C. R. Physique 19 (2018) 169–181

Contents lists available at ScienceDirect

Comptes Rendus Physique

www.sciencedirect.com

LEDs: the new revolution in lighting / Les LED : la nouvelle révolution de l'éclairage

LED-based white light

Les sources de lumière blanche à base de LED

Aurélien David [a,*], Lorne A. Whitehead [b]

[a] *Soraa Inc., 6500 Kaiser Dr., Fremont, CA 94555, USA*
[b] *Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada*

A R T I C L E   I N F O

Article history:
Available online 22 March 2018

Keywords:
Light emitting diode
Solid-state lighting
Lighting
Color science
Color rendering

Mots-clés :
Diode electro-luminescente
Eclairage a l'etat solide
Eclairage
Colorimetrie
Rendu des couleurs

A B S T R A C T

This article discusses the use of light-emitting diodes to generate white light – a research forefront in Physics and Ergonomics. We first present various technological approaches to white-light generation. After a general introduction to the human vision system, we discuss two key aspects of the quality of white light: the color of the light itself, and the color rendering of illuminated objects. We present the tools underlying modern color science, and review key color rendering metrics, from the well-known color rendering index to the latest improvements in the field.

R É S U M É

Cet article présente l'utilisation des diodes électroluminescentes comme sources de lumière blanche – un sujet à la pointe de la recherche en physique et en ergonomie. Nous présentons d'abord diverses approches technologiques à la génération de lumière blanche. Après une introduction générale au système visuel humain, nous discutons deux aspects cruciaux de la qualité de la lumière : la couleur intrinsèque de celle-ci et le rendu des couleurs des objets illuminés. Nous présentons les outils sous-jacents à la colorimétrie moderne, et examinons les principales métriques du rendu des couleurs : le CRI, ainsi que les derniers progrès dans le domaine.

1. Introduction

A significant portion of the world's energy consumption is used to generate white light to enable vision when daylight is not available. The first artificial illumination, firelight, predates history, and electric light has been common for over 100 years, so one might expect that all the relevant technical and physiological factors involved would by now be completely

* Corresponding author.
E-mail address: aurelien.david@polytechnique.org (A. David).

https://doi.org/10.1016/j.crhy.2018.02.004

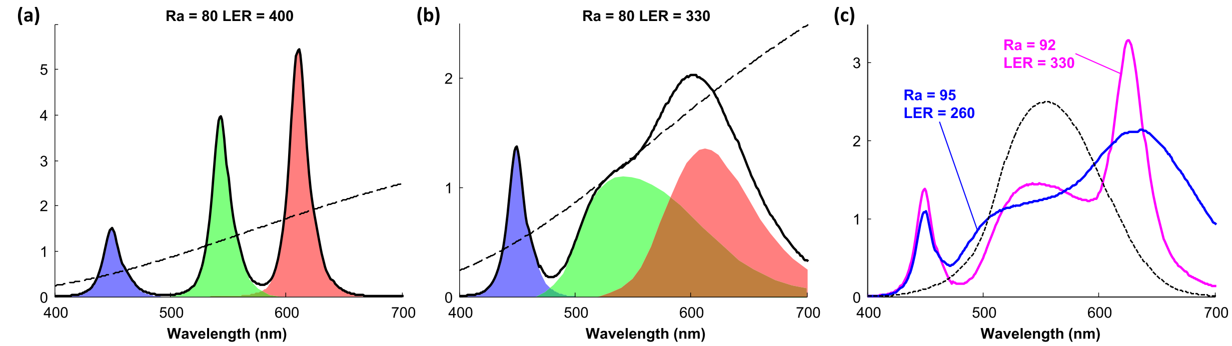

Fig. 1. SPDs of white LEDs, each with a correlated color temperature (CCT) of 3000 K and producing the same amount of lumens (the vertical axis is in arbitrary units). (a) Combination of blue, green and red LEDs: the black line is the composite LED SPD. The dashed line is the blackbody radiator of same CCT. (b) Same as (a), but for a blue-pump LED exciting green and red phosphors. (c) Dotted black line: photopic sensitivity curve $V(\lambda)$. Blue and magenta lines: two high-CRI LED SPDs (blue: using a standard red phosphor; red: using a narrow red quantum-dot). The two SPDs have similar R_a, but the latter has a much higher LER, since it emits little far-red radiation.

understood. Yet, that is not fully the case: researchers are not yet sure what light sources best meets the needs of users in various situations, and the design freedom offered by light-emitting diodes (LEDs) exacerbates this shortfall.

This article presents the status and recent developments of LED white light sources. We first discuss the technologies enabling LED-based generation of white light. We then introduce basic concepts in color science and discuss the challenges of producing white light that has high quality from an ergonomic perspective. We conclude with a discussion of needed future research on illumination.

2. How do we make white light?

LED chips emit radiation within a narrow wavelength range (10–25 nm) with energies close to that of the semiconductor bandgap. This contrasts with conventional light sources such as daylight and filament lights, which are approximately blackbody radiators with broadband emission. Therefore, LED spectra must be combined or modified to obtain white light. Two main technological approaches can be employed to this effect: combining multiple LED emitters, and down-conversion by phosphors.

In the following, we use the term Spectral Power Distribution (SPD) to describe a light source's distribution of intensity as a function of wavelength.

2.1. Combining LEDs

In this approach, several narrow-band LEDs have their light outputs blended to achieve white light. A common approach is to combine three kinds of LEDs, emitting in the blue, green and red ranges, as illustrated in Fig. 1(a). The relative power of these LEDs is selected to yield a desired chromaticity (see Section 3.2) for the blended light. An advantage of this approach is that it avoids creating light at unwanted wavelengths, since each emitter is narrow spectrally. Another is that the power for each LED color can be varied to tune the resultant chromaticity.

However, this approach creates several practical challenges. First and most importantly, materials are currently not available to make highly-efficient LEDs at all wavelengths. The III-Nitride material system excels in the blue range, but its efficiency is lower for green emission and very poor in the yellow range (in fact, the best green emitters today use a blue LED to pump a green phosphor). The "legacy" AlInGaAsP material system is also quite limited, primarily to the orange-red range. Therefore, some LEDs in a mixed system suffer from low-efficiency.

Second, this approach requires a complex system architecture: the different LED colors require separate driver channels; they also have different temperature-dependencies and aging characteristics, which must be compensated for to maintain the desired chromaticity. This complexity increases system cost. As a result, at present the combined-LED approach is mainly used in niche applications where active chromaticity control is required.

2.2. Phosphor down-conversion

In this approach, light is emitted at short wavelengths by a "pump" LED. Some of this pump light is absorbed by phosphor particles surrounding the LED, and down-converted to longer wavelengths. Most commonly, the pump LED is blue and is combined with a green/yellow and a red phosphor to form white light, as shown in Fig. 1(b). The concentration of these phosphors controls how much pump light is converted and how much is transmitted, in order to obtain a desired SPD.

From an efficiency standpoint, this approach is preferred today because good phosphors have high quantum yields (the ratio of emitted photons to absorbed photons), surpassing 90%. This is also a relatively simple technology, whose main manufacturing challenge is the precise control of the dispensed phosphor – which directly impacts the consistency of the chromaticity.

Nonetheless, the phosphor-conversion approach has two significant drawbacks – one fundamental and one practical. Fundamentally, even if a phosphor has a high quantum yield, energy is lost because photons are converted to a longer wavelength. The energy loss factor, known as Stokes loss, is substantial when converting blue photons to red photons. In practice, for a warm-white LED, about 25% of the energy is lost in this process. Furthermore, this lost energy is converted to heat in the phosphors, which are difficult to heatsink, with deleterious effects on reliability and performance.

A practical challenge is that many phosphors emit in a rather wide spectral range (with typical widths of 80–120 nm). Since suitable red phosphors ideally have a peak emission of about 620 nm, a significant portion of their emission is at longer wavelengths for which the human eye has very low sensitivity. This is wasteful, much like the infrared emission of traditional filament lamps.

This effect is quantified by the luminous efficacy η of a source, which is the ratio of emitted luminous flux to consumed electrical power. It is expressed in lumens per watts and given by: $\eta = WPE \times LER$. Here *WPE* is the source's wall-plug efficiency, a dimensionless measure of how many watts of radiation are emitted per watt of electrical drive power, and LER is the luminous efficacy of radiation, calculated by integrating the SPD weighted by the response curve of the human eye $V(\lambda)$, which represents the wavelength-sensitivity of our vision system at normal illumination levels. Fig. 1(c) shows $V(\lambda)$ and, for comparison, the SPDs of two phosphor-converted LEDs, using respectively a broad and a narrow red emitter. The two have similar color rendering properties, but the latter has a much higher LER, because it emits little long-wavelength radiation.

2.3. Technology outlook

Ongoing developments in light-emitter technologies may solve some of the issues described above.

On the front of direct LED systems, there is renewed emphasis on efficient longer-wavelength emitters – prompted by their anticipated usefulness in direct-view micro-LED displays and also because multi-primary systems have greater ultimate efficiency potential because they avoid Stokes loss. III-Nitrides remain a promising material for green emitters, despite important challenges in materials science. For red LEDs, various materials are being investigated, including improvements in the existing AlInGaAsP system. It is not yet clear how efficient yellow/orange emitters will be achieved; however, this wavelength range is less crucial for lighting.

Regarding phosphors, the quest for red converters with a narrow linewidth has been a major focus of research for more than a decade, owing to the previously-mentioned desire to minimize radiation at long wavelengths. There are several material systems under consideration, each with distinct advantages and challenges. Some are now coming into use in specific lighting applications, where they yield efficiency gains of 10–30% by virtue of the improved LER [1].

Yet another venue for research is laser-based lighting, where laser diodes are used as light sources [2]. Some theoretical arguments suggest laser diodes might surpass LEDs in efficacy. For example, a blue laser diode could efficiently pump a phosphor system. In the longer term, one may even envision a laser-based multi-primary emitter systems. However, significant progress would be required to make this approach sufficiently efficient and inexpensive for use general lighting.

Finally, a different field of research aims at developing light sources whose SPDs may provide additional benefits beyond enabling good vision. An important example, sometimes called "human-centric lighting", aims at influencing physiological functions by means of light – for instance by adding blue light to an SPD to stimulate the human circadian cycle [3], or conversely by minimizing this radiation to avoid circadian stimulation [4].

3. How do we make *good* white light?

3.1. Key aspects of human color vision

By definition, white light is perceived to be substantially free of color. White light is desirable for illumination because human color vision evolved to provide information about the color of surfaces illuminated by daylight, and an important purpose of artificial lighting is to enable human color vision to operate normally when daylight is not available. Interestingly, this simple-sounding requirement is not easily met. To understand why, it is helpful first to summarize the key relevant aspects of color vision.

We begin with the seemingly simple question: What *is* color? Interestingly, children often answer this question more easily and correctly than many scientists. The International Commission on Illumination (CIE) defines color in a fashion that sounds almost child-like: color is the perceptual experience that differentiates between different colors. Even though this definition seems vague, circular, and perhaps philosophical, it lies at the heart of a quantitative and accurate aspect of color science known as color matching:

Consider two adjacent flat surfaces, each facing and emitting light toward a viewer. Naturally, if the SPDs for the two surfaces are the same, their color appearances will match, but the converse is not always true: if the SPDs do *not* match, the perceived surface colors *may or may not match*. Importantly, most people with normal color vision will agree on whether or not the two colors match.

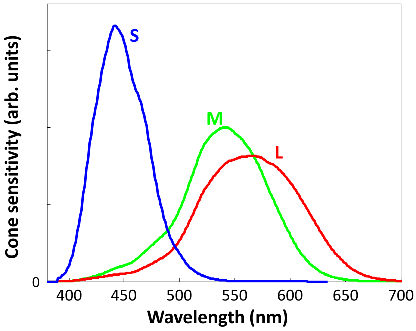

Fig. 2. Spectral sensitivity of the short (S), medium (M) and long (L) wavelength retinal cone cells (curves normalized to an equal area).

This arises because the retinal cells in the eye do not resolve the SPD of the reflected light. Rather, color information stems from the responses of three different kinds of retinal cone cells, each having a different spectral sensitivity, as shown in Fig. 2. Thus, the complex information carried by an SPD (in principle, of infinite dimensionality) is encoded in just three stimulus values [5].

These retinal response functions enable predictions of color matching: two SPDs "match" if they produce the same stimulation of each of the three short (S), medium (M) and long (L) wavelength (SML) retinal cone types. This can occur even if the SPDs themselves are different – a fact that lies at the heart of most modern color reproduction technology [6].

While the response functions shown in Fig. 2 are well-known today, they are actually difficult to determine directly from color-matching experiments [7]. That is true because if the S, M, and L cones stimulations match for two SPDs, then all linear combinations of those functions will also match. For this reason, if we perform a change of basis by multiplying the vector (SML) by a 3×3 matrix, to obtain functions ($S'L'M'$), these new functions will identify color matches equally well. For historical reasons that are now irrelevant, the pioneering work of color matching researchers early in the 20th century led to the selection of one particular basis for this purpose, known as the (XYZ) functions. These have no direct perceptual meaning, but their use is still commonplace to this day for legacy reasons [8].

In addition to identifying color matches, it has been found that people with normal color vision largely agree on observations of the *amount* of color mismatch. That is, if one observer perceives two surfaces to be nearly the same color (for example shades of green) most others will agree. Similarly, if two surfaces are seen to have very different color (for example red and blue), most others will also agree. An interesting question is whether it is possible to create a diagram depicting all surface colors, within which similar colors are placed close together and different colors are spaced far apart. This is possible in three-dimensional diagrams. Various such diagrams have been introduced; they are called color spaces. In essence, a color space is a mathematical model that assigns three-dimensional coordinates to any given color. (Perhaps not surprisingly, the need for three dimensions arises directly from the fact that the human retina cone cells have three different spectral sensitivity patterns.)

Numerous color spaces exist for various applications. They can be separated in two main families: source color spaces, which describe the color of light coming from a source, and object color spaces, which describe the color appearance of objects under illumination by a light source [9].

3.2. Chromaticity: the color of light

A first objective of color science is the quantification of the color of the light source itself, known as its chromaticity. Chromaticity can be experienced by looking at a source alone or its reflection off of a white surface, and it can be judged much more accurately through relative side-by-side comparison with other light sources. Depending on the relative amounts of short, medium and long-wavelength radiation in the SPD, its color may appear as having a pronounced tint or as substantially white. White light sources are of course of primary interest for lighting applications. In particular, when replacing an incumbent lighting technology with an upcoming one (for instance, incandescent with LEDs), an important challenge is to ensure that the two sources have the same perceived chromaticity.

Chromaticity is expressed in a source color space. Such spaces generally have a simple mathematical structure. They require one input: the SPD of the light source. Their output is a 3-dimensional quantity describing the chromaticity of the light. A well-known example is the CIE XYZ color space, whose coordinates are simply the values of the XYZ functions discussed above. For display purposes, this space is customarily projected into a 2-dimensional diagram, the so-called *xy* diagram – with the third dimension, related to the absolute intensity of the light, not shown for simplicity. The boundaries of the space correspond to the most saturated possible colors of light (the so-called optimal colors, corresponding to monochromatic radiation at a given wavelength). Just like the XYZ functions, the axes of the *xy* diagram have no direct intuitive interpretation; rather, the diagram is best thought of as an abstract space that is nevertheless useful for comparing and classifying chromaticity values [10].

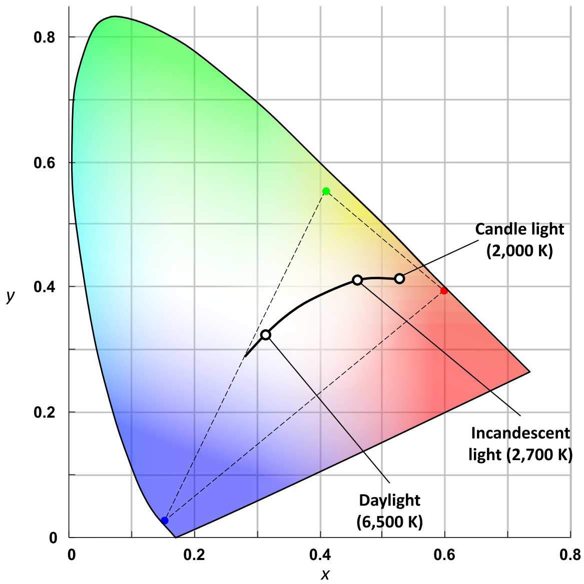

Fig. 3. The *xy* chromaticity diagram. The color shades are approximately indicative of the perceived color of light. The solid black line is the blackbody locus, with a few color temperatures pointed out. The locus of typical daylight at various times of the day spans a wide range of temperatures (5,000 K to more than 10,000 K); it is undistinguishable from the blackbody locus on the scale of this figure. The three colored dots correspond to the blue LED and the green and red phosphor of Fig. 1(b); the dashed triangle shows what chromaticities can be achieved by combining these emitters with different weights.

Fig. 3 presents the CIE *xy* diagram. White light resides near its center. Of particular interest are two curves: the locus of blackbody radiators at various temperatures, and the locus of typical daylight at various times of the day. These curves constitute guidelines for "natural" light sources. It is noteworthy that they span a wide distance across color space – in accordance with the fact that we encounter a large variation of light sources throughout a day.

Crucially, our vision system adapts to such changes: it is capable of perceiving this wide span of sources as being white, and to correct for the underlying variations in chromaticity through adaptive changes that occur at multiple levels within the human vision system [11]. This phenomenon, known as *chromatic adaptation*, is a critically-important property of our vision system. It makes color vision practical by ensuring that objects are perceived to have a stable color despite variations in the source's SPD, for instance at various times of the day or if a cloud temporarily obscures the sun. (Importantly, as will be discussed shorty, the accuracy of color constancy is sometimes reduced when light sources have highly-structured spectra.)

By design, the *XYZ* space is linear. Therefore, by combining three sources like a pump LED and two phosphors and varying their relative amplitudes, the blended light can have a chromaticity lying anywhere within the triangle defined by the three emitters' chromaticities (the so-called *color gamut* of the three-emitter set), as depicted in Fig. 3.

Most often, manufacturers seek to produce chromaticity values that lie on the locus of the blackbody radiator. In this case, the source can be said to have a *color temperature* equal to that of a blackbody radiator at that physical temperature. This idea can also be usefully extended to sources whose chromaticity lies somewhat off the blackbody curve. The so-called *correlated color temperature* (CCT) of a light source is thus defined as the temperature of the blackbody of closest chromaticity (this is calculated in the *uv* diagram, which is a derivative of the *xy* diagram) [12]. The CCT is an important parameter, which corresponds roughly to how "warm" a light source feels. Common residential light sources like filament bulbs are warm, with CCTs in the range 2700–3000 K. Cooler sources, such as many fluorescent tubes, have a CCT in the range 4000–6000 K and are more reminiscent of daylight. (Notice a sometimes-confusing convention: "warmer" color sources, having lower CCTs, correspond to *cooler* blackbody radiators.)

In practice, the achieved chromaticity can differ from the intended one, due to variations in LED-manufacturing processes. For instance, variations of a few % in the thickness a phosphor layer can have a noticeable impact on the chromaticity, by altering how much pump light is converted. Such variations in part-to-part chromaticity can be represented as a cloud of points in the *xy* diagram; the size and shape of this distribution is indicative of the chromaticity range for a given LED product.

Importantly, the *xy* diagram is perceptually non-uniform: chromaticity differences having similar perceived magnitudes may be separated by different distances in different parts of the *xy* diagram. This is akin to cartographic representations of the earth, where geographic distances are distorted across the map. To remedy this, more uniform color spaces have been proposed, such as the $u'v'$ diagram [13]; these offer improved, although still-imperfect, uniformity. Alternatively, the local color sensitivity across the diagram can be depicted as a series of ellipses whose dimensions indicate how a color difference

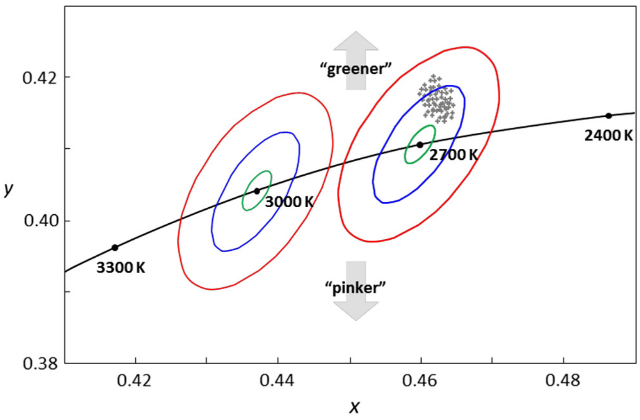

Fig. 4. *xy* chromaticity diagram around the blackbody locus (black line). Green/Blue/Red lines: 1-step/3-step/6-step MacAdam ellipses around 2700 K and 3000 K. Sources departing from the blackbody locus are perceived as relatively greener/pinker when the shift is above/below the locus, respectively. The gray crosses depict a hypothetical distribution of LEDs targeted near 2700 K: the cloud is smaller than a 3-step ellipse but is centered somewhat off the blackbody, so that only part of the distribution fits within the 3-step ellipse centered at 2700 K.

of constant perceived magnitude appears on the diagram (this is akin to the Tissot indicatrix on a map: a set of ellipses indicating the local distortion induced by the cartographic projection [14]).

Experimentally, these ellipses were first developed by MacAdam in the 1940s [15]. The resulting data suffers from limitations: it was only acquired from only one subject, in a specific viewing condition (small patches of light) and only at a few points across the color space. Nevertheless, it remains the best data available today on this topic, and it forms the basis for specifying chromaticity variations in modern light sources.

From these original MacAdam ellipses, ellipses for any location in the diagram can be determined by interpolation. These so-called one-step ellipses indicate chromaticity values which are within a just-noticeable difference (i.e. a chromaticity difference that is at the threshold of being perceptible) from the ellipse center. From that information, ellipses corresponding to larger differences can be derived. For instance, a 3-step ellipse has three times the size (and nine times the area) of the standard MacAdam ellipse. Many manufacturers today produce LEDs which conform to such a 3-step spread – meaning that the parts they sell fit within a 3-step ellipse. Part-to-part variations can still be noticed in such a case, and there is therefore a continued push for tighter distributions. Fig. 4 illustrates MacAdam ellipses at a few points on the blackbody locus.

As discussed above, most artificial light sources seek to reproduce the chromaticity of natural light, and are close to the blackbody locus. However, recent research suggests that, in some situations, sources can be perceived as white if their chromaticity lies "below" the blackbody locus [16]. The interpretation of these experiments is a subject of debate; however, there is general agreement that further research on the "white locus" is warranted.

3.3. Color rendering: the color of illuminated objects

3.3.1. The concept of color appearance

Although our modern world contains many deliberately colored light sources (such as traffic signals and theatrical lighting), it is interesting to consider that such stimuli were not present during the period over which color vision evolved. This raises the question of why color vision evolved: What was its evolutionary advantage?

The general consensus is that color vision enables a determination of the nature of the materials comprised in surfaces in our environment [17]. A familiar example is the color of a banana: as bananas ripen, they start out quite green, at which time they are neither soft nor sweet; as they turn yellow, they are becoming soft, sweet, and good to eat. The color change corresponds to changes in the underlying chemical constituents of the surface in question. Thus, color vision provides a way to safely evaluate surface chemistry without having to ingest a possibly hazardous substance. Clearly, this could be evolutionarily advantageous.

Indeed, there is a strong correlation between the shape of a graph of surface reflectance vs. wavelength, and the perceived color of a surface [18]. The reflectance of green surfaces peaks in the middle wavelengths and for purple surfaces there is instead a dip there; orange surfaces have reflectance values that increase with wavelength while for blue surfaces there is a decrease. Varying combinations of these patterns correspond to the many other distinguishable variations of surface color.

A number of mathematical models have been developed to predict the color coordinates for any given surface reflectance in a specified viewing condition. For example, the latest CIE model CIECAM02 has been widely accepted for providing engineering guidance for the design of color reproduction systems [19]. Color appearance has also been modeled in a mathematically simpler, physiologically-based treatment [11].

In summary, the human vision system does a wonderful job of detecting the "true color" of a surface, which can be thought of as the color observed under moderately bright daylight, providing it is neither too orange (as near dawn or dusk),

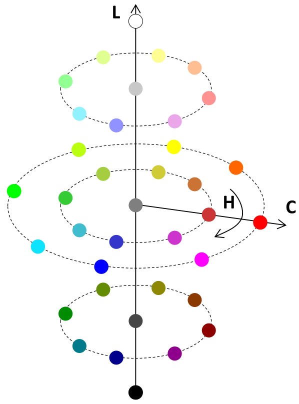

Fig. 5. Illustration of a generic object color space. The cylindrical coordinates are lightness (L), chroma (C) and hue (H). The dotted circles indicate colors of constant chroma and lightness, but varying hue. The space is bounded by the colors white at top and black at the bottom (maximum and minimum lightness, respectively).

nor too bluish (as in north sky illumination). The vision system is highly adaptive and responsive to changing conditions, so that accurate color vision works well over at least three orders of magnitude of illuminance and under widely varying color shifts in the ambient light.

With this background, we can now state a reasonable goal for white light sources: they should cause surfaces to appear essentially the same color as they do under daylight – a concept called color fidelity (see Section 3.3.2).

In that regard, it might seem plausible that as long as a light source appears white, then color surface should appear normal when illuminated by that source. Surprisingly, if anything, the opposite is the case. Sometimes surfaces can appear quite "natural" in color even if the light source has a clearly apparent tint. Conversely, some light sources that appear perfectly white can severely distort the color appearance of some objects.

This is because the human retina only has three types of color receptors. This means that people are effectively blind when it comes to assessing the detailed structure of the SPD of light. A light source may be seen as white, but it may nevertheless have a highly non-uniform SPD. Similarly, a surface may appear gray or only weakly colored, yet its reflectance may vary significantly over small wavelength regions. This issue is likely to be most pronounced when a highly structured SPD illuminates an object with a highly structured reflectance function: there can be an interference effect that causes significant color shift – in other words, the object may have an apparent color that is different from its color under daylight [20]. This effect can be problematic: if the purpose of color vision is to assess the nature of surface chemistry, and our database for such comparisons is our memory of the colors of objects under daylight, then these color shifts are distortions that should be minimized. Therefore, beginning in the 1950s, considerable effort has been devoted to understanding how to best quantify this color shift effect.

3.3.2. Quantitative assessment of color rendering

In the scientific literature, the term *color rendering* is defined very generally 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". We first review the technical tools used in color rendering calculations, before discussing some important color rendering metrics.

a) *The basic tools*

As previously mentioned, the perceived color of a surface under illumination can be quantified in an object color space. Such a space requires at least two inputs: the SPD of the white light source and the SPD of the light reflected from the object. The output is the 3-dimensional color coordinates of the object in the selected color space, which encodes the surface color that will be perceived when the object is illuminated by this light source.

Often, the dimensions of the object color space are related to perception. These are most easily discussed in polar coordinates. Typically, the vertical direction indicates lightness, a quantity describing whether the surface appears relatively light or dark, with zero indicating black, and the maximum lightness corresponding to white. The radial direction indicates chroma (also sometimes called saturation or colorfulness), with shades of gray positioned along the central vertical axis, and points becoming increasingly more colorful as distance from that axis increases. The polar angle corresponds to hue, which describes where the color lies on a circular continuum spanning the repeating pattern red/yellow/green/blue/purple. Fig. 5 illustrates these coordinates.

Different points in color space correspond to different colors. A desirable property of color spaces is color uniformity – whereby distances in color space are approximately proportional to perceived color differences. Modern color spaces seek to be as color-uniform as possible. It should be appreciated that object color spaces are empirical mathematical constructs, whose parameters are often derived by fitting data from experiments. In particular, experiments on color difference perception are used to determine the local metric of the space and make it approximately uniform. In a uniform color space, the color difference can conveniently be calculated as the Cartesian distance between points, giving the space an intuitive visual interpretation.

Color rendering calculations generally compare the colors of objects illuminated by a test source of interest with the colors under a well-known light source, called the *reference illuminant*. A frequent choice for the reference illuminant is the blackbody radiator having the same CCT as the test source (at high CCTs, daylight is sometimes used instead of a blackbody – a very minor change). This is justified not only by the fact that human vision evolved under such “natural” light sources (from daylight to flames), but also because those illuminants have smooth SPDs, under which the response of our vision system is continuous and well-behaved.

Given a reference illuminant and a test source, the color coordinates of an object can be computed under both illuminations and compared: a difference in coordinates, called a *color shift*, indicates a difference in perceived color.

b) *Color rendering metrics*

By using these technical tools, color rendering calculations seeks to answer a number of questions related to the perception of color. To list a few:

* Color fidelity quantifies objectively the degree to which a test source renders colors in the same way as the reference source. In practice, this is done by evaluating an average of color shift between the two sources. Fidelity is a key concept of color rendering.
* Gamut and chroma shifts: these various objective metrics quantify variations in chroma (indicating how much a source saturates or de-saturates colors), either as an average over all color or for specific colors.
* Discrimination: this involves examining whether a light source affects the degree to which different surfaces appear as having different colors. It is a subjective concept, often determined experimentally by having subjects sort or order objects of various colors. It is important, because the perception of color differences is sometimes necessary in tasks such as sorting produce.

In addition to the above metrics which are based only on color spaces, there is a more recent area of psychological investigation concerning how people feel about the overall pattern of color shifts caused by a source. Ideally, such investigations would occur in long-term experiments in commonplace settings. However, for reasons of practicality, so far most experiments have been short-term and involved comparing one light source to another either sequentially or side-by-side. This comparative method may be unrealistic in time frame, because it introduces a different effect from the one of interest – the powerful adaptive aspect of human vision. Despite these known limitations, this area of work has been popular because it has demonstrated a number of interesting and reproducible findings, related to the following metrics:

* *Naturalness* describes whether the subject feels that objects appear as expected. This is a subjective metric based on the user’s impression, and perceived naturalness tends to differ from objective fidelity. A related concept is “*color memory*”, which asks users to remember how they think objects usually appear.
* *Preference* describes the degree to which the subject feels that the appearance of objects is pleasing.

Although these metrics are more subjective, several strong and reproducible regularities have been found in these experiments, some of which correlate well with patterns in objective metrics, such as chroma changes for certain hues [21].

Many proposals for specific color rendering metrics have been introduced over the years. In the following, we discuss two important examples: the CRI (the dominant metric in the lighting industry), and the recently-introduced TM-30.

c) *The CRI*

The Color Rendering Index method [22] generates the average index value R_a, which is a color fidelity metric. It is unfortunate that the general term “color rendering” was used ambiguously, since the CRI only measures fidelity. The calculation of R_a is illustrated in Fig. 6, and proceeds as follows:

* the CCT of the test source is determined, and a reference illuminant of equal CCT is used as a comparison basis,
* the colors of eight pastel color patches – the so-called test color samples (TCS) – are calculated under the two sources,
* the color difference of each TCS between test and reference illuminations is evaluated, and the average color difference is scaled and subtracted from 100.

In summary, R_a quantifies how the colors of the eight TCS differ between the test source and the reference illuminant. A score of 100 indicates that all the sample colors match perfectly; a lower score indicates some color differences – but doesn’t convey which samples are affected or the nature of the color shifts they undergo. Fig. 7 illustrates possible appearances of a scene under low- and high-fidelity illuminations.

Despite its systematic use for decades, there are some limitations to the CRI.

The first is fundamental, and simply stems from the definition of fidelity: R_a only aims at describing whether a source causes color differences with respect to the reference illuminant. In spite of this, there has been a widespread mistaken belief that the value of R_a necessarily indicates a “more pleasing” light (a subjective and context-dependent concept). It should be noted that there is indeed often some correlation between the two concepts – in particular, most existing low-CRI sources also diminish color saturation, which is generally disliked, as portrayed in Figs. 7(a–b). Nonetheless, this correlation is not fundamental; Fig. 7(c) shows that high CRI and vivid colors are not systematically tied. Overall, color fidelity alone is unable to describe important and relevant aspects of color rendering [23].

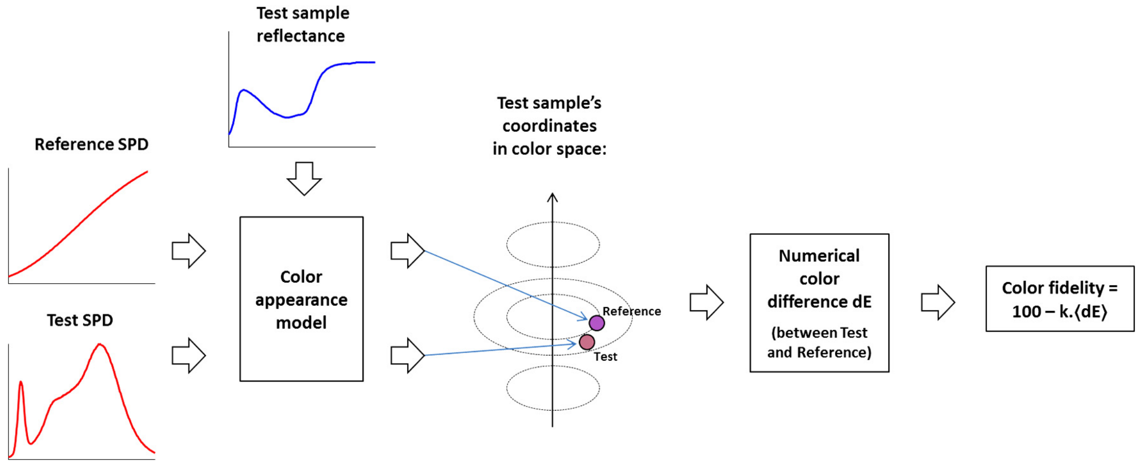

Fig. 6. Simplified calculation flow of the CRI R_a (and other fidelity metrics). For each test sample, the color coordinates are computed under the reference illuminant and test source. The Cartesian distance between the two coordinates is the color difference dE. This color difference is averaged over several test samples, and the average (scaled by a constant k) is subtracted from 100.

Fig. 7. Illustration of a retail store illuminated by three hypothetical light sources: (a) low-fidelity source which de-saturates most colors; (b) high-fidelity source; (c) low-fidelity source which saturates most colors. In practice, sources (a) and (c) might have a similar fidelity score but correspond to very different perceptual experiences. Notice that the high-fidelity source (e.g. $R_a \sim 100$) does not correspond to the most vivid colors.

There are also serious limitations of the CRI calculation in its role as a measure of color fidelity: it sometimes yields inaccurate predictions. In part this is because it was based on what is now decades-old color science. Since then, significant progress in color science has yielded much more accurate tools (including better models of color spaces, color differences, and chromatic adaptation). Furthermore, the eight TCS are an inadequate set of color samples, as will be discussed below. These flaws mean that the CRI cannot always be trusted as a good measure of fidelity [24].

d) TM-30

For these reasons, there has been a decades-long effort to replace the CRI with better metrics. The most recent alternative metric, which was recently developed by the Illuminating Engineering Society (IES), is the IES TM-30 method [25]. It has been gaining traction in the lighting world and has recently also been adopted as a calculation method by the International Lighting Commission (CIE) [26]. It addresses the major limitations of the CRI mentioned above, by improving the technical underpinning and by providing more output quantities in addition to a fidelity index.

On the technical front, TM-30 uses modern color science tools (including an improved color space) which ensure more accurate predictions. In addition, a key development is the improved set of test color samples. The CRI's TCS were limited to only a few pastel tones, unrepresentative of the variety of colors in our surroundings. In contrast, the TM-30 samples – the so-called Color Evaluation Samples (CES) – are uniformly-distributed across the color space and they all come from reflectance measurements for a variety of objects, both natural and human made. They are therefore expected to be as relevant as possible to a real-world visual experience.

The CES also overcome a previously-unrecognized problem with the TCS samples [27]: the TCS's reflectances are biased – meaning that their variations in reflectance tend to happen at a few specific wavelength ranges, with reflectance remaining comparatively uniform at other wavelengths (this is because the original, physical TCS samples were obtained by mixing only a few specific pigments). Consequently, the CRI calculation gives uneven importance to neighboring wavelengths, and the CRI score can be optimized by minute shifts in the peaks of an SPD; such optimization, however, isn't indicative of an actual improvement in color fidelity since there is no physiological basis for the wavelength bias. This bias, unfortunately, is also found in other attempts to improve color metrics that have been proposed over the years [28]. The TM-30 CES solve this problem: their spectral features are evenly-distributed across wavelengths such that, on average, no wavelength region is over-emphasized. Fig. 8 compares the two sample sets.

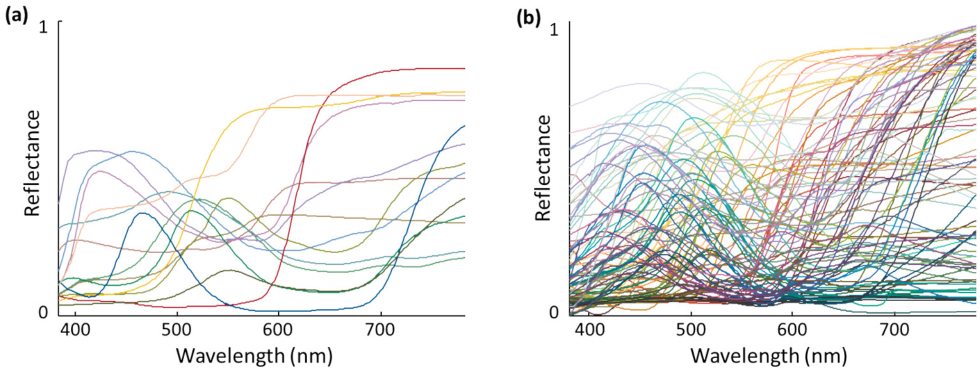

Fig. 8. Color test samples (a) TCS of the CRI method (b) CES of the TM-30 method.

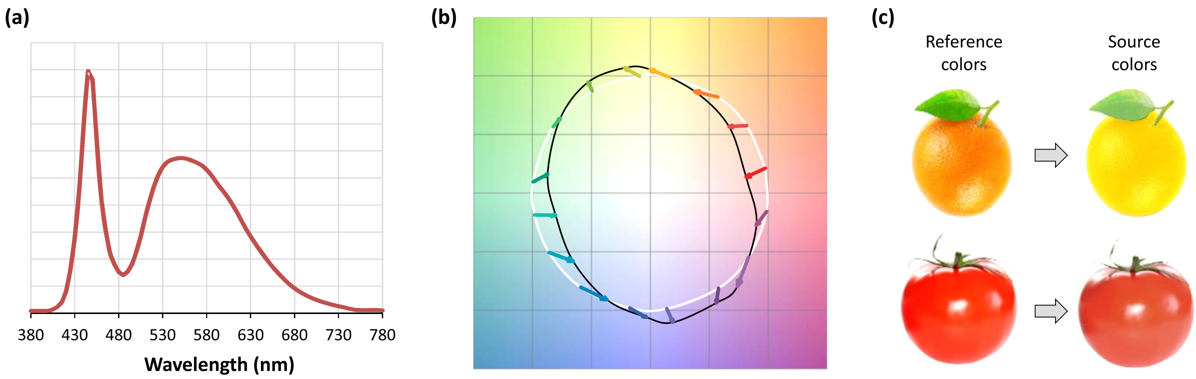

Fig. 9. (a) SPD for a low-fidelity LED. (b) Corresponding TM-30 color icon: each arrow shows the average color shift for a given hue. The white circle corresponds to colors under the reference SPD, and the black ellipse to colors under the test SPD. In this example, the source causes some hues to be desaturated (red, green), some to be saturated (yellow, blue) and some hues to shift to other hues (orange becomes more yellow). (c) Illustrative examples of color distortions of two objects under this source. The skin of the orange is hue-shifted. Its leaf, and the skin of the tomato, are de-saturated.

In summary, the CES are a set of 99 real-world color samples, uniform both in color space (thereby evenly spanning a variety of chromas and hues) and in wavelength space (thereby avoiding spectral bias), which makes them well-suited for accurate color-rendering calculations. Their selection is a key achievement of the TM-30 method.

Given these samples, the TM-30 calculation proceeds similarly to the CRI: for each color sample, the color shift between illuminations under the reference illuminant and the test source is evaluated; from these shifts, several metrics are derived.

The first metric is the fidelity index R_f. Similar to the CRI R_a index, R_f has a maximum score of 100 if no color shift occurs; otherwise, an averaged color shift is subtracted from 100, reducing the value of R_f. In some cases, the predictions of R_f and R_a agree very well (in particular for smooth SPDs), but in other case their predictions differ markedly. This happens especially for SPDs having very sharp spectral features – in part because these can probe the spectral bias of the CRI's test samples, of which the TM-30 samples are devoid.

R_f is most informative if its value is high, indicating that most color shifts are small. Conversely, in the case of lower fidelity, more substantial color shifts occur and R_f alone does not indicate what colors are affected nor what shifts they undergo. For this reason, a number of other metrics are also defined in TM-30; perhaps the most important is the color icon, illustrated in Fig. 9.

The color icon is a vectorial plot, which approximately shows the color space "seen from above" (i.e. projected in the chroma-hue plane of Fig. 5), and indicates the average color shift for various hues. The length of each vector indicates the relative magnitude of the shift, and its direction indicates the type of shift: inward for de-saturating, outward for saturating, and sideways for a hue shift. This icon is well-defined, thanks to the fact that the vectorial field of average color shift is a smoothly-varying quantity in color space [29].

The icon, together with the value of R_f, provides more useful information to the lighting practitioner, by indicating how samples of a given hue tend to be affected by a light source – keeping in mind, however, that these values are averages and do not inform on the exact behavior of a given object, due to individual variations in the details of object's reflectance values that are lost in the averaging process.

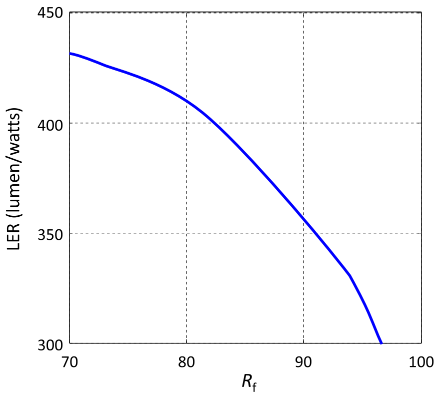

Fig. 10. Tradeoff between LED source's fidelity and efficiency: the line is an estimate of the highest value of LER which can be achieved for a given R_f.

3.4. Tradeoffs

Solid-state lighting offers an unprecedented degree of choice for spectral design. Existing technological limitations are gradually diminishing, and already we are facing an interesting and profound challenge – one that is generating considerable controversy world-wide: put simply, we do not yet know how to design the SPD for solid-state light sources in order to provide the optimum value in the emitted light, for the intended application.

As is often the case in product design, the issue is challenging because of conflicting design goals. As an analogy, consider the design of an automobile, for which there are two desirable characteristics – safety and fuel efficiency. Unfortunately, many of the things that could increase safety would create extra weight and thus reduce fuel efficiency, and vice versa. At some point there is a need to find the best balance between competing interests to yield designs that are best, overall, for people.

For instance, a well-known trade-off arises when trying to design a source with high color fidelity. A key issue arises from a feature in the long-wavelength region of the spectrum, where the spectral sensitivity of the L cone cell has a very long tail that extends weakly to a least 780 nm, even though it has already dropped to 10% relative sensitivity by 658 nm. In principle, to produce a perfect color match to colors under daylight, full illumination would be needed in this region of weak sensitivity, but this would reduce the luminous efficacy dramatically. This would be wasteful, because very few people if any would notice the very slight increase in color accuracy that this would achieve. The question then becomes: how much color distortion is acceptable to save energy? Some have argued that the best guide is the just noticeable difference – in other words, we should not aim for color fidelity perfection, but rather sufficient fidelity that most people would be oblivious to, and unaffected by, the small error produced.

This trade-off can be quantified in terms of the fidelity of a source and its LER, as illustrated in Fig. 10. It has been proposed that a value of R_f above 90 might be suitable to select high-fidelity sources, although values of R_f of about 75–80 are most commonly encountered in products today.

Some lighting practitioners also intentionally seek to distort color, rather than to provide high fidelity. This is used in applications where a specific type of object is illuminated, e.g. the well-known "butcher lamps" making meat look more red (although there is debate as to whether this is effectively misleading advertising). It has also been argued, based on short-term "color preference" studies, that general lighting could benefit from intentional color shifts; for instance, an increased chroma for some colors is sometimes proposed as desirable for long-term illumination. In such considerations, the trade-off becomes more complex and subjective: what magnitude of color shifts is acceptable? Is an increase in chroma for a color acceptable if it leads to a hue shift for another color?

Likely, the only one way to determine optimal functioning points for these various trade-offs is to carry out real-world human factor experiments in which various combinations of color fidelity and color distortion are provided as long-term illumination in controlled experiments, and the occupants are studied in multiple different ways to attempt to find the optimum lighting solution – which will probably depend on the characteristics of the test location and the occupants.

3.5. An application example: lighting in a museum

Modern color science offers the lighting practitioner a multitude of color metrics to guide design. However, what metrics are most relevant and how to best use them is a complex topic, drawing from rational arguments as well as personal and artistic judgment. The setting and application play a crucial role – the figures of merit for lighting a parking lot and a high-end retail store are obviously very different. Offering a guide of what metrics are relevant for what application is well beyond the scope of this article. Instead, below, we bring up simple considerations in the example use-case of museum lighting.

A first aspect is the color temperature of the light. Historically, museums have used a combination of cool-white diffuse daylight (for instance from ceiling skylights), and warm-white incandescent lighting. LED lighting gives access to any color

Fig. 11. (Left) Schematic example of the use of "color enhancement" to restore faded colors. The right half of the image shows the current colors of a tapestry. All colors are somewhat faded, with in particular greens showing as nearly-blue. The left half of the image shows a hypothetical rendition under a light source which saturates reds and blues somewhat strongly, and saturates and hue-shifts greens strongly towards their original tones. (Right) Example of a TM-30 color icon for a light source which might accomplish such an effect.

temperature in any form factor (diffuse or directional), and hence offers more freedom. This gives rise to various considerations. First, what color temperature is best suited for a given artefact? It might be the light originally used by the artist (daylight for an outdoor painting, candle light for an older studio painting...); or it might be a temperature selected to create a specific ambience in the art gallery [30]. Numerous "color preference" studies have been performed on this topic; they do not indicate systematic trends, with the preferred color temperature depending on the type of art piece and the cultural background of the observer. Clearly, the choice here must be guided by the curator's design objective. Second, how does the choice of color temperature impact the conservation of the art? Any light causes damage to colored pigments, and generally short-wavelength radiation is more damaging; beyond the obvious rule of avoiding ultra-violet radiation (which visible LEDs don't emit anyway), curators may prefer warm-white sources for fragile items (note, however, that damage also scales with the intensity of light, which ultimately offers a much larger lever arm on art conservation).

The freedom offered in the rendition of object's colors also gives rise to interesting choices: what should the light do to the colors of the art piece?

A widespread opinion is that colors should be rendered "faithfully", so that the viewer experiences the same colors that the artist intended. Given that most historical artwork was created under natural sources (daylight or blackbody radiators), it is therefore reasonable to request that an LED source offer high fidelity – this indicates that the colors suffer little distortion.

Over the past decade, there has sometimes been skepticism against LED lighting in the museum world, maybe in part because many products offer imperfect fidelity for the sake of slightly-higher efficiency. However, the availability of modern and accurate fidelity metrics, together with better education regarding the trade-off between color quality and efficiency, should help improve this perception and eventual adoption.

In addition, there are some limited cases where color fidelity may not be the most relevant metric. For instance, some art pieces have significantly faded over time. A high-fidelity source would merely render these faded colors as they are. Rather, and assuming the original appearance of these colors can be inferred, a light source can be design to "correct" for the fading, for instance by increasing saturation and shifting hue [31]. This is illustrated in Fig. 11, which also shows how the TM-30 color icon can provide an intuitive indication of how colors will be modified. In practice, designing a light source to accomplish this effect may be a demanding job, best accomplished by considering the actual reflectance of the artefact's colors.

Such "color-correcting" practice is currently very uncommon in museums. This may be in part because the enabling technology is quite recent, and may also be due to our cultural attachment to the present colors of familiar art pieces – as illustrated by the debate over the restoration of the Sistine Chapel's frescoes to their original colors. In this sense however, lighting offers a unique opportunity to modify an artefact's perception without physically altering it. One may therefore hope that future museum applications will make a broader use of the new possibilities offered by LED lighting.

More generally, while the present discussion centers on museum lighting, it is informative of the overall opportunities and challenges brought forth by LED lighting: as light can be controlled with great power to produce almost any effect, the onus is on the lighting practitioner to understand what color metrics are relevant for providing the right light in a given context.

4. Conclusion

LEDs are offering an unprecedented opportunity for improvement of quality of life in an area of profound importance – human vision. This is possible because LEDs have become efficient, affordable and practical, while also enabling a tremendous range of choice in the design of their spectral power distribution. In turn, this newfound freedom has shaken color science itself – leading to an awareness that previous methods for quantifying the color rendering characteristics of lamps were inadequate for LEDs. Perhaps more importantly, after successfully overcoming that problem through an international scientific consensus process, it has become evident that additional work is needed on the ergonomics of color rendering. More and better experiments are needed to establish the links that may exist between spectral engineering and long-term human well-being. In turn, this will inform the next generation of efficient and practical electric lighting systems.

References

[1] K. Shimizu, M. Bohmer, D. Estrada, S. Gangwal, S. Grabowski, H. Bechtel, E. Kang, K.J. Vampola, D. Chamberlin, O.B. Shchekin, J. Bhardwaj, Toward commercial realization of quantum dot based white light-emitting diodes for general illumination, Photon. Res. 5 (2017) A1–A6.
[2] K.A. Denault, M. Cantore, S. Nakamura, S.P. DenBaars, R. Seshadri, Efficient and stable laser-driven white lighting, AIP Adv. 3 (2013) 072107.
[3] A.U. Viola, L.M. James, L.J.M. Schlangen, D-J. Dijk, Blue-enriched white light in the workplace improves self-reported alertness, performance and sleep quality, Scand. J. Work Environ. Health 34 (4) (2008) 297–306.
[4] M. Krames, A. David, Circadian friendly LED light source, US Patent 9410664, 2016.
[5] E. Schrödinger, Grundlinien einer Theorie der Farbenmetrik im Tagessehen [Outline of a theory of colour measurement for daylight vision], Ann. Phys. 368 (1920) 397–426, 481–520.
[6] M.D. Fairchild, Color Appearance Models, 3rd ed., Wiley-IS&T, 2013.
[7] A. Stockman, L. Sharpe, The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype, Vis. Res. 40 (2000) 1711–1737.
[8] G. Wyszecki, Current developments in colorimetry, Proc. AIC Color 73 (1973) 21–51.
[9] R.G. Kuehni, A. Schwarz, Color Ordered: A Survey of Color Systems from Antiquity to the Present, Oxford University Press, 2007.
[10] CIE 15: 2004, Colorimetry 2004, International Commission on Illumination, 2004.
[11] K.A. Smet, M.A. Webster, L.A. Whitehead, A simple principled approach for modeling and understanding uniform color metrics, J. Opt. Soc. Am. A 33 (3) (2016) A319–A331.
[12] Y. Ohno, Practical use and calculation of CCT and Duv, Leukos 10 (1) (2014) 47–55.
[13] ANSI ANSLG C78.377–2011. American National Standard for Electric Lamps “Specifications for the Chromaticity of Solid State Lighting (SSL) Products”.
[14] K.A. Mulcahy, K.C. Clarke, Symbolization of map projection distortion: a review, Cartogr. Geogr. Inf. Sci. 101.28 (3) (2001) 167–181.
[15] D.L. MacAdam, Visual sensitivities to color differences in daylight, J. Opt. Soc. Am. 32 (5) (1942) 247–274.
[16] M.S. Rea, J.P. Freyssinier, White lighting, Color Res. Appl. 38 (2013) 82–92.
[17] K.S. Dulai, M. von Dornum, J.D. Mollon, D.M. Hun, The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates, Genome Res. 9 (7) (1999) 629–638.
[18] L.A. Whitehead, K. Smet, M. Webster, A physically based interpretation of the hue of surfaces, in: Proceedings of the CIE 2017 Midterm Meeting, October 2017.
[19] M.R. Luo, G. Cui, C. Li, Uniform colour spaces based on CIECAM02 colour appearance model, Color Res. Appl. 31 (2006) 320–330.
[20] K. Houser, M. Mossman, K. Smet, L.A. Whitehead, Tutorial: color rendering and its applications in lighting, Leukos 12 (1–2) (2016) 7–26.
[21] M.P. Royer, A. Wilkerson, M. Wei, K. Houser, R. Davis, Human perceptions of colour rendition vary with average fidelity, average gamut, and gamut shape, Light. Res. Technol. 49 (2016) 1477153516663615.
[22] CIE, Method of specifying and measuring color rendering properties of light sources CIE Publ. No. 13.3, Central Bureau of the CIE, Vienna, Austria, 1995.
[23] K.W. Houser, M. Wei, A. David, M.R. Krames, 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 (2013) 10393–10411.
[24] K.A. Smet, A. David, L.A. Whitehead, Why color space uniformity and sample set spectral uniformity are essential for color rendering measures, Leukos 12 (1–2) (2016) 39–50.
[25] A. David, P.T. Fini, K.W. Houser, Y. Ohno, M.P. Royer, K.A. Smet, M. Wei, L.A. Whitehead, Development of the IES method for evaluating the color rendition of light sources, Opt. Express 23 (12) (2015) 15888–15906.
[26] H. Yaguchi, A. David, T. Fuchida, K. Hashimoto, G. Heidel, W. Jordan, S. Jost-Boissard, S. Kobayashi, T. Kotani, R. Luo, Y. Mizokami, CIE 2017 Colour Fidelity Index for Accurate Scientific Use, CIE224:2017, ISBN 978-3-902842-61-9, 2017.
[27] K. Smet, L.A. Whitehead, Meta-standards for color rendering metrics and implications for sample spectral sets, SID Color Imaging Conf. 1 (2011) 76–81.
[28] W. Davis, Y. Ohno, Color quality scale, Opt. Eng. 49 (3) (2010) 033602.
[29] A. David, Color fidelity of light sources evaluated over large sets of reflectance samples, Leukos 10 (2014) 59–75.
[30] J. Schanda, P. Csuti, F. Szabó, A new concept of color fidelity for museum lighting, Leukos 12 (2016) 71–77.
[31] F. Viénot, G. Coron, B. Lavédrine, LEDs as a tool to enhance faded colours of museums artefacts, J. Cult. Herit. 12 (2011) 431–440.