

Perceptual Uniformity in Digital Image Representation and Display

Charles Poynton and Brian Funt
Simon Fraser University
Vancouver, B.C., Canada
cpoynton@sfu.ca, funt@sfu.ca

This is a preprint of a paper
published in Color Research and
Application 39 (1): 6-15
(2014, Feb).
Copyright © 2013-03-13
Charles Poynton

Abstract

Digital image representation is *perceptually uniform* if a small perturbation of a component value – such as the digital code value used to represent red, green, blue, or luminance – produces a change in light output at a display that is approximately equally perceptible across the range of that value. Most digital image coding systems – including sRGB (used in desktop graphics), BT.709 (used in high-definition television, HD), Adobe RGB (1998) (used in graphics arts), and DCI P3 RGB (used in digital cinema) – represent colour component (pixel) values in a perceptually uniform manner. However, this behaviour is not well documented, and is often shrouded in confusion. This paper surveys perceptual uniformity in digital imaging, and attempts to clarify some widely misunderstood aspects of image coding.

1 Introduction to perceptual uniformity

Many applications of digital colour imaging involve economic or technical constraints that make it important to limit the number of bits per pixel. Bits are most effectively used by perception if luminance values or tristimulus values are nonlinearly mapped, like CIE L^* , to mimic the lightness response of human vision. Digital imaging system engineers use vision’s nonlinearity to minimize the number of bits per colour component. Mappings based upon power functions are most common, though mappings based upon logarithms and other functions are sometimes used. The concept is fundamentally important to both the theory and practice of digital imaging, but it is widely neglected or misrepresented in the technical literature.

This paper addresses mainly image display. Many important issues related to processing in perceptually uniform space – for example, performing colour transformations in a manner that preserves hue, or coding that establishes a perceptually uniform chroma scale – are not covered in this work.

2 Luminance

Perceptual coding involves absolute luminance, relative luminance, and related quantities. These topics are generally well understood by colour scientists; however, in digital imaging more generally, much confusion surrounds these quantities. A detour into the nuances of luminance is necessary.

Absolute luminance, defined by the CIE [7], is proportional to optical power (flux) across the visible wavelengths, weighted according to a standardized spectral weighting that approximates the spectral sensitivity of normal human vision. Luminance is proportional to optical power, but derivatives are taken with respect to solid angle and with respect to projected area: Luminance relates to power in a certain direction, emitted from or incident upon a certain area. Absolute luminance has the symbol L_v (or just L , if radiometry is not part of the context); its units¹ are $\text{cd} \cdot \text{m}^{-2}$ (“nit,” or nt). The spectral weighting of luminance is symbolized $V(\lambda)$ or $\bar{y}(\lambda)$.

¹The foot-lambert unit [fL] once used for luminance was deprecated in 1948 [6] in favour of the SI unit, $\text{cd} \cdot \text{m}^{-2}$ (colloquially, “nit,” or nt). In our view, using foot-based units such as foot-lambert and foot-candle [fc] impedes the understanding of radiometry and photometry.

In applications of image capture, recording, and presentation – including photography, cinema, video, HD, digital cinema, and graphics arts – absolute luminance of the original scene is rarely important. Instead, scene luminance is characterized relative to an “adopted” scene white luminance associated with the state of visual adaptation of an actual or hypothetical person viewing the scene [16, 22].

Subsequent processing and display involves *relative luminance*, symbolized Y , whose value is a dimensionless quantity representing black at 0. A reference white value for relative luminance is chosen to correspond roughly to the luminance of a near-perfect diffuse reflector in the scene – that is, to a diffuse reflector having a luminance factor of about 0.9 or 1.0. Relative luminance of reference white was traditionally assigned the value 100, though many modern practitioners prefer to use a reference white value of 1. In digital imaging, reference black and reference white values correspond to integer values such as 0 and 255 (in sRGB, for desktop computing), 64 and 940 (in 10-bit studio digital video), and 0 and 4095 (in 12-bit digital cinema distribution). In some standards, such as studio digital video, codes are allowed to exceed the reference white level; codes above reference white are available to represent scene elements such as specular highlights.

Image scientists and engineers often omit the adjective *relative* in front of luminance; however, distinguishing absolute luminance and relative luminance – and symbolizing them differently, L and Y , respectively – is important because absolute luminance exerts a strong influence over colour appearance. Using relative luminance implicitly discounts that effect. Absolute and relative luminance are *linear-light* measures, directly proportional to light power.²

The term *relative luminance* and its symbol Y are well established in colour science; however, the term and the symbol are widely misused in the fields of video, computer graphics, and digital image processing. Workers in those fields commonly use the term “luminance” – or worse, the archaic term “luminosity” – to refer to a weighted sum of nonlinear (gamma corrected) red, green, and blue signals instead of the linear-light quantities defined by the CIE. The nonlinear quantity is properly termed *luma* and given the symbol Y' [37].

3 Tristimulus values

Three signals proportional to intensity, having physical or nonphysical chromaticities associated with three additive primaries, are called *tristimulus values* (or *tristimuli*). Tristimuli are dimensionless quantities – that is, they have no units [5, 17]. A colour scientist symbolizes tristimuli with capital letters and no primes; examples of tristimuli are RGB , LMS , and XYZ . Relative luminance is a distinguished, special case of a *tristimulus value*. A suitably-weighted sum of tristimuli yields luminance [46]; relative luminance can be augmented with two other linear-light components (having prescribed spectral composition) to yield tristimuli.

Cameras typically depart from the spectral sensitivities prescribed by CIE standards, consequently, tristimulus values with respect to the scene are usually estimated, not exact: The effect of the imperfect match to the CIE Standard Observer (i.e, camera metamerism) is embedded in the image data.

²Instead of using the informal term linear-light, some practitioners use the term *photometrically linear*. The adjective *photometric* properly refers to use of the CIE standard luminance spectral weighting; however, practical cameras typically do not closely approximate the CIE spectral weighting, so the term “photometrically linear” in this context is wrong. The term *radiometrically linear* is appropriate, because the adjective *radiometric* is not associated with any particular spectral distribution.

4 Picture rendering

The usual goal of digital imaging is to produce the intended presentation on the ultimate display device. Image data is typically referenced to a set of additive primaries. Once sensed and recorded, image data is associated with the colour representation defined in the interchange standard. For example, the sRGB standard applies to general computing; the BT.709 standard applies to HD. (Not coincidentally, these standards share the same set of primaries.) Faithful display is achieved on a display device that conforms to the intended colorimetric standard. In professional imaging, and in content creation, tristimulus values and luminance are then exact with respect to a reference additive *RGB display* (for example, a studio display).

All imaging applications involve non-ideal displays, and almost all applications involve image viewing in conditions different from those in effect at the time of image capture. The goal of most imaging applications is not to match relative luminance values between the scene and the display, but instead, to match the ultimate viewers' expectation of the appearance of the scene. Engineers and scientists unfamiliar with colour science are usually surprised to learn that the intended appearance is not achieved by matching relative luminance values between scene and display: Preserving appearance almost always requires manipulating the tristimulus value estimates between the scene and display. Manipulation is typically accomplished either algorithmically (for example, by firmware in a digital still camera) or manually, by a skilled specialist (such as a photographer or a colourist) [10, 12, 18].

Picture rendering of image data from a digital camera involves a complicated series of image processing operations, usually proprietary. The operations are often dependent upon exposure levels, and upon statistics derived from the image data. Picture rendering operation obscures any direct link to scene colorimetry.

An ISO standard [22] standardized the term *scene-referred* to describe image data having a colorimetric link to a scene (for example, “raw” sensor data), and *output-referred* to describe image data having a colorimetric link to an output device (such as a standardized display). The ISO standard was intended mainly for colour management for print, and “output” meant hard copy. Subsequent to adoption of the ISO standard, the term *display-referred* has come to describe image data having a colorimetric link to a digital display device [34, 13, 20].

In many applications of digital imaging, there is no camera. In many modalities of medical imaging (for example, CT scanning), image data is originated algorithmically and does not correspond to any optical image. In graphic arts it is common to use application software (such as Photoshop) to “paint” directly on the display screen, producing an image that has no direct counterpart in the physical world. Finally, in computer-generated imagery (CGI) for movies or games, attempts are made to compute physically plausible scenes that do not exist physically. In all of these applications, image data has no link whatsoever to a physical scene. In such cases, perceptual uniformity must be referenced to the display alone.

It is common for high-end professional digital single-lens reflex cameras (D-SLRs) to have the capability to record in *raw* mode, where image data from the sensor is recorded without any rendering operations. Such data is scene-referred. However, photographers typically process such data through the camera vendor's processing software, or through commercial software such as Lightroom (from Adobe) or Aperture (from Apple). These software packages read raw camera image data, perform picture rendering, and output display-referred image data.

Most industrial and scientific cameras do not incorporate complicated picture rendering operations; they simply transform sensor data through a linear-light 3×3 matrix to form *RGB* tristimuli estimates, then apply a power function having an exponent of around 0.4 (“gamma correction”). Provided that the parameters of matrixing and gamma correction are known, these cameras can be considered to be scene-referred.

In the remainder of this paper, we will discuss perceptual uniformity with reference to the display.

5 Visual response

Human vision has a nonlinear perceptual response to light power. Linearly quantizing a radiometric quantity such as luminance or tristimulus values is perceptually inefficient. *RGB* pixel values used in most commercial imaging systems – and in virtually all 8-bit imaging systems – are quantized having a nonlinear relationship to light power. It is a continuing serious source of confusion among computer graphics, imaging, and video practitioners that the term “intensity” is commonly used to refer to pixel component values even when the corresponding quantity is not proportional to light power. For example, *Mathematica* has a built-in function *GrayLevel* that “specifies ... gray-level intensity ...”; however, greyscale pixel values are implicitly coded nonlinearly (by virtue of display through a transfer function resembling that of sRGB) and *Mathematica*’s term *intensity* is therefore technically incorrect. As another example, the *MATLAB* system has four classes of images. Until version 5, one of the classes was called *intensity image*; however, its pixel values are implicitly coded nonlinearly, and again the term *intensity* is technically incorrect. The documentation for *MATLAB* was recently revised to use the more accurate term *grayscale image*.

The terms *luminance* and *lightness* apply directly to greyscale imaging. The terms are also used in colour imaging. However, most colour imaging systems encode a nonlinear transformation of red, green, and blue; neither luminance nor lightness is directly available. In what follows, the luminance and lightness of the individual primaries is addressed.

6 Logarithmic approximation

According to the historical Weber-Fechner model [15], lightness perception is very roughly logarithmic.³ Put briefly:

Vision cannot distinguish two luminance levels if the ratio between them is less than about 1.01 – in other words, the visual threshold for luminance difference is about 1 percent. [39, p. 12]

The ratio of 1.01 is the *Weber contrast*. A first approximation of perceptual uniformity is obtained by taking advantage of the Weber ratio, choosing a coding such that successive pixel component values are associated with a constant ratio of luminance from code to code across the tone range from some minimum representable luminance up to white. Such coding is effected by a logarithmic transform of relative scene luminance.

For a true logarithmic law having a 1.01-ratio between adjacent codes, the relative luminance difference between adjacent codes is 1% across the whole range. The number of codes (pixel values) required to maintain a 1.01 Weber ratio across a range of relative luminance values from 0.01 to 1 is as follows:

$$\frac{\log 100}{\log 1.01} \approx 462; \quad 1.01^{462} \approx 100$$

³In what follows, log denotes the base-10 (common) logarithm.

So, 462 codes cover a contrast ratio of 100:1.⁴ A photographer or cinematographer prefers to deal with light ratios expressed in “stops” (factors of two) of luminance.⁵ For pure logarithmic coding with a Weber fraction of 1%, there are 69 codes per stop – about six bits of data per stop:

$$\frac{\log 2}{\log 1.01} \approx 69; \quad 1.01^{69} \approx 2$$

Six bits covers about a stop, and three bits serve to enumerate eight stops (a 256:1 range); so $6 + 3 = 9$ bits covers a 256:1 luminance range with a Weber contrast of 1.01.

A logarithm to base b increments by one when the (positive) argument is multiplied by b . To map a relative luminance ratio of 100:1 (represented as signal values from 0.01 to 1) into a pure logarithmic code from 0 to 1, simply form the base-100 logarithm, then add one. (The base-100 logarithm is half the common base-10 log.) For the result to lie in the range 0 to 1, relative luminance values less than 0.01 must be excluded; in any event, zero must be excluded to avoid the singularity in the log function. The expedient method is to set the result to zero for any argument less than or equal to 0.01. The following equation encodes relative luminance T into normalized image data code value V :

$$V = \begin{cases} 0, & T \leq \frac{1}{100} \\ 1 + \frac{1}{\log_{10}(100)} \log_{10}(T), & \frac{1}{100} < T \leq 1 \end{cases}$$

In commercial systems that use log coding, it is usual to code R , G , and B components individually, instead of coding luminance; in these cases, T symbolizes a tristimulus value. The pure-logarithmic encoding just described – encoded into 8-bit components by multiplying V above by 255 and rounding to an integer – is one of two logarithmic encodings specified in the MPEG and H.264 video compression series of standards [23]. (The standards specify a handful of other, non-logarithmic encodings.) The first scheme has 127.5 steps per decade, corresponding to a Weber contrast of about 1.018. The second scheme covers a $10^{2.5}$ contrast ratio (about 316:1); it has 102 steps per decade, and has a Weber contrast of about 1.023. The second scheme can be implemented by the equation above by replacing each instance of 100 by $10^{2.5}$. As far as we are aware, neither of these schemes has been commercially deployed; one important reason is that clipping below $1/100$ or $1/316$ of relative luminance is very likely to produce artifacts. Quasilog coding schemes have been commercially deployed in digital cinema, as will be discussed later; however, they process radiometric values near black without clipping.

The Weber-Fechner Law is based upon the assumption that thresholds (*just noticeable differences*, JNDs) can be meaningfully integrated. S.S. Stevens criticized the Weber-Fechner law, declaring that “A power function, not a log function, describes the operating characteristic of a sensory system” [49]. Thresholds are defined by uncertainties; Stevens believed that integrating the thresholds would accumulate uncertainties. Stevens devised psychophysical experiments based upon magnitude estimation to obtain more direct measures of the relationship between physical stimulus and perceptual response. He concluded that lightness could be approximated by the 0.33-power – that is, the cube root – of relative luminance. His results agreed quite well with investigations made decades earlier by Albert E. O. Munsell [32] (son of Albert H. Munsell [33]).

⁴In the 1950s, the developers of colour television assumed that it was sufficient to cover a contrast ratio of 30:1 with a 1.02 ratio, yielding 172 steps, as documented by Fink [11, p. 201].

⁵An imaging scientist uses the term *optical density* to refer to the negative of the base-10 logarithm of reflectance or transmittance factor; both of these are proportional to relative luminance. The 100:1 contrast ratio mentioned above corresponds to 2 density units. Photography standards and textbooks define a stop in terms of \log_2 ; however, cameras have exposure time – or “shutter speed” – markings of $\frac{1}{1000}$, not $\frac{1}{1024}$ as would be the case if a stop was exactly a ratio of two. For scientific and engineering purposes, defining a stop as a ratio of exactly $10^{0.3}$, or about 1.995, is more realistic and more useful than the classic photographic definition: Consider a stop to be exactly 0.3 density units – that is, take a density unit as $3^{1/3}$ stops. To use this interpretation in photographic standards and textbooks, replace \log_2 by $0.3 \cdot \log_{10}$.

7 Lightness

In the context of the historical Weber-Fechner logarithm and Stevens' power function, an estimate of vision's lightness response, symbolized L^* , was eventually standardized by the CIE in 1976. The definition is essentially unchanged in today's colour science standards [7]. Given relative luminance, CIE L^* returns a value between 0 and 100; a "delta" (difference) of 1 is taken to approximate the threshold of vision for luminance differences. The L^* function is basically a power function with what we call an "advertised" exponent of $1/3$ – that is, a cube root. A linear segment is inserted near black, below relative luminance of about 1%; the power function segment is scaled and offset to maintain function and tangent continuity at the breakpoint. See Figure 1.

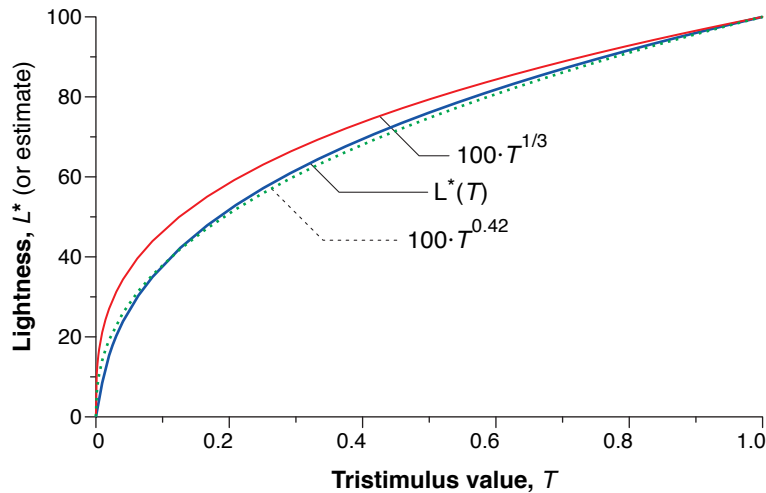


Figure 1: **CIE Lightness**, symbolized L^* , estimates the perceptual response to relative luminance. Here L^* is overlaid by a 0.42-power function, the pure power function that best fits L^* . The L^* function involves a cube root – that is, a $1/3$ -power function – but the power function is scaled and offset. A cube root is overlaid onto the plot: A pure cube root is a poor approximation to L^* .

The technical literature is rife with statements that L^* is a cube root [30, 42, 44]. In fact, the scaling and offset cause the function to approximate an "effective" 0.42-power over its entire range.

The linear segment at relative luminance less than about 0.01 was introduced for mathematical convenience [36]. What effect the linear segment has on perceptual uniformity is an open question.

8 Display characteristics and EOCF

In a properly adjusted historical CRT display, the electrostatic characteristics of the electron gun caused the CRT to impose an *electro-optical conversion function* (EOCF) that was well approximated by a power function from voltage input to light output. The symbol γ (gamma) represents the exponent at the display.

In computing, the sRGB standard [19, 50] establishes a 2.2-power function. In video and HD, gamma has historically been poorly standardized or not standardized at all. Recently, the ITU-R standardized the value 2.4 [24], carefully chosen to codify current practice: A studio reference display has gamma of about 2.4. A 2.4 power is a close match to the inverse of the L^* function; see Figure 2.

CRT displays are now obsolete. Virtually all non-CRT image display devices, including today's LCD and plasma (PDP) direct view displays, and DLP and LCoS projectors, are designed to mimic the historical behaviour of CRTs. In displays such as PDP and DLP that involve pulse-width modulation

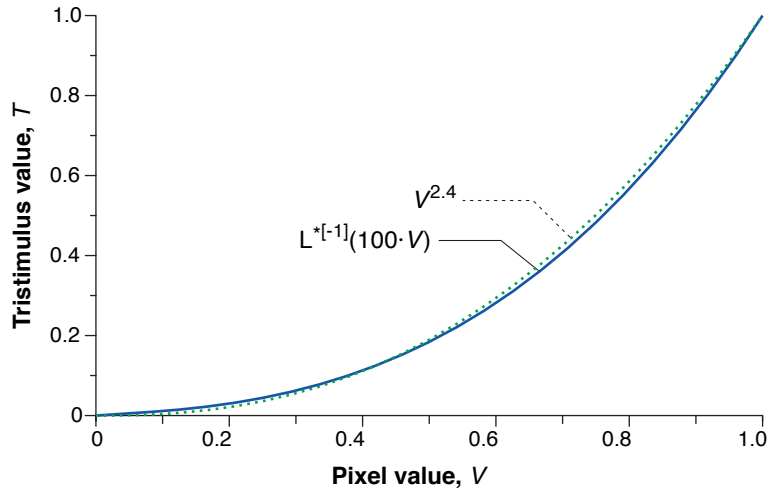


Figure 2: **EOCF of a standard studio HD display** is approximated by a 2.4-power function from video signal to light power. The gamma of a display system – for example, a CRT, or the reference sRGB EOCF – is the numerical value of the exponent of the power function. The inverse of the CIE L^* function is overlaid onto the plot: It is evident that a 2.4-power function is a very close match to the inverse of L^* .

that converts signal to light in a linear manner, a nonlinear function (“degamma,” or “inverse gamma”) is provided by signal processing, typically incorporating one or more lookup tables (LUTs). In displays such as LCDs that involve nonlinear physical transducers, signal processing incorporates a function that imposes the difference between the desired 2.2- or 2.4-power-law behaviour of the image exchange standard and the inverse of the native characteristic of the transducer.

9 Eight-bit pixel components

Eight-bit pixel components are very widely used in digital imaging. It is perceptually uniform coding, effected by the nonlinear characteristics of standard displays, that makes 8-bit components practical for continuous-tone imaging.

If 8-bit components were used to encode linear-light values, with black at 0 and white at 255, a Weber contrast of $\frac{255}{254}$ – about 1.004 – would be obtained at white, code 255. As pixel value drops below code 100, the Weber contrast would exceed 1.01; the boundary between adjacent pixel values would be susceptible to being visible (as “contouring” or “banding”). At pixel value 20, the Weber contrast would be 1.05, high enough that visible artifacts would be likely.

Figure 3 plots L^* as a function of code value for linear-light coding; for the 1.8-power coding typical of graphics arts (e.g., Macintosh prior to Mac OS X version 10.6); and for pure power functions having exponents of 2.2 (sRGB), 2.4 (studio video), and 2.6 (digital cinema, to be discussed). EOCF power function exponents of 2.2, 2.4, and 2.6 are all quite perceptually uniform, evidenced by their straight-line behaviour over most of the range of pixel values in the figure.

It is frequently claimed that 8-bit imaging has a “dynamic range” of 255:1 (or 256:1). To pick five of many, many examples in the literature:

“An 8-bit image has a dynamic range of around 8 stops.” [8]

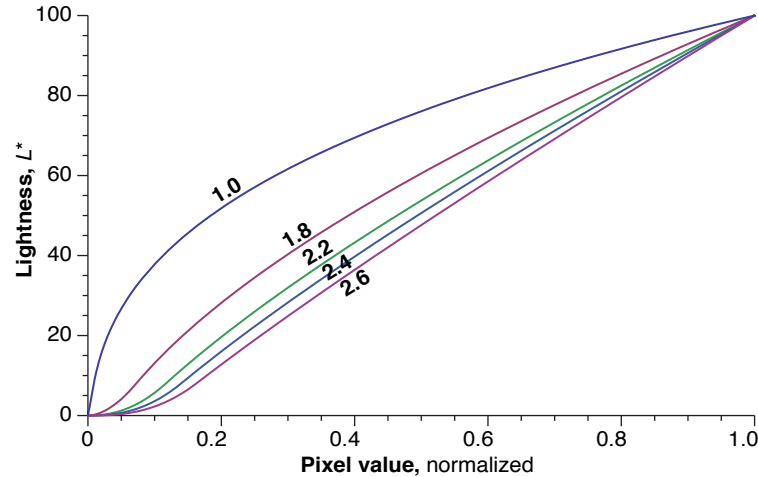


Figure 3: **CIE lightness (L^*) values** are plotted as functions of pixel values, for several pure power function EOCFs. Each curve is labelled by its power function exponent (“gamma”). Linear-light coding (exponent 1.0) exhibits poor perceptual uniformity: Above L^* value of 80, at least one bit is wasted compared to the other codes; below L^* 40, at least one additional bit would be necessary to achieve visual performance comparable to the other codes. The 1.8-power typical of graphics arts images exhibits good perceptual uniformity. Powers of 2.2 (sRGB), 2.4 (broadcast video and HD) and 2.6 (digital cinema) all exhibit excellent perceptual uniformity; the higher the power, the better the performance in very dark tones.

“Most of the images made for display on contemporary monitors have a dynamic range of only 256:1 per color channel, because that’s all that most monitors are built to support.” [29]

“A typical JPEG, TIFF, BMP image has 8 bits per color or a maximum dynamic range of 256 per color channel (256:1).” [4]

“A graphic image file with 8-bits signal depth in each channel has a dynamic range of 255:1, corresponding to a maximum density of 2.4.” [27]

“A range of 256 brightness steps is not adequate to cover a typical range from 0 to greater than 3 in optical density with useful precision, because at the dark end of the range, 1 part in 256 represents a very large step in optical density.” [43, p. 28]

Such claims arise from the implicit assumption that image data codes (pixel component values) are linearly related to light. For commercial imaging systems, that assumption is nearly always false: 8-bit image data is almost universally coded assuming that a 2.2- (sRGB) or 2.4-power (BT.1886) function will be applied at the display. Consequently, the dynamic range associated with code 1 is not 255:1 or 256:1, but about 200 000:1:

$$\left(\frac{1}{255}\right)^{2.2} \approx 0.000\,005 \approx \frac{1}{200\,000}$$

In the fourth quoted statement, 2.4 is the optical density corresponding to $1/255$. In the fifth statement, optical density of 3.0 corresponds to 1000:1 contrast ratio, typical of very high quality displayed imagery.

Covering a range of 3.0 in optical density with 8-bit coding using pure logarithmic pixel values yields a Weber contrast of $10^{3/255}$, about 1.027, and 85 pixel values per decade (or 25.5 pixel values per stop). Contrary to the fifth author's claim, quantizing a 3.0 density unit range into an 8-bit pixel value using pure log coding offers performance comparable to 8-bit coding of L^* .

Another aspect of claims commonly found in the literature, implicit in all quoted statements above, is that code 0 is disregarded – for no legitimate reason. In a simplistic, idealized system, you could take code 0 to produce luminance of zero, in which case the ratio of maximum to minimum luminance – the dynamic range – is infinity. In practice, physical factors cause minimum luminance to be greater than zero. The actual minimum luminance is an important aspect of the visual experience. If dynamic range is to be used to characterize the visual experience, it must be a ratio between physical quantities, not necessarily – and not usually – a ratio of image data values.⁶

10 Comparing 2.2- and 2.4-power EOCF to CIE L^*

As mentioned earlier, ΔL^* of unity is widely agreed to approximate the visual threshold between luminance levels. The ratio of luminance between L^* values of 99 and 100 is about 1.025 – that is, the relative luminance difference at threshold is 2.5% (the Weber fraction):

$$0.975 \approx L^{*[-1]}(99)$$

As relative luminance decreases, the luminance ratio between adjacent L^* values increases, as shown in Figure 4. At L^* of 8 (relative luminance just less than 0.01) the relative luminance ratio has reached 1.125, that is, a Weber fraction of 12.5%.⁷ The L^* scale assigns 92 levels – or 93, including the end-points – across a 100:1 range of relative luminance. Assuming that the visual threshold is 1 ΔL^* , seven bits suffice to encode L^* values.

Eight-bit sRGB has 255 steps, and assumes a 2.2-power EOCF [19]. Eight-bit digital studio video has 219 steps between black and white, and is standardized with a 2.4-power function at display [24]. These counts of possible integer pixel values, 255 and 219, are intermediate between the 462 codes required for pure log coding at a Weber contrast of 1.01 and the 92 codes of direct L^* coding. In Photoshop LAB coding [3], and in the LAB PCS of the ICC standard [21], L^* values are scaled by 2.55 for encoding into the range 0 through 255; that coding has 2.55 digital code values per L^* unit – that is, a Weber fraction of about 1% at white.

We have discussed the number of codes across 100:1 contrast ratio, or two decades of luminance. A particular imaging application may require a range less than or greater than 100:1. Also, typical photographic images have a certain amount of noise; visibility of contouring will be reduced by noise, and quantization will be less demanding.

For an example of perceptually uniform decoding in a different domain – that of medical imaging – see the DICOM standard [2].

⁶As a thought experiment, consider linear-light 8-bit greyscale imaging with pixel values from 1 to 220 driving a display having black at $1 \text{ cd} \cdot \text{m}^{-2}$ and white at $220 \text{ cd} \cdot \text{m}^{-2}$. Contrast ratio, or dynamic range, is 220:1. Now, modify the graphics subsystem driving the display to add an offset of +15. The offset values appear in image files, and appear across the hardware interface to the display; pixel values now range 16 to 235. Modify the display internals to subtract 15 from the incoming data values. Has the dynamic range now dropped from 220:1 to $\frac{235}{16}$, or about 15:1? Or is it unchanged?

⁷Take the first derivative of the inverse of L^* and divide by the inverse of L^* ; add one to get the Weber contrast.

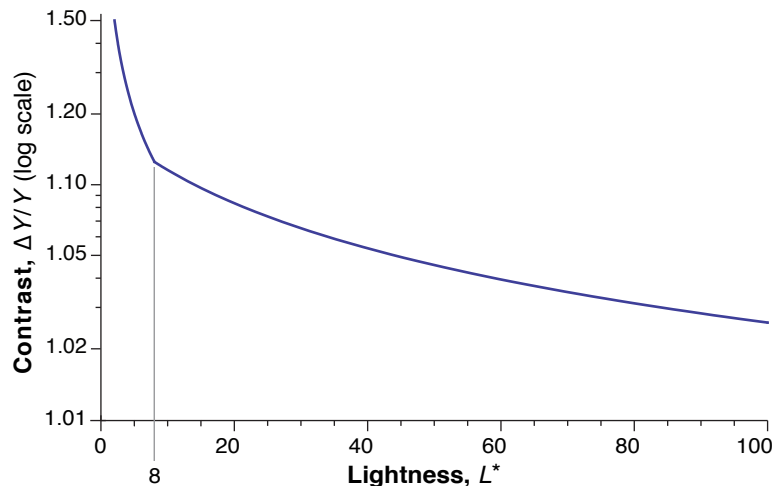


Figure 4: **Ratio of relative luminance values** for unit ΔL^* , across the L^* range from 1 to 100. Starting at the right, between L^* values 99 and 100, there is a 1.025 ratio (2.5%) between relative luminance values at the assumed threshold of unity ΔL^* . As L^* decreases, the “delta” increases. At L^* of 8 – corresponding to relative luminance of about 1%, or contrast ratio of about 100:1 – the difference has increased to 12.5%. As relative luminance decreases, image coding can get away with fewer and fewer pixel values per stop without the differences being visible.

11 Modern misconceptions

The NTSC monochrome (greyscale) television system was designed in the 1940s; its extension to colour was designed in the late 1940s and early 1950s. It is clear from the publications of the time [14, 26, 28, 31, 35] that the designers of those systems understood the importance of nonlinear coding to achieve good visual performance (although they did not give the concept the name *perceptual uniformity*).

Astonishingly, since about 1960 to the present, the significance of perceptual uniformity has been largely forgotten. Engineers are always desirous of linearity; video engineers apparently came to believe that the purpose of gamma correction was to overcome a supposed deficiency – that is, nonlinearity of the CRT. They realized that the sensible place to perform the “correction” was close to the transmitter, so as to avoid millions of nonlinear circuits in receivers; however, the link to perceptual uniformity was apparently forgotten. Widespread misunderstanding among television engineers of the fundamental reason for “gamma correction” remains rampant even today. As Poynton states [40, p. 316]:

If gamma correction were not already necessary for physical reasons at the CRT, we would have to invent it for perceptual reasons.

You can test your colleagues: Ask, “If television displays in 1953 had exhibited a linear relationship between voltage applied to the CRT and light output, would television standards have included gamma correction?” Anyone who answers “Of course not!” does not appreciate the importance of perceptual uniformity. It is clear from their writings that the historical authors cited earlier appreciated the perceptual advantage of the CRT’s characteristic.

Electrical engineers, video engineers, and digital image processing practitioners often claim that their systems are “linear.” However, “linearity” for engineering purposes just means that the properties $f(x) + f(y) = f(x + y)$ and $f(a \cdot x) = a \cdot f(x)$ are reasonably well approximated, completely independent of any potential link to optical or electrical power. If gamma correction has been imposed at image

capture or encoding, and an approximate inverse is imposed at decoding or display, then linearity in the R' , G' , and B' signal domain does not extend to luminance or tristimulus values. You can treat calculations in the tristimulus domain as linear; you can treat calculations in the $R'G'B'$ (video voltage, signal, or code) domain as linear. However, values in one domain are clearly not proportional to values in the other.

Poynton [38] reviewed several widely-held misconceptions concerning gamma, including these:

- The nonlinearity of a CRT display is a defect that needs to be corrected.
- The main purpose of gamma correction is to precompensate the nonlinearity of the CRT.
- Ideally, linear-intensity representations should be used to represent image data.

He then presented what he considered to be the facts of the situation:

- The nonlinearity of a CRT is very nearly the inverse of the lightness sensitivity of human vision. The nonlinearity causes a CRT's response to be roughly perceptually uniform. Far from being a defect, this feature is highly desirable.
- The main purpose of gamma correction in video, desktop graphics, prepress, JPEG, and MPEG is to code luminance or tristimulus estimates (proportional to intensity) into a perceptually-uniform domain, so as optimize perceptual performance of a limited number of bits (such as 8 or 10) in each of the RGB colour components.
- If a quantity proportional to intensity represents image data, then 12 bits or more would be necessary in each component to achieve high-quality image reproduction. With nonlinear (gamma-corrected) coding, just 8 bits usually suffice.

Poynton [38] referred to 8 bits per component being sufficient for video distribution purposes. In order to provide some measure of protection against roundoff error liable to be introduced by video processing, today's studio video standards – and most studio equipment devices – have 10 bits per component. CCD and CMOS sensors used in modern cameras are intrinsically linear-light devices;⁸ It is necessary to capture about 12 bits per linear-light component to maintain 10-bit accuracy once the signals are processed through picture rendering and gamma correction. Several digital cinema cameras offer 14-bit linear-light analog-to-digital converters, and thereby offer about 12 bits of quantization performance when coded perceptually (for example, by the $XYZ^{1/2.6}$ function specified in SMPTE/DCI standards for digital cinema [47]). Roughly speaking, representing colour components in a perceptually uniform manner saves 2, 3, or 4 bits per component compared to representation in linear-light form.

12 Gamma correction

We have mentioned scientific and industrial camera systems where an *opto-electronic conversion function* (OECF) – or loosely, “gamma correction” – is imposed at encoding, often immediately following transduction in the sensor. Gamma correction takes R , G , and B (radiometrically linear) tristimulus estimates from the scene, and forms (nonlinear) R' , G' , and B' quantities to represent those tristimulus values in a smaller number of bits. The primes signify the nonlinear relationship to light power. To achieve perceptual uniformity, the OECF roughly approximates vision's lightness sensitivity by imposing a function

⁸There are specialized sensors, typically used in surveillance cameras, that have nonlinear pixel circuitry – for example, logarithmic pixels – that enable high dynamic range capture. Such sensors have disadvantages for general purpose imaging, and they are not widely used.

comparable to L^* . Decoding and display of digital image data involves an EOCF that approximates the inverse of lightness sensitivity for each of the R , G , and B components.

Gamma correction is often described as a pair of inverse functions; for example, see Figure 5, taken from Wikipedia [51]. However, using an OECF at the camera that is the inverse of the display's EOCF fails to incorporate any picture rendering. If the scene is captured at light levels comparable to the ultimate display – say, with $80 \text{ cd} \cdot \text{m}^{-2}$ reference white – the resulting imagery may be visually correct. However, typical acquisition takes place in environments where a diffuse white reflector exhibits absolute luminance much higher than $80 \text{ cd} \cdot \text{m}^{-2}$, and failure to impose picture rendering is highly likely to produce displayed images that are judged as unsatisfactory.

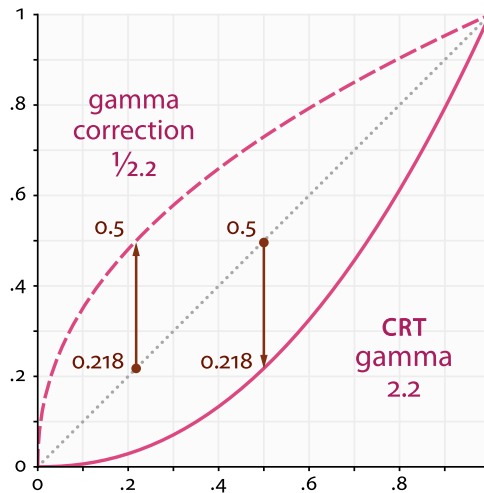


Figure 5: **Gamma function** as presented by Wikipedia shows a 2.2-power function, standardized for the EOCF of an sRGB display. The graph also shows the mathematical inverse, approximately a 0.45-power function. The suggestion is implicit that “gamma correction” should mathematically invert the display’s EOCF. When used with a 2.2-power EOCF, a 0.45-power OECF would not apply any picture rendering.

13 Modern practice in video and HD

As mentioned earlier, the ITU-R standardized the gamma value 2.4 [24]; this value is representative of today’s studio reference displays. Reference white luminance in HD is not standardized, but is typically between 80 and $120 \text{ cd} \cdot \text{m}^{-2}$. Contrast ratio is typically 800:1 or better. At program mastering, HD displays are viewed with a very dim surround, illuminated such that the surround luminance is about 1% of the reference white luminance.

Creative approval of program material in the studio environment causes not only the studio EOCF but also the studio viewing conditions to be implicit in the definition of the $R'G'B'$ exchange standard: It is implicit that the intended picture appearance at the consumers’ premises is obtained from a comparable EOCF in a comparable environment. Should a consumer’s display characteristics or viewing condition differ substantially from the studio – for example, if the consumer display is brighter, or has inferior contrast ratio, or is located in a lighter or darker surround than the studio – then image data should be altered at the consumer’s premises to yield a closer match to the intended appearance.

14 Modern practice in sRGB for desktop computing

The sRGB standard specifies an EOCF that is a pure 2.4-power function. Reference white luminance is specified at $80 \text{ cd} \cdot \text{m}^{-2}$, typical of a historical CRT. The sRGB standard specifies veiling glare of $0.2 \text{ cd} \cdot \text{m}^{-2}$ – that is, tristimuli equivalent to 0.0025 times reference white are expected to be added to the *RGB* tristimuli produced at the display surface. Contrast ratio is thereby expected to be 400:1. The veiling glare value in sRGB is realistic for LCD displays: The standard specifies ambient illuminance of 64 lx; a screen having 4% diffuse reflectance – a typical value for a modern LCD – would produce the $0.2 \text{ cd} \cdot \text{m}^{-2}$ veiling glare specified for sRGB.

Modern LCD displays typically deliver peak white luminance of 240 or 320 $\text{cd} \cdot \text{m}^{-2}$, three or four times sRGB’s reference luminance. A modern commodity LCD display can exhibit a contrast ratio as high as 800:1, twice times the value called for in the sRGB standard.

15 Modern practice in digital cinema

Standards for digital cinema mastering [47, 48] call for *R'G'B'* or *X'Y'Z'* components (at the reference projector interface, or the digital cinema distribution interface, respectively) to be raised to the power 2.6 for display. The 2.6-power is imposed to invert perceptually uniform encoding. Compared to the 2.4-power EOCF of studio video, the 2.6-power offers improved visual performance in the low luminance and dark surround situation of the cinema [9].

Digital cinema standards are completely display referred. There are no SMPTE/DCI standards for digital cinema acquisition; many techniques are in use. Acquisition is commonly accomplished using various quasilog codes; for a representative example – FilmStream – see SMPTE RDD 2 [45].

In classic photochemical film historically used in cinema, a rough approximation to picture rendering is evidenced by the standard *laboratory aim density* (LAD) practice [41] used in motion picture film laboratories: Relative luminance in the scene of about 0.18 produces optical density of about 1.06 in the print, and thereby produces relative luminance on-screen of $10^{-1.06}$, or about 0.087. Approximating the end-to-end function as a pure power function, the resulting effective end-to-end power function exponent is about 1.4:

$$0.18^g = 10^{-1.06} \approx 0.087; \quad g = \frac{-1.06}{\log 0.18} \approx 1.4$$

As a second point of reference, camera negative film stock has a film gamma of roughly 0.6, and print film stock has a film gamma of roughly 2.5; the product of these, 1.5, is a crude approximation of the exponent of the effective end-to-end power.

The end-to-end exponent of 1.4 or 1.5 for cinema is larger than the 1.2 typical of HD because the cinema display is darker (having a reference white of $48 \text{ cd} \cdot \text{m}^{-2}$ compared to $100 \text{ cd} \cdot \text{m}^{-2}$ of HD), and because the surround is black (0% in cinema) as opposed to very dim (1% in studio HD mastering).

This paper mainly concerns quantization of each colour component into a fairly small number of bits – say 8 or 10. Where that constraint is lifted, for example where 16 bits are available per component, then perceptual uniformity is still useful, but the ratios of luminance or tristimulus values between codes are lower than the visual threshold (even if components are coded in radiometrically linear, “linear-light” manner). In digital cinema acquisition and processing, OpenEXR [25] coding is fairly widely used. That coding uses 16 bits per component; pixel component values are represented in binary floating point, with a sign bit, five base-2 exponent bits, and ten fraction bits. The floating point encoding imposes a fixed Weber contrast (of about 0.1%, 1024 codes values per stop) over nearly the entire range of coded values.

Such coding is today most commonly used for scene-referred image data [1], but when used for display-referred data the logarithmic coding imposes a high degree of perceptual uniformity across a dynamic range up to 2^{30} .

16 Conclusion

Perceptual uniformity is a tremendously important aspect of digital image coding, particularly in video, HD, digital cinema, medical imaging, digital still photography, and desktop computer graphics. Without it, 10, 11, or 12 bits per component would be necessary, instead of the 8 bits common in consumer equipment or the 10 bits common in professional video and HD. If L^* is taken as a model for perceptual uniformity, gamma of 2.2 or 2.4 yields a remarkably good match. Perceptual uniformity was appreciated half a century ago, yet is either poorly understood or not recognized at all by a surprisingly large number of image scientists and engineers working today.

References

- [1] Academy of Motion Picture Arts and Sciences (AMPAS). Academy color encoding specification (ACES). S-2008-001, Aug. 12 2008. Version 1.0.
- [2] ACR/NEMA PS 3.14. Digital imaging and communications in medicine (DICOM) part 14: Grayscale standard display function, 2011.
- [3] Adobe Systems Inc. Adobe Photoshop TIFF technical notes, March 22 2002.
- [4] Aliaga, Daniel G. *Capturing and Rendering Real-World Scenes*, lecture notes for Purdue Univ. CS635-07. <http://www.cs.purdue.edu/homes/aliaga/cs635-07/lec-capture-devices.pdf>. Accessed, 2009-10-25.
- [5] Brill, Michael H. Do tristimulus values have units? *Color Research and Application*, 21 (4): 310–313, Aug. 1996.
- [6] Bureau International des Poids et Mesures (BIPM). The international system of units SI, 2006. Eighth edition.
- [7] CIE 15. Colorimetry, 3rd Edition, 2004.
- [8] Corke, Peter. *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*. Springer, Berlin, 2011.
- [9] Cowan, Matthew, Kennel, Glenn, Maier, Thomas O., and Walker, Brad. Contrast sensitivity experiment to determine the bit depth for digital cinema. *SMPTE Motion Imaging Journal*, 113 (9): 281–292, Sep. 2004.
- [10] Fairchild, Mark D. *Color Appearance Models*. Wiley, Chichester, U.K., second edition, 2005.
- [11] Fink, Donald G. *Color Television Standards*. McGraw-Hill, New York, 1955.
- [12] Giorgianni, Edward J. and Madden, Thomas E. *Digital Color Management: Encoding Solutions*. Wiley, Chichester, U.K., second edition, 2008.
- [13] Green, Phil J. *Color Management: Understanding and Using ICC Profiles*. Wiley, 2010.
- [14] Hazeltine Corporation. *Principles of Color Television*. Wiley, New York, 1956. By the Hazeltine Laboratories staff, compiled and edited by Knox McIlwain and Charles E. Dean.
- [15] Hecht, Selig. The visual discrimination of intensity and the Weber-Fechner law. *J. Gen. Physiol.*, 7 (2): 235–267, 1924.
- [16] Holm, Jack M., Tastl, Ingeborg, Hanlon, Larry, and Hubel, Paul. Color processing for digital photography. In Green, Phil and MacDonald, Lindsay W., editors, *Colour Engineering: Achieving Device Independent Colour*, chapter 9, pages 179–217. Wiley, 2002.
- [17] Hunt, Robert W.G. The heights of the CIE colour-matching functions. *Color Research & Application*, 22 (5): 335–335, Dec. 1997.
- [18] Hunt, Robert W.G. *The Reproduction of Color*. Wiley-IS&T Series in Imaging Science and Technology. Wiley, Chichester, U.K., sixth edition, 2004.

- [19] IEC 61966-2-1. Multimedia systems and equipment – Colour measurement and management – Part 2-1: Colour management – Default RGB colour space – sRGB, 1999.
- [20] International Color Consortium. Perceptual rendering intent use case issues. White Paper 2, Jan 2005.
- [21] ISO 15076. Image technology colour management – Architecture, profile format and data structure – Part 1: Based on ICC.1:2004-10, 2005. (Also published as ICC Specification ICC.1:2004-10).
- [22] ISO 22028. Photography and graphic technology – Extended colour encodings for digital image storage, manipulation and interchange – Part 1: Architecture and requirements, 2004.
- [23] ISO/IEC 14496-2:2004/Amd 3:2007. Coding of Audio-visual objects – Part 2: Visual, Amendment 3: Support for colour spaces, 2007.
- [24] ITU-R Rec. BT.1886. Reference electro-optical transfer function for flat panel displays used in HDTV studio production, Mar 2011.
- [25] Kainz, Florian. The OpenEXR image file format. In Fernando, Randima, editor, *GPU Gems: Programming Techniques, Tips and Tricks for Real-Time Graphics*, chapter 26. Addison-Wesley, 2004.
- [26] Kallmann, Heinz E. The gradation of television pictures. *Proc. IRE*, 28 (4): 170–174, Apr. 1940.
- [27] Kim, Min H. and MacDonald, Lindsay W. Rendering high dynamic range images. In *Proc. 17th Annual EVA Conference (EVA 2006)*, pages 22.1–22.11, Middlesex, England, July 2006. EVA Conferences International (ECI).
- [28] Maloff, Ioury G. Gamma and range in television. *RCA Review*, 3 (4): 409–417, Apr. 1939. Previously published by RCA Manufacturing Company in 1938 as a like-named 13-page monograph.
- [29] Mather, Jeff. Ask Dr. Color’s assistant: Tone-mapping in MATLAB. <http://jeffmatherphotography.com/dispatches/2007/09/ask-dr-colors-assistant-tone-mapping-in-matlab>, 2007. Accessed, 2009-10-25.
- [30] McCann, John J. and Stokes, Michael. Color spaces and image quality. In *PICS 1998: Image Processing, Image Quality, Image Capture, Systems Conference*, pages 140–144, May 1998.
- [31] Mertz, Pierre. Perception of television random noise. *J. SMPTE*, 54 (1): 8–34, Jan. 1950.
- [32] Munsell, Albert E. O., Sloan, L. L., and Godlove, I. H. Neutral value scales. I. Munsell neutral value scale. *J. Opt. Soc. Am.*, 23 (11): 394–402, 1933.
- [33] Munsell, Albert H. *Atlas of the Munsell Color System*. Wadsworth, Howland & Co., Inc., Malden, Mass. (Boston), 1907.
- [34] Myszkowski, Karol, Mantiuk, Rafał, and Krawczyk, Grzegorz. *High Dynamic Range Video*. Morgan & Claypool Publishers, 2008.
- [35] Oliver, Bernard M. Tone rendition in television. *Proc. IRE*, 38 (11): 1288–1300, Nov. 1950.
- [36] Pauli, Hartmut K. A. Proposed extension of the CIE recommendation on “uniform color spaces, color difference equations, and metric color terms”. *J. Opt. Soc. Am.*, 66 (8): 866–867, 1976.
- [37] Poynton, Charles. *YUV and luminance considered harmful*. http://www.poynton.com/papers/YUV_and_luminance_harmful.html. Accessed, 2009-10-25.
- [38] Poynton, Charles. The rehabilitation of gamma. In *Proc. SPIE, Human Vision and Electronic Imaging III*, volume 3299, pages 232–249, Bellingham, Wash., 1998. SPIE.
- [39] Poynton, Charles. *Digital Video and HDTV Algorithms and Interfaces*. Morgan Kaufmann, San Francisco, 2003.
- [40] Poynton, Charles. *Digital Video and HD Algorithms and Interfaces*. Elsevier/Morgan Kaufmann, Burlington, Mass., second edition, 2012.
- [41] Pytlak, John P. and Fleischer, Alfred W. A simplified motion-picture laboratory control method for improved color duplication. *SMPTE Journal*, 85 (10): 781–786, Oct. 1976.
- [42] Richter, Klaus. Cube-root color spaces and chromatic adaptation. *Color Research & Application*, 5 (1): 25–43, Apr. 1980.
- [43] Russ, John C. *The Image Processing Handbook*. CRC Press, Taylor & Francis Group, Boca Raton, Fl., fifth edition, 2006.
- [44] Schäfer, Ralf and Kauff, Peter. HDTV colorimetry and gamma considering the visibility of noise and quantization errors. *SMPTE J.*, 96 (9): 822–833, 1987.

- [45] SMPTE RDD-2. Use of Logarithmic Non-Linear Transfer Characteristic for Transmission of Linear Signals through Limited-Bit-Depth Image Representations and Interfaces, 2007.
- [46] SMPTE RP 177. Derivation of Basic Television Color Equations, 1993, reaffirmed 2002, archived 2010.
- [47] SMPTE ST 431-1. D-Cinema Quality – Screen Luminance Level, Chromaticity and Uniformity, 2006.
- [48] SMPTE ST 431-2. Reference projector and environment for display of DCDM in review rooms and theaters, 2006.
- [49] Stevens, Stanley S. To honor Fechner and repeal his law: A power function, not a log function, describes the operating characteristic of a sensory system. *Science*, 133 (13): 80–86, Jan. 1961.
- [50] Stokes, Michael, Anderson, Matthew, Chandrasekar, Srinivasan, and Motta, Ricardo. Proposal for a standard default color space for the internet – sRGB. In *Proc. IS&T/SID Fourth Color Imaging Conference*, pages 238–246, Nov. 1996.
- [51] Wikipedia. Gamma correction. http://en.wikipedia.org/wiki/Gamma_correction. Accessed, 2012-03-02.