388	38	58	108
G	Triphosphor Fluoro	3003	86	80	106
H	High Pressure Na	1923	15	29	61
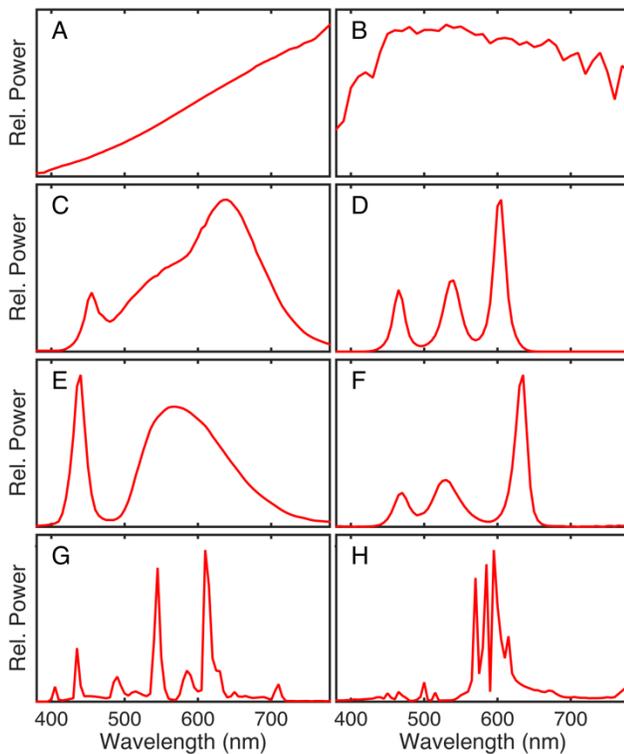
Figure 4: Relative spectral power distributions of selected light sources used in the model. Letter labels correspond to Table 1 and results: A is tungsten; B is CIE Daylight 5500K; C and E are different blue-pump phosphor LEDs; D and F are different RGB LEDs; G is triphosphor fluorescent; and H is high pressure sodium.

Results

The modeling results show how the natural variations in an observer population manifest themselves in terms of color rendering metrics. Looking first at the variation present in the first population, simulating the age distribution from the US census, we can see what degree of variation can be expected in the US population. In the TM30 Rg vs. Rf plot in Figure 5, clouds of points in red show the (Rf, Rg) values computed for the simulated observer population.

The clouds illustrate the variation and relative density surrounding the labeled black points, which are the TM30 values computed using the CIE 1964 10-degree standard observer. As a general trend, the clouds are smaller near the Rf=100 cusp, but their orientations and extents are not consistently patterned. Some (D, E, H) are essentially elliptical with various orientations, while others (C, F) appear to have an underlying "curl" to their shape. Results for the highest-Rf light sources (A, B) are essentially pinpoint, their extent exaggerated by the red plotted spot.
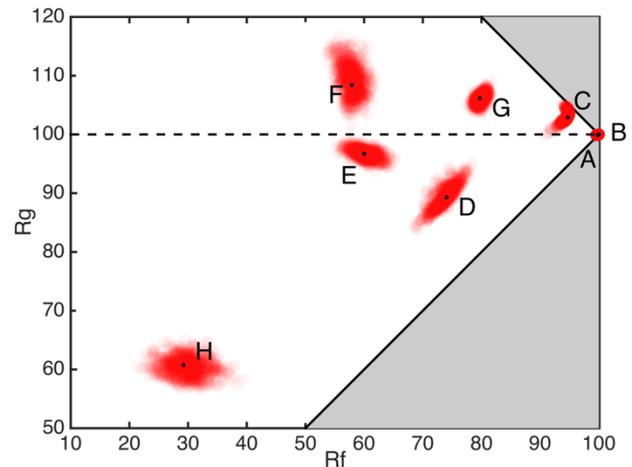
Figure 5: TM30 Rg vs. Rf plot showing clouds of points for each of the eight light sources due to modeled variability in observer color sensitivity. The labeled dots indicate the (Rf, Rg) values computed with the CIE 1964 standard observer. The gray regions of the plot are impossible because maximum fidelity (100) implies no change in gamut (100).
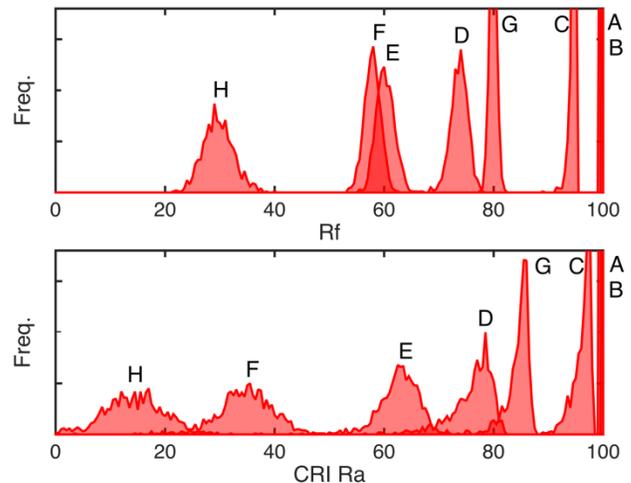
Figure 6: Histograms of the IES TM-30-15 Rf and legacy CIE CRI Ra values for each of the eight light sources. Visible distributions result from the differences among the individual colorimetric observers. Note that A and B have very narrow distributions in both metrics, close to 100.

Histogram distributions of TM30 Rf values and CIE Ra results computed for the individual colorimetric observers are shown in Figure 6. Admittedly, Rf and Ra are somewhat like apples and oranges, but comparing them is an obvious cross-check. Both

metrics show almost no variation for light sources A and B because they are spectrally nearly identical to their respective references. Interestingly, the new metric (Rf) shows narrower distributions than Ra for the remaining six light sources, with standard deviations ranging two to three times smaller.

TM30 emphasizes the use of color vector graphics, which indicate the distortion of binned hues in hue and saturation relative to the reference light source. Figure 7 shows color vector plots showing the relative distortion of four of the sources. In this graphic, the overall distortion due to the source is seen as the difference in shape between the shaded reference circle and the black polygon (computed using the CIE 1964 10-degree standard observer). The red polygons indicate the variation due to observer differences, which in some cases, like source F, is also hue dependent.
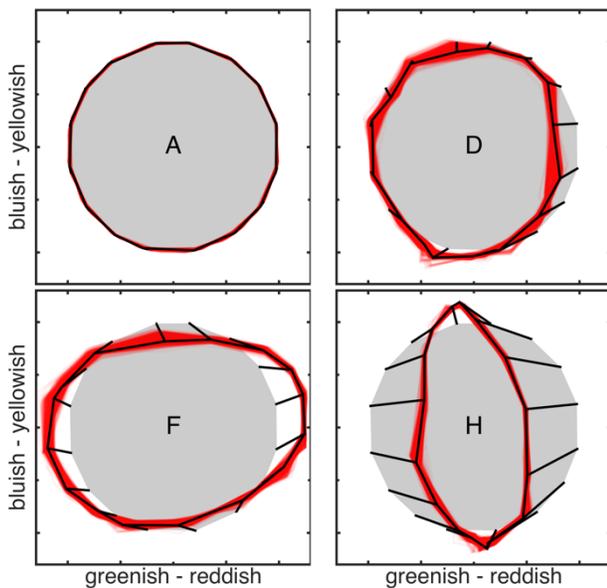
Figure 7: TM30 color vector graphics showing the relative distortion of hue and saturation in a normalized CIECAM02 UCS a'b' plane for four light sources. The resulting distortion using the standard observer is the difference between the shaded circle and the black polygon, highlighted with black vectors for each of 16 hue bins. Red polygons indicate the range of distortions seen by each of the individual colorimetric observers.

Turning to the second population with two different age subgroups, modeling results show distinct differences for some light sources. Figure 8 shows the TM30 Rg vs. Rf values computed for each of these ages, the younger group plotted in red and the older group plotted in blue. Where they largely overlap the clouds appear purple. Several of the sources in the figure show substantial disparities due to age, most notably D and F, the two RGB LED sources, and H, high pressure sodium. An interesting comparison can be seen with light sources E and F: younger observers would see the Rf of these sources very similar on average. However, the older observers would see an Rf difference of 5+ units. The older group would also see a larger apparent difference in Rg for these sources. In such a case, the equivalence of light sources' behavior, or sufficiency for Rf-based standards, would be directly affected by characteristics of different observer groups.
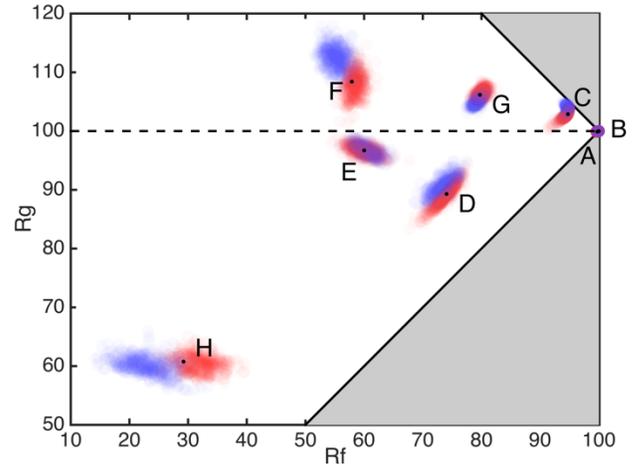
Figure 8: TM30 Rg vs. Rf plot for the different age populations. The 25 year old observers are plotted in red and 65 year old observers in blue. The biggest apparent differences due to observer age are seen in H (high pressure sodium), and D and F (both RGB LEDs).

Discussion

The present modeling results show how observer variability directly affects metrics used to describe color rendering. This reinforces the idea that while CIE standard observers provide a good indication of the mean of the population, they do not give any indication of the natural variation among individuals. One implication of this, for color rendering, is that the variability to be expected among people is in some cases greater than the nuanced differences between different light sources and technologies. In other words, based on the observed variations small differences (<5 units) in Rf or Rg might be "within the noise" and not meaningful.

Larger Differences for Lower Fidelity Sources

The highest fidelity light sources (approaching the maximum TM30 Rf value of 100) show the least susceptibility to inter-observer differences. Approaching this limit, light sources close to this Rf = 100 cusp are so spectrally similar to, and render color so similarly to, their reference sources that no individual observer sees the difference. Further from the cusp, the potential for observer differences increases. With larger rendering differences between sources and their references, the spectral details become important and mean individuals will see different things. Also, because Rg and Rf are not perceptually scaled to one another and may additionally be perceptually nonlinear, it is probably not appropriate to compare variation between the two axes.

Larger Differences for Narrowband Source Spectra

There appears a trend that differences between observers' responses to spectral power distributions are greater when the SPDs have steep transitions, narrow peaks, and/or little energy at some or many wavelengths. This is plainly manifested in the modeling results, where the clouds surrounding sources C-H are increasingly larger. The variability difference between D and F is quite large, and is surprising given overt similarity between their SPDs (and hinting at the importance of peak wavelength and minimum energy).

Variability in Ra and Rf

It would be a stretch to make a sweeping generalization based on limited modeling, but in the present investigation it was observed that inter-observer variability has a larger effect on CIE CRI Ra than on TM30 Rf. That being said, TM30's very existence is built upon the need for a more descriptive set of metrics than CRI ever offered. So while the standard deviations of Rf values are smaller than Ra values, the Rf values must be considered along with the Rg values and other information. Variability clouds for Rf and Rg apparently covary for most light sources as well, though they are not consistently elliptical, nor is there is an obvious pattern to their orientations.

Age Dependence

The age dependence visible in the modeling results is based on extremes, comparing 25 and 65 year old subgroups. It is expected that smaller age differences would correspond to smaller effects. Beyond the well-known age-dependent difference in visual sensitivity to short wavelengths, what is new in the present result is more clarity about how such sensitivity differences translate to color rendering differences. For some light sources, the differences in observed Rf and Rg values is quite large, and in some cases the older subgroup would see differences that the younger group would not.

Observer Metamerism Index

Quantifying the color rendering characteristics of light sources remains a tricky topic, and the IES has produced a rich toolset that continues to affect discussion throughout the industry. At the risk of complicating this discussion further, it would be interesting to consider a new kind of index that describes the susceptibility to or likelihood of inter-observer variability in color rendering. This could be based on simulation results as presented here, or perhaps modeled based on the correspondence of steep transitions to the regions of highest variability in color matching functions.

As scientific understanding of the natural variation in color sensitivity improves, reducing inter-observer variability may join the long list of goals that lighting manufacturers continue to try to optimize. An observer metamerism index could serve this goal. Additionally, illustrative statistics such as the expected covariance matrices for Rf and Rg over an observer population would enable uncertainty to be indicated in Rg vs. Rf plots and bar charts. Future work is suggested to create and validate a summary observer metamerism index for light sources.

Conclusion

Assessing the color rendering properties of light sources is a complicated matter that depends on the source SPD, the geometry of illumination, the colors evaluated, and as shown here the individual observer. Thus, it is not surprising that metrics designed to quantify the color rendition of light sources need to be more complex than the single number of CIE CRI Ra or even the two numbers and graphical tools of IES TM30. It has been shown that observer variability in color matching functions is real and predictable, and it can be used to compute the range of color rendition metrics in CRI and TM30 in a meaningful way.

Modeling results show a general trend with Rf and Ra such that light sources with a high color fidelity (high Rf and/or Ra) also show relatively little individual variation. Sources with lower fidelity show significantly more variation. This is somewhat to be expected since high fidelity sources tend to be spectrally similar to the reference light sources with continuous spectra that might help to minimize individual variation.

Examination of the simulated results indicates that the clouds of observer variation span distances in the Rg vs. Rf plot as large as the distances between some of the evaluated light sources. This suggests that users of TM30 should perhaps ignore differences smaller than the sizes of these clouds: five to ten units of Rf over much of the plot, slightly less at high Rf, and slightly more at low Rf. In addition, analysis of populations of different ages indicates that age dependency in the color rendition metrics is quite large, more than ten units of Rf in the lower Rf range.

As awareness of observer variation increases, further study of additional metrics is recommended to describe the variation in real populations of observers for different light source types.

References

[1] Y. Asano, M. D. Fairchild, L. Blondé "Individual Colorimetric Observer Model," PLoS ONE 11(2): e0145671. doi: 10.1371/journal.pone.0145671, 2016.

[2] T. Azuma et al. Method of Measuring and Specifying Colour Rendering Properties of Light Sources. CIE 13.3-1995, ISBN 978-3-900734-57-2, 1995.

[3] M. Royer et al. IES Method for Evaluating Light Source Color Rendition. IES TM-30-15, ISBN 978-0-87995-312-6, 2015.

[4] Y. Ohno; C. C. Miller; M. Fein, "Vision Experiment On Chroma Saturation for Color Quality Preference," in 28th Session of the CIE, Manchester, UK, 2015.

[5] Y. Lin, et al. "Colour preference varies with lighting application," Lighting Research and Technology, doi: 10.1177/1477153515611458, 2015.

Author Biographies

Michael J. Murdoch is an Assistant Professor in the Program of Color Science and Munsell Color Science Laboratory at RIT in Rochester, NY. He earned his PhD in Human-Technology Interaction from Eindhoven University of Technology (The Netherlands), MS in Computer Science from Rochester Institute of Technology, and BS in Chemical Engineering from Cornell University. His current research interests include visual quality of imaging and lighting systems and color perception in augmented and virtual reality. He is a member of IS&T.

Mark D. Fairchild is Professor and Associate Dean of Research and Graduate Education of RIT's College of Science and Director of the Program of Color Science and Munsell Color Science Laboratory. He received his B.S. and M.S. degrees in Imaging Science from R.I.T. and Ph.D. in Vision Science from the University of Rochester. Mark was presented with the 1995 Bartleson Award by the Colour Group (Great Britain) and the 2002 Macbeth Award by the Inter-Society Color Council for his works in color appearance and color science. He is a Fellow of the Society for Imaging Science and Technology (IS&T) and the Optical Society of America. Mark was presented with the Davies Medal by the Royal Photographic Society for contributions to photography. He received the 2008 IS&T Raymond C. Bowman award for excellence in education.