Charles Poynton
156 Bartlett Avenue
Toronto, ON M6H 3G1
CANADA
tel +1 416 535 7187
charles @ poynton.com
www.poynton.com

2010-06-16

Dear Colleagues –

I continue to be flummoxed by the absence of any viable, modern, realistic standard for the "gamma" – or properly, *electro-optical conversion function* (EOCF) – of studio reference displays for video and HDTV. With the demise of CRTs and with the introduction of reference-grade flat panel monitors, I believe that such a standard is now critically important. The first document attached here, *Picture rendering, image state, and BT.709*, is essentially a plea to standardize an EOCF that reflects current practice.

Poynton, Charles (2003), *Digital video and HDTV algorithms and interfaces* (San Francisco: Morgan Kaufmann).

Standardization of EOCF involves two intertwined topics, perceptual uniformity and picture rendering. During the last decade or so, I have investigated these issues as they relate to video, desktop graphics, consumer still photography, and digital cinema. My book, cited in the margin, documents these issues. In addition, I attach two recent survey documents, *Perceptual uniformity in digital imaging* and *Picture rendering in video* that summarize my conclusions about those topics. I also attach a fourth document, *Review of perceptual uniformity and picture rendering in video*. Any or all of the last three documents may eventually become conference presentations or journal papers.

Work to establish new standards is taking place within SMPTE, EBU, ARIB, and ITU-R. I welcome comments, corrections, additional information, and opinions.

Thanks,

Charles Poynton

Charles Poynton
tel +1 416 535 7187
charles @ poynton.com
www.poynton.com

Picture rendering, image state, and BT.709

The creators of video programs approve their work on studio reference displays. They expect to have their work displayed in the consumers' premises in a reasonable approximation of what they approve. To meet this goal, the reference display's conversion of *R'G'B'* signals to light must be approximated at the consumers' premises.

I use the term *electro-optical conversion function,* EOCF, in accordance with digital still camera terminology as exemplified in ISO 14524. What I call EOCF is commonly called *electro-optical transfer function*, EOTF, in video. What I call OECF is commonly called *opto-electronic transfer function*, OETF, in video.

Current HD video standards, including BT.709 and its various descendants such as SMPTE 274M, specify the camera's reference encoding – the opto-electronic conversion function (OECF). However, surprisingly, the electro-optical conversion function (EOCF, or "gamma") of studio reference displays has never been adequately standardized. Without standardization of the EOCF, creative intent cannot be reproduced with any certainty. Absent a studio standard, the consumer electronics (CE) industry has no reference. Without clear standards determining intended image appearance, the CE industry is effectively encouraged to interpret – even "enhance" – consumer imagery according to the judgements of its engineers and managers, often overriding the cinematographers' artistic intent.

ITU-R Rec. BT.709, *Parameter values for the HDTV standard for the studio and for international programme exchange*.

SMPTE 274M, *1920×1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates.*

SMPTE 170M, *Composite Analog Video Signal – NTSC for Studio Applications.*

In addition to their failure to effectively standardize EOCF, current video standards are inconsistent with modern understanding of image state and rendering as represented by published work in the colour management community and standards promulgated by organizations such as ISO and ICC. Image state and rendering in video today are intertwined with the reference OECF of the camera and the *de facto* EOCF of the display. I propose that BT.709 and related video standards be respecified so as to be explicitly output-referred, thereby bringing these standards into line with modern practice.

Finally, to enable predictable mapping of *R'G'B'* values to colour appearance, studio viewing conditions need to be standardized.

Introduction

There is no effective standard for EOCF of a studio video reference display. How that situation came about is a complicated story summarized in the sections *Perceptual uniformity in image coding* and *Picture rendering* below. My argument to standardize studio reference EOCF hinges on the preservation of creative intent, discussed in the correspondingly named section below.

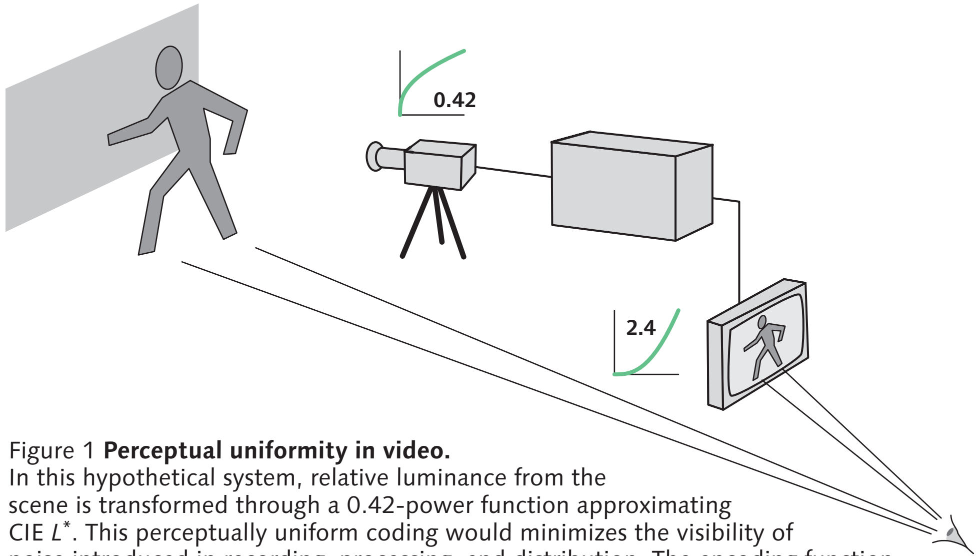


Figure 1 **Perceptual uniformity in video.**
In this hypothetical system, relative luminance from the scene is transformed through a 0.42-power function approximating CIE L^*. This perceptually uniform coding would minimizes the visibility of noise introduced in recording, processing, and distribution. The encoding function is inverted by a 2.4-power function at the display, thereby presenting the scene's relative luminance at the display: A hypothetical viewer could compare the two. The problem is that display and viewing conditions influence colour appearance, but the conditions rarely match between the scene and the reproduction. Faithfully presenting the *appearance* of the scene requires a nontrivial mapping of image data – *picture rendering*.

Perceptual uniformity in image coding

Poynton, Charles (2009), "Perceptual uniformity in Digital Imaging," in *Proc. Gjøvik Color Imaging Symposium* (GCIS 2009): 102–109.

An electronics engineer might expect image encoding and image decoding in video to be linear processes: The camera would produce an image signal proportional to intensity, and a display would produce intensity proportional to the image signal. However, (perceived) lightness is roughly the 0.42-power of (physical) intensity: 18% "mid grey" in the physical domain corresponds to about 50% on the video (or computer graphics) code scale. Compared to linear-light encoding, a dramatic improvement signal-to-noise performance can be obtained by using nonlinear image coding that mimics human lightness perception. Ideally, coding for distribution should be arranged such that each code step is proportional to a *just noticeable difference* (JND) in luminance. In practice, this situation is approximated in video systems. I discuss the details in my book cited in the margin. Virtually all commercial image systems incorporate perceptual coding.

Poynton, Charles (2003), *Digital video and HDTV algorithms and interfaces* (San Francisco: Morgan Kaufmann).

The L^* function was standardized in 1976 as the CIE's best estimate of the lightness sensitivity of human vision. Although its encoding equation incorporates a cube root, L^* is effectively a power function having an exponent of about 0.42 (as I describe in Chapter 20 of my book).

Emergent display technologies such as LCDs, PDPs, and DLPs don't have intrinsic 2.4-power functions. However, such displays incorporate approximately a 2.4-power function in their signal processing circuits.

The electrostatic characteristics of a CRT's electron gun cause a CRT to have nonlinear response from voltage to light – the EOCF. Since the earliest days of television, the display power exponent for studio video has been about 2.4, and this value remains representative of today's studio displays – even those using non-CRT technology.

Approximate inversion of the CRT's nonlinearity is accomplished by "gamma correction" at the camera: Encoding of video signals is thereby perceptually uniform. The situation is depicted in Figure 1.

Maloff, I.G. (1939), "Gamma and Range in Television," in *RCA Review* **3** (4): 409–417 (Apr.).

As early as 1939 – seventy years ago! – it was recognized that the EOCF of a CRT is reasonably close to the inverse of the lightness sensitivity of vision. The *de facto* 2.4-power function of today's reference studio displays almost perfectly inverts L^*. Consequently, "gamma correction" at the camera simultaneously performs two equally important tasks:

- Gamma correction encodes into a perceptually uniform space, so as to maximize perceptual performance from a limited number of bits per component; and
- Gamma correction precompensates for the nonlinearity of the CRT.

The second aspect of gamma correction is well understood by today's video engineers. The first aspect is not. As I argue in my survey document (cited above), perceptual uniformity was well understood in the 1950s when NTSC was standardized. However, the "undoing" of perceptual uniformity at a CRT display required no moving parts – in fact, it required no parts at all! Also, video engineers have historically not been well educated in aspects of perception and colour science. Perceptual uniformity was so unobtrusive, and worked so beautifully, that its primary justification was largely forgotten by video engineers.

Picture rendering

Poynton, Charles (2010), *Picture rendering in video* (unpublished).

Whether or not perceptually uniform coding is used, the engineer wishing for linearity encounters a surprise when image information is captured and displayed in different conditions: The environment in which images are viewed changes their appearance. My second survey paper summarizes the three main causes of appearance difference.

Perceptually correct reproduction is obtained by modifying image data, thereby altering the end-to-end relationship of scene luminance to reproduced luminance. In a greyscale system, a suitable correction can often be accomplished by arranging the system so that a gentle end-to-end power function acts upon relative scene luminance. Today, we are more interested in *colour* reproduction than in *grayscale* reproduction; a good starting point for the required colour image modification involves imposition, to each of the red, green, and blue tristimulus values, of a modest end-to-end power function. For a typical studio scene intended for display at about 100 $cd \cdot m^{-2}$ in a dim surround, a common "baseline" correction can be accomplished by using an end-to-end power function having an exponent of about 1.2. The power function increases contrast and colour saturation in the reproduced midtones, relative to the midtones of the scene.

Typical studio illumination is about 2000 lx.

The EOCF of today's studio displays closely approximates a 2.4-power function. The BT.709 reference OECF is essentially a 0.5-power function. Consequently, the end-to-end exponent implicit in BT.709 origination is 0.5 times 2.4, or 1.2. The perceptual significance of an end-to-end power of 1.2 is described in classic publications such as those from Bartleson and Breneman, Hunt, and DeMarsh that are cited in my survey paper. The situation is depicted in Figure 2.

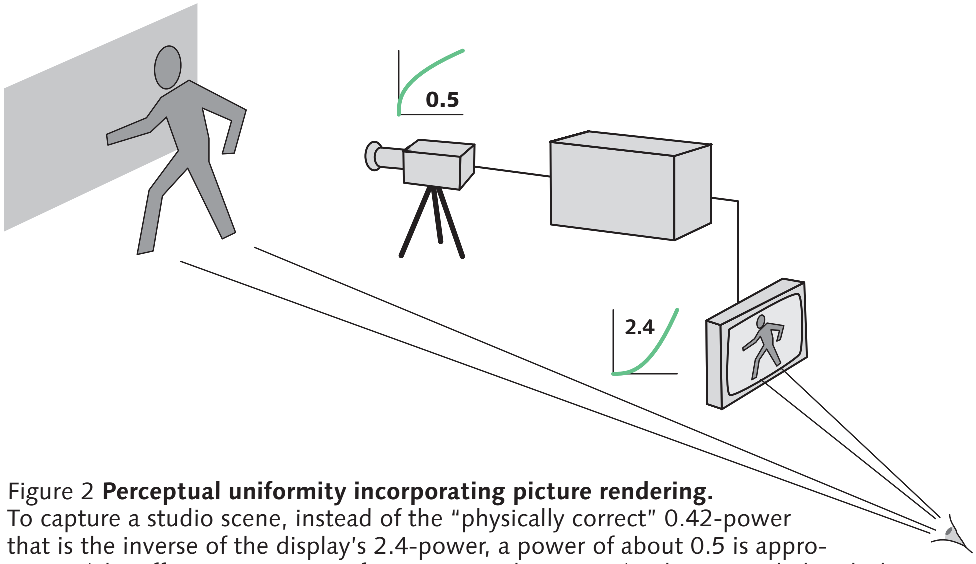


Figure 2 **Perceptual uniformity incorporating picture rendering.** To capture a studio scene, instead of the "physically correct" 0.42-power that is the inverse of the display's 2.4-power, a power of about 0.5 is appropriate. (The effective exponent of BT.709 encoding is 0.5.) When cascaded with the display's 2.4-power, and end-to-end exponent of 1.2 results. Reference display EOCF has never adequately been standardized; however, a power function having an exponent quite close to 2.4 is intrinsic in a CRT. In the absence of any viable EOCF standard, I believe that video engineers have come to consider the OECF to be paramount.

The electro-optical conversion function (EOCF) of a reference video display has never been adequately standardized. The *de facto* EOCF is a power function having an exponent between 2.3 and 2.4. Video engineers have come to consider the OECF to be the most important aspect of image coding. However, appearance of the reproduced image is utterly dependent upon the EOCF of the display. Secondarily, appearance depends upon viewing conditions at the display.

What I consider to the video engineer's view of the situation is depicted in Figure 3, below.

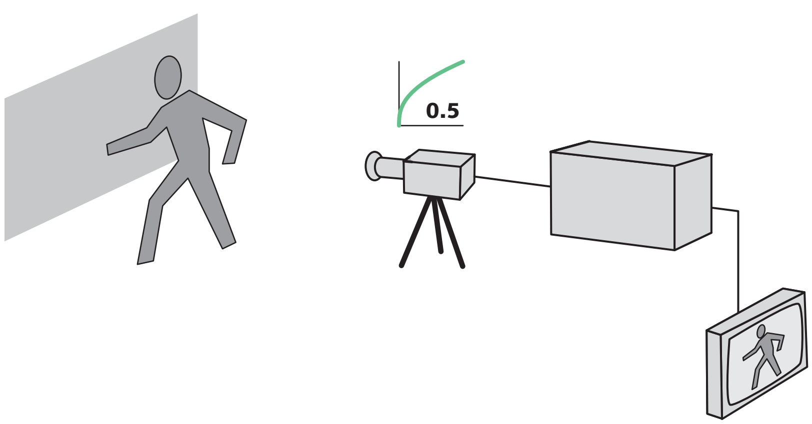


Figure 3 **Video engineer's view of BT.709.** No reference display EOCF is standardized by ITU, SMPTE, or other organizations, so I omit the EOCF curve from this sketch. (The *de facto* EOCF is a power function having an exponent between 2.3 and 2.4). I believe that because there is no standard EOCF, video engineers have come to consider the OECF – for example, that of BT.709 – to be the most important aspect of image coding for video.

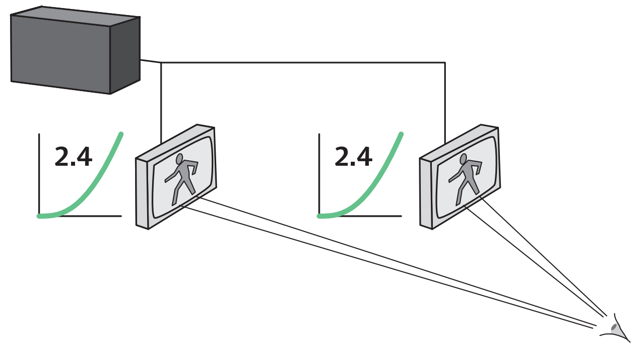


Figure 4 **Cinema engineer's view of image reproduction.** The cinematographer can originate the image on the studio reference display any way he or she wants: As far as the approval and distribution of the material is concerned, the entire production and postproduction chain can be considered to be a black box. What concerns the cinema engineer is that the image displayed and approved at the studio is presented faithfully at the consumers' premises.

Creative intent

The goal of video production is *not* to reproduce, at the viewer's premises, an accurate representation of the scene in front of the camera. Rather, the goal is to reproduce an accurate representation of what the director saw on his studio display upon approving the final product of post-production. Image data modifications are imposed for creative purposes at various stages of professional video production. Whatever image processing operations were used to create the final image – whether physically meaningful or not – are fair game. What I consider to be the cinematographer's goal is depicted in Figure 4.

In my survey paper, I describe the concept of image state: When picture rendering operations are interposed between capture and display, it becomes important to distinguish between image data representing *scene* tristimulus values, and image data representing intended *display* tristimulus values. Current video standards fail to make clear at which end of the system the standards apply: They fail to differentiate between *scene*-referred and *display*-referred image data. BT.709 needs to be recast in the framework of a display-referred image state.

BT.709 proposals

To achieve accurate representation of the director or cinematographer's visual experience, standardization of the reference display EOCF is necessary. I propose to codify current practice and standardize today's 2.4-power function as part of BT.709 and its derivatives. For creative purposes, there is no need to standardize OECF; however, retaining a reference OECF is sensible for engineering reasons.

I propose these improvements to BT.709 and its SMPTE and EBU derivatives. My recommendations essentially codify current practice:

1 Pertinent display characteristics and reference viewing conditions should be standardized. I propose that the studio reference display should have reference white luminance of 100 $cd \cdot m^{-2}$ at CIE D_{65}. Veiling glare should be specified at approximately 0.2% of reference white. The display should be viewed in a 50% diffuse neutral grey surround having 5% of the luminance of reference white. Standards groups should consider the manner in which viewing parameters have been specified in the sRGB standard, the opRGB (AdobeRGB) standard, and in ISO and ICC documents, and should consider discussions that have taken place within the colour management community.

2 EOCF of a studio reference display should be standardized based upon a 2.35-power function. (Other values such as 2.36 and 2.4 have been proposed; any value between 2.35 and 2.4 would serve.)

3 BT.709's current OECF should be retained as a reference for engineering purposes. BT.709 should make clear that its OECF is appropriate for studio scenes, and that modifications of the OECF for creative purposes – perhaps dramatic modifications – should be routinely expected. A statement is needed saying that encoding should be arranged such that the intended image appearance is obtained on the reference display in the reference viewing conditions.

4 Standards should discuss – or at a minimum, mention – image state as that term is used in the colour management community. In particular, BT.709 and its derivatives should be clarified to explain that the reference OECF included in the standard is meant to exemplify capture of a typical studio scene, and that the video signal (image data) is output (display) referred. ■

Perceptual uniformity in digital imaging

Charles Poynton
Simon Fraser University,
Vancouver, BC, Canada
charles@poynton.com

This paper is a lightly edited version of POYNTON, CHARLES (2009), "Perceptual uniformity in Digital Imaging," in *Proc. Gjøvik Color Imaging Symposium* (GCIS 2009): 102–109.

Abstract

Digital image coding is *perceptually uniform* if a small perturbation to a component value is approximately equally perceptible across the range of that value. Most digital image coding systems – including sRGB used in desktop graphics, and BT.709 used in HDTV – are perceptually uniform, but this fact is often shrouded in confusion. This document surveys perceptual uniformity in digital image coding and attempts to clarify some aspects of image coding that are widely misunderstood.

Luminance

Absolute luminance, defined by the CIE, is proportional to optical power across the visible wavelengths, weighted according to a standardized spectral weighting that approximates the spectral sensitivity of normal human vision. Luminance has units[1] of $cd \cdot m^{-2}$ ("nit," or nt); its symbol is L_V. The spectral weighting is denoted $V(\lambda)$ or $\bar{y}(\lambda)$.

The term *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, "luminosity" – to refer to a weighted sum of *nonlinear* (gamma corrected) red, green, and blue tristimulus signals instead of the linear-light quantities defined by the CIE [CIE 15]. The nonlinear quantity is properly termed *luma* and given the symbol *Y'* [Poynton 1999].

In image capture – including photography, cinema, video, HD, digital cinema, and graphics arts – we are rarely, if ever, concerned with the absolute luminance of the original scene. Instead, we characterize scene luminance relative to an "adopted" scene white luminance asso-

1 The foot-lambert unit [fL] once used for luminance is now deprecated. I use SI units, such as $cd \cdot m^{-2}$ ["nit," or nt], for light. In my view, using foot-based units such as foot-Lambert [fL] and foot·candle [fc] impedes the understanding of radiometry and photometry.

ciated with the state of visual adaptation of an actual or hypothetical person viewing the scene. Subsequent processing and display involves *relative luminance*, whose symbol is *Y,* and whose value according to CIE conventions is a pure number ranging 0 through 100. (Some practitioners, including me, prefer a range from 0 to 1.) Image scientists and engineers ordinarily call this quantity *luminance,* even though properly speaking it is relative luminance.

A set of three signals proportional to intensity, and having specific spectral weighting, are called *tristimulus values.* They are pure numbers with no units [Brill 1996]. *RGB, LMS,* and *XYZ* are all examples of tristimuli. A suitably-weighted sum of tristimuli yields luminance [Hunt 1997].

.Tristimulus values and luminance are what I call *linear-light* measures, directly proportional to light power. Cameras typically depart from the spectral sensitivities prescribed by CIE standards, so tristimulus values and luminance in video are usually estimated, not exact. Instead of using my 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. Practical cameras don't closely approximate the CIE spectral weighting, so the term *photometrically linear* shouldn't be used to describe them.

Introduction to perceptual uniformity

I introduce perceptual uniformity in Chapter 1 of my book [Poynton 2003]. 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.

The 1% value that I mention is the *Weber contrast.* Image coding whereby a constant ratio is maintained from code to code across the tone range from some minimum representable luminance up to white is effected by a logarithmic transform. Log transforms are rare in practical image coding.

For a true logarithmic law having a 1.01-ratio between adjacent codes, the relative luminance difference between codes is 1% across the whole range. There are 463 codes between relative luminance of 0.01 and 1 – that is, 463 codes cover a contrast ratio of 100:1. A photographer or cinematographer is interested in how many codes cover each "stop" (factor of two) of luminance. For pure logarithmic coding with a Weber contrast of 1%, there are 232 codes per decade, equivalent to 69 codes per stop – six bits of data per stop.

An estimate of vision's lightness response, denoted L^*, was standardized by the CIE in 1976 [CIE 15]: Given relative luminance, CIE L^* returns a value between 0 and 100; a "delta" (difference) of 1 lies approximately at the threshold of vision.[2] The L^* function is basically a power function with what I call an "advertised" exponent of $^1/_3$ – that is, a cube root. The technical literature is rife with statements that

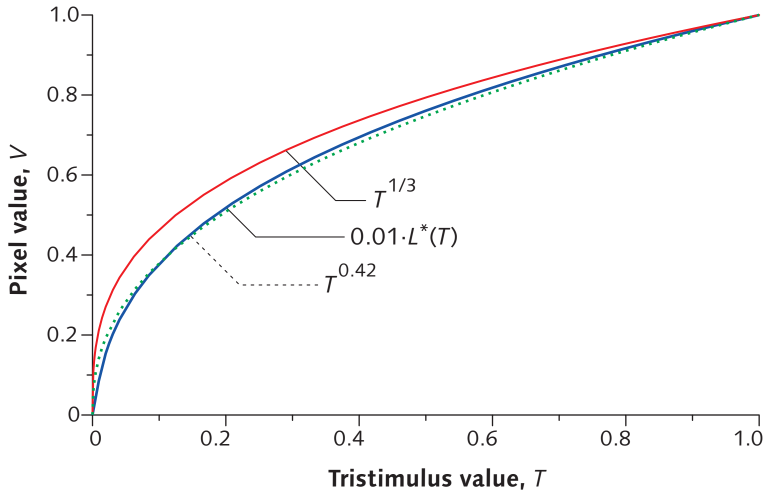


Figure 1 **CIE Lightness,** denoted L^*, estimates the perceptual response to light intensity (technically, relative luminance). Here L^*, scaled to the range 0 … 1, is overlaid by power function having an exponent 0.42, the exponent that best fits L^*. The L^* function involves a cube root – that is, a $^1/_3$-power function – but L^*'s power function is scaled and offset. I also overlay a cube root onto the plot: A pure cube root is a poor approximation to L^*.

L^* is a cube root. However, 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. The scaling and offset cause the function to approximate an "effective" 0.42-power over its entire range. See Figure 1.

In capturing, processing, storing, and transmitting image data, a limited number of bits are most effectively used by perception if coding of luminance values (or tristimulus values) is nonlinearly mapped, like L^*, to mimic the lightness response of human vision. Mappings based upon power functions are most common, though mappings based upon logarithms are sometimes used.

In nearly all commercial imaging systems, an *optoelectronic conversion function*[3] (OECF) – or loosely, "gamma correction" – is imposed at encoding. Gamma correction takes *R, G,* and *B* (linear) tristimulus estimates, and forms (nonlinear) *R', G',* and *B'.* The primes signify the nonlinear relationship to light power. To achieve perceptual uniformity, the OECF roughly approximates vision's lightness sensitivity (e.g., L^*). Decoding and display of digital image data involves an *electro-optical conversion function*[4] (EOCF) that approximates the inverse of lightness sensitivity.

2 Delta-L^* of 1 approximates a *just-noticeable difference* (JND), or equivalently, a just unnoticeable difference. L^* ranges 0 to 100, so it is implicit in the definition of L^* that vision can discriminate about 100 steps between black and white.

3 What I call OECF, in accordance with digital still camera terminology (as exemplified in ISO 14524) is commonly called *opto-electronic transfer function*, OETF, in video.

4 What I call EOCF is, in video, commonly called *electro-optical transfer function*, EOTF.

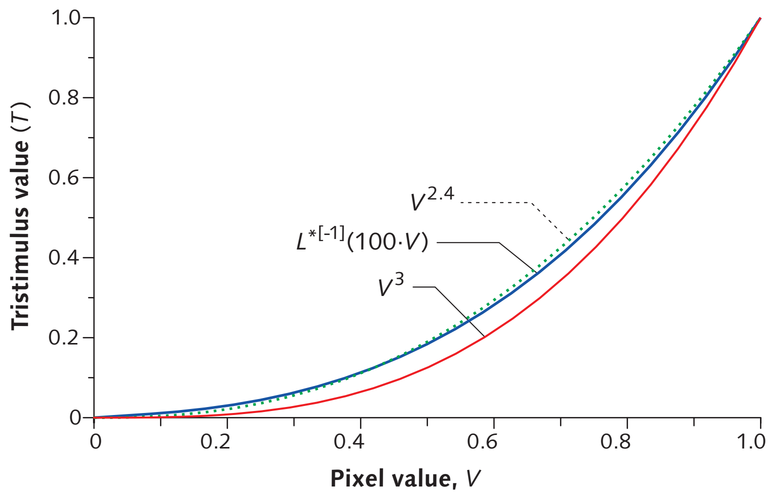


Figure 2 **EOCF of a typical CRT** is approximated by a 2.4-power function from video signal in to luminance. The *gamma* of a display system – for example, a CRT, or the reference sRGB EOCF – is the numerical value of the exponent of the power function. I overlay the inverse of the CIE L^* function: It is evident that a 2.4-power function is a very close match to the inverse of L^*. I also overlay a 3.0-power function; clearly, a cube function is a rather poor match to the inverse of L^*.

In a CRT display, the electrostatic characteristics of the electron gun cause the CRT to impose an EOCF that is approximately a 2.4-power function from voltage input to light output. The symbol γ (gamma) represents the exponent at the display: A studio reference display is said to have *gamma* of about 2.4. In non-CRT display devices, signal processing provides an equivalent nonlinear function. A 2.4 power is a near-perfect match to the inverse of the L^* function; see Figure 2.

It is frequently claimed that 8-bit imaging has a "dynamic range" of 255:1 or 256:1. Such claims arise from the assumption that image data codes are linearly related to light. However, nearly all 8-bit image data is coded perceptually, like sRGB, assuming a 2.2- or 2.4-power function at the display: The dynamic range associated with code 1 is close to a million to one, not just $^1/_{255}$. A related claim [Kim 2006] is that 8-bit imaging has an optical density range of about 2.4, where 2.4 is the base-10 log of $^1/_{255}$. This claim similarly rests upon the assumption of linear-light coding – an assumption which, for 8-bit coding, is nearly always false.

Figure 3 plots L^* as a function of code value for linear-light coding, a 1.8-power coding typical of graphics arts, and 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.

As I mentioned earlier, ΔL^* of unity is widely agreed to approximate the threshold of vision. 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 contrast*). The difference increases as relative luminance decreases; see Figure 4. At relative luminance of 0.01, L^* is about 8, and the relative luminance difference at threshold

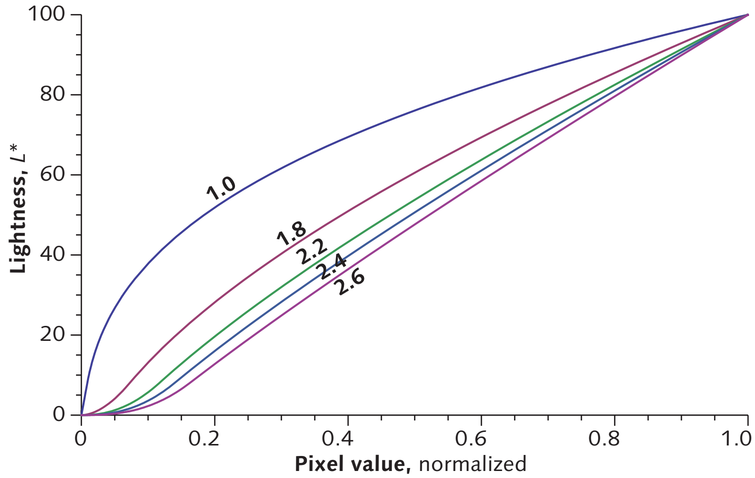


Figure 3 **Various pure power function EOCFs** are plotted as their CIE lightness (L^*) values against code values. The curves are labelled by exponent ("gamma"). Linear-light coding (exponent 1.0) exhibits poor perceptual uniformity below L^* value 60. The 1.8-power typical of graphics arts images exhibits good perceptual uniformity. Powers of 2.2 (sRGB), 2.4 (broadcast video and HDTV) and 2.6 (digital cinema) all exhibit excellent perceptual uniformity; the higher the power, the better the performance in very dark tones (as evidenced by the hockey-stick shape close to black).

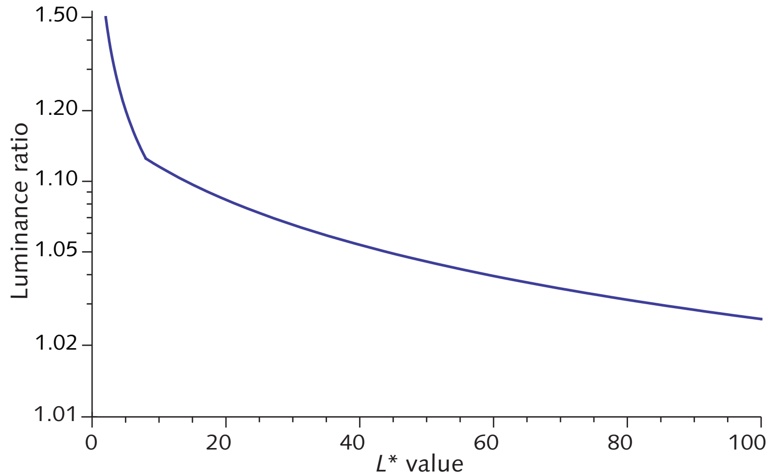


Figure 4 **Ratio of relative luminance values** for unit ΔL^*, across the L^* range from 1 to 100. Starting at the right, between L^* values 99 and 100, there is a 2.5% difference between relative luminance values at the assumed threshold of unity ΔL^*. As L^* decreases, the "delta" increases. At L^* of 8 – corresponding to relative luminance of about 1%, or contrast ratio of 100:1 – the difference has increased to 12.5%.

has reached 12.5%. The L^* scale assigns 92 levels – or 93, including the endpoints – across a 100:1 range of luminance. Seven bits suffice.

Coding L^* values produces considerably larger luminance ratios than logarithmic coding with a Weber contrast of 1%. Digital studio video has 219 steps over a comparable contrast ratio; sRGB has 255. These numbers are intermediate between the 463 codes of pure log coding (at a Weber fraction of 1.01) and the 92 codes of L^* coding. In Photoshop LAB coding, and in the LAB PCS of the ICC standard [ISO 15076], L^*'s range of 0 to 100 is coded digitally into the range 0 through 255: The coding has about 2.5 digital code values per L^* unit – that is, a Weber contrast of about 1% at white.

I have been discussing 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 this noise, and quantization will be less demanding.

If an imaging application were required to maintain relative luminance values from an encoder to a decoder, then the OECF (at encoding) should be chosen as the mathematical inverse of the EOCF that will be imposed at decoding and display. For the near-ideal 2.4 power used in studio video display, you would expect the encoder to have as its exponent the reciprocal of 2.4 – that is, 0.42. (For an example of perceptually uniform decoding in a different domain, medical imaging, see the DICOM standard [ACR/NEMA PS 3.14].)

Picture rendering

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. In most applications the goal is to not to match relative luminance values between the scene and the display, but to match 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 luminances between scene and display: Preserving appearance almost always requires manipulating the image data between the scene and display.

In many commercial imaging systems, including video and digital still photography, the intended appearance is often obtained by using an OECF that approximates a 0.5-power function, rather than the 0.42 that would perfectly invert a 2.4-power at decoding and display.

Perceptual encoding for distribution is performed in virtually all commercial image systems. In applications where image data is manipulated for creative purposes between capture and display – for example, in graphics arts, or in video post-production – perceptual uniformity is imposed at capture to the extent required for the image manipulation. Suppose that processing requires a linear-light gain of 4 to overcome poor lighting or incorrect exposure. Capture must then have quantization four times finer than the quantization required at the display. Where 8-bit *R′G′B′* components might suffice for distribution of consumer video or commodity JPEG imagery, to enable manipulation in post-production, 10 bit *R′G′B′* components might be required at capture. Video is typically processed in the camera to produce perceptually uniform signals; the recorded image data is quite close to the required final product, and not much processing headroom is needed. However, digital cinema capture typically involves downstream processing for creative purposes; more severe constraints are thereby placed on perceptual uniformity. Put simply, more bits per component are required.

Modern misconceptions

Astonishingly, since about 1960 to the present, the significance of perceptual uniformity has been largely forgotten! Engineers, always desirous of linearity, apparently came to believe that gamma correction was necessary 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 forgotten. Widespread misunderstanding among television engineers of the fundamental reason for "gamma correction" remains rampant even today. As I stated on page 258 of my book [Poynton 2003]:

> 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 applied voltage and light output, would television standards have included gamma correction?" Anyone who answers "Of course not!" does not, in my view, appreciate the importance of perceptual uniformity.

Electrical engineers, video engineers, and digital image processing practitioners often claim that their systems are "linear." However, 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! In other words, you can treat calculations in the tristimulus domain as linear, and you can treat calculations in the *R'G'B'* (video signal, code, or voltage) domain as linear, but values in one domain are clearly *not* proportional to values in the other.

In my paper "The rehabilitation of gamma" [Poynton 1998]
I 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 compensate the nonlinearity of the CRT.

- Ideally, linear-intensity representations should be used to represent image data.

My paper then presents what I consider 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 in each of the *RGB* 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 are sufficient.

In my 1998 paper, I 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 – have 10 bits per component. CCD and CMOS sensors used in cameras are intrinsically linear-light devices; it is necessary to capture at least 12 bits per component to maintain 10-bit accuracy once the signals are gamma-corrected [SMPTE 431-1]. Several digital cinema cameras offer 14 bit linear-light components, 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). 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.

Modern practice

Today's studio reference displays have gamma very close to 2.4, reference white luminance of between 80 and 120 $cd \cdot m^{-2}$, and a contrast ratio of about 250:1. They are viewed with a dim surround, illuminated such that the surround luminance is about 5% 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 the consumer's display characteristics or viewing conditions 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.

CRTs are now essentially obsolete, and several display technologies such as LCD, PDP, DLP, LCoS are vying to replace them. None of these technologies involves a physical 2.4-power law like that of a CRT. Some people argue that emergent display technology gives us a chance to adopt linear-light encoding; however, perceptual uniformity remains important for these reasons:

- Perceptually uniform coding maximizes the perceptual utility of a limited number of bits – usually 8, or 10, or 12 – per component;
- Nearly all commercially important digital image storage and exchange standards call for perceptual uniformity; and
- Billions of stored images incorporate perceptual uniformity.

Emergent, non-CRT display devices incorporate signal processing circuits that apply a transfer function to impose the difference between the device's native, physical response and the behaviour required to mimic the electro-optical conversion function (EOCF) implicit or explicit in exchange standards. DLP displays and PDP displays both have physical linear-light response; display systems incorporating these displays incorporate a power function, or a function approximating one, to convert *R'G'B'* signals (presented at the interface) to linear-light *RGB* that modulates the display itself.

Perceptual uniformity in D-cinema

SMPTE/DCI standards for digital cinema distribution [SMPTE 431-1, SMPTE 431-2] 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 OECF of studio video, the 2.6-power offers improved visual performance in the low luminance and dark surround situation of the cinema.

There are no SMPTE/DCI standards for digital cinema acquisition; many techniques are in use. The basic principles that I have outlined outlined apply when the cinematographer decides, based upon the scene being captured, upon a diffuse white reference near the top end of the digital coding scale. If specular highlights beyond diffuse white are to be accommodated, then the cinematographer may impose what an engineer might call a distortion of the code scale above diffuse white. The cinematographer may have reason to acquire a scene while deferring any decision about reference white – that is, the decision may be deferred until post-production. In that case there is an argument to have an acquisition standard that uses a pure logarithmic code, or a pseudolog code [SMPTE RDD 2], with an appropriate number of digital code values per stop of scene-space luminance ("exposure").

Conclusion

Perceptual uniformity is a tremendously important aspect of digital image coding, particularly video, HDTV, digital cinema, and digital still photography. Without it, we would need 11, 12, or 13 bits per component, instead of 8 or 10. 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

ACR/NEMA PS 3.14 (2006), *Digital Imaging and Communications in Medicine* (*DICOM*), Part 14: *Grayscale Standard Display Function*.

Brill, Michael H. (1996), "Do tristimulus values have units?", in *Color Research and Application* 21 (4): 310–313 (Aug.).

CIE 15 (2004), *Colorimetry* (Vienna, Austria: Commission Internationale de L'Éclairage).

Hunt, Robert W. G. (1997), "The Heights of the CIE Colour-Matching Functions," in *Color Research and Application*, **22** (5): 337 (Oct.).

ISO 15076-1 (2005), *Image technology colour management – Architecture, profile format and data structure – Part 1: Based on ICC.1:2004-10.*

Kim, Min H. and MacDonald, Lindsay W. (2006), "Rendering High Dynamic Range Images," in *Proc. 17th Annual EVA Conference* (*EVA 2006*): 22.1–22.11 (Jul.).

Poynton, Charles (1998), "The rehabilitation of gamma," in Rogowitz, B. E., and T. N. Pappas (eds.), *Human Vision and Electronic Imaging III*, Proc. SPIE/IS&T Conf. 3299: 232–249 (San Jose, Calif., Jan. 26–30).

Poynton, Charles (1999), *YUV and luminance considered harmful,* available at www.poynton.com.

Poynton, Charles (2003), *Digital video and HDTV algorithms and interfaces* (San Francisco: Morgan Kaufmann).

SMPTE 431-1 (2006), *D-Cinema Quality – Screen Luminance Level, Chromaticity and Uniformity.*

SMPTE 431-2 (2007), *Reference Projector and Environment for Display of DCDM in Review Rooms and Theaters.*

SMPTE RDD 2 (2007), *Use of Logarithmic Non-Linear Transfer Characteristic for Transmission of Linear Signals through Limited-Bit-Depth Image Representations and Interfaces.*

Charles Poynton
tel +1 416 535 7187
charles @ poynton.com
www.poynton.com

Picture rendering in video

POYNTON, CHARLES (2003), *Digital video and HDTV algorithms and interfaces* (San Francisco: Morgan Kaufmann).

This document surveys picture rendering in video, from its origins in the development of the NTSC colour television system in the early 1950s, to the present (2009). I assume that you are familiar with colour science, and with the basic concepts of video systems. An introduction to the technical issues is provided in my book *Digital Video and HDTV Algorithms and Interfaces* ("DVAI").

POYNTON, CHARLES, *YUV and luminance considered harmful*, available at www.poynton.com.

I assume that you are familiar with the term *luminance* and the symbol *Y* of colour science, and the term *luma* and the symbol *Y'* of video. These terms are discussed in the document cited in the margin. I use the term *luminance* (or *relative luminance*) when referring to greyscale reproduction. In additive colour systems, *tristimulus value* refers to a linear-light red, green, or blue component.

POYNTON, CHARLES (2009), "Perceptual uniformity in Digital Imaging," in *Proc. Gjøvik Color Imaging Symposium* (GCIS 2009): 102–109.

I also assume that you are quite familiar with perceptual uniformity, as outlined in the paper cited in the margin. Gamma correction at a video camera encodes the signal into a perceptually uniform domain; the display approximately inverts this coding.

Introduction to picture rendering

It is widely assumed that a reproduced image should have luminance values proportional to the corresponding values in the scene. To impose perceptual uniformity, the opto-electronic conversion function (OECF, "or gamma correction") at capture would then be the exact inverse of the display's transfer function (EOCF). However, if relative luminance were accurately maintained from the scene to presentation, several factors would conspire to alter the *appearance* of colours:

- The ambient conditions of viewing a reproduction are typically different from the conditions in which the scene was viewed.
- Typical displays have much lower luminance and lower luminance range (contrast ratio) than typical scenes.
- The reproduced image often has a dim surround (for example, television) or a dark surround (for example, cinema or home theatre), in contrast to the average surround typical of scenes being captured.

To overcome any or all of these effects, modifications must be made to reproduction of relative tristimulus values. In video, obtaining subjectively correct images is typically based upon modifying the power function of gamma correction from its mathematically-ideal value. Rather than encoding with a power $1/\gamma$ and decoding with γ, we encode with $1/\gamma_E$ and decode with γ_D, where γ_E and γ_D differ. Typically, an end-to-end power function having an exponent slightly greater than unity is called for, in which case $\gamma_E < \gamma_D$ (i.e., $1/\gamma_E \cdot \gamma_D > 1$).

GIORGIANNI, EDWARD J., and T.HOMAS E. MADDEN (2008), *Digital Color Management: Encoding Solutions* (Reading, Mass.: Addison-Wesley,).

Owing to the importance of the EOCF (characterized by γ_D) in perceptual uniformity, perceptually uniform coding and picture rendering are intertwined.

For creative purposes, *any* manipulation in image data is allowed if it achieves the intended appearance in the studio reference display! At capture, guidelines are useful, but if you believe as I do that art rules in the end, no capture standard is necessary. As a rough guide, a studio scene to be displayed on a studio reference display should have relative tristimulus values raised to an end-to-end power of about 1.2. For the 2.4-power EOCF of a typical studio reference display, the 1.2 end-to-end power implies encoding with an effective power function of about 0.5, the effective exponent of BT.709's OECF.

POYNTON, CHARLES (2009), *History of perceptual uniformity and picture rendering in video* (unpublished).

The necessity for picture rendering in video was appreciated almost three quarters of a century ago! In a separate document, I outline the history of picture rendering (and of perceptual uniformity) in video.

Image state

ISO 22028-1:2004, *Photography and graphic technology – Extended colour encodings for digital image storage, manipulation and interchange – Part 1: Architecture and Requirements.*

If perceptual encoding and decoding were standardized by an invertible function mapping scene tristimulus values to display tristimulus values, then an image coding system would be completely specified. In practice, however, high quality imagery requires various picture renderings for various scenes. In professional imaging, manual adjustments may be made at capture (for example, by a photographer) or in processing (for example, by a graphic arts technician). In professional video, manual adjustments are almost always made at capture.

No matter whether picture rendering is automatic, manual, or involves aspects of both, there are virtually no commercial imaging systems that do *not* involve some sort of picture rendering. Colorimetry can be applied in the scene and at the camera, or at the display; however, there is often no direct link between the two.

The possibility – or inevitability – of various automatic or manual adjustments causes a disconnect between image encoding and image decoding. You can have a colorimetric definition of each, but there is almost never a direct, fixed connection between the two.

The feature film production industry has adopted an image encoding system that is explicitly scene-referred: OpenEXR. See www.openexr.org.

The disconnect between encoding and decoding leads to image data existing in one of two states, termed *scene referred* and *display referred.* (In graphics arts colour management terminology, image data destined for hard copy output is said to be *output referred,* but there can be many kinds of output: I use the term *display referred* to emphasize image data destined for an electronic display.)

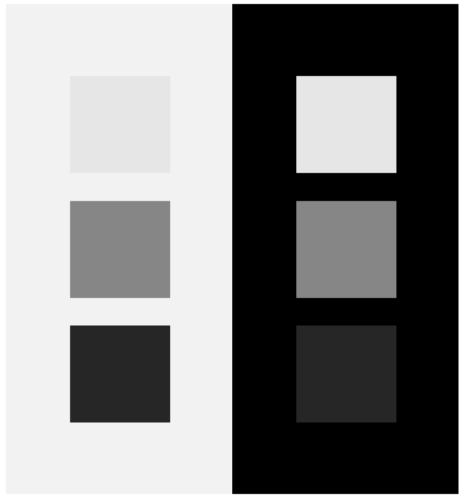


Figure 1 **Surround effect.** The three squares surrounded by light gray are identical to the three squares surrounded by black; however, each of the black-surround squares is apparently lighter than its counterpart. Also, the contrast of the black-surround series appears lower than that of the white-surround series.

DEMARSH, LEROY E., and EDWARD J. GIORGIANNI, "Color Science for Imaging Systems," in *Physics Today*, Sept. 1989, 44–52.

Picture rendering in the modern era

DVAI's chapter entitled *Rendering intent* of gives a reasonably accurate account of the picture rendering for modern video. However, I confess that I titled that chapter inappropriately: I should have titled the chapter *Picture rendering.* I plan to make that change in subsequent editions.

As I have outlined, in the early days of television engineering it was appreciated that subjectively correct pictures could be obtained through imposing an overall power function. That idea was largely discounted or forgotten from about 1960 to the present. During the last decade, colour management system (CMS) developers working primarily in graphics arts have studied how to modify image data to achieve subjectively acceptable reproduction across different media and different viewing conditions. The colour management community refers to these adjustments as *picture rendering.* I adopt that term to describe what we have been doing in video for about 50 years.

Modern understanding of colour appearance, developed over the last 15 years, identifies three main effects that need to be compensated:

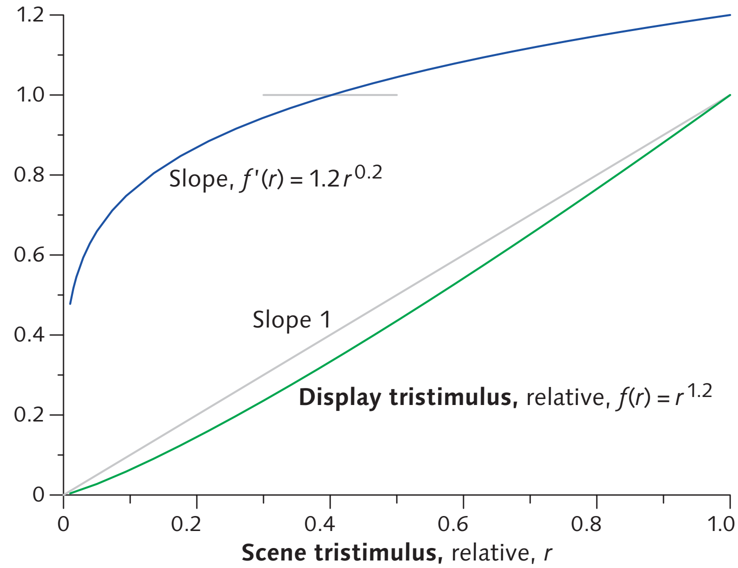
HUNT, R.W.G., *The Reproduction of Colour,* Sixth Edition Chichester, U.K.: Wiley, 2004). See Chapters 6 and 11.

- First, colourfulness decreases as illumination decreases – the *Hunt effect.* Consider this example: Flowers viewed in daylight (perhaps 30,000 $cd\cdot m^{-2}$) appear much more colourful than the same flowers viewed at twilight (perhaps 300 $cd\cdot m^{-2}$). If an image is captured in daylight and its linear-light *RGB* values are linearly scaled then displayed at 300 $cd\cdot m^{-2}$, the image will look like it was captured at twilight. To make the image look like it represents a daylight scene, the colourfulness needs to be increased by altering the image data.

- Second, apparent contrast decreases as illumination decreases – the *Stevens effect.*

- Third, image appearance is affected by its surround. Consider Figure 1. When a tone scale – in this example, having just three patches – is displayed surrounded by black, the tones are visually different from the same tones surrounded by white. To achieve the appearance of a light-surround scene when reproduced and viewed in a dim surround, the image data must be altered. Scenes are most often captured with an "average" surround whose characteristics resemble

Figure 2 **An end-to-end 1.2-power function,** graphed here in the lower (green) curve, is typical of the transfer of relative luminance in video systems. (Compare the 1.2-power with the identity function, graphed in light grey.) The upper (blue) curve graphs the slope of the power function. Where the slope is less than unity (below relative scene tristimulus of about 0.4), scene luminance is compressed; above that, it is expanded.



the scene itself, but television is ordinarily viewed in a dim surround. Correction for the subjective effect is necessary.

The first requirement – compensation for the Hunt effect – could potentially be accomplished by an increase in colour saturation. In a $Y'C_BC_R$ colour difference system, chroma gain of perhaps 1.2 could be applied to C_B and C_R. However, the second and third effects above involve the tone scale, where chroma gain would have no effect.

Although it isn't strictly an appearance effect, some form of image data manipulation is necessary in the usual case that the scene spans a different ratio of luminances than the contrast ratio of the display. Ordinarily, the scene exhibits a larger ratio of luminances than the display contrast ratio: Relative luminance values have to be compressed into the luminance range of the display. Less frequently, a scene may span a narrower range of luminances than the display – for example, a scene captured in haze or fog. In such cases, the scene luminance range may need to be expanded to achieve a convincing reproduction.

Picture rendering in video standards

Decades of experience have taught that all three appearance effects can be compensated in one step by imposing a modest end-to-end power function through the whole system. Picture rendering in video is based upon a combination of power functions at encoding and decoding; picture rendering is accomplished by adjusting the power function exponents. Encoding and decoding involve perceptual uniformity, so picture rendering and perceptual uniformity are intertwined.

Typical studio illumination is about 2000 lx.

For capture of video in the studio, encoding typically incorporates an effective 0.5-power in the camera. Video decoding typically incorporates a 2.4-power (e.g., a studio reference display). The lower curve of Figure 2 shows a graph of the 1.2-power function that is typical of the resulting end-to-end reproduction. The 1.2-power function looks very gentle in this portrayal; nonetheless, its effect is significant.

See Giorgianni and Madden's book, cited in the margin of page 2.

In the typical case that the scene spans a larger luminance range than the display is capable of reproducing, relative luminance in the scene must be compressed through the system: Portions of the end-to-end function must have slope less than unity. However, to overcome the other effects that I have mentioned, some expansion of contrast – in particular, the midtones – is almost always necessary. The slope of a 1.2-power function is graphed in the upper (blue) curve of Figure 2. The slope is less than unity below a relative luminance of about 0.4; below that value, the function compresses contrast. The slope exceeds unity beyond a relative luminance of about 0.4; above that value – in the midtones and highlights – the function expands contrast.

The greyscale is defined by equal red, green, and blue tristimulus values: a 1.2-power function applied to each component does not introduce any colour. Consider the case where R and G have equal values, but B departs. The power function accentuates the colour difference, thereby increasing chroma. Maximum chroma "boost" is introduced at midscale; there is no boost at $B=0$, and no boost at $B=1$.

An S-shaped function is characteristic of photographic film. Film characteristics are usually plotted in log-log coordinates, but the S-curve is also evident when plotted on linear axes. See page 8.

For some scenes, it may be appropriate not only to compress shadows and expand midtones, but also to compress highlights. Highlight compression can be achieved through function having an S-shape. Optimization of the individual R, G and B tone curves to a particular scene can be accomplished either automatically or manually.

Holm, Jack M., US Patents 6628823 and 6249315.

Consumer digital still cameras (DSCs) have no photographer or videographer to alter the tone function to suit the scene. Instead, algorithms in the camera perform the mapping. See Holm's patents for examples of such techniques.

What I call *ratio of diffuse white* is the ratio of absolute luminance of an ideal diffuse (Lambertian) reflector in the scene to the absolute luminance of its representation on the display.

Published work concerning rendering in video tends to emphasize the effect of the surround; the ratio of scene luminance to display luminance is hardly mentioned. Modern work – by Hunt and by Giorgianni, for example, both cited earlier – suggests that what I call the "ratio of diffuse white" (the first point on page 3) dominates.

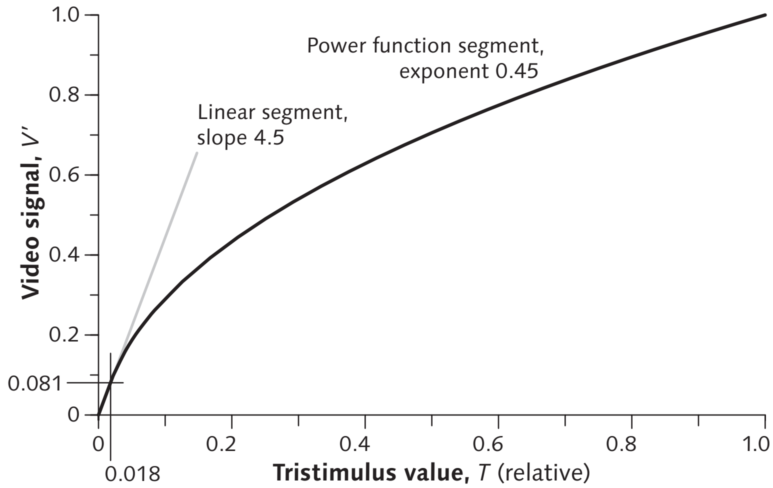
As I have mentioned, studio reference displays today have exponents of about 2.4. BT.709's effective 0.5-power at encoding, combined with a 2.4-power function at the display, yields an end-to-end exponent of 1.2. An end-to-end exponent of 1.2 is appropriate for reproduction, in the studio control room environment, of a studio scene.

OECF standards

ITU-R Rec. BT.709, *Parameter values for the HDTV standard for the studio and for international programme exchange* (Geneva: ITU).

BT.709 standardizes a reference camera transfer function (OECF), graphed in Figure 3. BT.709's encoding equation incorporates a power function exponent of 0.45 – what I call the "advertised" gamma. However, BT.709's equation incorporates a linear segment near black; the power function portion is scaled and offset to achieve function and slope continuity at the breakpoint (which lies at about 2% rela-

Figure 3 **BT.709 OECF** is standardized for HDTV. Although the "advertised" exponent is 0.45, the curve is scaled and offset, and a linear segment is inserted near black; the effective exponent is about 0.5. If this function were fixed in every camera, then BT.709 could be considered to be scene-referred.



tive luminance). The scaling and offseting cause the effective exponent to be higher than the exponent "advertised" in the equation. If BT.709's curve is approximated as a single pure power function, the best exponent is about 0.5: BT.709's encoding is effectively a square root.

It is obvious how to invert BT.709 encoding. However, applying the inverse function directly would serve to recover approximate relative *scene* tristimulus values from the reference camera; to use such a function prior to display would disregard the necessity for picture rendering. BT.709 itself makes no mention of picture rendering, and in normal viewing conditions the tristimulus values obtained by inverting the OECF would be unsuitable for use as *display* tristimulus values in a reproduced image.

Creative intent

In high-end video production, image data modifications are imposed at various stages. A cinematographer typically adjusts various controls on the camera to achieve the desired reproduction on a reference display. A colourist may alter the image data when grading material that has just been captured. A colourist may also make modifications to the edited program at the approval stage.

Figure 4 shows the effect on the OECF of five controls commonly found in professional video cameras. In addition to these five, overall exposure can be controlled (through the lens iris and/or through VIDEO GAIN); there are also PEDESTAL controls. All of these controls are used routinely in professional video production to establish an OECF that renders the scene in accordance with the desired creative intent.

In scientific and industrial applications of imaging, standardization of a camera's OECF may be appropriate or even necessary. However, in professional video and film production the goal is *not* to reproduce, at the viewer's premises, an accurate representation of the scene in front

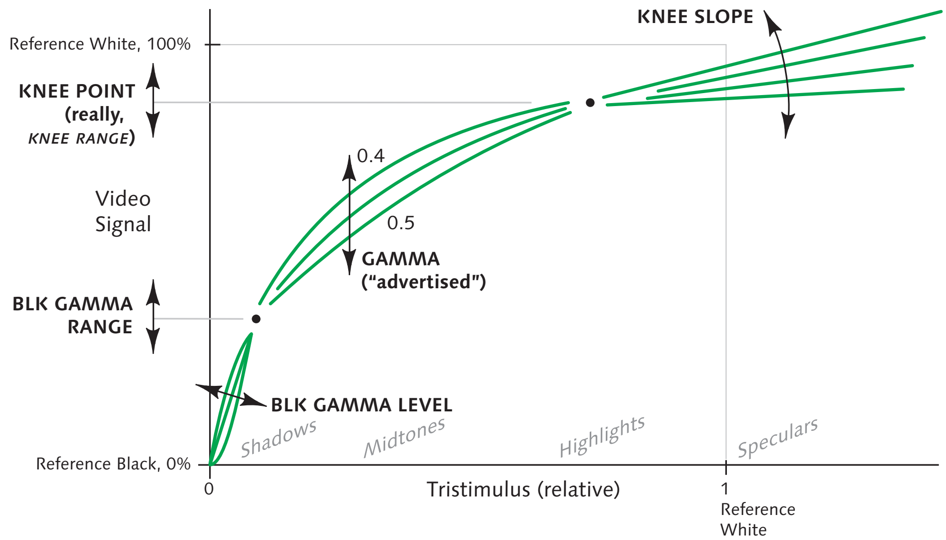


Figure 4 **Typical camera OECF controls** are used by the videographer or cinematographer to impose the desired creative intent. Video cameras default to BT.709; however, as soon as any control is adjusted off-detent, the video signal can no longer be considered to be scene-referred.

of the camera – rather, the goal is to reproduce what the director, cinematographer, or colourist saw on his or her studio display upon approving the final stage of post-production. To preserve creative intent, it is not the camera's OECF that needs to be standardized but rather the reference display's EOCF!

ITU-R, SMPTE, and other bodies have standardized video OECF but have failed to effectively standardize display transfer functions. In a separate document, I make a plea for EOCF standardization: Poynton, Charles (2009), *Picture rendering, image state, and BT.709* (unpublished).

Consumer displays are expected to approximate the characteristics of studio displays upon which the final artistic decisions of production are based. However, standards cannot be enforced at consumer displays, and no transfer function is standard in that arena. The best that we can hope for is a studio display standard that provides guidance to consumer electronics manufacturers.

Viewing conditions in the studio

Several assumptions about the reference display and its viewing conditions are implicit in current studio video practice:

- The reference display incorporates a power function (EOCF) with an exponent between 2.35 and 2.4.
- It has reference white luminance of between 80 $cd \cdot m^{-2}$ (EBU standard) and 120 $cd \cdot m^{-2}$ (SMPTE standard).
- It is viewed in diffuse illumination of about 100 lx in a dim surround having reflectance of about 18%, and surround luminance of about 5% of reference white.
- It has off-state reflectance of about 0.5%, which – combined with the ambient illumination in the viewing environment – gives a veiling glare of about 0.25 $cd \cdot m^{-2}$.

If video originated with these assumptions is viewed on a display with different characteristics (say, maximum luminance of 300 cd·m^{-2}) or in a different viewing environment (say, an "average" surround instead of a dim surround), then image data should be altered at the display to achieve the intended appearance.

Viewing conditions in cinema

Film gamma is measured as the slope, in log-log coordinates, of the characteristic curve. A straight line having slope *n* on a log-log plot is characteristic of a power function having the exponent value *n*.

Standards for D-cinema call for reference white (or D_{MIN}) luminance of 48 cd·m^{-2}. However, commercial cinemas are rarely that bright; maximum luminance is likely to be around 40 cd·m^{-2}. The surround in cinema is completely dark, and the contrast ratio is excellent. Reproduction of a studio scene in these conditions requires an end-to-end exponent of about 1.5, typically obtained through the combination of the nominal 0.6-gamma of camera negative film and the effective 2.5-power of release print film.

DCI standards for digital cinema have no provision for – and no need for – picture rendering. DCI standards call for reproduction, in the cinema, of the colorimetric values displayed by the reference projection. The reference projection viewed by creative staff in a screening theatre, and the commercial projection viewed by the consumer have essentially identical viewing conditions; consequently, no adjustment for appearance effects is required. Rightly, in my opinion, DCI standards make no mention of the scene and no mention of the camera.

DCI standards for digital cinema call for a display EOCF that is a pure 2.6-power function; the EOCF serves the purpose of perceptual uniformity. Owing to the DCI's system design, the EOCF has no relationship to picture rendering.

Consider video material that is originated with the assumption of viewing in a studio environment. When such material is viewed in a cinema environment, according to film practice, the end-to-end exponent should be increased from 1.2 to about 1.5. In principle, that increase – a ratio of 1.25 – should be effected by raising the display's gamma from 2.4 to 3.0. Practical experience suggests that gamma of 2.6 or 2.7 suffices. The reasons for the discrepancy are not yet clear.

Viewing conditions in the office

In an office environment, maximum luminance could be as high as 300 cd·m^{-2}; the contrast ratio is poor (perhaps 20:1); and the usual surround is "average" – that is, the surround luminance approximates that of the displayed image. In these circumstances, an end-to-end exponent of about 1.1 is suitable for a studio scene. For a studio scene originated with a 0.5-power OECF, the 1.1 end-to-end exponent can be achieved through display gamma of about 2.2.

IEC 61966-2-1, *Multimedia systems and equipment – Colour measurement and management - Part 2-1: Colour management – Default RGB colour space – sRGB*.

The sRGB standard specifies a viewing situation comparable to the one I have just described. The specification calls for a reference white luminance of 80 cd·m^{-2} and ambient illuminance of 64 lx; however, typical reference white levels and typical ambient illuminance values in today's sRGB applications are several times those values. Also, the sRGB specification anticipates veiling glare of 0.2 cd·m^{-2}, but practical veiling glare approaches 1 cd·m^{-2}. The EOCF of sRGB calls for a pure2.2-power function EOCF, which I consider appropriate for decoding and display of video material in viewing conditions typical of sRGB's application today. For viewing under the conditions written in

See IEC TC100 US TAG, *Recommendations to enhance the IEC 61966-2-1 sRGB standard and support best implementation practices*, IEC/TC 100 document 100/AGM(Secretariat)375.

the sRGB standard itself, in my view "gamma" of about 2.3 would be appropriate. Several participants in the colour management community argue that sRGB's parameters should be changed to bring them into line with common practice; see Holm's paper.

Equivalence of video and sRGB imagery

As I mentioned earlier, image data carries – either implicitly, or rarely, explicitly – an image state that incorporates assumptions about display characteristics and viewing conditions. OECFs and EOCFs have no image state of their own, but are intimately related to the transforms between image states. We can compare BT.709 and sRGB OECFs:

- BT.709's reference OECF is effectively a 0.5-power function. The BT.709 OECF takes scene-referred linear-light *RGB* data – typically, directly from the sensor – and imposes picture rendering to transform to display-referred *R'G'B'*. The reference OECF incorporates picture rendering suitable for encoding a studio scene for reproduction on a 2.4-power EOCF in studio viewing conditions.

- The OECF of sRGB takes display-referred linear-light *RGB* data and transforms to display-referred *R'G'B'*. No picture rendering is imposed. The sRGB standard specifies an OECF, but the gamma is different from the BT.709 OECF: sRGB's OECF has an advertised gamma of $^1/_{2.4}$ – that is, about 0.42 – and an effective gamma of 0.45. The OECF specified in the sRGB standard is effectively just the inverse of the sRGB EOCF, adjusted to include a linear segment near black: The OECF of sRGB therefore incorporates no picture rendering. sRGB's OECF is appropriate when no rendering is necessary, as would be the case in a flatbed scanner capturing an image (such as a photograph or offset print reproduction) that has already had rendering applied.

Although the "official" sRGB web site is defunct, a one-page summary remains available: *sRGB Gamma Calculation*, www.srgb.com/srgbgammacalculation.pdf [accessed 2009-04-12].

Despite the superficial similarity between BT.709's OECF and sRGB's OECF, the two are functionally quite different. Although they are both display-referred, BT.709's OECF includes picture rendering, but sRGB's does not.

"BT.709" image data is intended for reproduction through a 2.4-power function on a display in studio viewing conditions.

Practical sRGB displays have higher display luminance than studio displays, perhaps 300 $cd \cdot m^{-2}$ instead of 80. Typical environments in which sRGB is used have more veiling glare than studio environments (perhaps 2% instead of 0.5%), and an average (bright) surround instead of a dim surround. The differences between practical sRGB conditions and studio conditions necessitate a picture rendering power function exponent for sRGB about $^1/_{1.1}$ times that of the studio. sRGB's EOCF is a 2.2-power function, compared to the 2.4-power used in the studio. The ratio between these exponents is $^1/_{1.1}$. For a given set of image data, I conclude that approximately the same appearance will result when that data is displayed at 2.2-gamma in sRGB's practical environment and at 2.4-gamma in studio conditions. ■

Charles Poynton
tel +1 416 535 7187
charles @ poynton.com
www.poynton.com

Review of perceptual uniformity and picture rendering in video

Digital image encoding is *perceptually uniform* if a small perturbation to a component value – such as the digital code value used to encode luminance, red, green, or blue – is approximately equally perceptible across the range of that value. Perceived lightness follows approximately a 0.42-power function with respect to luminance (which is, by definition, physical). The most effective use is made of *a* limited number of bits per component when coding approximates lightness.

Picture rendering refers to modifications to image data that are made in order to obtain subjectively correct reproduction under display and viewing conditions different from capture conditions.

Poynton, Charles (2009), "Perceptual uniformity in Digital Imaging," in *Proc. Gjøvik Color Imaging Symposium* (GCIS 2009): 102–109.

Although the principles of perceptual uniformity and picture rendering are used in virtually all commercial imaging systems, their use in video is widely unknown or misunderstood. This document surveys the development and deployment of these concepts in video, from their origins in the 1930s to their contemporary use at present. In the paper cited in the margin, I survey contemporary use of these concepts.

Poynton, Charles (2003), *Digital video and HDTV algorithms and interfaces* (San Francisco: Morgan Kaufmann).

I assume that you are familiar with colour science and with the basic concepts of video systems. I also assume that you are familiar with video terminology. An introduction to the technical issues of perceptual uniformity is provided in my book *Digital Video and HDTV Algorithms and Interfaces* ("DVAI") cited in the margin.

History of perceptual uniformity

Maloff, I.G. (1939), "Gamma and Range in Television," in *RCA Review* **3** (4): 409–417 (Apr.).

Perceptual uniformity in electronic imaging was appreciated almost three quarters of a century ago! In 1939, Maloff wrote this about black and white television:

The phrase in square brackets is mine.

> A small increment in light intensity is more noticeable to the eye in dark parts of the picture than is the same increment in bright parts of the picture. When the picture at the receiver is expanded [by the CRT's power function], the highlights are over-emphasized and shadows are under-emphasized, and in this way a greater amount of interference may be tolerated. … The expansion or increase in contrast may be applied either at the transmitter or the receiver, but

> the interference consideration makes it more desirable to expand at the receiver.

Fink, Donald G. (1940), *Principles of Television Engineering* (New York: McGraw-Hill). See Chapter VIII. The phrase in square brackets is mine: In modern terminology, the term brightness refers to a perceptual attribute, not a physical one.

In his 1940 book about monochrome television, Fink summarized the relationship of image coding and perceptual uniformity. The NTSC monochrome standard, established in 1941, stated:

> The transmitter output shall vary in substantially inverse logarithmic relation to the brightness [now, luminance] of the subject.

The "logarithmic relation" reflects the Weber-Fechner law, which was detailed by Fink. The reference to "inverse" concerns the transmitter's negative modulation polarity, which is irrelevant to our purposes.

Mertz, Pierre (1950), "Perception of Television Random Noise," in *J. SMPTE* **54** (1): 8–34 (Jan.).

In 1950, during development of the NTSC colour system, Mertz published in the *SMPTE Journal* a comprehensive description of the issue. He concluded (using the symbol *n* instead of today's γ):

> The characteristic for $n = 1$ gives … the greatest susceptibility to additive noise. … As *n* is increased, the susceptibility to additive noise is reduced … The changes are slow beyond $n = 2$.

Had Mertz explored values of γ much beyond 3, he would have found noise in the whites!

In 1950, the noise was analog – what Maloff called "interference," and what Mertz called "random." Nowadays, in digital systems, the noise arises from quantization; however, whether noise arises from analog or digital processes, the effects are similar. What Mertz describes as "characteristic for $n = 1$" we would today term *linear-light.*

Oliver, B.M. (1950), "Tone Rendition in Television," in *Proc. IRE* **38** (11): 1288–1300 (Nov.). See page 1294.

CRTs of Mertz's day had power laws of about 2.5, and were appreciated in his time to be well matched to perception. In his 1950 paper "Tone Rendition in Television," Oliver gave quite a lucid summary of the Weber-Fechner law applied to luminance, then stated:

> With a 2.5-root transmitter, a 2.5-power receiver, properly adjusted, would give linear reproduction … any receiver exponent lying between 2 and 3 is not too bad a match [to perception].

Applebaum, Sidney (1952), "Gamma Correction in Constant Luminance Color Television Systems," in *Proc. IRE* **40** (11): 1185–1195 (Oct.).

Applebaum clearly expressed the desirability of nonlinear processing in terms of vision:

> The brightness-transfer characteristic of the average picture tube compresses the shadow tones and expands highlight tones. However, it has been shown that the nonlinear characteristic of the average picture tube is almost ideal for minimizing noise sensitivity. This indicates that the precorrections for the nonlinear characteristic of the receiver should be applied at the transmitter …

U.S. Reg. Title 47 (47 CFR Ch. I), Part 73.682, page 212; see www.fcc.gov/searchtool.html

The FCC standard for NTSCUS 47 CFR refers to *R, G,* and *B* signals

> … having a transfer gradient (gamma exponent) of 2.2 associated with each primary color.

The parenthesized words are in the original. The standard isn't explicit regarding whether the 2.2 value refers to the reciprocal of the effective power function exponent at the encoder (what I term γ_E), or to the effective power function exponent at the decoder (γ_D).

The FCC avoided placing a tolerance on gamma:

> NOTE: At the present state of the art it is considered inadvisable to set a tolerance on the value of gamma ...

ITU-R Rep. BT.624-4 (1990), *Characteristics of television systems*.

Apparently the issue was not sufficiently well understood in 1953 to standardize a firm number. The FCC has never revisited the issue. The 2.2 figure – again without reference to whether it is intended for encoding or decoding – is documented in ITU-R Report 624.

NTSC was thereby standardized with decoding having an exponent of somewhere around 2.2, expecting encoding at the camera through a power function having an exponent of somewhere around $^{1}/_{2.2}$ (gamma correction). In practice, decoding must have used the power function intrinsic in CRTs of the day, which I suspect had exponents around 2.3 or 2.4; encoding would have been done with whatever function made the best looking pictures on such a display.

HAZELTINE CORPORATION (1956), *Principles of Color Television*, by the Hazeltine Laboratories staff, compiled and edited by MCILWAIN, KNOX and DEAN, CHARLES E. (New York: Wiley). See page 273.

The famous Hazeltine Labs book states:

> … it is very benificial from the standpoint of combatting radio-path noise to have the receiver compress the dark shades and expand the light shades. The usual picture tube, either monochrome or color, has a characteristic in the right direction in this regard.

The passage refers to "radio-path noise": additive white Gaussian noise (AWGN) in the RF channel was the dominant noise source at the time. The perceptual coding that was effected by the power functions at the encoder and decoder caused noise to be distributed approximately uniformly across the tone scale from black to white.

ROBERTS, ALAN (1993), "Measurement of display transfer characteristic (gamma, γ)," in *EBU Technical Review* **257**: 32–40 (Autumn).

Many European video engineers expect the decoder gamma (of a studio display or a television receiver) to be 2.8: That value was enshrined in EBU standards, and is documented in ITU-R Rep. 624. I have found no evidence that such a high value has ever been used. To the contrary, Alan Roberts – recently retired from BBC Research and Development – found values of the exponent in Europe between 2.2 and 2.4, quite consistent with the 2.4 value found in North America.

The seminal documents of the NTSC used the term *luminance signal.* Judging from their published work, the participants in the first decade clearly understood that the signal was not linearly related to colour science luminance. However, over the following several decades the distinction was lost to almost everyone involved in video engineering. Despite its rigid definition in the colour science community as a linear-light quantity, the term *luminance* came to be used by the video engineering community to reflect the nonlinear quantity representing the achromatic signal. Confusion resulted. I discuss the confusion in Appendix A of *Digital video and HDTV algorithms and interfaces* (cited earlier). The sloppy nomenclature made its way into ostensibly authoritative video references, such as Pritchard's SMPTE paper.

PRITCHARD, D.H. (1977), "U.S. Color Television Fundamentals – A Review," in *SMPTE Journal*, 86 (11): 819–828 (Nov.).

History of perceptual uniformity in computer graphics

Stockham, Thomas G. (1972), "Image processing in the context of a visual model" in *Proc. IEEE* **60** (7): 828–842 (Jul.).

Catmull, Edwin (1979), "A tutorial on compensation tables," in *Proc. SIGGRAPH 1979:* 1–7.

Computer graphics pioneers recognized early on the importance of perceptual uniformity (although they did not give it that term – or indeed, any particular term). Stockham, at the University of Utah, analysed the situation very thoroughly in a 1972 paper, and detailed the advantages of logarithmic coding. Shortly afterward, that university became a hotbed of computer graphics development. Edwin Catmull – later to found Pixar – attended that university, and in 1979 characterized the transfer of computer images to film with demonstrating a good understanding of the perceptual requirements (but using the word "intensity" quite loosely).

Smith, Alvy Ray (1978), "Color Gamut Transform Pairs," in *Computer Graphics* **12** (2): 12–19 (Aug., *Proc. SIGGRAPH 78*).

Alvy Ray Smith joined Catmull at NYIT. In his quest to adapt video principles to computer graphics, Smith apparently encountered the word *luminance* (presumably from Pritchard's paper). Smith apparently correlated Pritchard's use of the term *luminance* with his own knowledge of the term as used in colour science. Understandably – though wrongly – he concluded that video "luminance" and colour science luminance were identical. His 1978 paper clearly presented "NTSC luminance" as a linear combination of *RGB.* So, perceptual uniformity was appreciated, but it was mistakenly not understood to extend to video. Smith's conclusions were published in the highly influential proceedings of the SIGGRAPH conference.

Foley, James D., and Andries van Dam, (1984), *Fundamentals of Interactive Computer Graphics* (Reading, Mass.: Addison-Wesley).

It took only a few years for Smith's interpretation to pervade computer graphics. The seminal Foley and van Dam textbook states without attributing any primary sources,

> The Y component of YIQ is not yellow but luminance, and is defined to be the same as the CIE Y primary.

Foley and van Dam discuss the desirability of an exponential relationship between *RGB* digital code values and the associated tristimulus values ("intensities"), thereby suggesting a logarithmic relationship between "intensities" and code values. However, like Smith, they fail to extend this concept to video, and wrongly describe video as having linear-light coding. Foley and van Dam cite Pritchard's 1977 paper. Clearly, Pritchard's failure in his SMPTE paper to properly describe video's "luminance" was crucial.

Foley and van Dam subsequently revised their book into several editions (adding co-authors Feiner and Hughes), and unknowingly propagated the error. The die was cast. Confusion was to reign for the next few decades. With the emergence of colour management systems around 1995, colour science concepts and terminology reached a broad audience among computer professionals, and – perhaps a decade later – a broad audience among video professionals. Only recently has the confusion begun to subside.

History of picture rendering in video

Maloff, I.G. (1939), "Gamma and Range in Television," in *RCA Review* **3** (4): 409–417 (Apr.).

Well before the invention of the NTSC colour system, the necessity of picture rendering – then termed "modification to the tone scale" – was appreciated for monochrome television. Maloff stated,

The phrase in square brackets is mine.

> Unity gamma is not sufficient for transmitting studio and outdoor pickup and for such occasions the contrast should be raised [to achieve a] resultant overall contrast of 1.6.

The high value of 1.6 is presumably due to displays of the time being rather dim.

Fink, Donald G. (1940), *Principles of Television Engineering* (New York: McGraw-Hill). See page 337.

Fink's 1940 book contains this passage:

> An over-all value of gamma of between 1.2 and 1.7 is used in commercial motion pictures … similar values of gamma should serve equally well for television work.

Maloff's terms "unity gamma" and "resultant overall contrast" and Fink's term "over-all value of gamma" all refer to the end-to-end power function exponent – the product of the gamma exponents at each stage. Today, some people call this "system gamma." Decades ago, the "system" comprised just a few subsystems between the original scene and eventual presentation; in that era, the term "system gamma" was unambiguous. Now, many subsystems intervene between capture and display, and many people use the term "system gamma" to refer to the power function imposed by an individual *subsystem*, thereby introducing ambiguity and confusion. Lacking a clear delineation of what constitutes the "system," I avoid the term.

U.S. Reg. Title 47 (47 CFR Ch. I), Part 73.682, (page 212); see www.fcc.gov/searchtool.html

The 1953 U.S. FCC standard for NTSCUS 47 CFR refers to *R*, *G*, and *B* signals

> … having a transfer gradient (gamma exponent) of 2.2 associated with each primary color.

Hazeltine Corporation (1956), *Principles of Color Television*, by the Hazeltine Laboratories staff, compiled and edited by Knox McIlwain and Charles E. Dean (New York: Wiley). See page 233.

Describing the NTSC colour system, the famous Hazeltine book cited in the margin states:

> … typical conditions might include a [gamma] corrector adjusted to an exponent of 0.64 ($= 1/_{1.6}$), which will reduce the 2.2 of the picture tube to 1.4 for the system.

Bingley, Frank J. (1954), "Transfer Characteristics in NTSC in Color Television," in *Proc. IRE*, **42** (1): 71–78 (Jan.).

I find Hazeltine's quoted CRT exponent of 2.2 somewhat low. Some accounts of the time quote CRT exponents as high as 2.75. Bingley's 1954 paper on transfer characteristics discusses, on page 75, an encoder having an exponent of $1/_{2.2}$ ($1/_{\gamma_E}$) mated to a decoder having an exponent of 2.75 (γ_D), yielding an end-to-end exponent of 1.25.

I find it impressive that the NTSC researchers identified the necessity of perceptually uniform coding, and that they understood the necessity of applying an end-to-end power function to impose what we now call picture rendering. However, colour appearance phenomena were not sufficiently well understood to permit exact encoding and decoding exponents to be standardized.

BARTLESON, C.J. and E.J. BRENEMAN (1967), "Brightness Reproduction in the Photographic Process," in *Photographic Science and Engineering* **11** (4): 254–262.

During the decades following the introduction of NTSC broadcasting, a deeper understanding of the implications of nonlinear encoding and decoding developed. In 1967, Bartleson and Breneman published the results of some experiments in photographic reproduction. Later, DeMarsh extended their results to television. By 1975, DeMarsh was aware that many video engineers were under the mistaken impression that end-to-end linearity was a goal, and he stated with confidence,

DEMARSH, LEROY E. Citation to follow.

> It is often assumed that television should have a system gamma of unity. … This assumption is wrong.

DeMarsh continued,

Today, "dim surround" would be $1/_{20}$ of reference white, i.e., 5%.

> When we look at television pictures in a dark surround, we prefer a television system gamma of 1.5. If we view the same pictures with a dimly lit surround ($1/_{10}$ of picture highlight luminance), we prefer a gamma of 1.2; if we view these pictures with a bright surround (equal to highlight luminance), we want a gamma of 1.0.

SPROSON, W.N. (1983), *Colour Science in Television and Display Systems* (Bristol, U.K.: Adam Hilger). See page 108.

Sproson's book refers to encoding with $1/\gamma_E = 1/2.2$ and decoding with $\gamma_D = 2.75$. This achieves an end-to-end exponent of 1.25, in good agreement with DeMarsh's 1.2 value for a dim surround. (However, I find Sproson's quoted exponent of 2.75 to be unreasonably high.)

ROBERTS, ALAN (1993), "Measurement of display transfer characteristic (gamma, γ)," in *EBU Technical Review* **257**: 32–40 (Autumn).

According to Roberts' 1993 paper, typical display exponents 15 years ago were between 2.3 and 2.4. I have concluded that professional studio displays in use today for SDTV and HDTV, as configured for studio control room environments, have exponents very close to 2.4.

POYNTON, CHARLES (2009), *Picture rendering, image state, and BT.709* (unpublished).

My conclusions about the current state of perceptual uniformity and picture rendering in video and digital cinema can be found in the companion documents that I cite on page 1 of this document. My recommendations for addressing deficiencies in current standards can be found in the document cited in the margin. ■