Development of the IES method for evaluating the color rendition of light sources

Aurelien David,[1,*] Paul T. Fini,[2] Kevin W. Houser,[3] Yoshi Ohno,[4] Michael P. Royer,[5] Kevin A. G. Smet,[6] Minchen Wei[3] and Lorne Whitehead[7]

[1]*Soraa Inc., 6500 Kaiser Dr, Fremont California 94555, USA*
[2]*Cree Inc., 340 Storke Rd, Goleta, California 93117, USA*
[3]*Dept. of Architectural Engineering, Penn State University, State College, Pennsylvania 16801, USA*
[4]*National Institute of Standards and Technology, 100 Bureau Dr, Gaithersburg, Maryland 20899, USA*
[5]*Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, Washington 99354, USA*
[6]*Light&Lighting Laboratory/ESAT, KU Leuven, 3000 Ghent, Belgium*
[7]*Department of Physics and Astronomy, University of British Columbia, Vancouver V6T 1Z4 BC, Canada*
[*]*aurelien.david@polytechnique.org*

Abstract: We have developed a two-measure system for evaluating light sources' color rendition that builds upon conceptual progress of numerous researchers over the last two decades. The system quantifies the color fidelity and color gamut (change in object chroma) of a light source in comparison to a reference illuminant. The calculations are based on a newly developed set of reflectance data from real samples uniformly distributed in color space (thereby fairly representing all colors) and in wavelength space (thereby precluding artificial optimization of the color rendition scores by spectral engineering). The color fidelity score R_f is an improved version of the CIE color rendering index. The color gamut score R_g is an improved version of the Gamut Area Index. In combination, they provide two complementary assessments to guide the optimization of future light sources. This method summarizes the findings of the Color Metric Task Group of the Illuminating Engineering Society of North America (IES). It is adopted in the upcoming IES TM-30-2015, and is proposed for consideration with the International Commission on Illumination (CIE).

References and links

1. Commission Internationale de l'Eclairage, "Method of measuring and specifying colour rendering properties of light sources," Technical Report CIE **013.3** (1995).
2. J. A. Worthey, "Color rendering: asking the questions," Color Res. Appl. **28**(6), 403–412 (2003).
3. Commission Internationale de l'Eclairage, "Color rendering of white LED light sources," Technical Report CIE **177**, 2007 (2007).
4. D. L. DiLaura, K. W. Houser, R. G. Mistrick, and G. R. Steffy, eds., *The Lighting handbook: reference and application*, Xth Edition (Illuminating Engineering Society, 2011).
5. X. Guo and K. W. Houser, "A review of colour rendering indices and their application to commercial light sources," Lighting Res. Tech. **36**(3), 183–197 (2004).
6. K. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, and P. Hanselaer, "Correlation between color quality metric predictions and visual appreciation of light sources," Opt. Express **19**(9), 8151–8166 (2011).
7. K. W. Houser, M. Wei, A. David, M. R. Krames, and X. S. Shen, "Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition," Opt. Express **21**(8), 10393–10411 (2013).
8. The term "color rendition" does not appear in the CIE International Lighting Vocabulary. We use its general definition as "the effect of a light source on the color appearance of objects illuminated by the light source".
9. Commission Internationale de l'Eclairage, "Colour Rendering, CIE TC 1-33 closing remarks," CIE **135/2** (1999).
10. K. W. Houser, "If not CRI, then what?" Leukos **9**(3), 151–153 (2013).
11. ANSI, "Specifications for the chromaticity of solid state lighting products," standard **C78.377–2008** (2008).
12. E. E. Dikel, G. J. Burns, J. A. Veitch, S. Mancini, and G. R. Newsham, "Preferred chromaticity of color-tunable LED lighting," Leukos **10**(2), 101–115 (2014).

13. M. S. Rea and J. P. Freyssinier-Nova, "Color rendering: a tale of two metrics," Color Res. Appl. **33**(3), 192–202 (2008).
14. N. Sandor and J. Schanda, "Visual colour rendering based on colour difference evaluations," Lighting Res. Tech. **38**(3), 225–239 (2006).
15. W. Davis and Y. Ohno, "Approaches to color rendering measurement," J. Mod. Opt. **56**(13), 1412–1419 (2009).
16. K. Smet and L. Whitehead, "Meta-standards for color rendering metrics and implications for sample spectral sets," Proceedings of the 19th Color and Imaging Conference, San Jose USA (2011).
17. M. Rea and J. P. Freyssinier, "Color rendering beyond pride and prejudice," Color Res. Appl. **35**(6), 401–409 (2010).
18. M. Wei, K. W. Houser, G. R. Allen, and W. W. Beers, "Color preference under LEDs with diminished yellow emission," Leukos **10**(3), 119–131 (2014).
19. E. de Beer, P. van der Burgt, and J. van Kemenade, "Another color rendering metric: do we really need it, can we live without it?" Leukos (2015).
20. W. Davis and Y. Ohno, "Color quality scale," Opt. Eng. **49**(3), 033602 (2010).
21. K. Smet, J. Schanda, L. Whitehead, and M. R. Luo, "CRI2012: A proposal for updating the CIE colour rendering index," Lighting Res. Tech. **45**(6), 689–709 (2013).
22. K. Smet, L. Whitehead, J. Schanda, and M. R. Luo, "Toward a replacement of the CIE color rendering index for white light sources," Leukos (2015).
23. C. Li, M. Ronnier Luo, C. Li, and G. Cui, "The CRI-CAM02UCS colour rendering index," Color Res. Appl. **37**(3), 160–167 (2012).
24. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. **31**(4), 320–330 (2006).
25. A. Zukauskas, R. Vaicekauskas, F. Ivanauskas, H. Vaitkevicius, P. Vitta, and M. S. Shur, "Statistical approach to color quality of solid-state lamps," IEEE J. Sel. Top. Quantum Electron. **15**(6), 1753–1762 (2009).
26. P. van der Burgt and J. van Kemenade, "About color rendition of light sources: the balance between simplicity and accuracy," Color Res. Appl. **35**(2), 85–93 (2010).
27. L. Whitehead and M. Mossman, "A Monte Carlo method for assessing color rendering quality with possible application to color rendering standards," Color Res. Appl. **37**(1), 13–22 (2012).
28. A. David, "Color fidelity of light sources evaluated over large sets of reflectance samples," Leukos **10**(2), 59–75 (2014).
29. Commission Internationale de l'Eclairage, "Division 1: vision and color, meeting minutes," CIE Division 1 Meeting, Warsaw (2012).
30. R. Hunt and M. Pointer, *Measuring colour,* IVth edition (Wiley Ed, 2011), Chap. 15.
31. Lightness shifts are also possible but are not shown for simplicity. Further, Fig. 1 only shows average color shifts; however in practice metameric colors generally undergo different shifts.
32. D. Judd, "A flattery index for artificial illuminants," Illum. Eng. **62**, 593–598 (1967).
33. C. W. Jerome, "Flattery vs color rendition," J. Illum. Eng. Soc. **1**(3), 208–211 (1972).
34. C. W. Jerome, "The flattery index," J. Illum. Eng. Soc. **2**(4), 351–354 (1973).
35. W. A. Thornton, "A validation of the color-preference index," J. Illum. Eng. Soc. **4**(1), 48–52 (1974).
36. K. A. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, and P. Hanselaer, "Memory colours and colour quality evaluation of conventional and solid-state lamps," Opt. Express **18**(25), 26229–26244 (2010).
37. M. Wei, K. Houser, A. David, and M. Krames, "Perceptual responses to LED illumination with colour rendering indices of 85 and 97," Light. Res. Tech. (2014).
38. Y. Ohno, "Vision experiment on chroma saturation for color quality preference," to be published (2015).
39. K. Smet and P. Hanselaer, "Memory and preferred colours and the colour rendition evaluation of white light sources," Light. Res. Tech. (2015).
40. Y. Ohno, "Spectral design considerations for white LED color rendering," Opt. Eng. **44**(11), 111302 (2005).
41. The shaded area of Fig. 2 is an approximate boundary; some SPDs far off-Planckian can in fact slightly enter the grayed-out zone.
42. Specifically, the small color difference sets which are most relevant for color rendition calculations.
43. The chromaticity of samples only weakly depends on CCT in CAM02-UCS, so that color space uniformity at a given CCT is translated to other CCTs.
44. A. David, "Colour fidelity evaluated over large reflectance datasets," Proceedings of the CIE meeting, Kuala Lumpur (2014).
45. A. David, K. A. G. Smet, and L. Whitehead, to be published.
46. Y. Ohno, "Practical use and calculation of CCT and Duv," Leukos **10**(1), 47–55 (2014).
47. In past research, use of the RMS mean has been proposed. In our case however, the large number of samples makes the arithmetic mean a safe and simple choice. Besides, in practice, arithmetic and RMS means yield nearly identical results for most SPDs.
48. These boundaries were obtained by considering the 5,000 SPDs of Appendix A, together with a second library of 550,000 four-laser-line spectra with a CCT of 3000 K, whose peak wavelengths were systematically varied.
49. A. David, to be published.
50. M. Rea and J. P. Freyssinier, "White lighting," Color Res. Appl. **38**(2), 82–92 (2013).
51. Y. Ohno and M. Fein, "Vision experiment on acceptable and preferred white light chromaticity for lighting," Proceedings of the CIE conference on lighting quality and energy efficiency, Kuala Lumpur (2014).
52. K. Smet, G. Deconinck, and P. Hanselaer, "Chromaticity of unique white in illumination mode," Opt. Express **23**(10), 12488–12495 (2015).

53. K. W. Houser, M. Wei, A. David, and M. Krames, "Whiteness perception under LED illumination," Leukos **10**(3), 165–180 (2014).
54. C. Li, M. R. Luo, M. Pointer, and P. Green, "Comparison of real colour gamuts using a new reflectance database," Color Res. Appl. **39**(5), 442–451 (2014).
55. International Organization for Standardization, "Graphic technology–standard object colour spectra database for colour reproduction evaluation (SOCS)," ISO Norm **16066** (2003).
56. M. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," Color Res. Appl. **19**(1), 4–9 (1994). Database at ftp://ftp.eos.ncsu.edu/pub/eos/pub/spectra/ (accessed 2013).
57. University of Eastern Finland, "Spectral database" at http://www.uef.fi/spectral/spectral-database (accessed 2013).
58. S. Arnold, S. Faruq, V. Savolainen, P. McOwan, and L. Chittka, "Fred: the floral reflectance database—a web portal for analyses of flower colour," PLoS ONE **5**(12), e14287 (2010). database @ http://reflectance.co.uk/tou.php (accessed 2013).
59. Z. M. Kovacs-Vajna, "Rs2color database" at http://www.ing.unibs.it/zkovacs/color/rs2color/rs2colore.html (accessed 2013).

1. Introduction

Limitations of the general color rendering index (CRI) [1], developed by the International Commission on Illumination (CIE), have been extensively documented [2–7]. Despite many past efforts to develop complimentary or alternative ways for evaluating light sources' color rendition [5–8], a suitable alternative has not yet been widely adopted; efforts by the CIE to revise the CRI in the 1980s-90s did not succeed [9]. Yet, with the proliferation of solid-state lighting—a family of light sources with tremendous opportunities for spectral engineering and optimization—the need for a method to evaluate color rendition is greater than ever [10]. The system must be scientifically defensible, yet easy to interpret and practical. The accuracy of such a method is also critical in terms of energy efficiency due to the complex trade-offs between aspects of color rendition. The method described in this article was developed with these goals in mind.

2. Background

Below, we summarize a few important characteristics of a light source specific to its spectral power distribution (SPD).

- •**Luminous efficacy of radiation** (LER) describes the energy efficiency of an SPD, and is calculated as the ratio of luminous flux (lumens) to radiant flux (watts). LER generally increases if the SPD contains less short- and long-wavelength radiation, but this can negatively impact aspects of color rendition.

- •**Chromaticity** describes the coordinates of an SPD in a diagram such as CIE (*xy*). Blackbody radiators fall upon a curve in this diagram and can be parameterized by their color temperature. Other light sources are characterized by their **correlated color temperature** (CCT) and their distance from the Planckian locus D_{uv} [11], which may also affect color rendition and preference [12].

- •**Color fidelity** quantifies the accuracy with which the color appearances of illuminated objects match their appearances under a reference illuminant (such as daylight or blackbody radiation). The CRI is an example of a fidelity measure. Although it has usefully guided the design of light sources for decades, it has well-known limitations (as summarized in Section 3). There is a widespread desire to improve upon the CRI and one of the purposes of this article is to do so.

- •**Color gamut** quantifies the average increase or decrease of the chroma of objects (relative to those under a reference illuminant), and approximately describes how vivid objects appear. We note that "gamut" in this context has a different meaning from its common use in display applications: here we use the term to designate an *overall change* in chroma of objects induced by a light source. The well-known Gamut Area Index (GAI) [13] brought recent popularity to the use of the word

"gamut". GAI suffers from some flaws, and another purpose of this article is to provide an improved gamut measure.

There are unavoidable tradeoffs between the aforementioned concepts; therefore it is important that their measures be accurate to optimize against these tradeoffs. These four concepts are *objective* measures that do not depend on the personal opinions of observers. They should be distinguished from an often-discussed *subjective* concept called **color preference** of a light source—which has connections to color gamut, as will be discussed.

This article presents the synthesis of efforts to develop an improved system to characterize two key aspects of color rendition: color fidelity and color gamut. The work presented here builds upon important concepts introduced over the past decade and integrates them into a consistent framework, while improving their accuracy. In the following, we review the previous relevant research, especially the important concepts motivating this work. Then we propose a new two-measure system and discuss some of the characteristics of the resulting information.

3. Discussion of previous work

3.1 Intentional limitations and unintentional shortcomings of the CRI

The CRI, by design, is a fidelity measure: it produces an average number (R_a) quantifying how closely a test source renders objects' colors like a reference illuminant does. In many situations, fidelity is an important characteristic; for instance, when people are consciously or unconsciously comparing the color appearance of objects to that in their previous experiences. However, fidelity does not always equate to desirability [7,14]—a confusion commonly made in interpreting R_a values. While this is intentional (fidelity is an objective measure and thus cannot quantify subjective desirability), additional measures related to desirability would be useful.

Additionally, the CRI also suffers from deficiencies in its purpose as a fidelity measure, which fall into two categories:

- Calculation procedure: the calculation of object colors and their color shifts are all carried out using outdated tools and formulas (such as the CIE *U*V*W** space and the von Kries chromatic adaptation) [15]. Thus, there is a need to update calculation procedures with the most up-to-date methods.

- Test Color Samples (TCS): the eight TCS used in calculating R_a are low-chroma Munsell colors. These few samples are not fully representative of our common environment [15]. In particular, R_a is poorly predictive of the fidelity of saturated red colors and often has to be supplemented with the special index R_9 for deep red. Additionally, the TCS are more sensitive to some wavelengths than others because they are made by combining only a few pigments whose spectral features are not uniformly distributed across visible wavelengths. As a result, the CRI can be "gamed" by selectively optimizing an SPD in ways that do not improve color fidelity, but nevertheless boost R_a [16].

In summary, the CRI is limited in its ability to assess color fidelity, and alternative measures are needed to assess other important effects of light sources on color rendition.

3.2 Past work on color rendition measures

Extensive research activity has sought to supplement the CRI in past decades. Below, we summarize a few key conceptual contributions that form the basis upon which the current proposal was built.

In [6,13,17], it was proposed to describe color rendition using **two measures** [13]. proposed using R_a for fidelity and the newly-introduced GAI for gamut. The addition of a gamut measure brings additional information on color rendition: a source may make colors more dull or more vivid than does a reference source, which respectively corresponds to a decrease or increase in gamut. User perception—especially subjective preference—is

influenced by chroma changes, and thus GAI provides valuable information complementary to that of R_a. It is important to note that R_a and GAI are single numbers, and thus only provide average information about color rendering [18,19].

Regarding the concept of pairing two measures [8], recently reviewed a large number of color rendition measures and analyzed correlations in their predictions. The main conclusion of the study was that all reviewed measures can be grouped (in terms of statistical correlation) into one of two clusters, respectively identified with **fidelity and gamut/preference**. From this perspective, the two-measure proposal has general bearing: combining a fidelity measure with a gamut measure is the most general way to convey information about color rendition with only two values (within the context of the reviewed measures). Of course, the choice of the specific fidelity and gamut measures is important, as not all measures are equally accurate.

The Color Quality Scale (CQS) [20] proposed various improvements to the accuracy of color rendition measures. It consists of three indices: overall color quality (Q_a), fidelity (Q_f) and gamut (Q_g). Among the features included in the CQS are a new set of 15 color samples with higher chroma and uniform hue coverage, a more accurate color space (CIELAB), and visualization tools. The use of **high-chroma samples** addresses the shortcoming of the CRI for saturated colors. Another important feature of Q_a is the saturation factor (in which moderate increases in chroma are not penalized) to account for the influence of chroma on preference. The **visualization tools** convey intuitive information on color distortion and supplement the gamut value Q_g; especially the color-saturation icon, which indicates how various colors are distorted and complements the single-valued measures.

The CRI2012 [16,21,22] introduced the new concept of **spectral uniformity**. As mentioned previously, the CRI is susceptible to selective optimization; that is, its TCS induce more sensitivity to some wavelengths than others. To address this [16,22], proposed employing a set of color samples for which no wavelength is privileged—a property termed spectral uniformity. This was achieved by carefully choosing reflectance spectra, namely either a large set of 1,000 samples or a smaller set of 17 samples. Both of these sets were specially designed for spectral uniformity and have a more varied chroma than the CRI's TCS. Following [23], the CRI2012 also adopted one of the most **uniform color spaces** known to date, CAM02-UCS [24].

Finally, various groups have noticed the availability of databases of reflectance samples over which color rendition calculations can be performed [16,25–28]. While large sample sets can improve accuracy, they can also introduce additional bias and averaging must be done carefully [28]. discussed this and showed that variations in color fidelity may occur anywhere in color space, suggesting that a thorough coverage of color space is desirable—in contrast to the few probe points offered by small sample sets. This **'color space uniformity'** approach extends the concept of uniform hue coverage [20,21,26] to all color dimensions (hue, chroma, lightness). The use of large sample sets, however, also brings up the **tradeoff between complexity and accuracy** [6,26]: ideally, one wants a sample set of minimal size that yields sufficient accuracy.

Unfortunately, there has been no consensus thus far on integrating the important features highlighted above in a consistent framework [29]. This has been hindered by their seemingly incompatible requirements; for instance, how many measures are desirable and acceptable, and what samples should be used (few for simplicity or many for completeness, real or artificial samples, etc). The present work is a consensus reached by a diverse group of academic and industrial stakeholders, combining these features without significantly compromising any of them.

4. Two-measure system based on an improved sample set

This proposal integrates the key improvements described above:

•a two-measure system describing color fidelity (R_f) and color gamut (R_g)

•a set of 99 real test samples with color space uniformity and spectral uniformity

•a calculation engine based on the state-of-the-art uniform color space CAM02-UCS

•convenient visualization tools including a color distortion icon

The reader is reminded that CAM02-UCS is a color appearance space with three dimensions: J' (lightness), a' (red-green opponent channel) and b' (yellow-blue opponent channel). It has been optimized to correlate well with experimental observations of the color appearance of objects under a wide variety of visual settings [24,30].

4.1 Use of a two-measure system

It is by now widely accepted that no single measure can adequately describe all aspects of color rendition. Color fidelity, which quantifies distortions to objects' color appearance with respect to that under a reference illuminant, is well-defined but, by design, conveys limited information. When color appearance is distorted by a light source, the color shift can occur in various directions; it can be a de-saturating shift (making the color duller), a saturating shift (making it more vivid), and/or a hue shift. A fidelity measure considers all such shifts on equal footing (any shift causes a decrease in fidelity); however, they correspond to different perceptual experiences. Color gamut measures assess this distinction by computing the average change in chroma caused by a source: saturating/de-saturating shifts respectively correspond to a gamut increase/decrease. Figure 1(a) illustrates this [31].
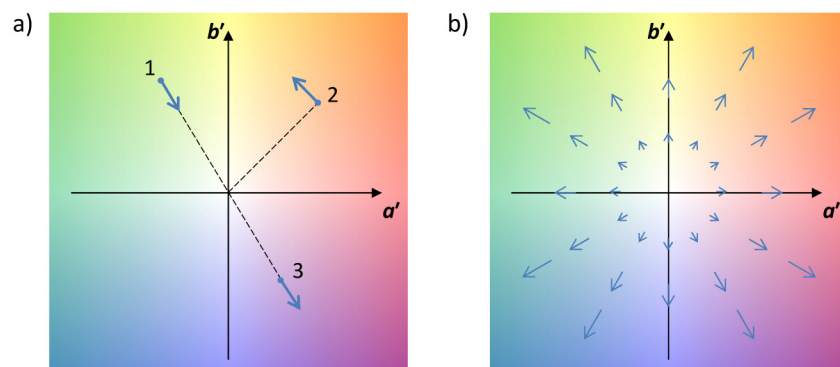

Fig. 1. (a) A light source can induce various color shifts, as shown here in the (a', b') plane, such as (1) desaturating (chroma decrease), (2) hue shift, and (3) saturating (chroma increase). These shifts form a vector field that varies through color space [25,28]. (b) Hypothetical light source causing purely saturating shifts; loss of fidelity is then associated with a maximal gamut increase.

As shown in [8], nearly all existing color rendition measures are well correlated to fidelity, gamut, or a combination of the two. Therefore, among these existing rendition measures, quantifying fidelity and gamut conveys the maximum independent amount of information that can be expressed by two numbers, and is expected to bring useful information about the light source.

As mentioned previously, there is also a connection between the objective concept of color gamut and the loosely-defined but important subjective concept of color preference. Indeed, in various studies [32–39], light sources that generally increase chroma are described as more pleasant by most observers. For sources rendering colors with a lower chroma than the reference illuminant, preference and fidelity are strongly correlated. This correlation, however, is reduced for sources that render colors with a higher chroma than the reference, which is possible with narrow-band sources [40]—excessive increase of chroma can cause a negative perception effect, so the nature of relationship of color gamut to user preference is not simple. Currently, there is no widely-accepted method for predicting color preference; experimental research is still ongoing on this complex topic. For this reason, in this article we take a conservative position in employing a gamut measure when attempting to predict user satisfaction, but we make no general recommendations for gamut values. In other words, the

gamut value is informative, providing the user with additional knowledge about what color distortions to expect. This information can then be considered within the context of each specific application.

There is a fundamental limiting relationship between fidelity and gamut: perfect fidelity (a score of 100) can only be obtained when colors exactly match those under the reference illuminant, thus yielding no variation in chroma (a gamut score of 100). Color shift in any direction will decrease fidelity, and may either increase or decrease chroma. There is a maximum amount of chroma that can be gained (or lost) for a given color shift—this is the case if all shifts are in the radial direction, as illustrated in Fig. 1(b). Therefore, there are a maximum and a minimum gamut achievable for a given fidelity. This limiting relationship is illustrated on Fig. 2. When using our metrics R_f and R_g, the shape of this boundary is a diagonal line [41]. The use of accurate measures is essential to properly describe this important limiting relationship; for instance, R_a and GAI do not provide such clear insight due to the fixed reference illuminant of GAI [8,19] (this yields misleading results for some SPDs, where GAI values can be increased without a concurrent decrease in Ra).
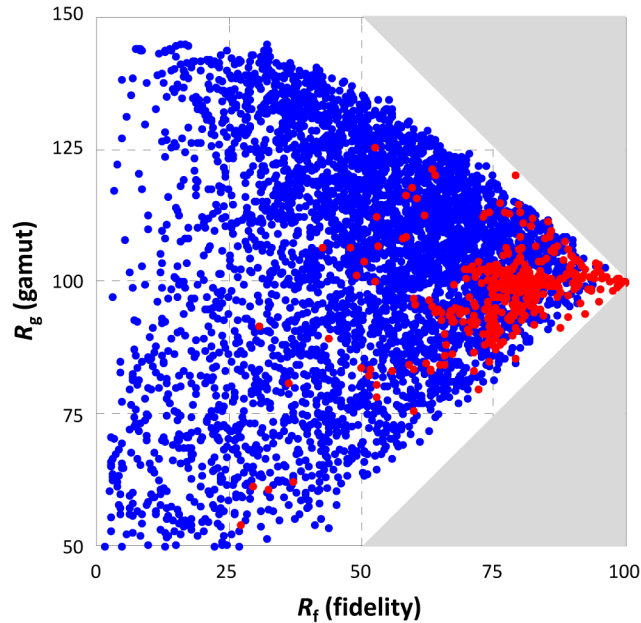

Fig. 2. Tradeoff between fidelity and gamut; light sources can only reside in the non-shaded area. Each dot corresponds to an SPD as described in Appendix A (red dots: 'real' SPDs from [8]; blue dots: synthetic SPDs).

Finally, there are cases where single averaged numbers are not enough to provide sufficient information (for instance, if a user needs to know which specific colors are distorted and what kind of distortion to expect). We address this need by also computing a color distortion icon, which draws from the color vector maps of [26] and the icon of [20].

4.2 Sample set

The derivation of the sample set was the subject of particular care. As a starting point, a large collection of reflectance data was gathered, as described in Appendix B. The resulting *Large Set* contained about 105,000 reflectance spectra from various types of objects: flowers and other natural objects, skin tones, textiles, paints, plastics, printed materials, and color systems (*i.e.,* Munsell, Natural Color System [NCS], German Institute for Standardization [DIN]). These objects are representative of materials found in interiors and in nature; they also

possess a variety of surface finishes (some are glossy, purely diffuse). Since we seek to quantify *average* color errors, we employ the total diffuse reflectance of these objects.

In order to generate a usable set from the Large Set, we first obtained a *Reference Set* that fulfilled all desired properties; specifically, it is composed of real reflectance spectra representative of common objects, and is uniformly distributed in color space while having spectral features that are evenly distributed in wavelength space. Subsequently, a *Final Set* with far fewer samples was selected to reproduce the predictions of the Reference Set with a faster calculation time and ease of programming. The workflow followed in generating the sample set is summarized in Fig. 3. The selection procedure is described below, and the method for correlating predictions between sets is described in Appendix A.
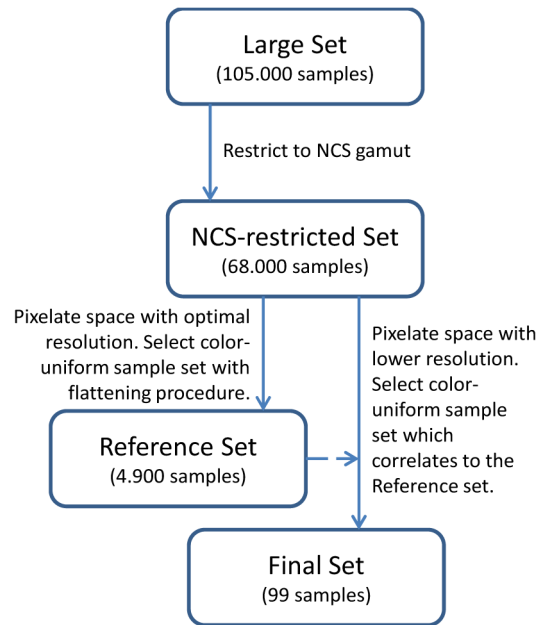
```
              ┌─────────────────────┐
              │     Large Set       │
              │  (105.000 samples)  │
              └─────────────────────┘
                        │
                Restrict to NCS gamut
                        │
              ┌─────────────────────┐
              │ NCS-restricted Set  │
              │  (68.000 samples)   │
              └─────────────────────┘
```

Pixelate space with optimal resolution. Select color-uniform sample set with flattening procedure.

Pixelate space with lower resolution. Select color-uniform sample set which correlates to the Reference set.

```
   ┌─────────────────────┐         ┌─────────────────────┐
   │   Reference Set     │ ──────▶ │     Final Set       │
   │   (4.900 samples)   │         │    (99 samples)     │
   └─────────────────────┘         └─────────────────────┘
```

Fig. 3. Workflow of sample sets generation.

4.2.1 Reflectance data and gamut restriction

As discussed in [28], the Large Set covers a large gamut in color space—much larger than that of typical samples used by other color rendition measures. Therefore, it was necessary to decide whether this whole gamut should be employed in calculations. On one hand, it was desirable to account for as many colors as possible. On the other, samples with extreme color coordinates (very dark or saturated) might be uncommon and less relevant in a typical interior environment. Finally, the range of validity of color error formulas had to be considered. All modern color error formulas, including CAM02-UCS, are derived from a common set of experiments on test samples called *color-difference samples*. Outside the gamut where this experimental data is fitted, there is no certainty that color error formulas are accurate.

Figure 4 shows the gamut of all samples in the Large Set compared to the gamut of the color-difference samples [42]. Clearly, the Large Set extends to regions of the color space where color error formulas have not been tested. Figure 4 also shows the gamut of the NCS color atlas. Interestingly, the latter agrees very well with the gamut of the color-difference samples (except for very dark samples). Therefore, we made the conservative choice of only considering samples if they were inside the NCS gamut. This way, we eliminated extremely saturated and dark samples, and only operated in regions where color error is accurate. It is also likely that such extreme samples are not commonly encountered, and it is proposed that the restriction to the NCS gamut better represents common interior situations. Another

consequence of this restriction is to reduce the difficulty to obtain a given fidelity score for a given LER, as this tradeoff is generally more pronounced for very saturated colors.

After this cropping procedure, we obtained a set of 68,000 samples. This set was quite exhaustive, however it was uneven both in color space (some colors were over-represented) and in wavelength space (some types of objects containing a few specific dyes were prevalent). Therefore, we still needed to extract samples evenly distributed in color space and having evenly-distributed features in wavelength space.
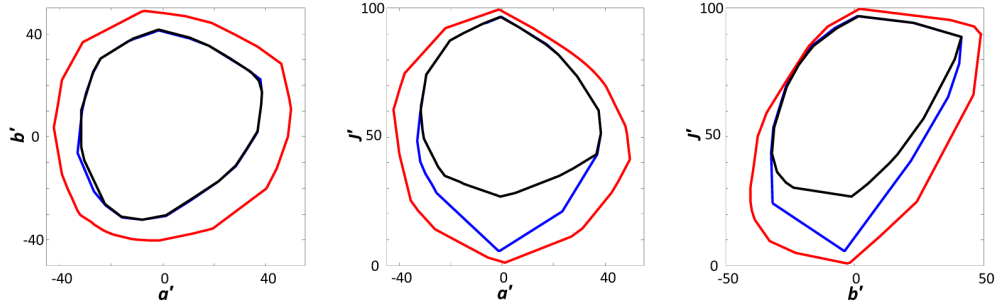
Fig. 4. Gamut of three sample sets (cross-sections of CAM02UCS space). Red: large set; Blue: *color-difference samples*; black: NCS atlas.

4.2.2 Color space uniformity

Color space uniformity was achieved following the method described in [28]. We partitioned the (J', a', b') color space into cubic pixels with a small side length d (with $d = da' = db' = dJ'$). For each reflectance sample, we computed the color coordinates under a 5000 K blackbody [43] and binned the samples in each pixel. Given a test spectrum, the RMS mean of the color errors for each sample in the pixel was determined, yielding an average color error for the pixel; this is illustrated in Fig. 5(a).
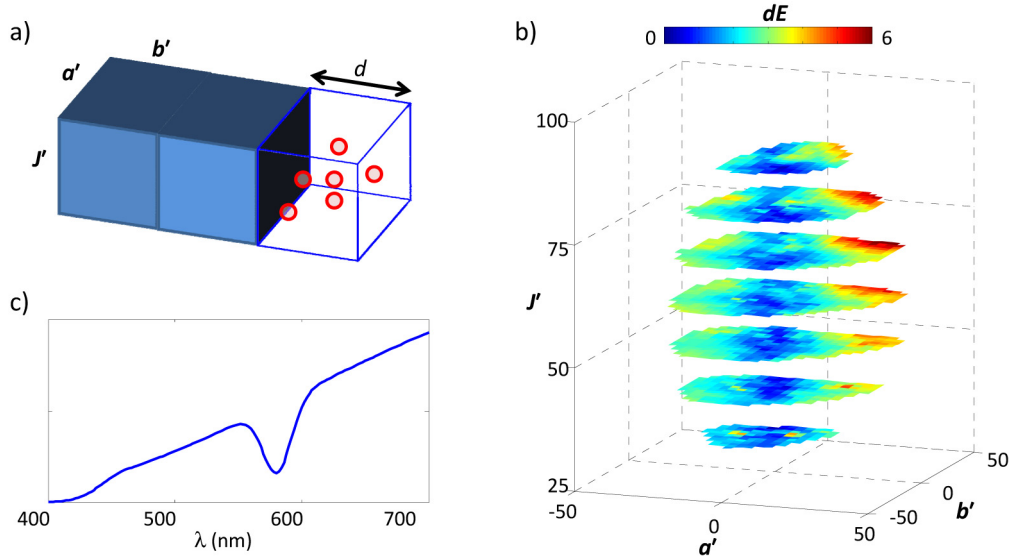

Fig. 5. (a) Pixelation approach. All the samples within a pixel are considered (approximately) metameric. Values are averaged over each pixel, then across pixels. (b) Map showing the local value of color error dE in (J', a', b') space. The calculation is performed over the Reference Sample Set. The corresponding SPD (here a neodymium incandescent source) is shown in (c).

This procedure effectively removed selection bias in our reflectance sample set. Indeed, if many nearly-identical reflectance samples were included in the Large Set, they all contribute to the same pixel and their outsized representation in the sample set does not dominate the final results. The underlying argument justifying this pixel approach is as follows. Assume for simplicity that the pixel size d is on the order of a just-noticeable color difference (say $d = 2$). In this case, all samples in a pixel can be considered metameric, and each pixel represents a distinguishable color. Therefore, the average color error for a pixel characterizes the average rendition of this specific color.

Thus, a pixelated map of color error in color space can be created, which can be used to perform fidelity calculations. Likewise, other quantities can be computed besides color error (chroma, hue, etc.) to perform a gamut calculation. Obviously, the argument is only valid if the color space is sufficiently uniform. In this respect, the excellent uniformity properties of CAM02-UCS have been well documented [24]. As an illustration, Fig. 5(b) shows a three-dimensional color error map for a pixel size $d = 3$.

Further, it can be checked [44] that it is in fact sufficient to keep only one of the metameric samples in each pixel (rather than all the samples) without significantly altering the calculated measures, provided the pixel size d is small enough. The sample can be randomly picked among all samples in the pixel; this random down-selection procedure yields a smaller sample set of a few hundred to a few thousand samples, depending on the pixel size.

In summary, the down-selection procedure yielded a set of real samples that was uniform in color space and could be used to compute color rendition measures. However, this set still lacked spectral uniformity because many of the manmade objects in the Large Set contain only a few dyes—a bias which will be addressed in the following section.

4.2.3 Spectral uniformity

Achieving spectral uniformity required further efforts. The key aspects of the procedure are summarized below; more details can be found in [45].

Spectral uniformity is based on the hypothesis that reflectance features should be equally likely for all wavelengths of light, when averaged over a variety of objects. Its need was proposed in [16,21] which introduced two sample sets displaying this property. However, despite the care taken in generating these samples, it was difficult to ascertain that their behavior was representative of real-world objects; in fact, the resulting fidelity predictions differed significantly from predictions based on other sample sets for some SPDs, and it was unclear whether this was a desired outcome of the spectral uniformity or an artifact of their synthetic nature. In order to address this issue, in the present work we developed a method for selecting a set of *real samples* exhibiting spectral uniformity.

To introduce the approach, we begin with a brief discussion of spectral sensitivity in a simplified framework. Let us consider an equal-energy SPD (constant at all wavelengths) as a reference, and add hypothetical spectral perturbations to this SPD that conserve chromaticity. Such perturbations remove radiation at some wavelengths and add radiation at others; they also induce color appearance shifts (versus those rendered by the equal-energy SPD). Perturbations of any shape can be decomposed on a basis of local perturbations such as those illustrated on Fig. 6—centered, say, on a specific wavelength λ_p.

Spectral uniformity amounts to requiring that the color appearance shifts induced by these local perturbations be equally weighed by the sample set, no matter what the value of λ_p. If this property is enforced for local perturbations, which are a basis of functions, it is also enforced for general perturbations. Thus, this property means that the sample set does not favor perturbations that remove or add radiation at particular wavelengths.

It can be shown that for a test sample (i) of reflectance function r_i, the color error response to the first-order perturbation of Fig. 6(b) scales with $(r_\mathrm{i}')^2$, where r_i' is the derivative of r_i versus wavelength (its local slope). Therefore, the average response of the test sample set is the average of this quantity over all samples, which is denoted $(r')^2$. The criterion for spectral flatness imposes that $(r')^2$ be constant at all wavelengths. This constant is then given by the

average of (r')^2 over wavelengths; therefore a practical way to implement spectral flatness is to numerically minimize the following function:

$$\int \left| (r')^2 - \left\langle (r')^2 \right\rangle \right| d\lambda \tag{1}$$

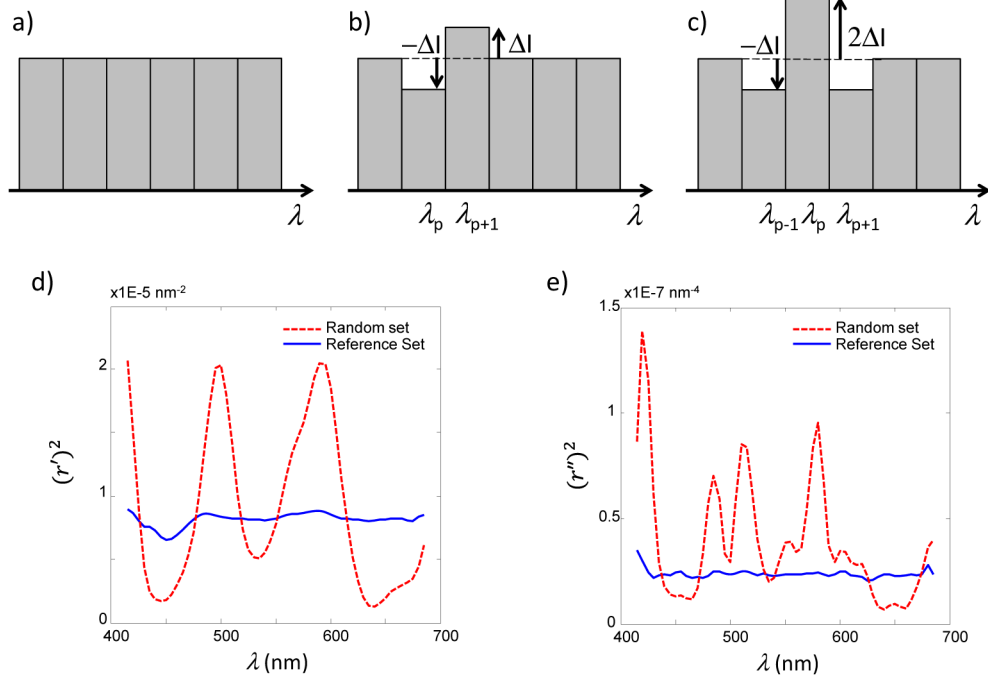where the bracket $\langle \cdot \rangle$ represents averaging over wavelength.

Fig. 6. (a) Equal-energy SPD. (b) and (c) First- and second-order perturbations to the SPD with maintained chromaticity. The corresponding color errors scale with $(r')^2$ and $(r'')^2$. d-e) Flatness figures of merit $(r')^2$ and $(r'')^2$ for two sample sets. Solid lines correspond to the Reference Set (which has undergone the flattening procedure) and are approximately constant. Dashed lines correspond to a random set (where samples were selected without consideration of wavelength uniformity) and show large peaks and valleys.

Likewise, the average response to the second-order perturbation of Fig. 6(c) scales with $(r'')^2$, where r'' is the second-order derivative operator (or curvature), and so on with higher-order derivatives for subsequent higher-order perturbations.

In practice, we implemented these criteria for spectral flatness by including a "flattening" step in the sample selection procedure. As already discussed, the color space was partitioned into small pixels, with one of the Large Set samples selected from each pixel. However, rather than randomly picking the sample in each pixel, we selected them in order to minimize the following *Flatness* figure of merit (F):

$$F = k_1 \int \left| (r')^2 - \left\langle (r')^2 \right\rangle \right| d\lambda + k_2 \int \left| (r'')^2 - \left\langle (r'')^2 \right\rangle \right| d\lambda \tag{2}$$

F contains the two terms described previously. In principle, higher-order terms should also be included for higher-order perturbations; however such terms did not further affect the predictions of the calculations. This definition of F updates and improves the approach of [16]. The scaling constants k_1 and k_2 were introduced to make the two terms of equal magnitude, so that they were equally improved by the flattening procedure.

By performing this flattening procedure, we obtained a *Reference Set* of 4,900 samples, which fulfilled all desired properties and constituted, in our opinion, a "gold standard" for

color rendition calculations. Note that enforcing uniformity both in wavelength space and in color space is important: if only the former was achieved without precaution, the resulting set may favor uncommon colors or over-emphasize some colors; in contrast, the present approach ensures that reasonable colors are considered evenly.

Figures 6(d) and 6(e) illustrate the results of the flattening procedure, showing the quantities $(r')^2$ and $(r'')^2$ for the Reference Set. For comparison, a random set (*i.e.,* a set where one sample was picked randomly in each pixel, without considering spectral uniformity) was also generated. Both sets have the same sample size and cover the color space uniformly. However, the random set shows significant spectral non-uniformity in $(r')^2$ and $(r'')^2$, which stem from the prevalence of samples containing a few types of dyes. This can be seen in the large peaks and valleys for the random set in Figs. 6(d) and 6(e), which indicate that $(r')^2$ and $(r'')^2$ are far from constant across wavelengths. Such peaks and valleys are typical of most sample sets employed for color rendition, and correspond to wavelengths with higher/lower sensitivity. In contrast, uniformity is better by an order of magnitude or more in the flattened Reference Set, with $(r')^2$ and $(r'')^2$ being near-constant across wavelength, thanks to the minimization of F. It should be stressed that this smoothness only holds for the *average* of the set: some individual samples possess sharp variations in reflectance at a given wavelength, and our procedure does not lead to a preferential selection of smooth samples—it is only the cumulative variation that is constant.

4.2.4 Reduction of sample number to a Final Set

At 4,900 samples, the Reference Set is fairly large. While it is tractable by modern computers, it was still desirable for practicality to reduce the number of samples—provided this had no bearing on the accuracy of the calculations—to obtain a so-called *Final Set* for general use.

In principle, a simple way to decrease the size of a set is to merely increase the pixel size in color space and repeat the generation procedure described above. However upon doing this, the predictions of the color rendition calculations were altered by the coarser resolution in color space, and no longer matched those of the Reference Set. Rather, the Final Set was selected more cautiously, so that it was maximally correlated to the Reference Set.

This was achieved by building the Final Set iteratively, as follows. We selected a large enough pixel size and divided the color space. Then, each pixel was considered in a random order. For each pixel, we considered all available samples from the Large Set and selected the one that minimized a correlation figure of merit C. C is the sum of four terms comparing the predictions of the Final Set to that of the Reference Set: fidelity, gamut, spectral flatness, and color icon shape, defined as follows:

$$C = c_1 \cdot E\left(R_f^{ref}, R_f^{fin}\right) + c_2 \cdot E\left(R_g^{ref}, R_g^{fin}\right) + c_3 \int \left| (r')^2 - \left\langle (r')^2 \right\rangle \right| d\lambda + c_4 \oint \left(I^{ref} - I^{fin}\right)^2 d\theta \ (3)$$

Here, E is the Spearman error (*i.e.,* one minus the Spearman correlation coefficient); the superscript '*ref*' correspond to quantities evaluated over the Reference Set; the superscript '*fin*' corresponds to quantities evaluated over the Final Set we are building; the term $(r')^2$ was introduced in Section 3.2.3; I is the radius of the color icon (see Section 4.4) at hue angle θ; and finally, c_1–c_4 are scaling constants to weigh all terms in the minimization of C. The quantities R_f, R_g, $(r')^2$ and I are calculated for the 5,000 SPDs described in Appendix A.

The first and second terms of C ensure that the Reference Set and the Final Set have similar predictions for R_f and R_g. The third term ensures that the Final Set retains spectral uniformity. The fourth term ensures that the color distortion icons (see Section 4.4) obtained from the two sets have the same shape. Coefficients c_1–c_4 were selected to give most weight to the first two terms, as they are the most important (namely: $c_1 = 1$, $c_2 = 1$, $c_3 = 200$, $c_4 = 0.2$)

In summary, by minimizing C, the sample in each pixel that best correlated to the Reference Set was identified. After all pixels had been considered, the procedure was repeated (cycling through each pixel again to further minimize C) several times until no change occurred, thus indicating a local minimum.

Finally, as an additional criterion in generating the smaller set, we imposed that two of the samples be skin reflectance spectra (a fair and a dark skin), which were specifically chosen to be highly predictive for the R_f and R_g values of the thousands of other skin tones in the Large Set. This does not influence the overall predictions, but it enables an explicit estimation of the rendition of skin tones.

By applying this procedure with a pixel size $d = 15$, we obtained a Final Set of 99 color evaluation samples (CES) with excellent correlation to the Reference Set: as show in Appendix A, we found that the two sets agreed within ± 1 point, both for fidelity and gamut predictions. Like the Reference Set, the Final set is uniform in color and wavelength spaces.

We believe that the derivation of the Final Set represents an important achievement, as it meets all requirements for highly predictive calculations while maintaining a small sample size. Figure 7 shows the reflectance and CAM02-UCS coordinates of the CES. It can be seen that for any given wavelength some of the reflectance spectra of Fig. 7(a) display variations, in accordance with the spectral uniformity criterion.
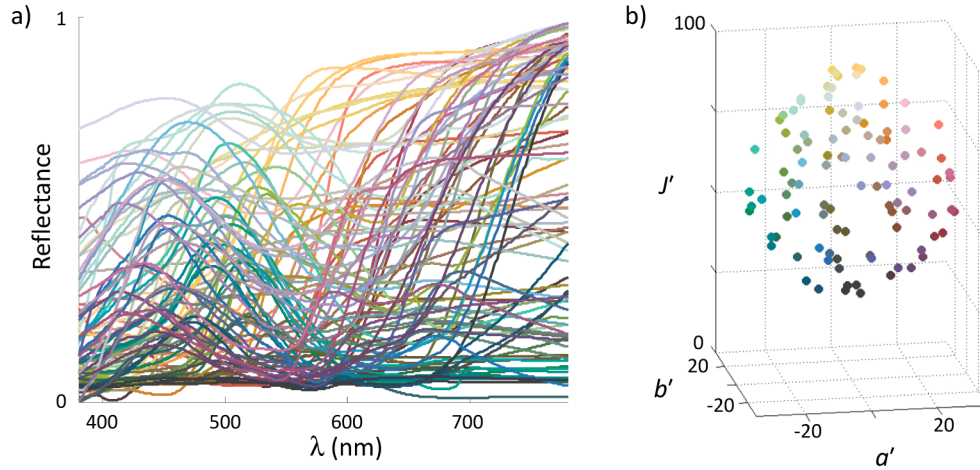

Fig. 7. (a) Reflectance of the 99 CES of the Final Set. Reflectance variations are distributed evenly with no privileged wavelength, in accordance with the spectral uniformity requirement. (b) Corresponding coordinates in CAM02-UCS, when the CES are illuminated by a 5000 K blackbody radiator; the CES are uniformly distributed within the NCS gamut. The colors are indicative of the samples' perceived colors.

4.3 Fidelity measure

The calculation of the color fidelity measure R_f using the 99 CES is straightforward and follows previous work. The workflow of the calculations is shown on Fig. 8.

First, given a test SPD, the CCT is computed in the standard way [46]. Next, the reference illuminant of the same CCT is determined. In the CRI, the reference illuminant jumps from a blackbody radiator to a phase of daylight at a CCT of 5000 K—causing a small but unwanted discontinuity in calculated values. The proposed method addresses this shortcoming by using a linear combination of blackbody and daylight spectra for CCTs in the range 4500–5500 K; specifically, the illuminant is a blackbody at 4500 K, D55 at 5500 K and a linear combination of blackbody and daylight in-between. The impact on fidelity values is very small and the discontinuity is thus avoided.

All the ensuing calculations use 1964 10° color-matching functions. The CAM02-UCS color coordinates of the CES under the reference and test illuminants are calculated, along with the resulting color error for each CES. The arithmetic mean [47] of the errors is determined to obtain an average error ΔE for the SPD and an intermediate fidelity score:

$$R_f{}' = 100 - k \times \Delta E \tag{4}$$

Since this formula can yield $R_f' < 0$ (for very large values of ΔE), the procedure of [20] is used to rescale the score to a range of 0 to 100, and finally obtain:

$$R_f = 10 \times \ln\left(\exp\left(R_f'/10\right)+1\right) \qquad (5)$$

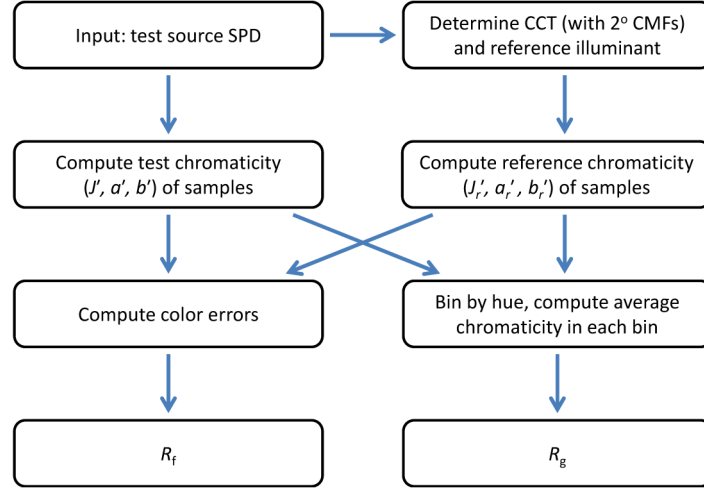The scaling constant k is obtained as discussed in [20], yielding $k = 7.54$.

Fig. 8. Workflow of R_f and R_g calculation. Calculations are in CAM02-UCS and use 10° color matching functions (CMFs) (except for CCT).

It bears mentioning that fidelity values are sometimes misinterpreted as percentage points (for instance, as a percentage of colors which are correctly rendered). Rather, R_f describes an average color error. Thus a same score can be obtained from many small errors or a few large errors—this limitation of a single-valued measure will be addressed by the additional information brought by the color icon, discussed in Section 4.4.

As an illustration of R_f calculations, Fig. 9(a) shows the correlation between R_f and the CRI R_a for the 401 SPDs compiled in [8]. For a value of $R_a = 80$, R_f spans the range 70–87, a sizeable variation. For SPDs with lower fidelity, the width of the distribution widens further.
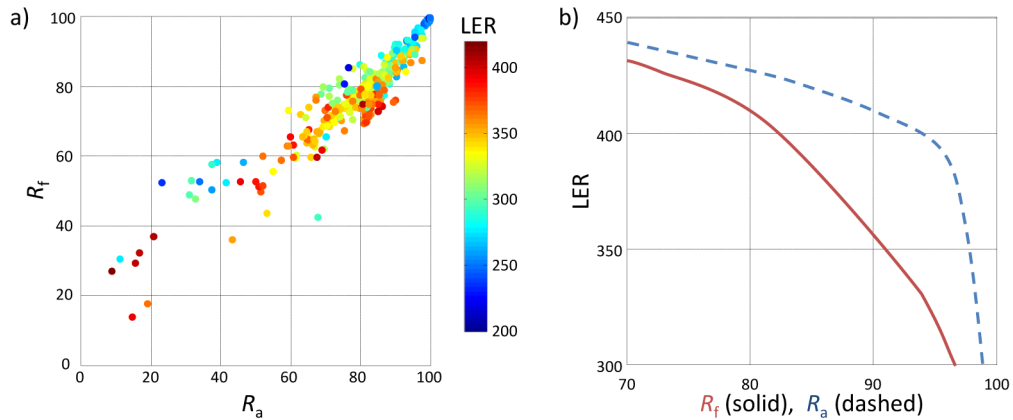

Fig. 9. (a) Correlation between R_a and R_f for the 401 SPDs of [8]. The color of each point indicates its LER. (b) Approximate Pareto boundary for (LER-R_a) (solid line) and (LER-R_f) (dashed line).

In general, the difference between R_a and R_f is moderate for smooth spectra and more significant for discontinuous or spikey spectra—in agreement with the conclusions of [28]. This has implications for the tradeoff between color fidelity and LER. Indeed, SPDs with sharp features offer opportunities for maximizing LER by removing radiation at short and long wavelengths; in some cases, such SPDs can be designed to also optimize the value of R_a, but this optimization is no longer valued by the spectral uniformity enforced in R_f. This is readily apparent on Fig. 9(a), where a cluster of SPDs with $R_a \approx 85$ and high LER (≈ 350–400) obtain $R_f \approx 70$–75; these correspond to triband fluorescent lamps and some narrowband LEDs.

This is further illustrated in Fig. 10, which shows the same data as Fig. 9(a) but focuses on fidelity values above 70, and classifies SPDs by type of light source. Correlation is typically good for phosphor-converted LEDs and broadband fluorescents, which have smooth SPDs. There is more scatter for color-mixed narrowband LEDs. Finally, most narrowband fluorescents have lower R_f than R_a, presumably due to their selective optimization of R_a.
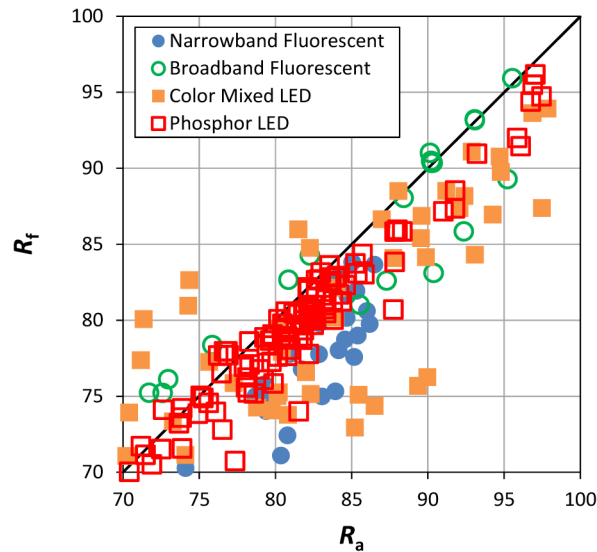

Fig. 10. Detail of Fig. 9(a), showing the correlation between R_f and R_a. The SPDs are classified by type of light source.

We note that agreement between fidelity measures is sometimes analyzed in terms of a correlation coefficient such as R^2. Figures 9(a) and 10 show why such an approach can be misleading: the SPD library comprises a sizeable number of smooth SPDs, for which correlation is indeed good; however, the more important result is the existence of other SPDs —common in practice—for which the scores disagree strongly (sometimes corresponding to a factor of two in average color error). Therefore the spread of the cloud in Fig. 9(a) is a better indicator of the differences in predictions between R_a and R_f. Also note that some computed SPDs based on narrow-band sources (not shown here) display an even larger disagreement.

The tradeoff between fidelity and LER is further illustrated in Fig. 9(b), which shows the approximate Pareto boundaries (*i.e.*, the boundaries maximizing the tradeoff) of (LER-R_a) and (LER-R_f) [48]. For a given LER, the maximum value of R_f is always lower than that of R_a. Again, we attribute this to the fact that the CRI's test samples are less sensitive to radiation at some wavelengths, leading to an artificially high score for some SPDs; in addition, the use of more saturated samples generally lowers the R_f score versus R_a. Most importantly, R_a and R_f lead to somewhat different optimal SPDs for a given LER; therefore, the use of R_f is of direct practical importance when designing optimized lighting.

4.4 Gamut measure

The reader is reminded that our use of the term *gamut* is different from the common use in display applications, and designates an average variation in the chroma of illuminated objects.

Various color gamut measures have been proposed in the past. In general, they consist of computing the area spanned by set of samples in a color space, and normalizing it to a reference area. Interestingly however, there has been limited discussion of the relative merits of these various measures. In part, this is because gamut is not as well-defined as fidelity. However, we believe there are some features to be considered for a good gamut measure:

- •Proper chroma calculation: The change in chroma can be computed in various ways (for instance, averaging chroma directly over the whole space or taking the average across an area or a volume). As argued in [49], area-based calculations are most sensible.

- •Appropriate color space: Probably the best-known gamut measure is GAI, which measures the relative area encompassed by the CRI's TCS in the $(u'v')$ chromaticity diagram; however, $(u'v')$ is not appropriate as it is only a chromaticity diagram (rather than an object color space), where chroma enhancements in some directions (especially blue enhancements) are greatly exaggerated. Computing the area in an object color space is therefore desirable. This was first done in the CQS with the use of CIELAB; the present use of CAM02-UCS further addresses the non-uniformities of CIELAB.

- •CCT sensitivity: In some gamut measures (including GAI), the reference illuminant is at a fixed CCT, whereas the test source has a varying CCT. This tends to favor SPDs with high CCTs, especially if a non-uniform color space is used. However the CCT is a given specification in most lighting applications, and one wants to optimize the gamut under this constraint. Therefore, it makes more sense to use a reference source with the same CCT as the test source, just like is done for fidelity calculations [8].

- •Color samples: The same concerns about color samples expressed in the context of fidelity calculations also apply to gamut calculations. We therefore propose to use the Final Set of 99 CES for this purpose as well.

With these prescriptions in mind, the gamut measure proceeds as follows. Given an SPD, the CCT and reference illuminant are determined as for fidelity. The (J', a', b') color coordinates of the test samples under the reference illuminant and test source are computed and grouped into 16 hue bins of equal width, based on their chromaticity under the reference illuminant. In each bin we compute the average values of a' and b', resulting in 16-point polygons in the (a', b') plane for the reference and test samples. The relative gamut is then:

$$R_g = 100 \times A_{test} / A_{ref} \qquad (6)$$

where A_{test} and A_{ref} are the areas of the test and reference polygons. This procedure is similar to that employed by the CQS Q_g, but averaged over many samples rather than only one sample per hue, thereby improving statistical accuracy.

Figure 11(a) compares the values of R_g and GAI for the 401 SPDs of [8]. A large scatter is observed, which is in part due to the CCT-sensitivity of the GAI calculation. However, Fig. 11(a) shows that even within a narrow CCT range, large variations of GAI (as much as 50 points) can occur for a given value of R_g due to the use of $(u'v')$.

This calculation also allows the definition of a color icon which shows the relative average distortion of hue and chroma for the 16 bins used in the gamut calculation, as illustrated in Fig. 11(b). This icon conveys important information besides the values of R_f and R_g—here, it illustrates the dulling of warm colors which explains the poor subject assessments of standard white LEDs [18,37]. As is clear from Fig. 11(b), color distortion is usually hue-

dependent—an aspect that is lost using a single number like R_g, but revealed by the shape of the color icon [19].
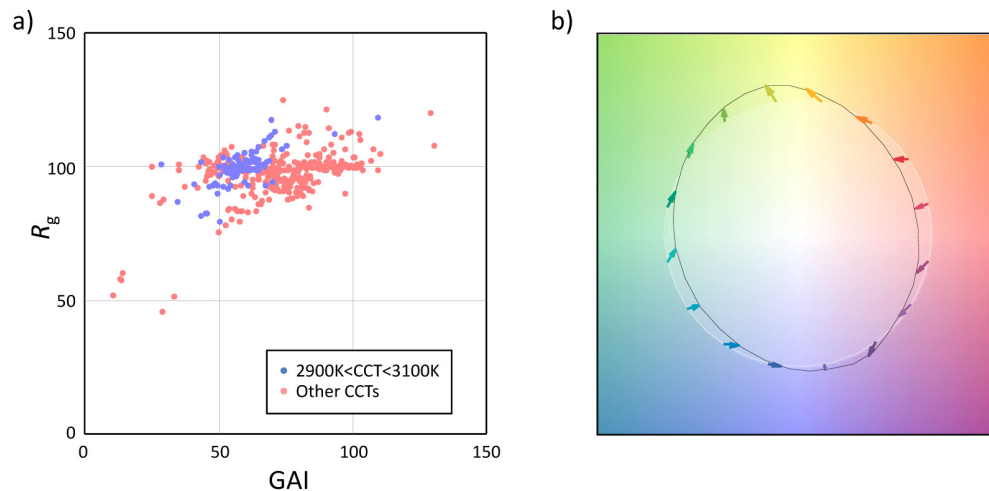

Fig. 11. (a) Correlation between GAI and Rg. Scatter is large, even over a narrow CCT range of 2900–3100 K (blue dots). (b) Color icon. The icon draws the relative change in (a', b') coordinates between the reference illuminant and test source, averaged in each hue bin. The white circle represents reference colors, and the black line and vectors show the color distortion of the test source (here, a typical blue-pumped white LED with Ra = 82).

5. Outlook and future work

We conclude with brief comments on possible future improvements.

- Gamut and preference: R_g describes gamut but makes no prescriptions regarding a recommended value. Future research on color preference should help clarify the tradeoff between color fidelity and color gamut, which could lead to recommended zones in the R_f-R_g diagram for specific applications. Additionally, further research may lead to a measure with better correlation to preference than R_g. For instance, R_g has little sensitivity to pure hue shifts, and it values chroma shifts of different colors equally—both effects stand in contrast to our visual perception.

- The optimal chromaticity of light sources has attracted significant attention recently, with results suggesting that sources far off-Planckian may be desirable [12,50–52]. The present work allows for evaluation of off-Planckian sources but does not favor or penalize any chromaticity. Further research on this topic could lead to such recommendations.

- Fluorescence/whiteness effects: the present framework only deals with non-fluorescent samples; however, some fluorescent objects are commonplace in our environment (especially white objects containing whitening agents) and play a large role in visual perception [37,53]. Additional work is ongoing to define a measure for these effects.

6. Conclusions

We introduce a two-measure system bringing significant progress in quantifying color rendition. Both measures employ a set of Color Evaluation Samples that, for the first time, represent real samples uniformly spanning color space and giving equal importance to all visible wavelengths, thus enabling modern color calculation procedures to yield accurate results. The first measure, R_f, assesses color fidelity and is an improved version of the CIE CRI; the second, R_g, is an improved color gamut measure for assessing the variation in the chroma of illuminated objects. Both measures are based on the new sample set and updated

calculation methods. Thus, they can be used together to provide more useful predictions of the value of light in various lighting situations. In addition, a color icon provides a visual description of color distortions. The improved accuracy underlying the computations will help in designing future light sources that more properly optimize the complex tradeoffs and interactions between efficacy, chromaticity and color rendition. In turn, this should lead to greater value per watt of radiation, greater acceptance of energy saving measures and, ultimately, improved human well-being. This method is adopted in the IES Technical Memorandum TM-30-2015, and is proposed for consideration with the CIE.

Appendix A. SPD library and correlation measure

The color rendition measures were evaluated over a library of 5,000 SPDs, which were obtained as follows. First, we included the 401 real SPDs of [8]. These corresponded to various lighting technologies, including filament, fluorescent, discharge, and LED. They are mostly measured data, with some of the LED spectra coming from simulations. These SPDs were supplemented by generating additional synthetic SPDs. The synthetic SPDs are composed of four Gaussian functions with randomly varying peak wavelengths (in the range 420–650 nm) and widths (in the range σ = 2–30 nm). The amplitudes of the Gaussian functions are balanced so that the SPDs are on the Planckian locus, with a random CCT (in the range 2500—6500 K). Each Gaussian function has an individual width σ, so that the SPDs mix sharp and smooth features, as found in real SPDs.

Next, we needed a method to compare the predictions of various color measures evaluated over this SPD library. Conventional correlation measures (Pearson, Spearman) are not ideal for this, because their values are not easily related to the magnitude of color measures. Instead, any two measures (for instance R_a and R_f) were compared by first computing the values of R_a and R_f over the SPD library, then calculating the histogram of $\delta R = (R_a - R_f)$, and finally computing the width $\varDelta$ of the histogram's 95% confidence interval. The confidence interval indicates the level of agreement of two measures, in units of the measures. Because some of the synthetic SPDs have very sharp features, this test of correlation is quite stringent (indeed, $\varDelta$ is halved if only the 401 real SPDs are considered). Figure 12 illustrates the procedure, showing a comparison of R_f computed over the Reference Set (4,900 color samples) and the Final Set (99 color samples). The result is a value of $\varDelta=\pm1.2$, indicating that the two measures agree within ±1.2 point for 95% of the SPDs and thus validating the reduction in number of samples.
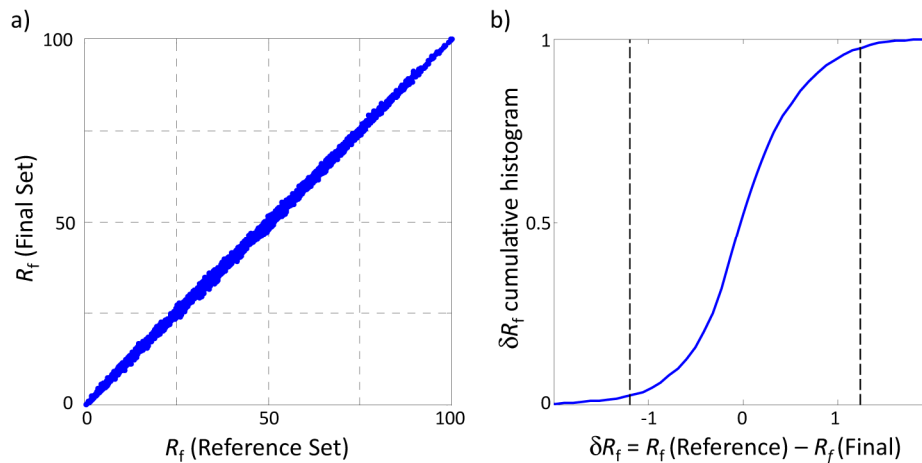

Fig. 12. (a) Correlation between R_f computed over the Reference Set and the Final Set. Each point is an SPD. (b) Cumulative histogram of the difference δR_f between the two measures. Dashed lines: 95% confidence interval.

Appendix B. Reflectance samples database

Our initial reflectance set (the Large Set) contained about 105,000 measured reflectance samples. The largest contribution came from the University of Leeds database [54], which is itself a meta-base with various origins: textiles, plastics, skin tones, and color systems. The Leeds database also includes the SOCS database [55], which contains printed materials, skin tones, natural objects, paints, and textiles. The Leeds samples were complemented with additional data: natural objects [56,57], flowers [58], and paints [56,59].

Some of the samples only have data in the range 400–700 nm, whereas SPDs are often specified in the range 380–780 nm. Therefore, for samples with missing data, we extrapolated the reflectance in missing ranges using the following procedure:

- Map reflectance value r from [0..1] to real numbers by defining $r' = \ln(r/(1-r))$

- Perform a linear extrapolation of r' at short- and long- wavelength to obtain r'_{ex}

- Map back from real numbers to [0..1] to obtain $r_{ex} = \exp(r'_{ex})/(1 + \exp(r'_{ex}))$

The mapping to real numbers avoids reflectance values below 0 or above 1. We checked that this procedure had minimal impact on the results of R_f and R_g (less than 0.1 point on average), and that it produced reasonable results by testing it on samples with known data at short and long wavelengths.