https://doi.org/10.2352/CIC.2022.30.1.06
©2022 Society for Imaging Science and Technology

Luminance, brightness, and lightness metrics for HDR

Charles Poynton, Independent consultant, Toronto, Canada, charles@poynton.ca

Abstract

High dynamic range (HDR) technology enables a much wider range of luminances – both relative and absolute – than standard dynamic range (SDR). HDR extends black to lower levels, and white to higher levels, than SDR. HDR enables higher absolute luminance at the display to be used to portray specular highlights and direct light sources, a capability that was not available in SDR. In addition, HDR programming is mastered with wider color gamut, usually DCI P3, wider than the BT.1886 ("BT.709") gamut of SDR. The capabilities of HDR strain the usual SDR methods of specifying color range. New methods are needed.

A proposal has been made to use CIE LAB to quantify HDR gamut. We argue that CIE *L** is only appropriate for applications having contrast range not exceeding 100:1, so CIELAB is not appropriate for HDR. In practice, *L** cannot accurately represent lightness that significantly exceeds diffuse white – that is, *L** cannot reasonably represent specular reflections and direct light sources. In brief: *L** is inappropriate for HDR. We suggest using metrics based upon ST 2084/BT.2100 PQ and its associated color encoding, IC_TC_P.

Lightness metrics

Figure 1 sketches *L**, sRGB, BT.1886, and PQ as functions of absolute luminance. We will analyse the perceptual performance of these EOTFs.

*L**, sRGB, and BT.1886 are all based upon relative, not absolute, luminance. Here, we have chosen suitable absolute reference points. CIE *L** and BT.1886 are both referenced to 100 nt peak/reference white. The sRGB function is referenced to 320 nt; although 80 nt is standardized, few if any commercial displays today meet this standard, and 320 nt is a representative white luminance of today's sRGB displays.

There is only one standard metric for the perceptual response to luminance: CIE metric lightness, *L**. It is based upon a 2-part function: a linear segment near black, and a scaled and offset power function. It is effectively a power function having an exponent of about 0.42. The relation between absolute luminance and relative luminance is established by normalizing to the absolute luminance of diffuse white. The linear segment is in effect at relative luminance below about 1% of the reference white. Figure 2 sketches the ratio of luminance of adjacent *L** units (in the upper graph, red), along with *L** units scaled by 2.55 (in the lower graph, blue, as in Photoshop LAB coding), both as functions of *L**.

Figure 3 sketches the ratio of luminance of adjacent *L** units (for CIE *L** and 2.55 · *L**, as above), as functions of relative luminance on a log axis. Because the Fechner fraction involves taking the derivative, the offset term of *L** vanishes; the linear segment and the cube-root function both plot in these log-log coordinates as a straight lines. The red marker line extending up from 1 on the *x*-axis represents the reference point of *L**; At this point, CIE *L** has a Fechner fraction of about 2.5%; 2.55 · *L** has a Fechner fraction of about 1%. The breakpoint between the linear segment and the power-function segment is about two decades of $\log_{10}$ relative luminance below the reference point; another decade below that, the Fechner fraction of 2.55 · *L** has reached 1, where unit incre-

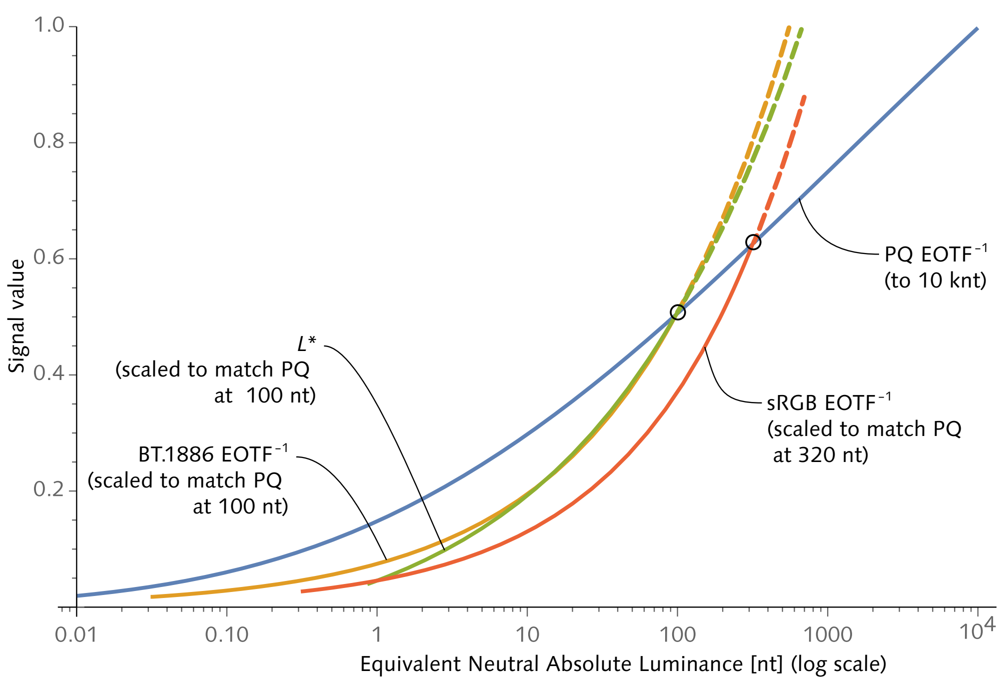

Figure 1 Several perceptual-based functions (comparable to inverse EOTFs) are plotted as a function of absolute luminance over six decades from 0.01 nt to 10^4 nt (10 knt): CIE metric lightness *L**; BT.1886 (for HD); sRGB; and PQ (for HDR).

The *x*-axis is *equivalent neutral absolute luminance* (ENAL), the quantity appropriate to characterize individual *R'G'B'* components. If *R* = *G* = *B*, then this quantity is equal to absolute luminance (*L*).

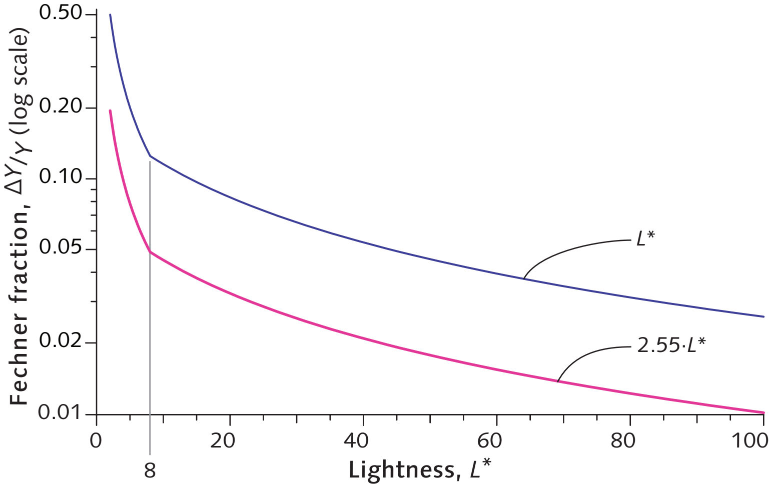

Figure 2 Ratio of relative luminance values for unit ΔL^* (in the upper, blue, graph), and L^* coding scaled by2.55 (in the lower, red, graph).

ment of L^* corresponds to doubling of relative luminance. Clearly, L^* is not representative of perception below this point.

We will argue here that classic L^* is limited to a luminance range of 100:1; L^* is not designed or intended to estimate visual performance outside that range. For a metric applicable to HDR, we have to look outside the CIE. We suggest ST 2084 PQ.

SDR and HDR

High-definition video (HD) in the studio covers a range of absolute luminance levels from about 0.032 to 100 nt, for a contrast range of 3200:1. HD consumer equipment covers about 0.32 nt to 320 nt, for a contrast range of 1000:1.

For the first 20 years of HD (1990 to 2010), the industry-standard EOTF was a 2.4-power function. The BT.1886 EOTF, established in 2011, ensconced the 2.4-power; that is reasonably perceptually uniform [7]. However, the BT.1886 function (and the related OETF, BT.709) have poor perceptual uniformity (that is, poor codeword efficiency) for the top $^1/_3$ or so of its range. At acquisition, the desire for better performance for higher code values led to techniques such as Hypergamma and Cinetone, essentially compressing the top end of the BT.709 OETF to enable acquisition of image data having higher dynamic range. However, no comparable adjustment was made at the display EOTF. That these techniques were possible and effective is evidence that BT.1886 overstates perceptual discrimination at higher code values. We will return to this point.

For today's HDR in the studio, portrayal of diffuse white is typically around 200 nt, with peak (or in IDMS terminology [9], *extreme*) luminance of about 1000 nt (subject to power limiting, or in IDMS terminology, *loading*). For consumer presentation, in dim ambient light, black is presented at about 0.064 nt, yielding a contrast range of about 16 000:1.

Lightness metrics for HDR

There are three historical traditions of functions that relate physical light intensity (formally, luminance) with perceptual performance: Weber/Fechner; de Vries/Rose; and Stevens. Based upon these theories, there are two relevant modern industry standards: DICOM and PQ/IC_TC_P. All of these approaches seek to identify the physical magnitude of amount of light associated with a perceptual *just-noticeable difference* (JND). In the context of the current issue, we want to subdivide three-dimensional color space into elements that represent the threshold of noticeability (or un-noticeability). It is established that human vision does not admit a Euclidean metric; the best we can hope for is a useful engineering approximation. Here, we will focus on the one-dimensional interpretation of the relationship between luminance and brightness/lightness.

The Weber/Fechner "law" states that human vision cannot distinguish ratios of luminance values smaller than about 1.01. Fechner established that Weber's law implies a logarithmic relationship between physical stimulus magnitude and sensation. Assigning unit increments to successive ratios of 1.01 is accomplished by the log

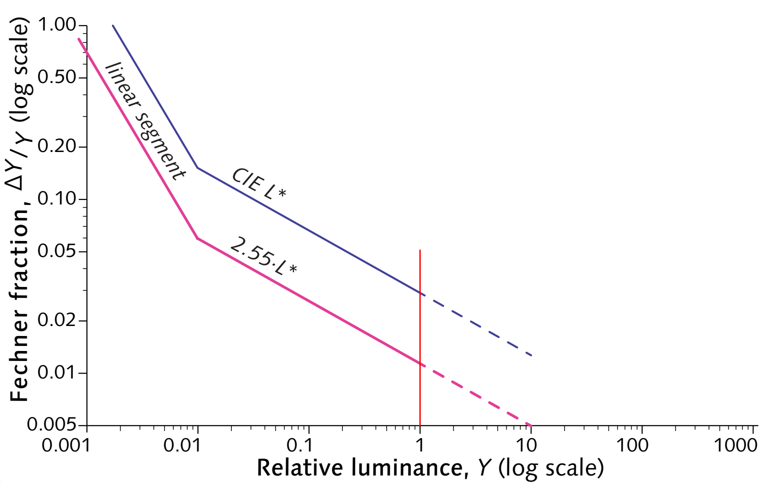

Figure 3 Ratio of relative luminance values for unit ΔL^* as in Figure 2, now plotted as a function of relative luminance on a logarithmic axis. Both the linear segment and the power-law segment plot as straight lines, with slopes –1 and $-^1/_3$ respectively. The white reference has relative luminance of 1; dashed lines are shown above this luminance.

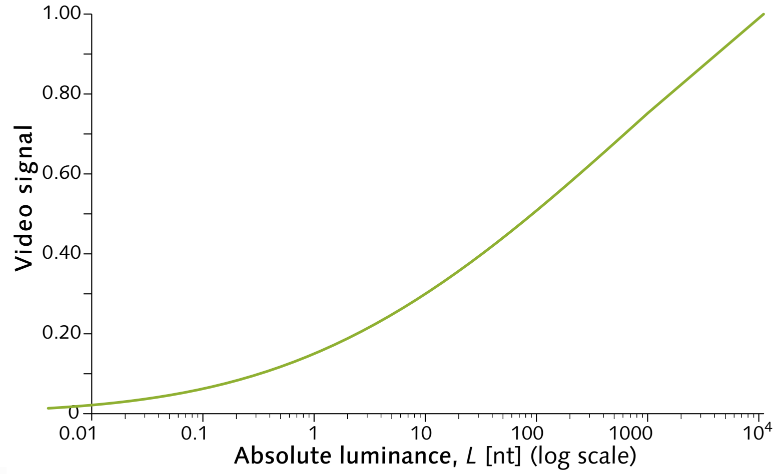

Figure 4 The PQ inverse EOTF is plotted as a function of absolute luminancefrom 0.001 nt to 10 knt, on a log scale.

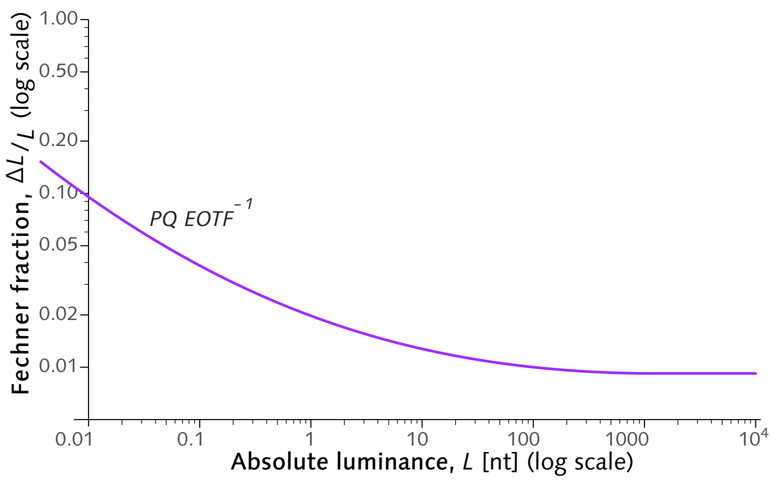

Figure 5 The PQ EOTF^{-1} function has de Vries/Rose behaviour at luminance below 0.1 nt, and Weber/Fechner behaviour at luminance above about 10 nt, remaining smooth throughout. The PQ function is defined for absolute luminance from 0 to 10^4 nt; here the six-decade range 0.01 to 10000 is plotted. This plot uses 876 PQ codes, as specified in BT.2020 "narrow" range, across the luminance range from 0 to 10^4 nt.

to base 1.01. Figure 6 graphs log of Fechner fraction $^{\Delta L}/_{L}$ against log of absolute luminance, for luminance above 1 nt where the relationship is found to hold in practice. The log to base 1.01 of 10 is about 232; that is, 1.01 raised to the power 232 is about ten, so the $\log_{1.01}$ function is equivalent to $232 \cdot \log_{10}$.

De Vries and Rose independently established the photon-noise-limited character of vision, where sensation (at least at low levels of light) is proportional to the square root of physical stimulus magnitude. Figure 7 shows a graph of the de Vries/Rose relationship, in terms of Fechner fraction plotted against luminance on a logarithmic axis. In these coördinates, the relationship has a slope of −0.5; that is, for each two decade increase in luminance, the Fechner fraction drops one decade. The square root function has been scaled by 200, leading to a 1.01 ratio of successive codes at absolute luminance 1 nt.

Modern research and practice has shown that the de Vries/Rose function is applicable at low absolute luminance, at average absolute luminance (L) levels up to about 1 nt. Above that, perceptual discrimination is better predicted by Weber's law [8]. Figure 8 sketches the concept of a two-segment relationship between absolute luminance and the contrast discrimination of vision. At the left, up to 1 nt, the behaviour is according to the de Vries/Rose square-root relationship. On the right of the graph, above 1 nt, behaviour is according to Weber's law, here shown as a 1.01 ratio.

Peter Barten of Philips applied these historical concepts to establish a parametric function that estimates visual sensitivity as a function of absolute luminance across a wide range of conditions [1, 2]. We will return to Barten's work in a moment.

Absolute and relative metrics

Stevens analyzed the Weber/Fechner relationship, and disagreed. Although Stevens was apparently unaware of the work of de Vries and Rose, he estimated the lightness sensitivity of vision as proportional to a power function of stimulus having an exponent between about 0.33 and 0.42. Subsequently, this relationship became known as "Stevens' Law." Stevens' exponent is comparable to the $^1/_3$ exponent established in the CIE L^* definition from 1976.

Both the Stevens "law" and CIE L^* are based upon relative luminance, not absolute: Both assume that vision is adapted to the average absolute luminance of the central visual field. Color scientists consider the average relative luminance of a typical viewed scene to be about 0.2 relative to a perfect diffuse white.

Figure 3, shown earlier, graphed CIE L^* and $2.55 \cdot L^*$ against relative luminance, taking 1.0 as the L^* reference point. If those functions were plotted against absolute luminance, the functions would slide left and right along the x-axis depending upon the chosen reference point (which we hope would align to a reference luminance, most commonly adapted white).

We described the de Vries/Rose function in terms of absolute luminance; Figure 8's two-segment function made a transition from the de Vries/Rose model to the pure-log model at an absolute luminance of 1 nt. If the diameter of the iris of the viewer's eye closes by a factor of 3.2, say from 6.4 mm to 2 mm, light reduces

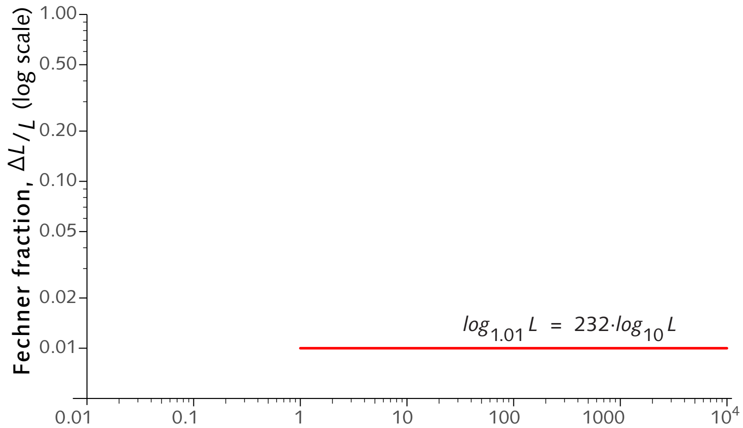

Figure 6 Assuming a Weber/Fechner relationship, taking a ratio of luminance of 1.01 at the threshold of discriminability leads to a $\log_{1.01}$ relation between luminance and JND steps. On a log-log plot with Fechner fraction on the *y*-axis, the relationship plots as a horizontal line. (Axis labels are those of Figure 9, below.)

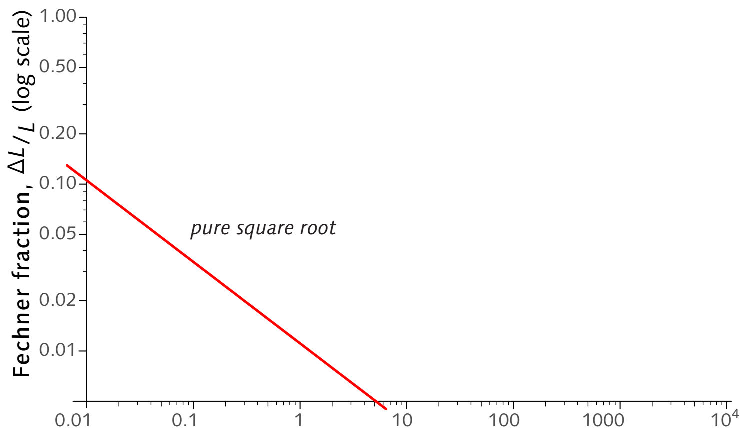

Figure 7 The de Vries/Rose relationship leads to a square root relation between luminance and JND steps. On a log-log plot with Fechner fraction on the *y*-axis, the relationship is a line having slope $-{}^1/_2$. Here, the square root is scaled by 200. A three-decade (1000:1) increase in absolute luminance leads to a 1.5-decade (32:1) reduction in Fechner fraction (from 0.32 to 0.01).

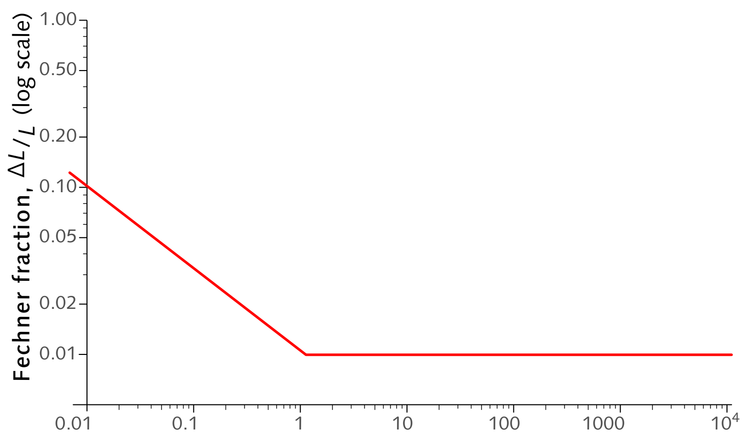

Figure 8 This hypothetical function has de Vries/Rose behaviour at luminance below 1 nt, and Weber/Fechner behaviour at luminance above 1 nt.

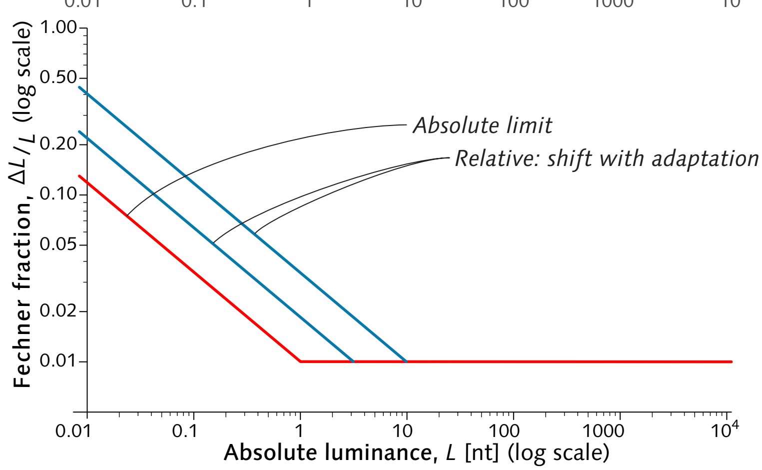

Figure 9 The absolute and relative nature of the de Vries/Rose relationship is sketched. With adaptation to higher average absolute luminance, the square-root region shifts to the right.

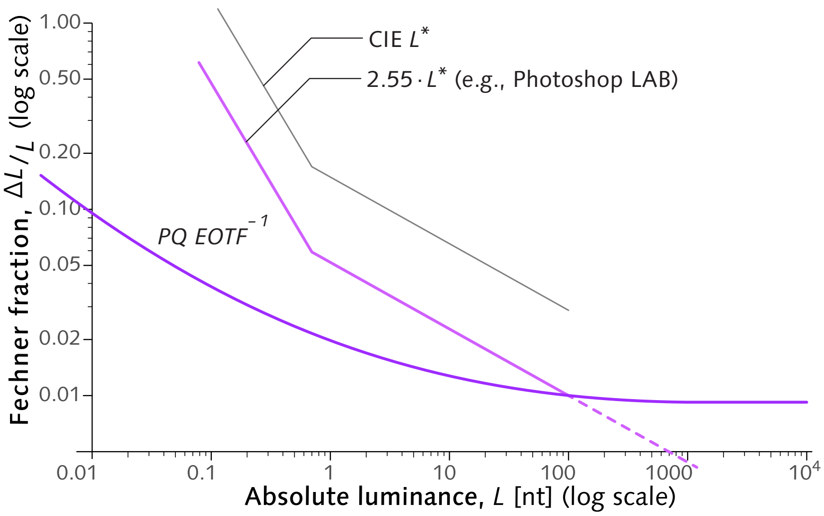

Figure 10 The Fechner fraction of two functions is graphed, as a function of luminance over six decades: CIE metric lightness L^* (referenced to white at 100 nt and scaled by 2.55), and PQ. At about 100 nt, the performance is comparable. Below 100 nt, $2.55 \cdot L^*$ underestimates perceptual performance compared to PQ: there are more colors in this region than L^* or LAB will estimate. Above 100 nt, $2.55 \cdot L^*$ overestimates perceptual performance: A color area or volume metric based upon L^* or LAB will count more colors in this region that can be distinguished by vision.

by a factor of ten. The de Vries/Rose behaviour of vision will then slide one log unit to the right when plotted with luminance on a log scale. So, with adaptation, the de Vries/Rose function is also applicable to relative luminance. Figure 9 repeats the conceptual two-segment function of Figure 8, but now showing the de Vries/Rose characteristic shifted to several higher absolute luminance levels. CIE L^* is based upon relative luminance, so the L^* curve shifts similarly.

Industry standards

The concepts of Weber, Fechner, de Vries, Rose, Stevens, and Barten have evolved into standards for image encoding and perceptual estimation. The medical imaging community adapted Barten's work, and established the DICOM standard grayscale display function (GSDF).

Later, Barten's work (possibly influenced by the DICOM GSDF) was adapted by Dolby as the *perceptual quantizer* (PQ) [5]. PQ was subsequently standardized by SMPTE as ST 2084, and later incorporated into the ITU-R BT.2100 standard for UHD/HDR. Applied to absolute luminance (L), PQ maps luminance from 0 to 10 000 nt to a ten-bit quantity in what a video engineer would call "narrow" range of 876 code values: PQ accommodates 876 steps in the perceptual response from 0 to 10 000 nt. At the top end, there are about 70 PQ codes per stop (factor of two) of absolute luminance. The graphs in Figure 1 include the PQ inverse EOTF; that function is shown again in Figure 4, from 0.01 nt to 10,000 nt. If that function is scaled by the number of digital signal codes in use (eg, 876 or 1023), then differentiated, then divided by absolute luminance, the graph of Figure 10 results, where the Fechner fraction (on a log scale) is on the y-axis.

PQ was conceptualized in terms of absolute luminance. In its engineering application to HDR video encoding, the PQ function is applied to the so-called intensity (I) component of IC_TC_P, where the C_T and C_P components encode chroma. The I component is $0.5 \cdot L' + 0.5 \cdot M'$, formed by applying the PQ EOTF^{-1} function to L and M, where L and M are estimated tristimuli according to the signal processing specified in BT.2100. In its engineering application to HDR display interfaces, the PQ function is applied to R, G, and B tristimuli having BT.2020 primary chromaticities. In all three of these cases – application of PQ to relative luminance, to *LMS*, and to *RGB* – perceptual uniformity is exhibited.

Perceptual estimation

The IC_TC_P components of HDR color image encoding have been recognized by ITU-R as providing a metric suitable for assessing perceptual differences not too much larger than a few JND: The BT.2124 standard provides a definition of ΔE_{ITP}, where Euclidean distance is computed, and unit ΔE is taken to approximate perceptual uniformity. (The C_T component is scaled prior to Euclidean distance being computed; see BT.2124 for details.)

It is evident from Figure 1 that up to about 10 nt, all four functions have comparable slope. However, above 10 nt, the L^*, sRGB, and BT.1886 functions all exhibit slope significantly greater than the slope of the PQ function. This behaviour indicates failure of perceptual uniformity – in particular, a signal value excursion in excess of what is required for vision. (We referred earlier to BT.1886 having too many codes in the top $^1/_3$ or so of its code range, an aspect that enabled modification of encoding functions [OETFs] to accommodate higher dynamic range).

Interpreted in terms of Fechner fraction, the behaviour of L^*, sRGB, and BT.1886 at the top their ranges indicates that they all overestimate perceptual differences by a factor of about two – that is, increasing L^*, sRGB, or BT.1886 values will have about half of the expected perceptual effect. In this example, CIE L^* has been referenced to 100 nt, a typical portrayal of diffuse white. If instead CIE $L^*$is normalized at a higher luminance, performance will be even worse.

Figure 10 graphs, in Fechner fraction form, $2.55 \cdot L^*$ (referenced to 100 nt white) and PQ. At 100 nt, the two functions have equivalent performance. However, the straight-line nature of the L^* function when plotted in this form makes evident two facts: First, as luminance decreases below the reference point, L^* has increasingly coarse quantization. Second, as luminance increases above the reference point, L^* has increasingly fine quantization. Below 100 nt, $2.55 \cdot L^*$ underestimates perceptual performance compared to PQ: there are more colors in this region than L^* or LAB will estimate. Above 100 nt, $2.55 \cdot L^*$ overestimates perceptual performance: A color area or volume metric based upon L^* or LAB will count more colors in this region that can be distinguished by vision.

Discussion

We have presented various luminance/lightness metrics. If a color *area* metric is based upon any of the L^*, sRGB, or BT.1886 quantities – for example, the a^* and b^* chroma components of CIE LAB used to represent area directly, or the a^* and b^* components used to estimate area used in a gamut ring calculation [4, 10] – then the overestimation by a factor of two in the lightness dimension is squared, resulting in an overestimation of perceptual effect by a factor of four. If a color *volume* metric is based upon any of these quantities – for example, cubic delta-E (ΔE^3) – then the overestimation by a factor of two in the lightness dimension is cubed, resulting in an overestimation by a factor of eight. These metrics greatly overemphasize the light areas of the color area or volume, compared to metric such as PQ that has a better perceptual foundation. A LAB-based metric with a white reference chosen a factor of five below HDR peak white might estimate a space as having 64 million colors, where a PQ-based metric might more realistically estimate 8 million.

Some researchers have used an L^*-based metric taking the reference point of L^* as peak white. Such an approach causes the linear segment of L^* to intrude into the normal 100:1 black-to-diffuse white range of L^*. If peak white is five times the luminance of the portrayal of diffuse white – a common situation – then the power-function range of L^* occupies only a 25:1 ratio of luminance, instead of the expected 100:1.

If instead the L^* normalization is referenced to the portrayal of diffuse white, then at peak-to-diffuse ratio of 5, L^* values extend to about 180, well beyond the range generally agreed to have reasonable perceptual performance. In this case, the slope discrepancy compared to PQ is amplified even further, leading to an even greater overestimation of the perceptual effect of higher luminance colors.

Conclusion

CIE L^* is only appropriate for contrast range not exceeding 100:1. In practice, this limitation excludes portrayed luminance that significantly exceeds diffuse white (ie, limited portrayal of specular reflections and direct light sources) – that is, this limitation excludes HDR: CIE L^* and CIE LAB are unsuitable for estimation of perceptual parameters of HDR systems.

The PQ function applied to absolute luminance, or IC_TC_P applied to color, aligns much better to human visual performance, and thereby yields a color metric that more closely reflects visual performance.

Acknowledgements

Thanks to Scott Daly, Timo Kunkel, David A. LeHoty, and Dale Stolitzka for their helpful comments.

References

[1] Barten, Peter G. J. (1999), *Contrast Sensitivity of the Human Eye and its Effects on Image Quality,* PhD diss., T U Eindhoven (HV Press Knegsel).

[2] Barten, Peter G. J. (2004), “Formula for the contrast sensitivity of the human eye,” *Proc. SPIE* **5294:** 231 – 238.

[3] LeHoty, David A. and Poynton, Charles (2020), “Comparing Displays Using Luma Contours and Color Volumes,” in *Information Display* **36** (5): 16–21.

[4] Masaoka, Kenichiro; Jiang, Fu; Heckaman, Rodney L.; and Fairchild, Mark D. (2018), “2D Representation of Display Color Gamut,” in *SID Symposium Digest of Technical Papers* **49** (1): 1048–1051.

[5] Miller, J. Scott; Nezamabadi, Mahdi; and Daly, Scott (2013, May). “Perceptual signal coding for more efficient usage of bit codes”, in SMPTE Motion Imaging Journal **122** (4): 52–59.

[6] Poynton, Charles (2012), *Digital Video and HD Algorithms and Interfaces,* Second edition (Burlington, Mass.: Elsevier/Morgan Kaufmann).

[7] Poynton, Charles (2018), *Colour Appearance Issues in Digital Video, HD/UHD, and D-cinema,* PhD Dissertation (Burnaby, B.C.: Simon Fraser University).

[8] Schreiber, William F. (1991), *Fundamentals of Electronic Imaging Systems,* Second edition (Springer-Verlag). Chapter 3.

[9] SID/VESA ICDM International Committee for Display Metrology (2021, June), IDMS – Information Display Measurement Standard, Version 1.1.

[10] Smith, Euan; Heckaman, Rodney L.; Lang, Karl; Penczek, John; and Bergquist, Johan (2020), “Measuring the color capability of modern display systems,” in *J. Society for Information Display* **28** (6): 548–556.