International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BT.2390-1

(10/2016)

High dynamic range television for production and international programme exchange

BT Series

Broadcasting service (television)

International Telecommunication Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at http://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2016

REPORT ITU-R BT.2390-1

High dynamic range television for production and international programme exchange

(2016)

Summary

Recommendation ITU-R BT.2100 – Image parameter values for high dynamic range television for use in production and international programme exchange, specifies parameters for High Dynamic Range television (HDR-TV) signals to be used for programme production and international programme exchange. This report provides background information on HDR in general, and for the perceptual quality (PQ) and hybrid log-gamma (HLG) HDR signal parameters specified in the Recommendation.

As HDR-TV is at a formative stage of research and development as presented in this Report, a call for further studies is made, in particular on the characteristics and performance of the recommended HDR-TV image parameter values, for use in broadcasting.

1 Introduction and design goals for HDR television

HDR-TV enables more natural images that contain wider variations in brightness. While HDR-TV does allow the picture average brightness to increase, the expectation is that indoor scenes produced in HDR will generally be at a similar brightness as with legacy TV systems. The brightness range available with HDR enables outdoor sunlit scenes to appear noticeably brighter than indoor scenes, thus providing a more natural look. All scenes, especially outdoor, will be able to produce small area highlights such as specular reflections or emissive light sources at much higher brightness. There is also an improvement in the ability to show details in dark areas; this feature is dependent on the black level of the display and the viewing environment.

1.1 Common misconceptions on HDR

HDR for video and display is an entire ecosystem that encompasses much more than the words underlying the acronym. Before discussing system issues, there are number of frequent misconceptions about HDR video, such as: ‘It is all about brighter pictures’, ‘It is all about dynamic range’ , ‘It is all about bit-depth’, ‘It is primarily an image capture issue’, ‘It is primarily a display capability issue’, ‘It makes images look like paintings’.

Of these, we will only address the first one here. The misconception about HDR being **simply brighter**[1] pictures arises from the fact that the maximum luminance *capability* is indeed much higher than standard dynamic range (SDR) television. However, this higher maximum is primarily used by the highlight regions of images. While the highlights will indeed appear brighter [1], they are nearly always small in region, and the overall image may not necessarily appear brighter. This is because the overall appearance of an image’s brightness is dominated by the average brightness, not the small regions usually occupied by highlights. One type of highlight is the specular reflection. The advantages of having more accurate specular reflections enabled by HDR include better surface material identification [2] as well as in depth perception, even with 2D imagery [3] [4].

---

1 Brightness is technically a perceptual measure, and not linear to luminance. However, in the majority of consumer TV literature, ‘brightness’ is used to convey either overall luminance, or the maximum luminance. We will use the term in that sense here.

By comparison, in the process of making the SDR content (whether colour grading in post-production or selection of the camera settings in live broadcast), human decisions are invariably made to fit the higher dynamic range of the scenes into the standard range.

In typical practice, highlights are processed through a shoulder operation or simply clipped. This loses not only the amplitudes of the highlights, but also the details within and around the highlights. Similarly, shadow detail is lost. Colour emissive highlights result in the colour component going through different portions of the shoulders such that the colour shifts towards white. These different aspects resulted in the realization that a new HDR signal format needed to be developed to allow for the HDR display to truly deliver an HDR experience.

There *is* another way to utilize the new range capabilities than to utilize it solely for highlights. This is to allow for more realistic scene-to-scene luminance variations. In current SDR, with a range of less than three log10 luminance, it was always difficult to render evening scenes, and nearly impossible to render the luminance differences of indoor and outdoor scenes. Acknowledging this limitation with SDR, some creatives like to use the increased dynamic range of HDR to have larger scene-to-scene variations in mean luminance. So for this particular approach, HDR may result in brighter images for some scenes.

However, despite these variations in intent for invoking increased brightness, HDR also allows for lower black levels than traditional SDR, which was typically in the range between 0.1 and 1.0 cd/m^2 for cathode ray tubes (CRTs), and is now in the range of 0.1 cd/m^2 for most standard SDR liquid crystal displays (LCDs). So a key design question is how low should the black level be?

1.2 System black level determination

In order to determine the system black level, the state of light adaptation[2] is central. The classic psychophysical study on dark adaptation was by Hecht *et al* [5], which corresponds to the top data line of the plot in Fig. 1, which is a compilation of more recent studies [6]. The left branch of the curve corresponds to the cones, while the right branch of the curve corresponds to rod vision. While threshold values of less than 0.00001 cd/m^2 can be obtained, they can take significant durations of dark adaptation, which are not likely in entertainment media. If one restricts consideration to cone vision's left branch of the uppermost curve, we can see visibility doesn't go as low, but it still can be below ~0.02 cd/m^2.

However, detectability as low as 0.02 cd/m^2 seems to require minutes of dark adaptation time, which in traditional entertainment media is considered unrealistic[3] . Often, the early part of the curve (< 1 minute) is used to conclude that black levels of between 0.3 and 1.0 cd/m^2 are sufficient, and in previous years display capability has been limited to be greater than 0.1 cd/m^2 (e.g. for fixed backlight LCD). Using data such as those presented in Fig. 1 to conclude that the human eye cannot see black level differences below 0.1 cd/m^2 overlooks that the curves depend on the initial adaptation condition. The other curves shown in the figure show that as the initial adaptation level is lowered, the ability to see lower luminance levels improves. While the plotted time scale does not allow for determination of adaptation ranges on the order of video scene cuts (3-5 s), the leftmost data points are enough to show that visual detectability of black level can be close to 0.001 cd/m^2 for the 25 cd/m^2 initial level, close to SDR average luminance levels (i.e. average picture level (APL)). Thus from Fig. 1, one would easily conclude that the black level of video should allow levels as low as 0.001 cd/m^2.

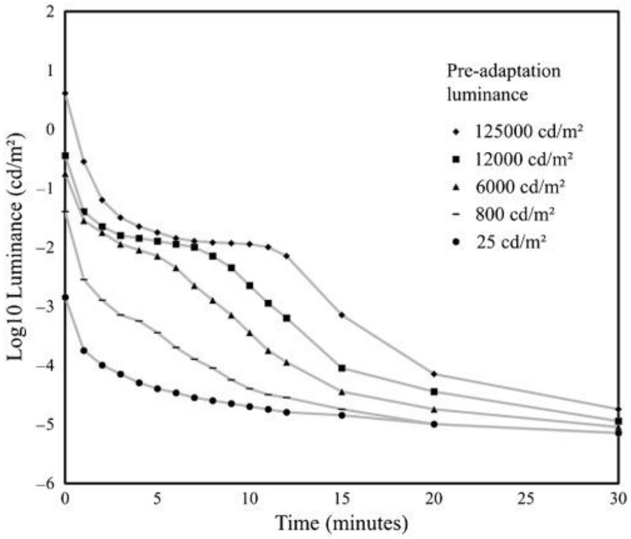
---

[2] Sometimes called dark adaptation when adapting toward dark.

[3] Creatives in production and post have desired to allow for longer periods of dark adaptation in their content.

FIGURE 1

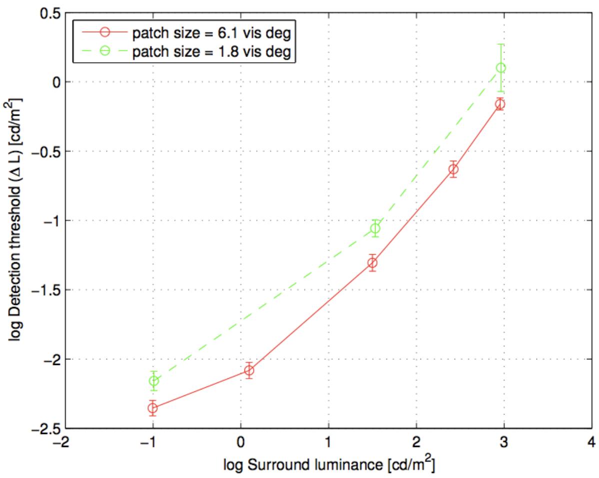
Black level detectability as a function of duration for different initial adaptation levels. From Stokkerman [6]



However, system design by the use of data as in Fig. 1 leans toward the most demanding cases, where the entire image may be dark. Other approaches consider that images generally do not consist of all-dark regions; there is a mixture of different luminance levels. The general approach is to treat the image as a surround around a possible black area. Using rectangular patches with a white surround, Mantiuk *et al* [7] studied black level threshold as a function of the size of the black region. The area outside of the patch was termed the surround, and the surround serves as a surrogate for an actual image with average image luminance level. The results in Fig. 2 show the lowest black level that can be discriminated from zero luminance is ~−2.4 log10 cd/m^2 (0.0039 cd/m^2), at least for the darkest surround that they studied, which was 0.1 cd/m^2. Lower thresholds would be expected from darker surrounds, such as might occur in home theatre, or some evening viewing situations.

FIGURE 2

Detectability of black level differences for a rectangular patch of either 6.1 or 1.8 visual degrees, both as a function of surround luminance level



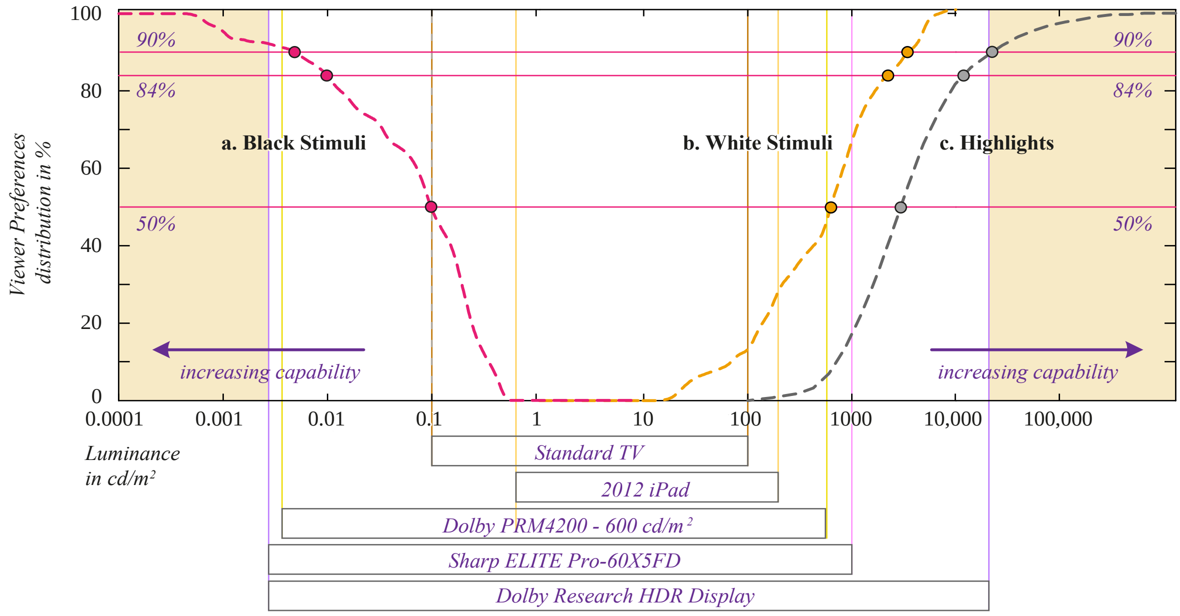
Two things are clear. As the surround luminance decreases, the detectable black level decreases. That is, the expected surround luminance that results from practical imagery can determine the necessary black level to achieve a pure black perception, as well as finding the level where dark detail is no longer distinguishable. The other effect is that thresholds for the larger black region are lower than for the smaller. Thus in designing a system black level, the expected size of the black region is a key factor. Note that the largest region studied in this work was 6 degrees, whereas the image size for HDTV viewed at 3H is approx. 35 degrees (UHDTV @ 1.5 H is ~70 degrees).

Another approach for determining system black level is to not base it on psychophysical detection tasks with abstract geometric stimuli, but rather use preferences while viewing more natural imagery. Rempel at al. [8] measured preference for display black level and brightness in short video clips (a sitcom) and found all participants consistently set the black level to the lowest possible setting, which was about 0.3 cd/m^2 for their display. So the only conclusion from this was that 0.3 is not low enough. A more recent study using an experimental HDR display with very low black level capability [9] [10] [11] found levels near *its* minimum capability, which was 0.004 cd/m^2. In order to meet the preferences of 90% of the viewers, a level of 0.005 cd/m^2 was needed. The typical current black level LCD TVs of 0.1 cd/m^2 would meet the preferences of only half of the viewers. Results are shown in Fig. 3.

The plot in Fig. 3 demonstrates the results of psychophysical experiments designed to understand the preferred dynamic range [9] [10] [11]. The experiment was based on a two-alternative forced choice paradigm using static images shown sequentially for average shot durations (2-5 s) and trial durations of around 20 s to include response times, for an experiment lasting a total of 40 minutes per participant. The stimuli were drawn from three classes of images, containing shadow detail, reflective white stimuli, and highlight stimuli. A dual modulation display was used using an LCD panel backlit by a digital cinema projector, allowing a luminance range between 0.004 and 20 000 cd/m^2. Separate experimental sessions were conducted for the black level scenes vs. the white and highlight level scenes; the results of all the experiments are plotted on the same figure but this should not be interpreted as indication that both extremes can be perceived simultaneously.

FIGURE 3

Cumulative distribution functions for a. black stimuli, b. reflective white stimuli and c. emissive and highlights. For comparison, the dynamic ranges of common displays are given



Regarding the black level, there are a number of studies that found detectability as well as preferences well below the level of 0.1 cd/ m^2, which was common for SDR displays. Values in the range of 0.001 to 0.005 cd/ m^2 could be deduced from the studies described here, and regarding preferences there may be upward biases due to the smaller field of view used in [9] than occurs with UHDTV.

1.3 System white and highlight level determination

In video, the system white is often referred to as reference white, and is neither the maximum white level of the signal nor that of the display. When calibration cards are used to set the reference white, it is a diffuse white (also called matte) that is placed on the card, and measured. The ideal diffuse white has a Lambertian reflection. The luminances that are higher than reference white are referred to as highlights. While there are several key quality dimensions and creative opportunities opened up by HDR (e.g. shadow detail, handling indoor and outdoor scenes simultaneously, and colour volume aspects), one of the key differentiators from SDR is the ability for more accurate rendering of highlights. These can be categorized as two major scene components: specular reflections[4] and emissives (also referred to as self-luminous). They are best considered relative to the maximum diffuse white luminance in the typical image. Most scenes can be broken down into two key ranges: object's diffuse reflectances and the highlights. (Some scenes would defy such categorization, e.g. fireworks at night.) The object's reflectance is important to convey its shape due to shading and other features, and the visual system has strong ability to discount the illuminant to be able to estimate the reflectance [12].

However, the human ability to perceive both types of highlights is much less accurate and less computationally sophisticated as the ability perceive reflectances [12]. Illustrations of emissives and specular highlights are shown in Fig. 4.

FIGURE 4

Emissive light sources, specular reflections, and diffuse white

In traditional imaging, the range allocated to these highlights was fairly low and the majority of the image range was allocated to the diffuse reflective regions of objects. For example, in hardcopy print the highlights would be 1.1x higher luminance than the diffuse white maximum [13]. In traditional video, the highlights were generally set to be no higher than 1.25x the diffuse white. Of the various display applications, cinema allocated the highest range to the highlights, up to 2.7x the diffuse white.

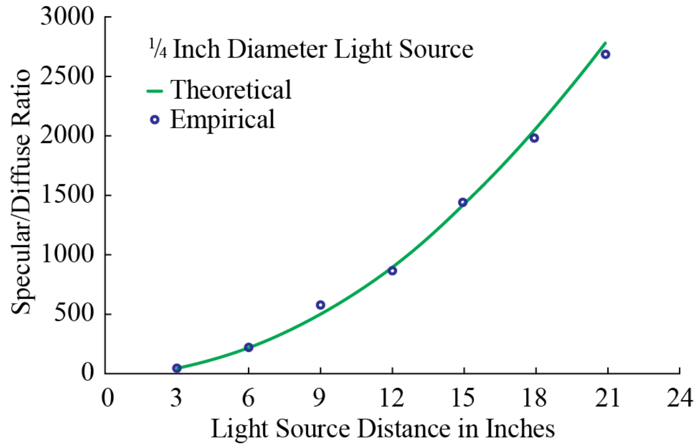
Actual measurements show the specular regions can be over 1 000x higher than the underlying diffuse surface [2], which is presented in Fig. 5. This means the physical dynamic range of the specular reflections vastly exceed the range occupied by diffuse reflection. If a visual system did not have specialized processing as previously described, and saw in proportion to luminance, most objects would look very dark and the visible range would be dominated by the specular reflections. Likewise, emissive objects and their resulting luminance levels can have magnitudes much higher

[4] In traditional photography, the term 'highlights' is sometimes used to refer to any detail near white, such as bridal lace, which may entirely consist of diffuse reflective surfaces. In HDR literature, the use of 'highlights' is intended for the specular or emissive regions in an image since that is a key feature opened up by HDR.

than the diffuse range in a scene or image. The most common emissive object, the disk of the sun, has a luminance so high (~1.6 billion cd/m^2), it is damaging to the eye to look at more than briefly, and exceeding even the speculars. A more unique aspect of the emissives is that they can also be of very saturated colour (sunsets, magma, neon, lasers, etc.).

FIGURE 5

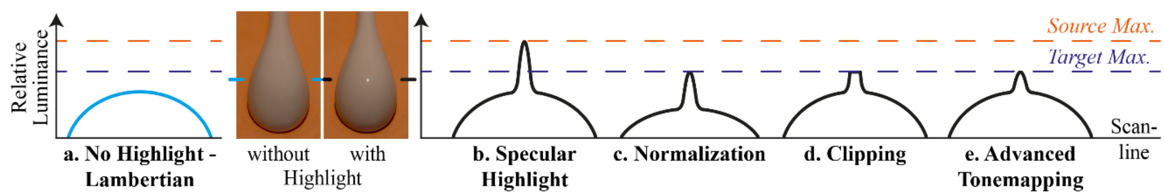
Measurements showing that the specular regions can be over 1 000x higher in comparison to the underlying diffuse surface. After Wolff (1994)



With traditional imaging's under-representation of highlight ranges, the question arises: what happens to the luminances of highlights? Figure 6 shows example scanlines of common distortions from a specular highlight from a glossy object, (b). It exceeds the maximum luminance of the display (or the signal), indicated as the dashed line titled 'Target Max.'. Illustration (c) shows a distortion that is seldom selected, that is, to renormalize the entire range. Another approach, (d) preserves diffuse luminances, and the highlight is simply truncated (hard-clipping). Details within the highlight region are replaced with constant values, giving rise to flat regions in the image, looking quite artificial. Typical best practices (e), have been referred to as soft-clipping, or a knee. Here the shape and internal details of the highlight are somewhat preserved, without flattened regions. HDR allows for a result closer to scanline (b). The more accurate presentation of specular highlights, (assuming the entire video pathway is also HDR), is one of the key distinctions of HDR. A number of perceptual papers have looked closely at specular reflection, as mentioned in the beginning of this section. Preferences of luminances for diffuse white and highlights are shown in Fig. 3.

FIGURE 6

Effects of highlight rendering, clipping and (tonescale) compression



2 Television system architecture

2.1 The relationship between the OETF, the EOTF and the OOTF

This Report makes extensive use of the following terms:

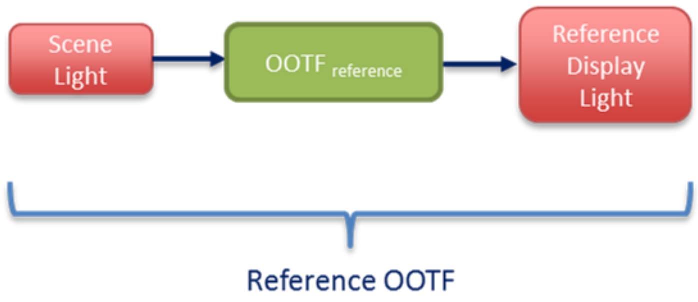
OETF: the opto-electronic transfer function, which converts linear scene light into the video signal, typically within a camera.

EOTF: electro-optical transfer function, which converts the video signal into the linear light output of the display.

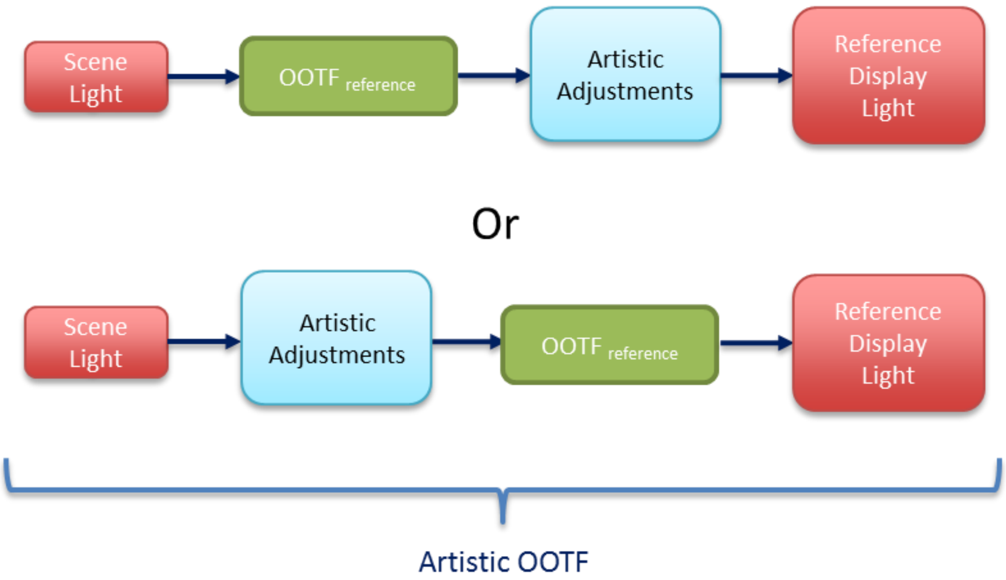
OOTF: opto-optical transfer function, which has the role of applying the "rendering intent".

These functions are related, so only two of the three are independent. Given any two of them the third one may be calculated. This section explains how they arise in television systems and how they are related.

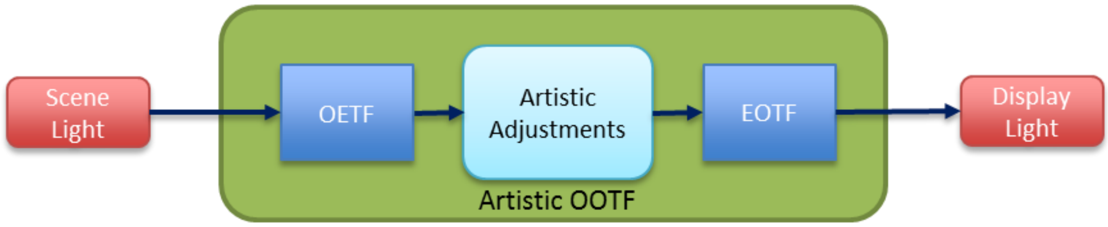
In television systems the displayed light is not linearly related to the light captured by the camera. Instead an overall non-linearity is applied, the OOTF. The "reference" OOTF compensates for difference in tonal perception between the environment of the camera and that of the display. Specification and use of a "reference OOTF" allows consistent end-to-end image reproduction, which is important in TV production.



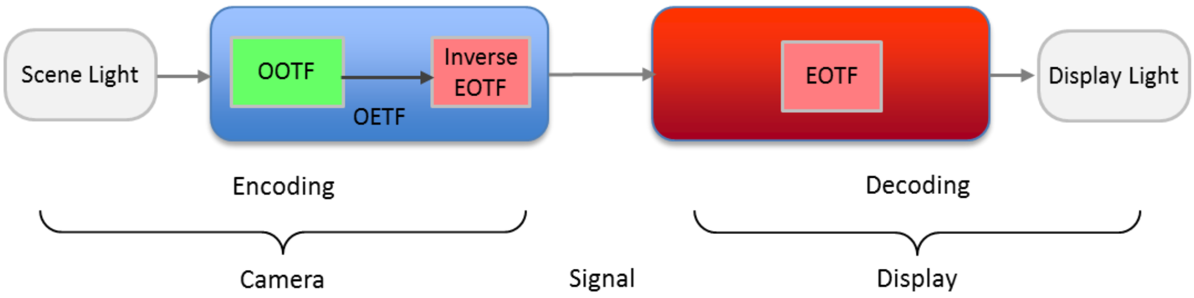
Artistic adjustment may be made to enhance the picture. These alter the OOTF, which may then be called the "artistic OOTF". Artistic adjustment may be applied either before or after the reference OOTF.



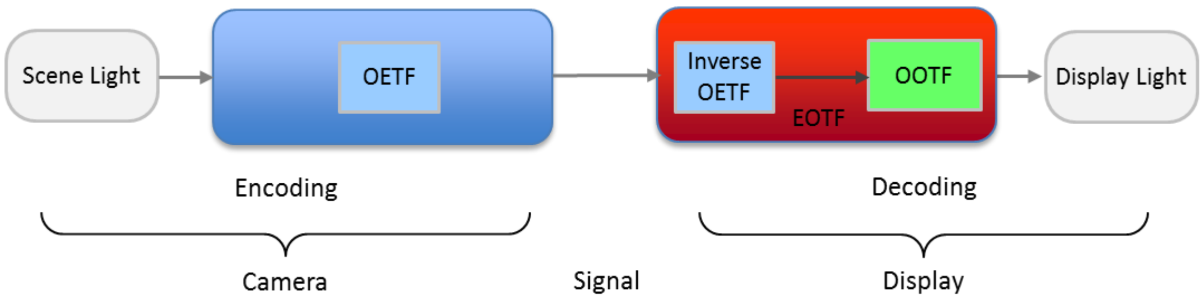
In general the OOTF is a concatenation of the OETF, artistic adjustments, and the EOTF.



The PQ system was designed with the model shown below, where the OOTF is considered to be in the camera (or imposed in the production process):



The HLG system the system was designed with the model shown below, where the OOTF is considered to be in the display:



Only two of three non-linearities, the OETF, the EOTF, and the OOTF, are independent. In functional notation (where subscripts indicate the colour component):

$$\mathrm{OOTF}_R(R,G,B) = \mathrm{EOTF}_R(\mathrm{OETF}_R(R,G,B))$$
$$\mathrm{OOTF}_G(R,G,B) = \mathrm{EOTF}_G(\mathrm{OETF}_G(R,G,B))$$
$$\mathrm{OOTF}_B(R,G,B) = \mathrm{EOTF}_B(\mathrm{OETF}_B(R,G,B))$$

This is clearer if we use the symbol $\otimes$ to represent concatenation. With this notation we get the following three relationships between these three non-linearities:

$$\mathrm{OOTF} = \mathrm{OETF} \otimes \mathrm{EOTF}$$
$$\mathrm{EOTF} = \mathrm{OETF}^{-1} \otimes \mathrm{OOTF}$$
$$\mathrm{OETF} = \mathrm{OOTF} \otimes \mathrm{EOTF}^{-1}$$
$$\mathrm{OOTF}^{-1} = \mathrm{EOTF}^{-1} \otimes \mathrm{OETF}^{-1}$$
$$\mathrm{EOTF}^{-1} = \mathrm{OOTF}^{-1} \otimes \mathrm{OETF}$$
$$\mathrm{OETF}^{-1} = \mathrm{EOTF} \otimes \mathrm{OOTF}^{-1}$$

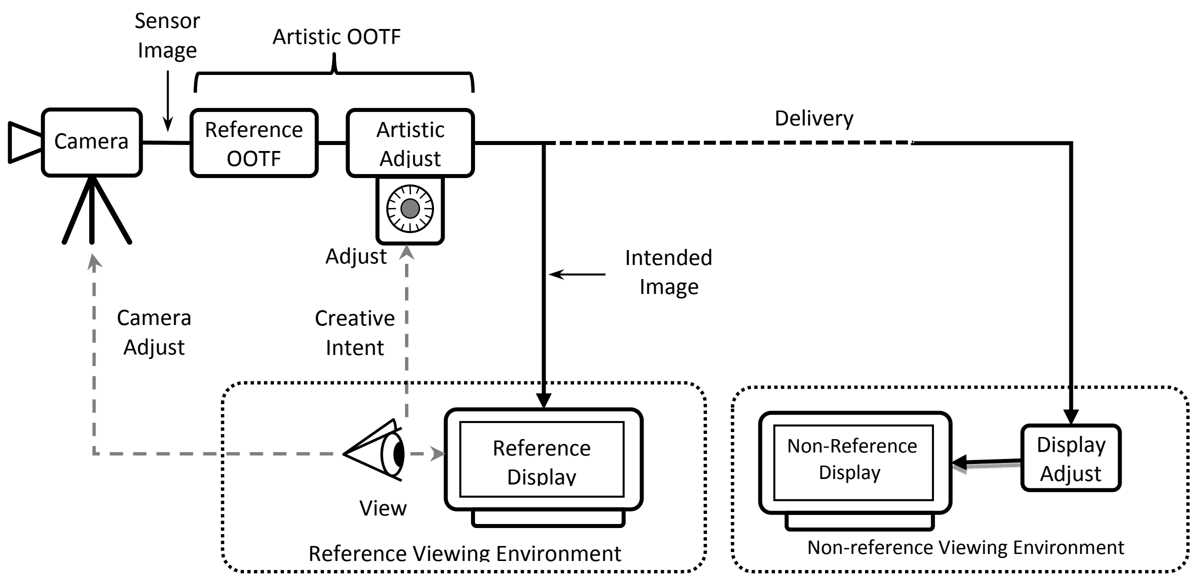
The PQ approach is defined by its EOTF. For PQ the OETF may be derived from the OOTF using the third line of the equations above. In a complementary fashion the HLG approach is defined by its OETF. For HLG the EOTF may be derived from the OOTF using the second line of the equations above.

2.2 Conceptual TV system showing basic concepts

Figure 7 is a high level conceptual flow of a simplified television system that does not employ a non-linearity (such as gamma) in order to reduce the bit depth needed to represent the baseband signal; such a non-linearity is needed in signal pipelines that have limited bit depths (e.g. limitations to 8-12 bit values), but these pipelines will be considered later and the conceptual system described here is considered to have no such restrictions. In Fig. 7, the camera outputs a linear light signal, which is representative of the scene in front of the lens. Exposure controls (camera iris and filters) perform a global scaling so the camera output is proportional to absolute scene light. The signal can be represented by high bit-depth integers, or for more efficiency, as 16-bit floating point. Non-reference viewing includes consumer viewing, as well as much TV production which often takes place in non-reference environments.

FIGURE 7

The conceptual TV system



A linear display of the scene light would produce a low contrast washed out image as illustrated in Fig. 8. Therefore, the signal is altered to impose rendering intent, i.e. a *Reference OOTF* (opto-optical transfer function) roughly like that shown in Fig. 9. The sigmoid curve shown increases contrast over the important mid-brightness range, and softly clips both highlights and lowlights, thus mapping the possibly extremely high dynamic range present in many real world scenes to the dynamic range capability of the TV system.

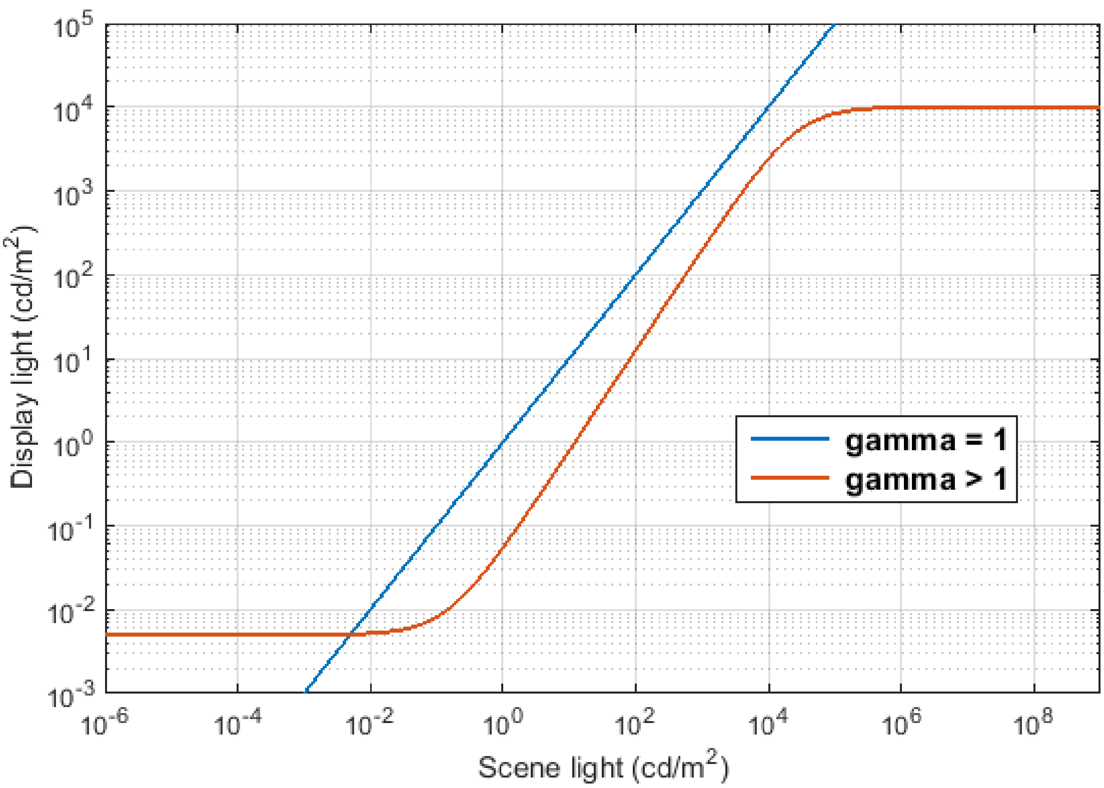
FIGURE 8

The left image has a system transfer function (or greyscale) of unity slope. The right image has a system transfer function consistent with ITU broadcast practices. From [Giorgianni2009]

A reference display in a reference viewing environment would, ideally, be used for viewing in production, and adjustments (e.g. iris) are made to the camera to optimize the image. Use of the *Reference OOTF* to produce images, with viewing done in the reference viewing environment, allows consistency of produced images across productions. If an artistic image “look” different from that produced by the reference OOTF is desired for a specific programme, “Artistic adjust” may be used to further alter the image in order to create the image “look” that is desired for that programme. Artistic adjustments may be made through the use of camera settings or after image capture during editing or in post-production. The combination of the reference OOTF plus artistic adjustments may be referred to as the “Artistic OOTF”.

FIGURE 9

Typical sigmoid used to map scene light to display light; extreme highlights and dark areas are compressed/clipped, the mid-range region employs a contrast enhancing gamma>1 characteristic



On the receive side where the consumer will view the image, if the consumer display is capable, and the consumer viewing environment is close to that of the reference viewing environment (dim room), then the consumer can view the image as intended. There may be limitations on both the viewing environment and the display itself. The viewing environment may be brighter than the reference environment, and the display may be limited in brightness, blackness, and/or colour gamut. Figure 7 shows “display adjust” as an alteration made to accommodate these differences from the reference condition. To compensate for a brighter environment, display adjust may lift the black level of the signal. To accommodate limited brightness capability of the display, system gamma may be changed or a “knee” may be imposed to roll off the highlights. To accommodate a limited colour gamut, gamut mapping would be performed to bring the wide gamut of colours in the delivered signal into the gamut that the display can actually show.

In practice television programmes are produced in a range of viewing environments using displays of varying capabilities. Thus similar adjustments are often necessary in production displays to achieve consistency.

3 The legacy television architecture

Since its beginning, television has employed restricted signal pipelines. Limited signal-to-noise ratios in the analogue days have transitioned to limited bit depths in the digital age. A non-linearity in the basic video signal was required in order to improve the visible signal-to-noise ratio in analogue systems, and the same non-linearity helps to prevent quantization artefacts in digital systems. This is the typical “gamma” curve that is the natural characteristic of the CRT, and that is documented in Recommendations ITU-R BT.709, BT.1886, and BT.2020.

Until recently all displays were based on the CRT which, based on the common physics, all had a similar characteristic function converting the electrical signal to light, the so-called “electro-optical transfer function” or EOTF. The camera characteristic of converting light into the electrical signal, the “opto-electronic transfer function” or OETF, was adjusted to produce the desired image on the reference CRT display device. The combination of this traditional OETF and the CRT EOTF yielded the traditional OOTF. The non-linearity employed in legacy television systems (Recommendations ITU-R BT.601, BT.709 and BT.2020) is satisfactory in that 10-bit values are usable in production and 8-bit values are usable for delivery to consumers; this is for pictures with approximately 1 000:1 dynamic range[5], i.e. 0.1 to 100 cd/m^2.

3.1 HDTV as specified in Recommendations ITU-R BT.709 and BT.1886

Recommendation ITU-R BT.709 explicitly specifies a reference OETF function that in combination with a CRT display produces a good image. Creative intent to alter this default image may be imposed in either the camera, by altering the OETF, or in post-production, thus altering the OOTF to achieve an “artistic” OOTF. As the CRT is no longer manufactured, it became impractical to rely on the inherent CRT characteristic in order to achieve uniformity in reference displays. In the year 2011 Recommendation ITU-R BT.1886 was approved; this new Recommendation specified the EOTF of the reference display to be used for HDTV production; the EOTF specification is based on the CRT characteristics so that future monitors can mimic the legacy CRT in order to maintain the same image appearance in future displays. A reference OOTF is not explicitly specified for HDTV.

---

[5] This definition of dynamic range refers to the luminance ratio between the dimmest and brightest possible pixels presented on the display. However quantization artefacts, known as banding, may be visible, particularly in low lights, at luminance levels substantially brighter than the dimmest pixel. Quantization artefacts may, therefore, limit the “effective” dynamic range that is free from banding.

Nevertheless, as shown in Fig. 10, in practice it exists as the cascade of the specified OETF (BT.709) and EOTF (BT.1886).

FIGURE 10

The BT.709 HDTV television system architecture

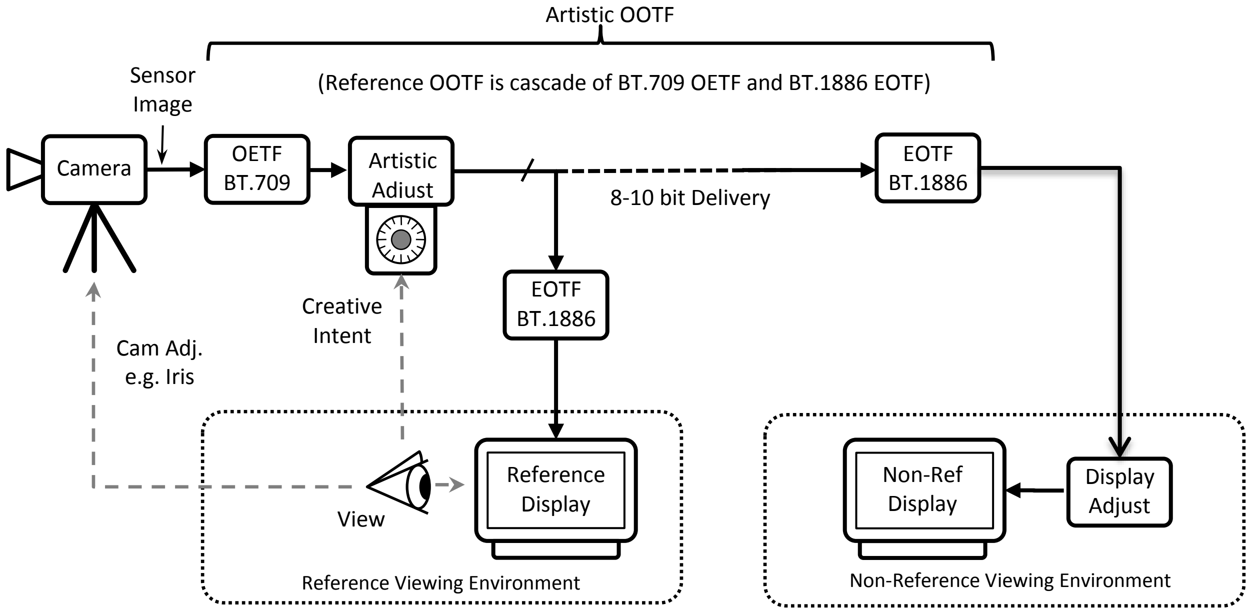


Figure 10 shows the HDTV system. The linear light is encoded into a non-linear signal using the OETF specified in Recommendation ITU-R BT.709. Creative intent may be imposed by altering this encoding or in post-production by adjusting the signal itself; this can be considered as an alteration outside of the Recommendation ITU-R BT.709 OETF (e.g. as “artistic adjust” in the diagram). Recommendation ITU-R BT.1886 specifies the conversion of the non-linear signal into display light. This drives the reference display in the reference viewing environment. The image on the reference display drives adjustment of the camera iris/exposure, and if desired, artistic adjust can alter the image to produce a different artistic look. At the receiver (ideally a reference display in a reference viewing environment) the non-linear signal is converted to display light using the Recommendation ITU-R BT.1886 specified function. There is typically further adjustment (display adjust) to compensate for viewing environment, display limitations, and viewer preference; this alteration may lift black level, effect a change in system gamma, or impose a “knee” function to soft clip highlights. (In practice the EOTF gamma and display adjust functions may be combined in to a single function.)

In a typical TV system the soft clipping of the highlights (sometimes known as the “shoulder”), described earlier and illustrated in Fig. 3, is implemented in the camera as a camera “knee”. This is part of the artistic adjustment of the image. Part of the low light portion of the characteristic (sometimes known as the “toe”) is implemented in the display as a black level adjustment. This adjustment takes place in the display as part of the Recommendation ITU-R BT.1886 EOTF and implements soft clipping of the lowlights.

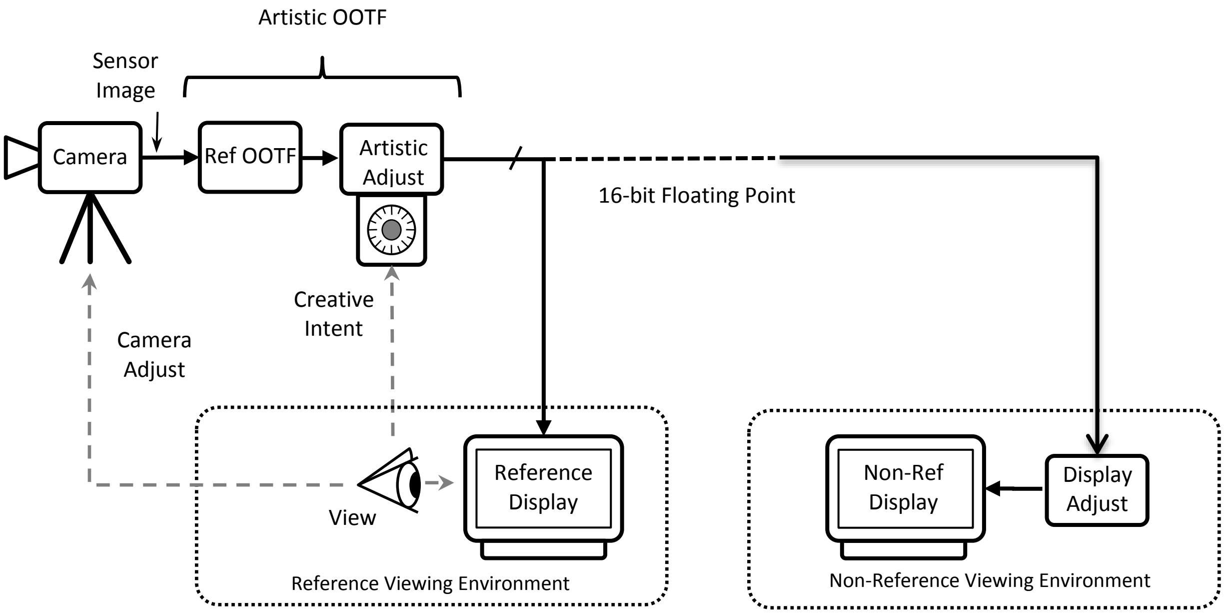
There is no clearly defined location of the reference OOTF in this system. The reference OOTF is the cascade of the OETF and the EOTF, and the actual OOTF is the cascade of those plus the artistic and display adjustments. Any deviation from the reference OOTF for reasons of creative intent must occur upstream of delivery. Alterations to compensate for the display environment or display characteristics must occur at the display by means of display adjust (or a modification of the EOTF away from the reference EOTF).

4 RGB floating point HDR-TV system

A 16-bit RGB HDR system is defined for use when 48-bit/pixel pipelines are available. This architecture is shown in Fig. 11.

FIGURE 11

HDR floating point system



The reference OOTF is implemented directly after camera capture of the scene, and Adjust 1 may be used to make additional changes as desired for creative intent. The output of the OOTF block is 16 bit floating point format which allows for adequate precision even for large colour volumes. The floating point values directly represent light values on the display, i.e. R = G = B = 1.0 means 1.0 cd/m^2 of white for a pixel. As before, display adjust is used to compensate as much as possible for limitations of displays, and for environments that may differ from the reference viewing environment that was (ideally) used during programme production.

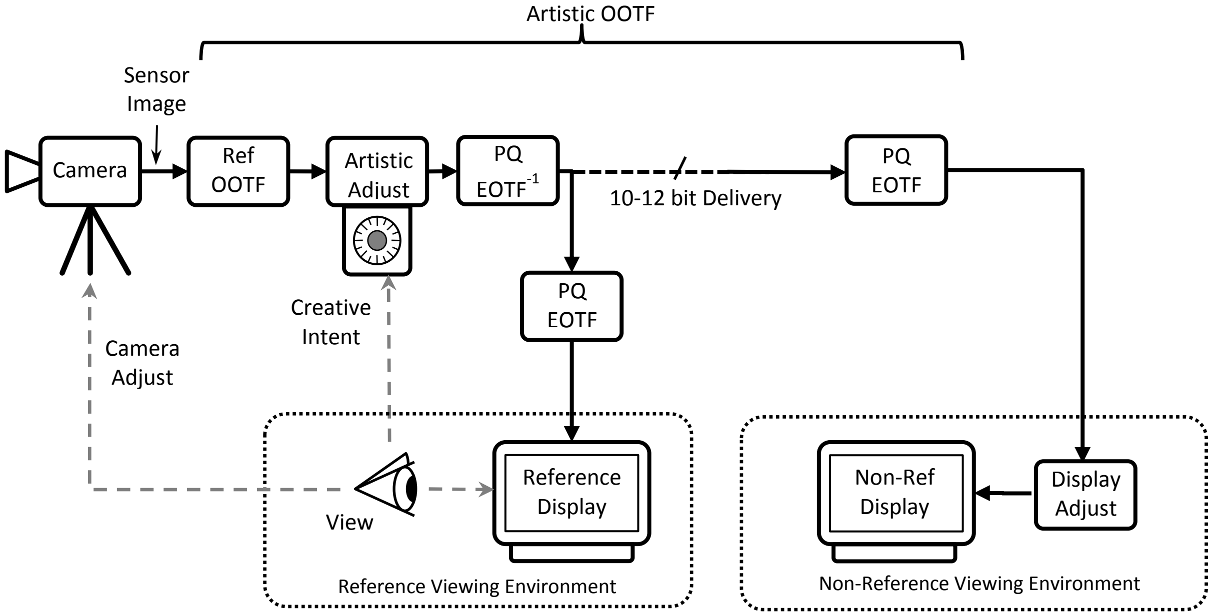
5 PQ HDR-TV

5.1 PQ system architecture

When bit-constrained pipelines are required for television production systems, then an HDR implementation very similar to the current HDTV system of Fig. 10 can be constructed. This implementation is shown in Fig. 12.

FIGURE 12

PQ HDR-TV system with 10 or 12 bit integer values



An optimized non-linear signal representation is used so that 10-12 bit depth values can accommodate the larger colour volume of HDR; otherwise this system is very similar to the HDTV system in use today. The PQ EOTF replaces the Recommendation ITU-R BT.1886 function of SDR HDTV, and the corresponding PQ OETF replaces the Recommendation ITU-R BT.709 OETF as the default camera capture curve. Once again Adjust 1 may be used to further modify the creative intent of the image, and Adjust 2 is used to adapt the signal for different consumer displays and display environments. No use of metadata is shown or required.

5.2 Design of the PQ non-linearity

As described in [14] the traditional gamma nonlinearities of Recommendations ITU-R BT.709 and BT.1886 are unsatisfactory when stretched to the much larger dynamic ranges desired for future television productions.

FIGURE 13

Contrast step size vs. display luminance for 12 bit signals

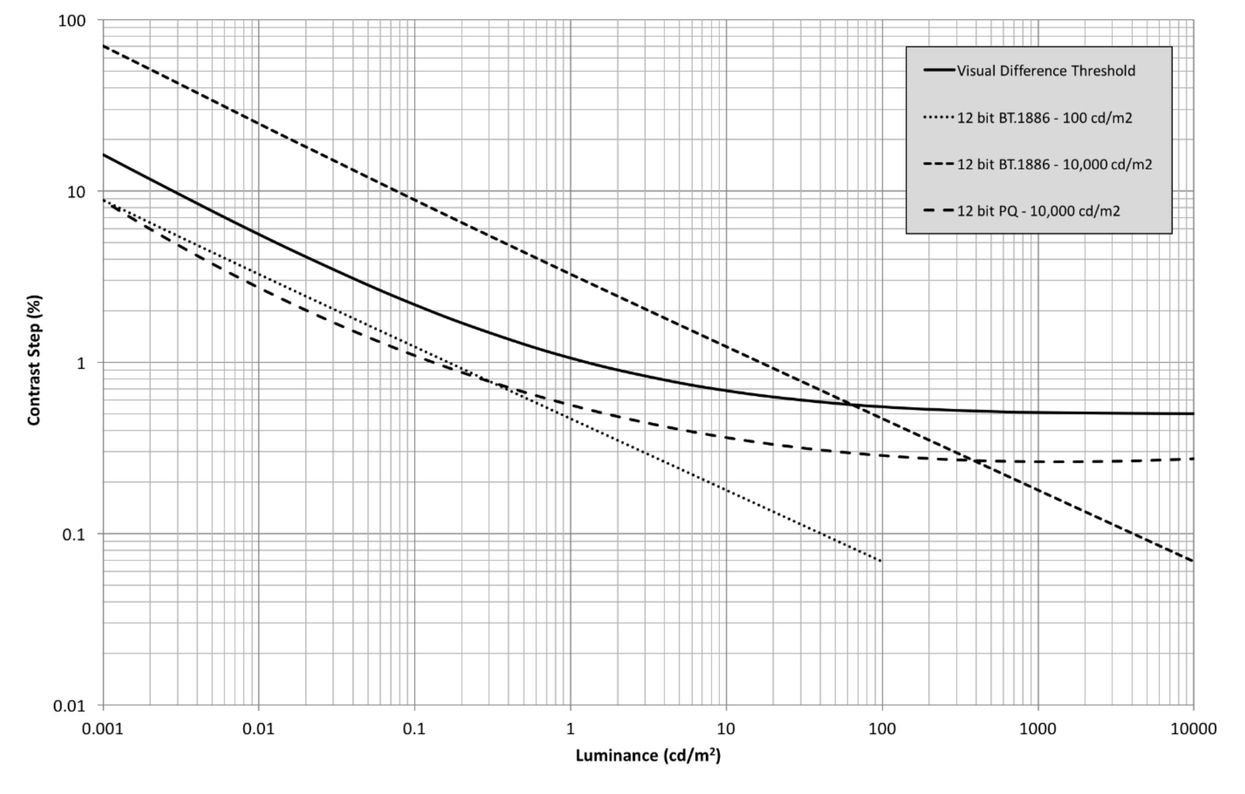


Figure 13 shows the approximate visual difference threshold as a solid black curve on a log-log plot with luminance on the x-axis and contrast step size (due to bit depth limitation) in % on the vertical axis. This threshold is based on the detailed Barten model of the human visual system. Lines which fall below this threshold curve will not exhibit any visible quantization artefacts such as image banding, while lines above the threshold curve may exhibit visual artefacts. While the legacy Recommendation ITU-R BT.1886 operating with a peak level of 100 cd/m^2 is comfortably below the threshold curve when using 12 bits, it rises substantially above the visual threshold when operating with a 10 000 cd/m^2 peak. A traditional "gamma" power function is not a good approximation for human vision over an extended range of luminance values (too many code words allocated to very bright regions and not enough allocated to dark regions). This inefficiency was not a serious problem with SDR systems due to their limited dynamic range, but when trying to represent HDR luminance ranges, an improved curve is required. By using the same Barten model as the visual threshold calculation itself, an optimized nonlinear function was developed for the PQ signal, which can operate over the entire range from 10 000 cd/m^2 down to less than 0.001 cd/m^2 without any visible quantization artefacts using 12 bit coding precision.

FIGURE 14

Contrast step size vs. display luminance for 10 bit signals

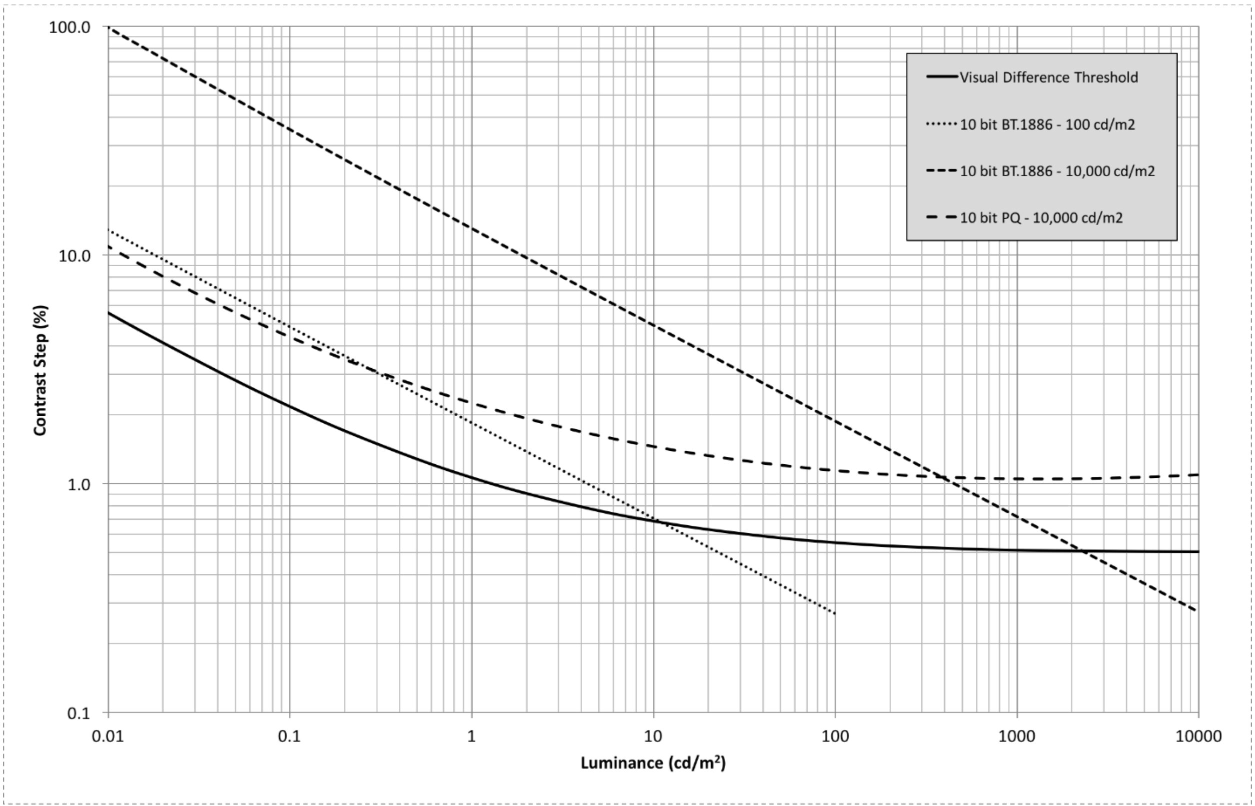


Figure 14 shows the same plots as Fig. 13 but with all three systems using 10 bit quantization. Though the signal lines all come above the threshold curve to some extent, experience has shown that with realistic camera noise levels, the slight quantization artefacts predicted for 100 cd/m^2 Recommendation ITU-R BT.1886 or 10 000 cd/m^2 are masked and thus do not present real problems in television production.

5.3 OOTF and OETF

This subsection describes the PQ opto-optical transfer function (OOTF) and the resulting opto-electronic transfer function (OETF). The PQ opto-optical transfer function is normatively specified in Recommendation ITU-R BT.2100, which is intended to be compatible with existing SDR Recommendation ITU-R BT.709 signal sources and Recommendation ITU-R BT.1886 compliant displays. This maximizes compatibility for mixed source applications wherein some sources are HDR and some are SDR. We want the image from an SDR source and that from an HDR source to match everywhere the HDR image brightness overlaps the range of the SDR source (the HDR OOTF extends up to the maximum PQ displayed light level of 10 000 cd/m^2).

5.3.1 Generalized OOTF from Recommendation ITU-R BT.1886 in combination with Recommendation ITU-R BT.709

In order to maximize compatibility with existing SDR signals we desire an OOTF consistent with the effective OOTF of existing practice which is:

$$OOTF_{SDR} = EOTF_{1886}\left[OETF_{709}\right] \qquad (1)$$

We only need to extend the range of $OETF_{709}$ and $EOTF_{1886}$ for HDR.

The extension factor for displayed light is 10 000 / 100 = 100.

As the SDR OOTF has a roughly gamma = 1.2 characteristic at the high end, the extension relative to scene light (the input to OOTF) is approximately $100^{1/1.2} = 46.42$. When the exact equations for Recommendations ITU-R BT.709 and BT.1886 are used, the extension for HDR is 59.5208.

To expand the range of $OETF_{709}$ to G_{709} for HDR the equation is therefore (HDR E normalized to range of 0 to 1):

$$E' = G_{709}[E] = \begin{matrix} 1.099(59.5208E)^{0.45} - 0.099 & \text{for } 1 > E > 0.018/59.5208 \\ 4.5(59.5208E) & \text{for } 0.018/59.5208 \geq E > 0 \end{matrix} \tag{2}$$

Consequently, the range of E' is [0, 6.813] for HDR while it remains [0,1] for SDR. To expand the range of $EOTF_{1886}$ to G_{1886} for HDR no change to the equation is necessary, we simply allow the argument to extend to 6.813 (from 1) and hence the range increases from 100 to 10 000:

$$G_{1886}[E'] = 100(E')^{2.4}$$

These extensions satisfy the boundary conditions:

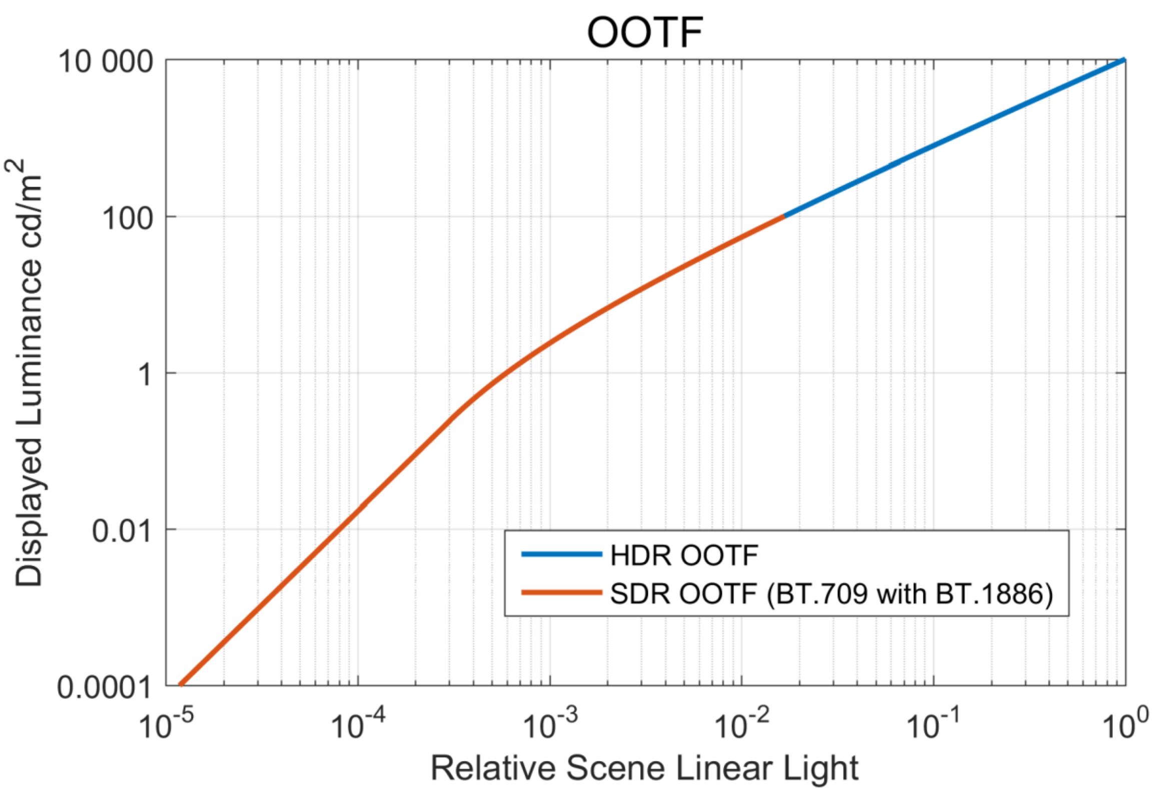
a) $E = 1$ produces a displayed luminance of 10 000 cd/m^2

b) $E = 1/(59.5208)$ produces a displayed luminance of 100 cd/m^2

The resulting OOTF is shown in Fig. 15. The x-axis, relative scene light is the same as E for SDR while for HDR it is 59.5208*E since the domain of E is [0,1]:

FIGURE 15

PQ and SDR OOTF

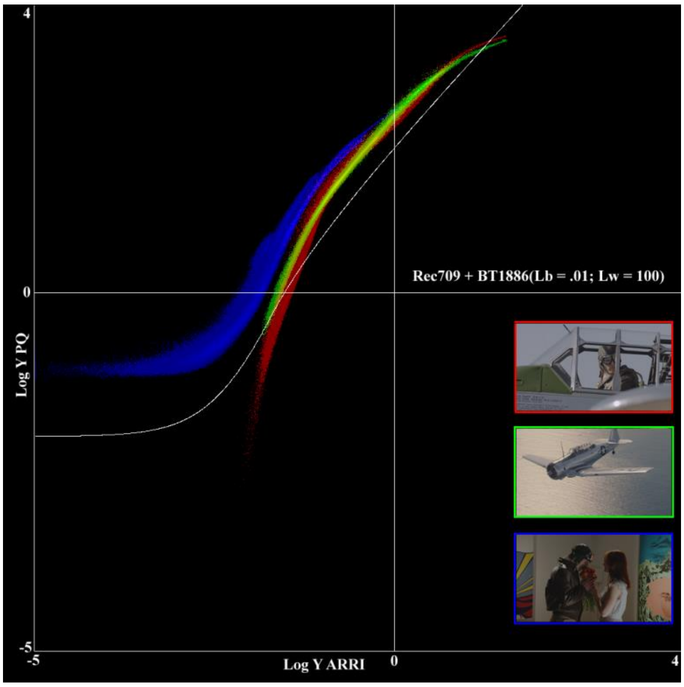


5.3.2 Actual OOTFs from manually graded content

It is instructive to compare this proposal with the actual OOTFs that are imposed when manually grading camera RAW output. The OOTF is the ratio of the graded linear output to the RAW linear input. Figure 16 shows several examples from the HDR sequence “Fantasy Flights”:

FIGURE 16

Extracted OOTFs from Fantasy Flights (3 of 3)



These Figures show scatter plots of the log of the output luminance derived from the PQ grade versus the log of the relative input luminance derived from the ARRI RAW camera output. These scatter plots are colour-coded (RGB) to match the images shown in the lower right corner of each figure. For comparison, we have plotted in white the OOTF from the combination of Recommendations ITU-R BT.1886 and BT.709. This shows that the extracted OOTFs are, as one would expect, a bit brighter than SDR. We can draw some preliminary conclusions from this experimental data:

1 For this manually graded content, the OOTF is not a straight line, and thus the actual OOTF does not correspond to an overall “system gamma”.

2 Darker indoor scenes tend to be noise limited at the bottom end and the OOTF exhibits a very clear toe.

3 The extracted OOTFs appear to have roughly the same curvature in the mid-tones as the proposed model.

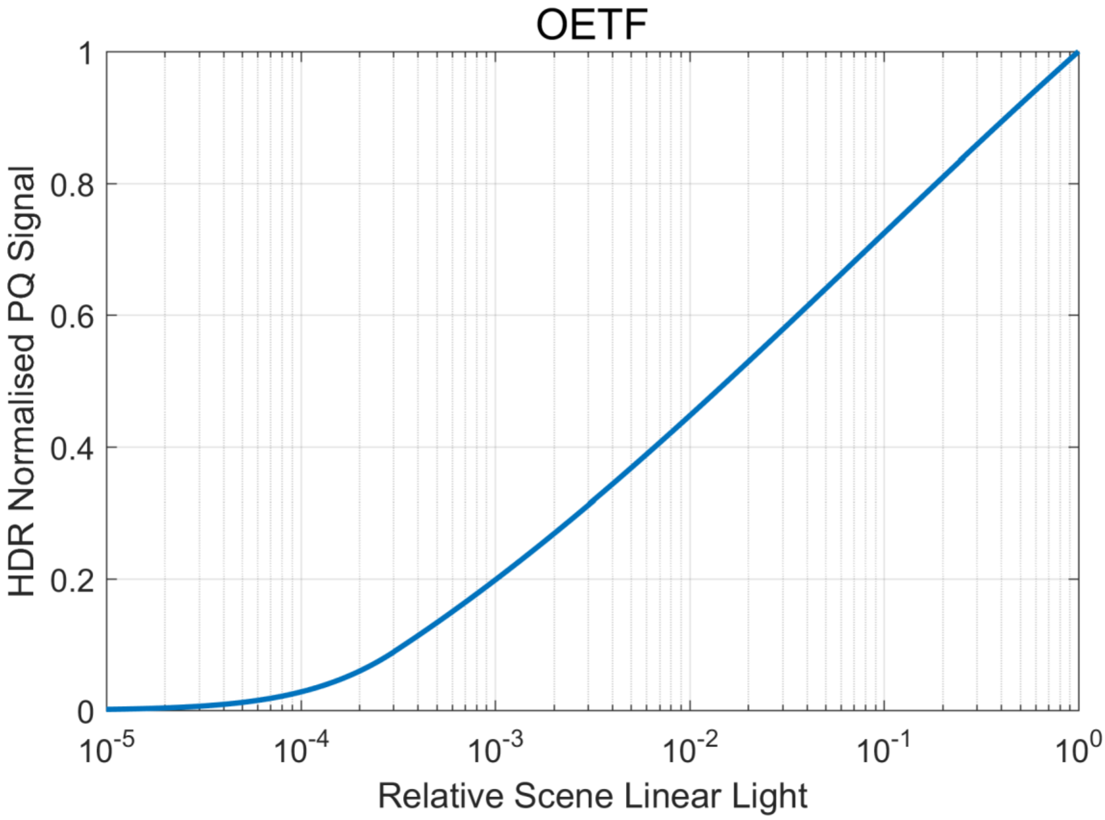
5.3.3 Resultant OETF

This OOTF can be combined with the inverse of the EOTF to produce an OETF. That OETF is shown in Fig. 17.

In actual cameras there is noticeable noise at low signal levels, and in practice the OETF slope at low levels is limited so as to “crush” the noise in black, thereby putting a “toe” into the response. The reference OETF does not have such a “toe”, but one is apparent in the OOTF plot for the indoor scene of “Fantasy Flights” shown above.

FIGURE 17

HDR OETF



This OETF:

– emulates the "look" of Recommendation ITU-R BT.709 plus Recommendation ITU-R BT.1886 for display light up to the limit of SDR;

– facilitates mixing of legacy Recommendation ITU-R BT.709 signals and PQ HDR signals;

– offers reasonable behaviour for levels above those of SDR.

5.4 Display mapping

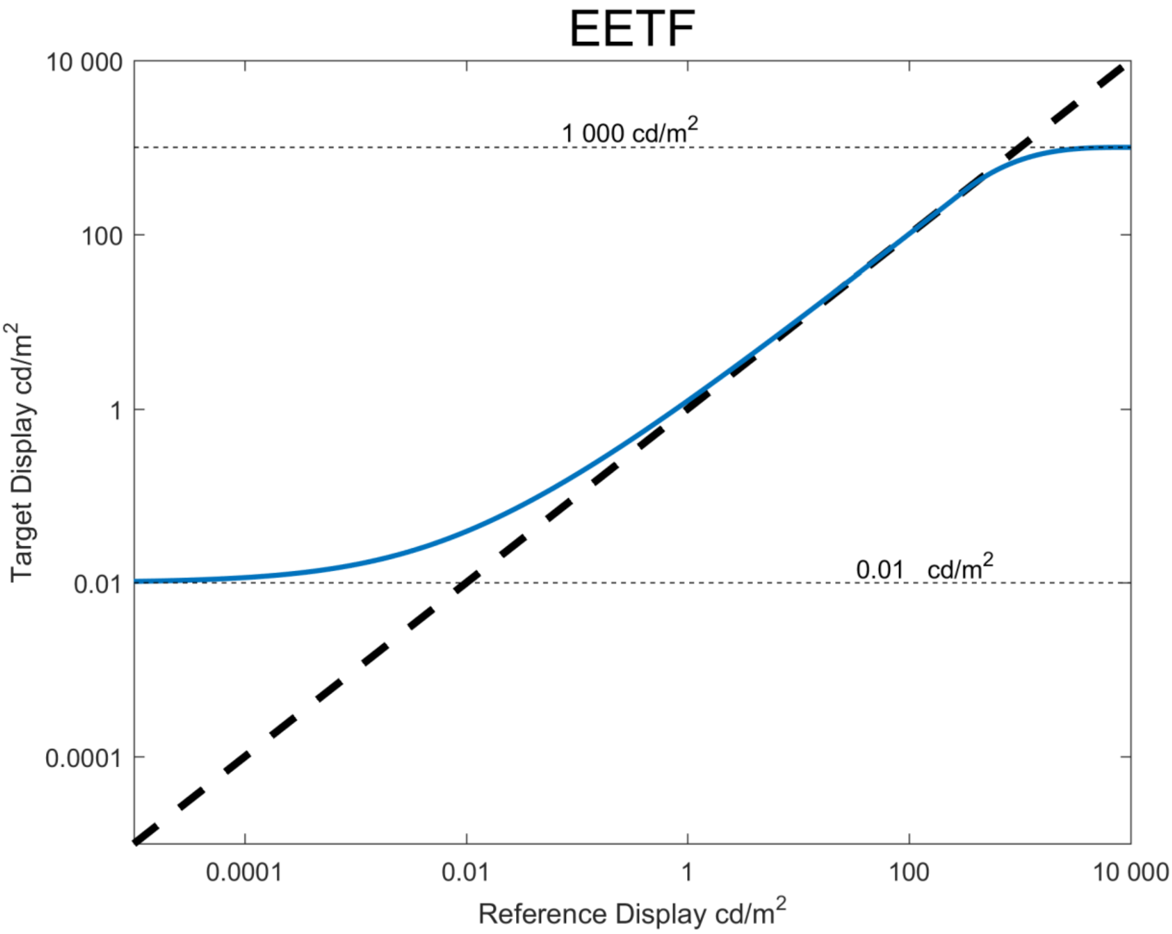
The PQ HDR system generates content that is optimum for viewing on a reference monitor in a reference viewing environment. The reference monitor would ideally be capable of accurately rendering black levels down to or below 0.005 cd/m^2, and highlights up to 10 000 cd/m^2. Also, the ideal monitor would be capable of showing the entire colour gamut within the Recommendation ITU-R BT.2020 triangle. The viewing environment would ideally be dimly lit, with the area surrounding the monitor being a neutral grey (6 500 degree Kelvin) at a brightness of 5 cd/m^2. However, content often must be viewed or produced in environments brighter than the reference condition, and on monitors that cannot display the deepest blacks or brightest highlights that the PQ signal can convey. In these cases the display characteristic needs to be changed in a process often referred to as display mapping (DM).

5.4.1 Mapping to display with limited brightness range

High dynamic range content may be viewed on displays that have less dynamic range than the reference display used to master the content. In order to view HDR content on displays with a lower dynamic range, display mapping should be performed. This can take the form of an EETF (electrical-electrical transfer function) in the display. This function provides a toe and knee to gracefully roll off the highlights and shadows providing a balance between preserving the artistic intent and maintaining details. Figure 18 is an example EETF mapping from the full 0 - 10 000 cd/m^2 dynamic range to a target display capable of 0.01 – 1 000 cd/m^2. The EETF may be introduced into the PQ signal; the plots show the effect of the mapping, i.e. how the intended light is changed into actual displayed light. In practice the mapping is done on the PQ signal.

FIGURE 18

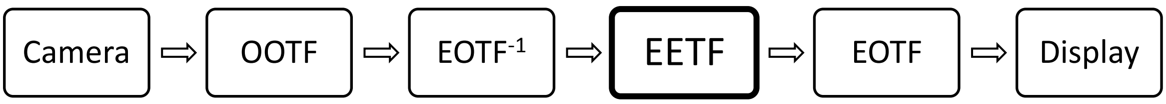
Example EETF From 0 - 10 000 cd/m^2 to 0.01 - 1 000 cd/m^2



Below are the mathematical steps that implement this tone mapping function for displays of various black and white luminance levels. The EETF may be applied in the non-linear domain to either the intensity (I) channel in IC_TC_P or to the luma (Y') channel of $Y'C'_BC'_R$. Applying the curve to individual R'G'B' components may introduce colour and saturation shifts so is not recommended. Figure 19 shows the block diagram of where the EETF should be applied.

FIGURE 19

Block diagram of signal chain showing location of EETF application



Calculating the EETF

The central region of the tone mapping curve is defined as a 1:1 mapping. A "knee" roll off may be calculated using a hermite spline to create a mapping that will reduce the luminance range to the capability of the display. The black level lift is controlled by an offset, b, which would be determined by a PLUGE adjustment. The difference between this proposal and the black level adjustment per Recommendation ITU-R BT.1886 is the addition of a tapering factor $(1 - E_2)^4$. Without such a tapering factor, a constant offset throughout the entire signal range has the effect of increasing the brightness at the high end. With Recommendation ITU-R BT.1886 this effect was limited and not problematic due to the large number of code values at the high end of the gamma curve. The perceptual uniformity of the PQ EOTF causes this effect to be unacceptable. The tapering function allows fine-tuning the lift without a significant impact on mid-tones or highlights.

The 1:1 mapping and knee are calculated first (E_2). The turning point (KneeStart or KS) for the spline is the point where the roll off will begin [15]. The luminance values, *minLum* and *maxLum*, are the maximum and minimum luminance values of the target display including ambient.

$$KS = 1.5maxLum - 0.5$$

$$\mathrm{b} = minLum$$

Solve for the EETF (E_3) with given end points. E_1 is the reference input signal in normalized PQ code words.

STEP 1:

$$E_2 = E_1 \qquad for\ E_1 < KS$$

$$E_2 = P(E_1) \qquad for\ KS \leq E_1 \leq 1$$

STEP 2:

$$E_3 = E_2 + \mathrm{b}(1 - E_2)^4 \qquad for\ 0 \leq E_2 \leq 1$$

Hermite spline equations

$$P(B) = (2T(B)^3 - 3T(B)^2 + 1)KS + \left(T(B)^3 - 2T(B)^2 + T(B)\right)(1 - KS) + (-2T(B)^3 + 3T(B)^2)maxLum$$

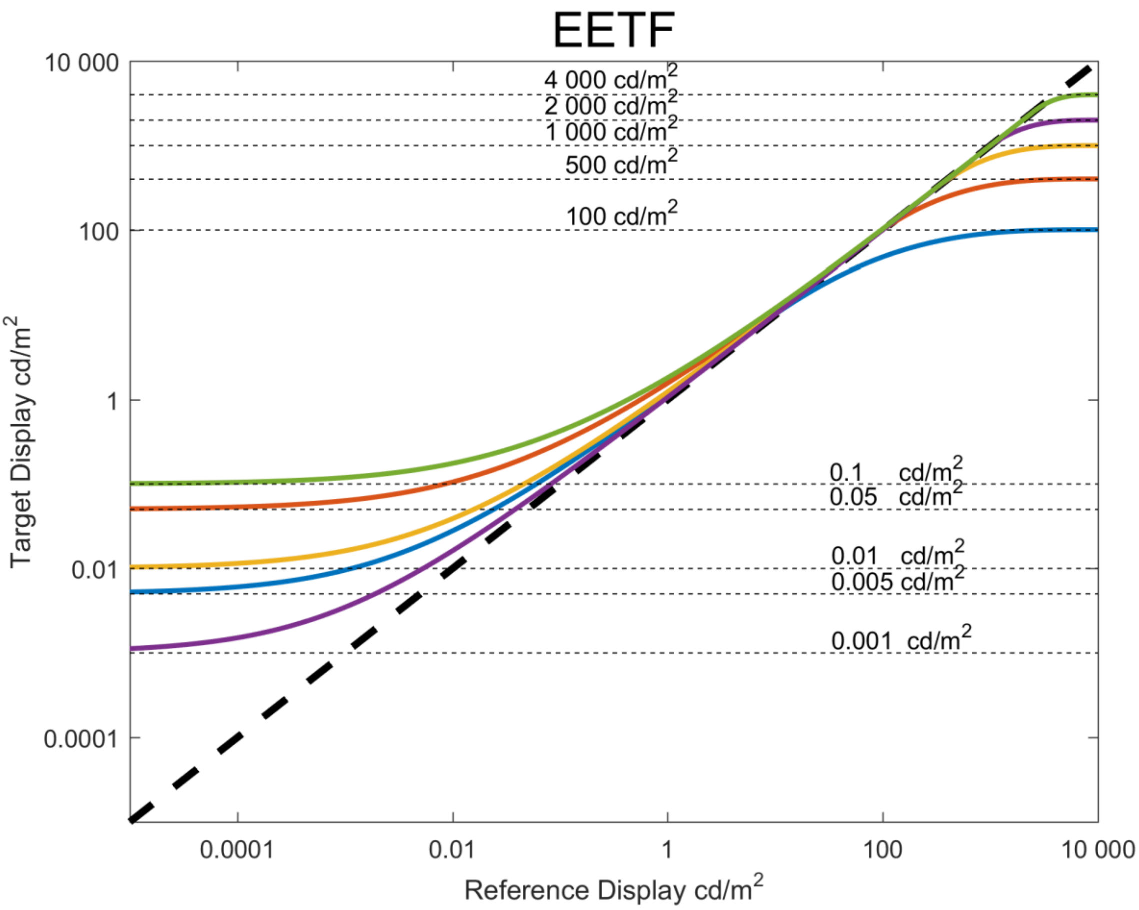
$$T(A) = \frac{A - KS}{1 - KS}$$

Practical application

The sample curves shown in Fig. 20 are designed for tone mapping to display black level up to 0.1 cd/m^2 and display white level as low as 100 cd/m^2.

FIGURE 20

Example EETFs of various target displays



The resulting EETF curve can be applied to either the intensity I channel of IC_TC_P or the luma Y channel of $Y'C'_BC'_R$. Here are the notable options:

1) I of IC_TC_P – process the intensity (I) channel of IC_TC_P though the EETF

$$I_2 = EETF(I_1)$$

– More accurately adjusts grayscale

– No colour shifts

– Changes in saturation will be needed and should be applied to the C_T and C_P channels using this equation:

$$C_{T2}, C_{P2} = min\left(\frac{I_1}{I_2}, \frac{I_2}{I_1}\right) \times (C_{T1}, C_{P1})$$

2) Y' of $Y'C'_BC'_R$ – run the luma Y' channel of $Y'C'_BC'_R$ though the EETF

$$Y'_2 = EETF(Y'_1)$$

– More accurately adjusts grayscale.

– Limited colour shifts.

– Changes in saturation will be needed and should be applied to the C'_B and C'_R channels using this equation:

$$C'_{B2}, C'_{R2} = min\left(\frac{Y'_1}{Y'_2}, \frac{Y'_2}{Y'_1}\right) \times (C'_{B1}, C'_{R1})$$

6 HLG HDR-TV

The hybrid log-gamma (HLG) HDR-TV signal parameters were designed from the outset to offer broadcasters and programme producers an evolutionary approach to HDR production and distribution. The signal characteristic is similar to that of a traditional standard dynamic range camera with a "knee" and requires no production metadata. It is therefore compatible with conventional standard dynamic range production equipment, tools and infrastructure. Furthermore, the HLG HDR-TV signal parameters were designed to provide a significant degree of compatibility on Recommendation ITU-R BT.2020 colour SDR displays (see § 6.4). Thus HDR monitors are only necessary in critical monitoring areas. The design of the HLG HDR signal parameters is intended to allow distribution networks to provide a single HEVC Main 10 bitstream that can target both SDR and HDR receivers, where those SDR receivers support the Recommendation ITU-R BT.2020 colour container (e.g. DVB and ARIB HEVC UHD receivers).

6.1 The hybrid log-gamma opto-electronic transfer function (OETF)

In the brighter parts and highlights of an image the threshold for perceiving quantization is approximately constant (known as Weber's law). This implies a logarithmic OETF would provide the maximum dynamic range for a given bit depth. Proprietary logarithmic OETFs are in widespread use. But in the low lights it becomes increasingly difficult to perceive banding. That is, the threshold of visibility for banding becomes higher as the image gets darker. This is known as the De Vries-Rose law. The conventional gamma OETF used for SDR comes close to matching the De Vries-Rose law, which is perhaps not coincidental since gamma curves were designed for dim CRT displays. So an ideal OETF would, perhaps, be logarithmic in the high tones and a gamma law in the low lights, which is essentially the form of the hybrid log-gamma OETF.

The dynamic range of modern video cameras is considerably greater than can be conveyed by a video signal using a conventional OETF gamma curve (e.g. Recommendation ITU-R BT.709 or Recommendation ITU-R BT.2020). In order to exploit their full dynamic range conventional video cameras use a "knee" characteristic to extend the dynamic range of the signal. The knee characteristic compresses the image highlights to prevent the signal from clipping or being "blown out" (overexposed). A similar effect is also a characteristic of analogue film used in traditional movie cameras. When a hybrid log gamma HDR video signal is displayed on a conventional SDR display the effect is similar to the use of a digital camera with a knee or using film. It is not surprising therefore, that the HLG video signal is highly compatible with conventional SDR displays, because what you see is very similar to the signal from an SDR camera. Indeed the knee characteristic of the HLG characteristic, defined in Table 5 of Recommendation ITU-R BT.2100 (and shown below), provides an extended range that is conservative compared with current SDR practice.

A HLG signal is defined as:

OETF:

With E is normalized to the range [0:1] then the equation for the OETF is:

$$E' = \mathrm{OETF}[E] = \begin{cases} \sqrt{3E} & 0 \le E \le 1/12 \\ a \cdot \ln(12E - b) + c & 1/12 < E \le 1 \end{cases}$$

where:

E: signal for each colour component $\{Rs, Gs, Bs\}$ proportional to scene linear light and scaled by camera exposure, normalized to the range [0:1].

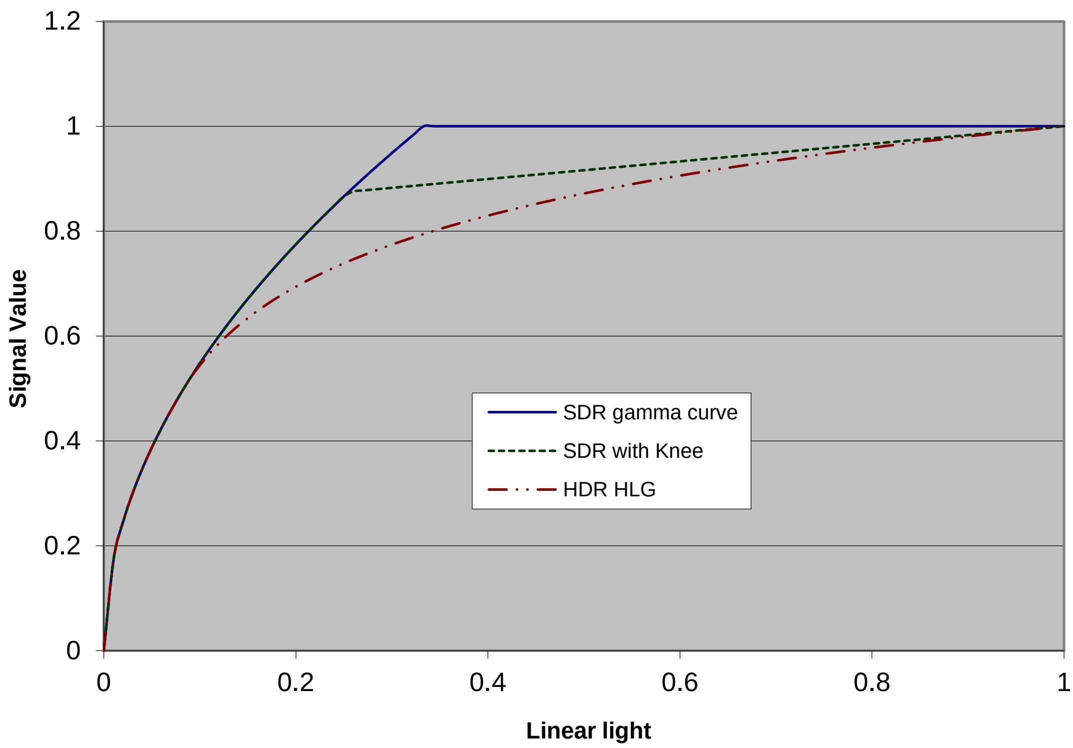
E': resulting non-linear signal $\{R', G', B'\}$ in the range [0:1].

$$a = 0.17883277,\ b = 1 - 4a,\ c = 0.5 - a \cdot \ln(4a)$$

The HLG OETF is shown in Fig. 21 alongside the conventional SDR gamma curve and a knee characteristic. Note that the horizontal axis for the SDR curve, defined in Recommendation ITU-R BT.2020, has been scaled to emphasize the compatibility of the HLG curve. Furthermore, because the HLG signal only describes the relative light representing the scene, it is independent of the display. Consequently, with a suitable EOTF, it may be used with any display.

FIGURE 21

Comparison of SDR and HLG HDR OETFs



6.2 System gamma and the opto-optical transfer function (OOTF)

As is well known, and explained in § 2.2, the light out of a television display is not proportional to the light detected by the camera. The overall system non-linearity, or “rendering intent” is defined by the opto-optical transfer function, or OOTF. The OOTF maps relative scene linear light to display linear light. Rendering intent is needed to compensate for the psychovisual effects of watching an emissive screen in a dark or dim environment, which affects the adaptation state (and hence the sensitivity) of the eye. Traditionally movies were, and often still are, shot on negative film with a gamma of about 0.6. They were then displayed from a print with a gamma of between 2.6 and 3.0. This gives movies a system gamma of between 1.6 and 1.8, which is needed because of the dark viewing environment. Conventional SDR television has an OOTF which is also a gamma curve with a system gamma of 1.2. But, for HDR, the brightness of displays and backgrounds/surround will vary widely, and the system gamma will need to vary accordingly.

Colour images consist of red, green and blue components and this affects how the OOTF should be applied. Simply applying a gamma curve to each component separately as is done for SDR television distorts the colour; in particular it distorts saturation but also to a lesser extent the hue. As an illustration, suppose the red, green and blue components of a pixel have (normalized) values of (0.25, 0.75, 0.25). Applying a display gamma of 2, (i.e. squaring the value of the components) we obtain (0.0625, 0.5625, 0.0625). In this example, the pixel has got slightly darker and the ratio of green to blue and red has increased (from 3:1 to 9:1). This means, a green pixel would have appeared as a discernibly different shade of green. This approach is far from ideal if we wish to avoid distorting colours when they are displayed.

Instead of the current SDR practice of applying a gamma curve independently to each colour component, for HDR it should be applied to the luminance alone. The luminance of a pixel is given by a weighted sum of the colour components; the weights depend on the colour primaries and the white point. According to Recommendation ITU-R BT.2100, luminance is given by:

$$Y_s = 0.2627R_S + 0.6780G_S + 0.0593B_S$$

where Y_S represents normalized linear scene luminance and R_S, G_S and B_S represent the normalized, linear scene light (i.e. before applying OETF) colour components. By applying rendering intent (OOTF) to the luminance component only it is possible to avoid colour changes in the display.

The HLG reference OOTF is therefore given by:

$$F_D = \mathrm{OOTF}[E] = \alpha Y_S^{\gamma-1}E + \beta$$
$$R_D = \alpha Y_S^{\gamma-1}R_S + \beta$$
$$G_D = \alpha Y_S^{\gamma-1}G_S + \beta$$
$$B_D = \alpha Y_S^{\gamma-1}B_S + \beta$$

where:

F_D: luminance of a displayed linear component {R_D, G_D, or B_D}, in cd/m^2

E: signal for each colour component {Rs, Gs, Bs} proportional to scene linear light and scaled by camera exposure, normalized to the range [0:1].

α and β are given by,

$$\alpha = (L_W - L_B)$$
$$\beta = L_B$$

γ: = 1.2 at the nominal display peak luminance of 1 000 cd/m^2

L_W: nominal peak luminance of the display in cd/m^2

L_B: display luminance for black in cd/m^2.

In order to determine the appropriate system gamma for a 1 000 cd/m^2 reference display, NHK conducted a series of experiments with an indoor test scene. Lighting was adjusted so that the luminance level of the diffuse white was 1 200 cd/m^2. The subjects were requested to adjust the system gamma and camera iris with reference to the real scene so that a tone reproduction similar to the scene could be obtained on the display. It was found that personal preference has an impact in determining the optimum system gamma for a given brightness display. But for a 1 000 cd/m^2 OLED display (Sony BVM-X300) the average optimum system gamma was found to be 1.18. Similar tests were repeated using a 2 000 cd/m^2 peak luminance LCD display (Canon DP-V3010), where it was found that the average preferred system gamma was 1.29.

Similarly, the BBC conducted subjective tests to determine the value of system gamma that delivers the best compatible SDR image. For those tests two Sony BVM-X300 OLED displays were used, one in its SDR mode (Recommendation ITU-R BT.1886, 100 cd/m^2 peak luminance) and the other a running prototype HLG HDR firmware (1 000 cd/m^2 peak luminance). In those tests the BBC found that the value of system gamma that delivers the best SDR compatible picture with a 1 000 cd/m^2 display was 1.29. A value of 1.18 was found to be the best value when the peak brightness of the display was reduced to 500 cd/m^2.

Notably both NHK and the BBC reported values of 1.29 and 1.18 independently, albeit at different peak brightness values.

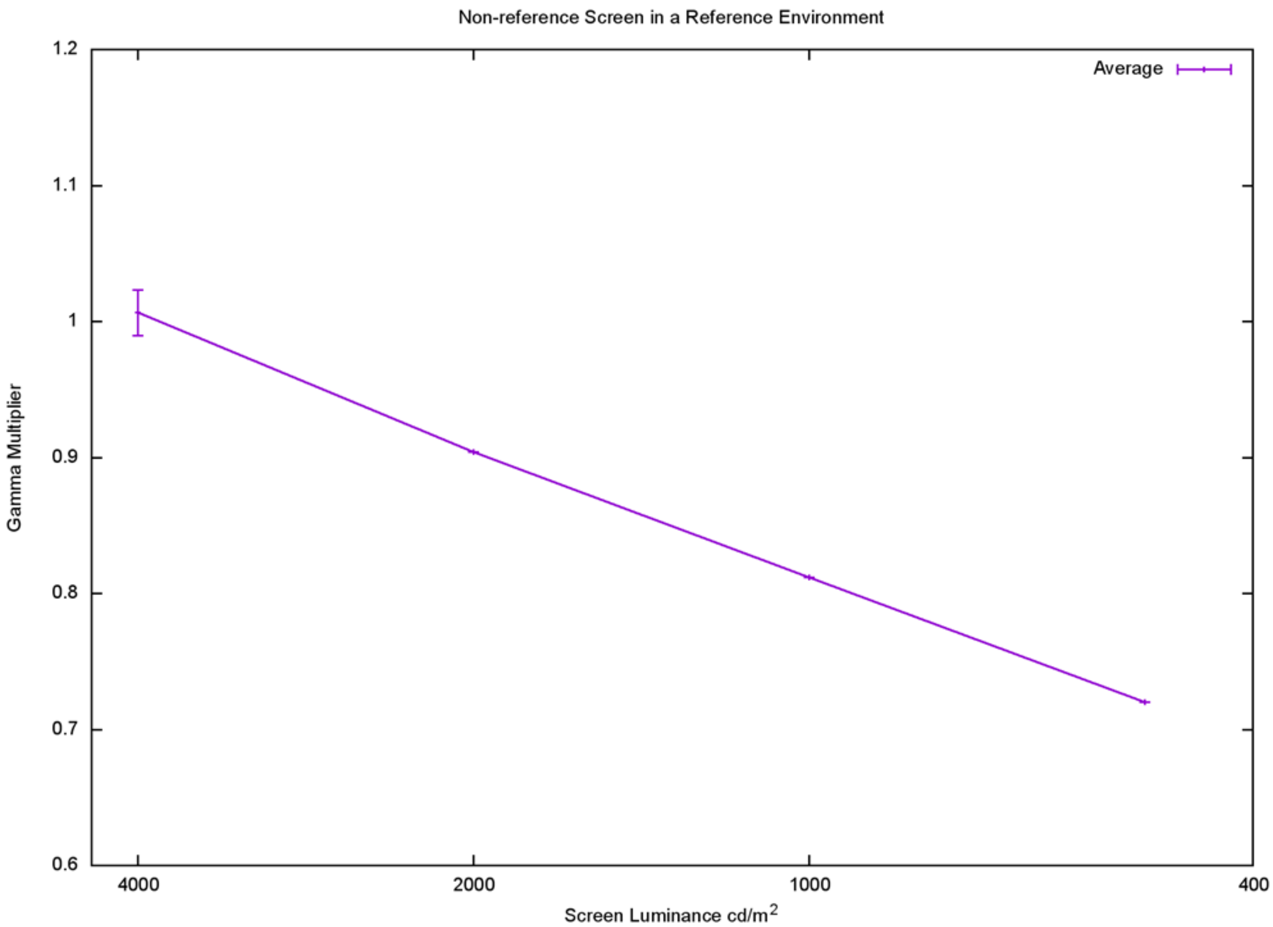
When designing the HLG HDR system, it was considered more important to weigh the choice of gamma value in favour of HDR production, rather than backwards compatibility with SDR displays. So a value of 1.20 was adopted for the reference 1 000 cd/m^2 display.

The clear indication from both of these studies is that system gamma needs to vary according to display peak brightness. In order to establish a more precise relationship between the gamma and display brightness, the BBC conducted further subjective tests where images were viewed with different gammas at different luminances (and with a fixed background luminance of 5 cd/m^2). The pictures were derived from HDR linear light images selected from Mark Fairchild's HDR Photographic Survey. Test subjects were asked to perceptually match as closely as possible an image displayed with a reference peak brightness to the same image with a non-reference peak brightness by adjusting the system gamma applied to the non-reference brightness image. The images were displayed on a calibrated SIM2 HDR47E display using its LogLuv input. The minimum black level viewable in the test environment was determined using an HDR PLUGE test signal, and an appropriate "brightness" offset added to the test images.

The results, illustrated in Fig. 22, provide a means to determine how system gamma should change for displays of different peak brightness. In these experiments the "reference" display was assumed to have a peak brightness of 4 000 cd/m^2. So the gamma "adjustment" is plotted as a multiplier to be applied to the appropriate system gamma for a 4 000 cd/m^2 display.

FIGURE 22

Average gamma multiplier for different screen peak brightness



The BBC confirmed the relationship in further tests for a 1 000 cd/m^2 to 500 cd/m^2 change using a prototype Sony BVM-X300 OLED display. These results are also consistent with the ratio of gamma values found by NHK for a 2 000 cd/m^2 LCD display and a 1 000 cd/m^2 OLED display, and similarly consistent with the ratio of values determined by the BBC for optimum SDR compatibility at 1 000 cd/m^2 and 500 cd/m^2.

Bringing together the results of all studies, it is found that the appropriate system gamma (γ) for different brightness displays, in the reference environment, can be determined using the following equation:

$$\gamma = 1.2 + 0.42\,\mathrm{Log}_{10}\left(L_W/1000\right)$$

where L_W is nominal peak luminance of the display in cd/m^2.

Using the HLG OOTF with gamma values determined by the formula, displays with the different values of peak luminance L_W are considered to be perceptually equivalent. That is, all displays are considered to be sufficiently perceptually similar that no substantial difference would have occurred had material been graded at one or other setting. The differences between settings are likely to be less than the difference between reference displays from different manufacturers. Therefore all settings may be regarded as perceptually equivalent.

6.3 The hybrid log-gamma electro-optical transfer function (EOTF)

In order to specify the complete television system we need an EOTF as well as the OETF defined in § 6.1. The HLG EOTF maps the HLG signal representing the scene to the light emitted from the display.

The EOTF mapping should:

1) preserve the artistic intent of the programme maker (and provide a suitable rendering intent),
2) allow for the dynamic range of the display from black level to peak white, and
3) minimize quantization artefacts.

The EOTF defined in Table 5 of Recommendation ITU-R BT.2100 and described below is similar to the conventional display gamma curve, thereby maximizing backward compatibility, whilst also meeting the three preceding requirements. As described above, for HLG the OOTF forms part of the EOTF, thus:

$$F_D = \mathrm{OOTF}\left[E\right] = \mathrm{OOTF}\left[\mathrm{OETF}^{-1}\left[E'\right]\right]$$

where,

F_D: luminance of a displayed linear component {R_D, G_D, *or* B_D}, in cd/m^2

E': non-linear signal {R',G',B'} as defined for the OETF.

Thus,

$$F_D = \mathrm{OOTF}\left[E\right] = \alpha Y_S^{\gamma-1} E + \beta$$

And,

$$R_D = \alpha Y_S^{\gamma-1} R_S + \beta$$
$$G_D = \alpha Y_S^{\gamma-1} G_S + \beta$$
$$B_D = \alpha Y_S^{\gamma-1} B_S + \beta$$

where:

R_S, G_S, B_S: scene linear light signals, E, for each colour component normalized in the range [0:1], and derived by applying the inverse OETF to the non-linear signal components, *R',G',B'*.

$$E = \mathrm{OETF}^{-1}[E'] = \begin{cases} E'^2/3 & 0 \le E' \le 1/2 \\ \{\exp((E'-c)/a)+b\}/12 & 1/2 < E' \le 1 \end{cases}$$

and:

R_D, G_D, B_D: displayed light for each colour component, in cd/m^2.

The values of parameters a, b, and c are as defined for the OETF.

The values of γ, L_B and L_B are as defined for the OOTF.

The nominal signal range of E, Rs, Gs, Bs, and Y_S is [0:1].

The reference display shall not display values greater than $E' = 1.0$. Such values should be clipped to 1.0 prior to display.

6.4 Compatibility with SDR displays

Both PQ and HLG provide limited compatibility when directly connected to legacy SDR displays with BT.709 colorimetry. In the absence of additional processing HLG has a degree of compatibility when shown on SDR UHDTV displays that have been designed to accept signals in the Rec. ITU-R BT.2020 colour space.

Concerning the degree of compatibility achieved by HLG, hue changes can be perceptible on the SDR display should images contain bright areas of highly saturated colour or very high code values. Generally such high code values would be used for specular highlights and thus constitute a small proportion of the picture. The acceptability of the degree of compatibility of HLG might be a commercial decision by specific broadcasters or for a specific application.

When PQ or HLG HDR signals are converted for use in SDR ITU-R BT.709 facilities, the conversion process is expected to perform the colour space, HDR to SDR and any video format conversion in such a way as to minimise perceptible changes in colour for all types of HDR content, regardless of the code value ranges in use.

7 Conversion between HLG and PQ signals

The following diagram illustrates conversion from the PQ signal to the HLG signal. The signal processing is that the PQ signal is decoded by the PQ EOTF to yield a signal that represents linear display light. This signal is then encoded by the HLG inverse EOTF to produce an equivalent HLG signal. When this HLG signal is subsequently decoded by the HLG EOTF in the display, the result will be the same display light that would be produced by decoding the original PQ signal with the PQ EOTF. The HLG inverse EOTF is the HLG inverse OOTF followed by the HLG OETF. For the HLG inverse OOTF, black level should be zero, and the gamma parameter is determined by the peak level of the PQ signal.

The following diagram illustrates conversion from the HLG signal to the PQ signal. The signal processing is that the HLG signal is decoded by the HLG EOTF to yield a signal that represents linear display light. This signal is then encoded by the PQ inverse EOTF to produce an equivalent PQ signal. When this PQ signal is subsequently decoded by the PQ EOTF in the display, the result will be the same display light that would be produced by decoding the original HLG signal with the HLG EOTF. For the HLG EOTF, black level should be zero, and the gamma may be set to the value specified in Table 5 of Recommendation ITU-R BT.2100 (presuming peak luminance of 1 000 cd/m^2)

8 Colour representation for chroma sub-sampling

The legacy $Y'C'_BC'_R$ non-constant luminance format is a colour-opponent based encoding scheme (in which signals are interpreted based on colour differences in an opposing manner) intended to separate luma from chroma information for the purposes of chroma subsampling (i.e. 4:2:2 and 4:2:0). High dynamic range and wide colour gamut content reveal the limitations of existing colour encoding methods. Errors that were previously small with standard dynamic range can become magnified. Recommendation ITU-R BT.2020 provides an alternative to $Y'C'_BC'_R$, i.e. the $Y'_CC'_{BC}C'_{RC}$ constant luminance format. This format resolves the issue of chroma leakage into the Y' luma signal, but does not solve the problem of luminance contamination of the C'_{BC} and C'_{RC} components. Recommendation ITU-R BT.2100 provides an alternative method for colour difference encoding called constant intensity, which is based on IPT colour space [17] developed by Ebner and Fairchild.

8.1 Non-constant luminance (NCL) $Y'C'_BC'_R$

$Y'C'_BC'_R$ is widely used for standard dynamic range content and requires a specific conversion based on the primaries being encoded and decoded. Recommendation ITU-R BT.2100 specifies PQ as a non-linearity to be used with the Recommendation ITU-R BT.2020 colour primaries. While $Y'C'_BC'_R$ performs satisfactorily in many cases, some limitations have emerged for its use in high dynamic range wide colour gamut scenarios.

Limitations of $Y'C'_BC'_R$ with wide colour gamut and high dynamic range

– Quantization distortions due to bit depth limitations with the increased colour volume.

– Chroma subsampling distortions due to a perceptually uneven distribution of code words.

– Colour volume mapping distortions due to incorrectly predicted hue and luminance.

– Error propagation from chroma to luma channels.

The constant luminance method specified in Recommendation ITU-R BT.2020 helps reduce the last of these, but this solution is not being widely adopted because the benefits are considered modest and entail some additional complexity.

8.2 Constant intensity IC_TC_P encoding

An alternative to constant luminance (CL) $Y'_CC'_{BC}C'_{RC}$ is the constant intensity (CI) IC_TC_P colour representation. Like $Y'C'_BC'_R$, IC_TC_P is a colour-opponent based encoding scheme intended to separate luma from chroma information. CI offers the same benefit as CL in that the chroma channels are lacking luminance, but IC_TC_P has the advantage that the lines of constant hue are straighter, and the MacAdam's ellipses are more circular. The CI neutral (grey) axis is encoded with the PQ or HLG non-linearity to match the human visual system, and to optimize it for high dynamic range signal encoding. The alternative 3x3 colour matrices used to generate the colour difference channels have been optimized [18] for the human visual system perception of HDR and WCG. The in-camera encoding and in-display decoding steps for IC_TC_P are identical to those for NCL $Y'C'_BC'_R$, so IC_TC_P is compatible with that hardware.

8.2.1 Constant intensity IC_TC_P encoding

Below are the conversion steps needed to get from camera linear RGB sensor signals into $Y'C'_BC'_R$ and into IC_TC_P [19]. Note that the matrix coefficients are decimal values that differ very slightly from the values shown in Recommendation ITU-R BT.2100; the values shown in in the Recommendation should be used in actual implementations.

FIGURE 23

Camera RGB conversion To $Y'C'_BC'_R$

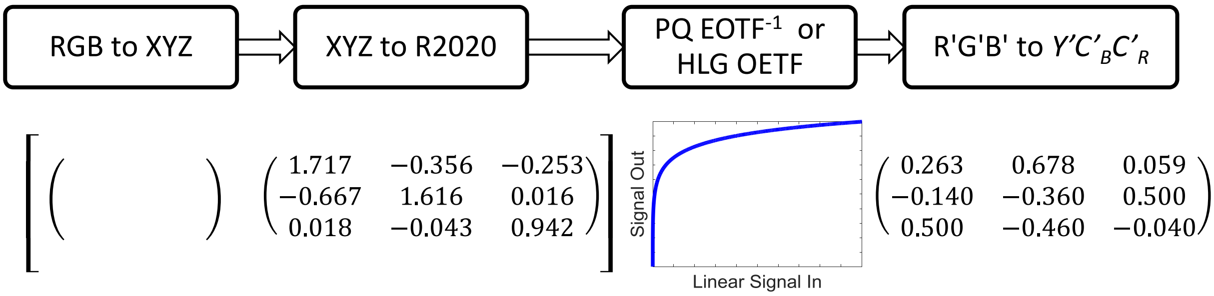
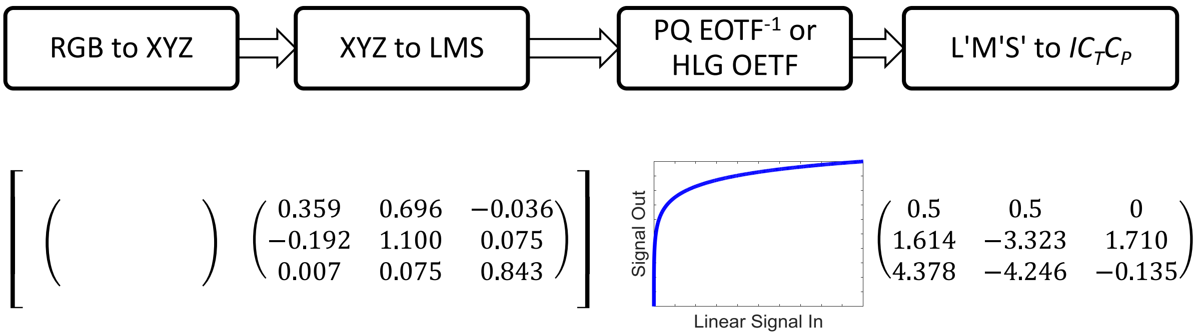


FIGURE 24

Camera RGB conversion to IC_TC_P



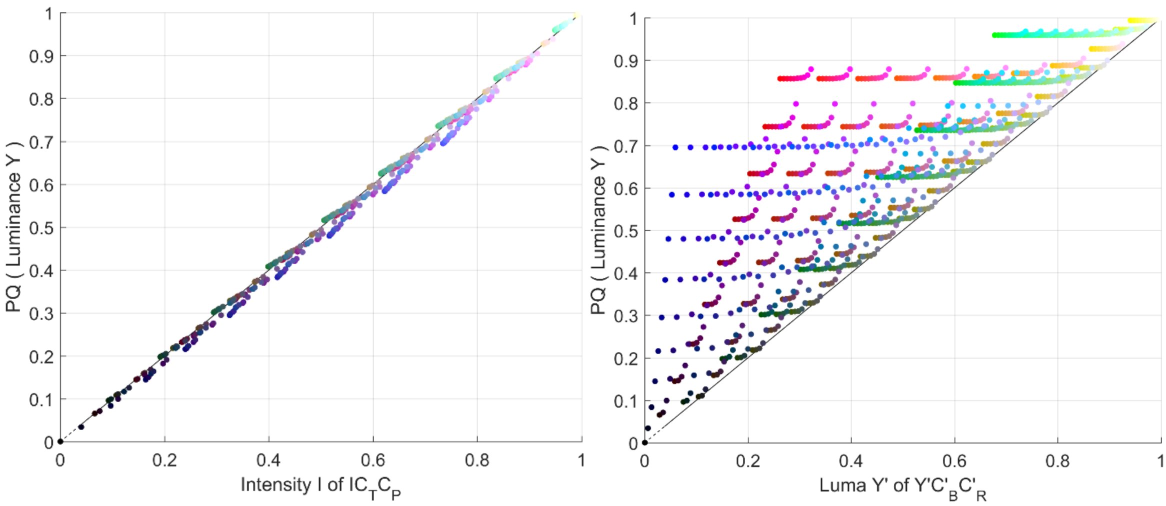
8.2.2 Advantages of constant intensity IC_TC_P

The specific design of the constant intensity colour space provides several benefits versus the Non-Constant Luminance colour space when used with the PQ or HLG non-linearity to provide HDR.

Achromatic channel: The achromatic axis of $Y'C'_BC'_R$ (Y' encoded in PQ or HLG) does not fully decorrelate luminance from colour. Therefore distortions introduced into the chroma channels can propagate to luminance where they become much more noticeable. As shown in Fig. 25, the achromatic axis of IC_TC_P (I) corresponds very closely with luminance (where luminance is a weighted sum of linear R,G,B). This is an indicator of how well IC_TC_P separates luma from chroma information. This reduces errors that can be introduced when spatially sub-sampling the chroma components compared to conventional non-constant luminance encoding. The axes in Fig. 25 are from zero to full scale in PQ space. (The luminance errors shown for $Y'C'_BC'_R$ are not as large for legacy systems using standard dynamic range with gamma encoding.)

FIGURE 25

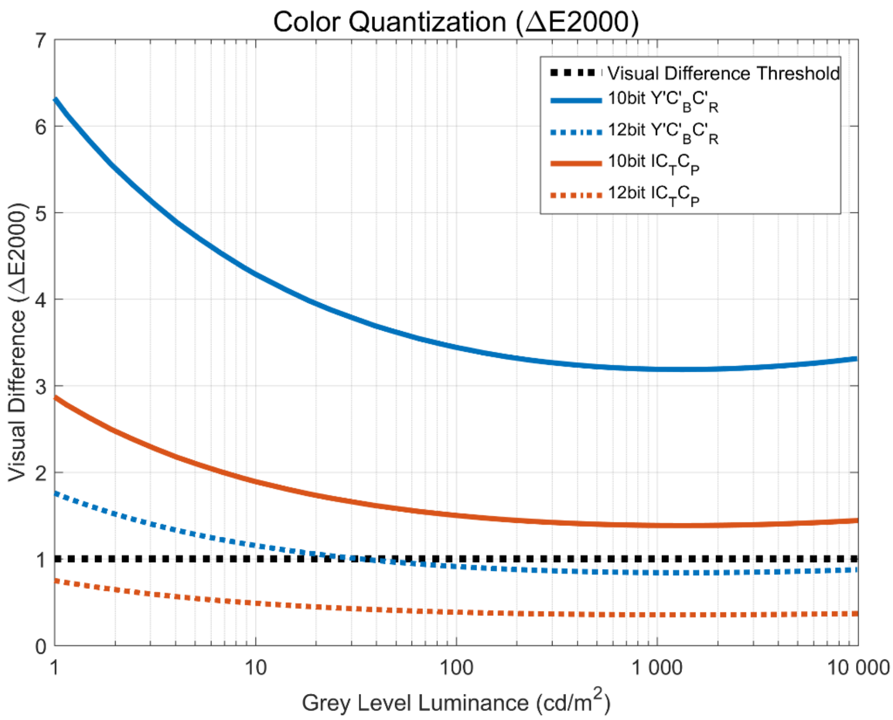
Luminance correlation



Quantization to limited bit-depth: Figure 26 shows the worst case visual colour difference between chroma channel code values (using $\Delta E2000$) at various luminance levels. 10-bit IC_TC_P provides an approximately 1.5 bit colour difference improvement over 10-bit $Y'C'_BC'_R$. At less than an average of 1.0 ΔE above the visual difference threshold, use of IC_TC_P significantly decreases visible distortions thus enabling excellent colour performance with 10-bit encoding.

FIGURE 26

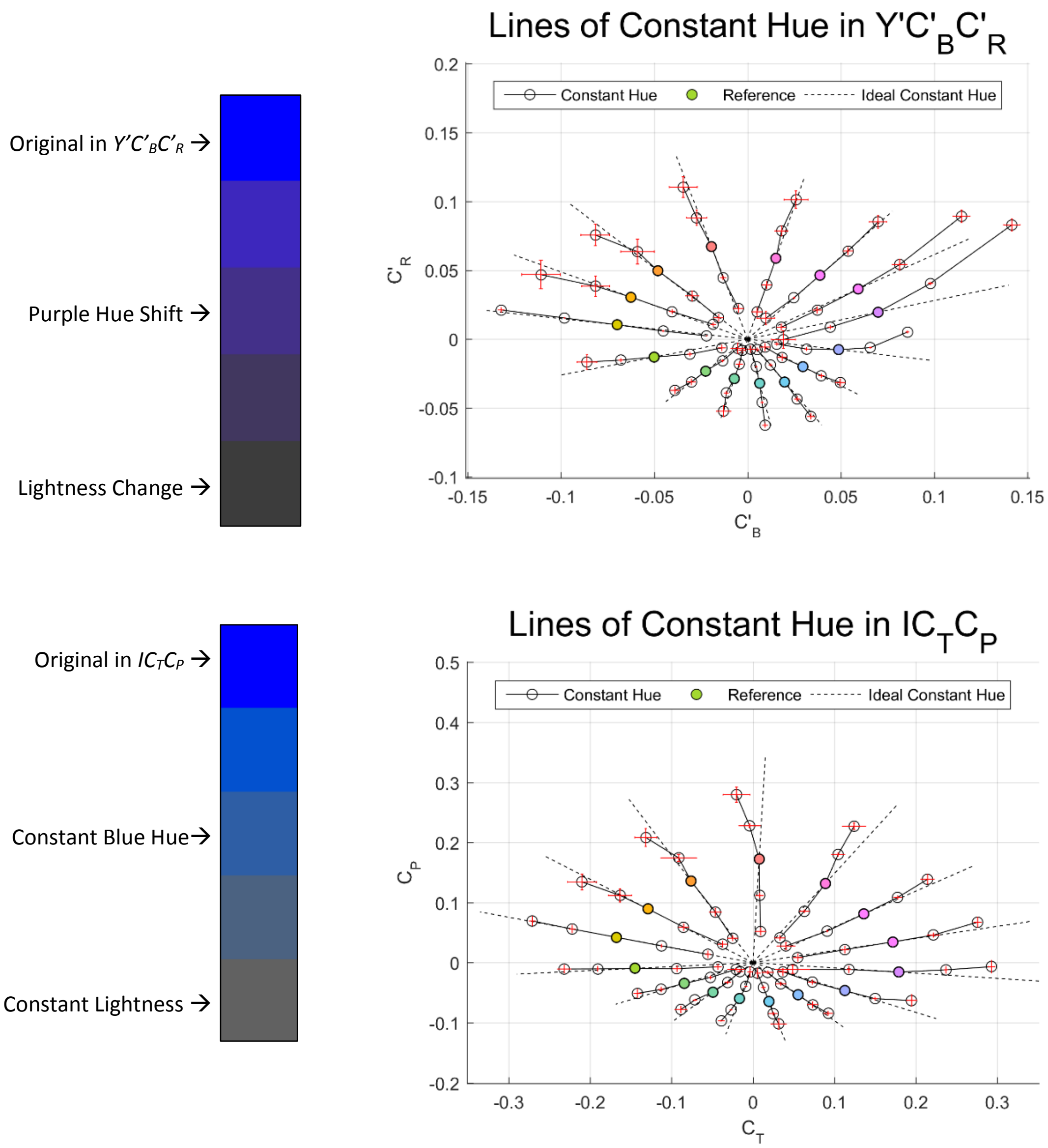
Maximum colour deviation at various bit-depths



Uniformity and hue linearity: A colour space is hue linear when the hue remains constant as saturation or intensity are changed. Hue linearity is important during any interpolation such as colour volume mapping, chroma subsampling, and blending/fading. Y’C’BC’R has large deviations (see Fig. 27) that cause hue shifts with highly saturated colours. ICTCP was designed to minimize deviation from lines of constant hue thereby reducing hue shifts. In addition, ICTCP has a more uniform distribution of colours. This improves efficiency, reduces worst case quantization and interpolation errors.

FIGURE 27

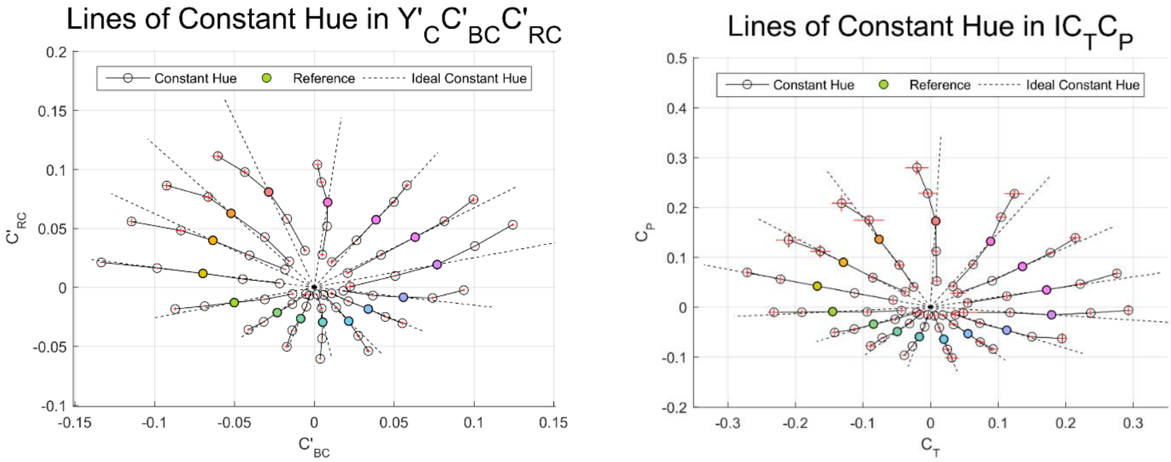
Blue Hue Linearity Comparison (using PQ)



If the CL format specified in Recommendation ITU-R BT.2020 is applied to HDR, the $Y'_CC'_{BC}C'_{RC}$ representation introduces additional (over NCL) errors in skin tones. The blue is significantly improved versus NCL (but still contains errors) and CL has significantly worse errors in the red and green regions (see Fig. 28). (The Recommendation ITU-R BT.2020 CL coefficients were designed for use with the SDR camera characteristic, and thus were not optimized for use in HDR.)

FIGURE 28

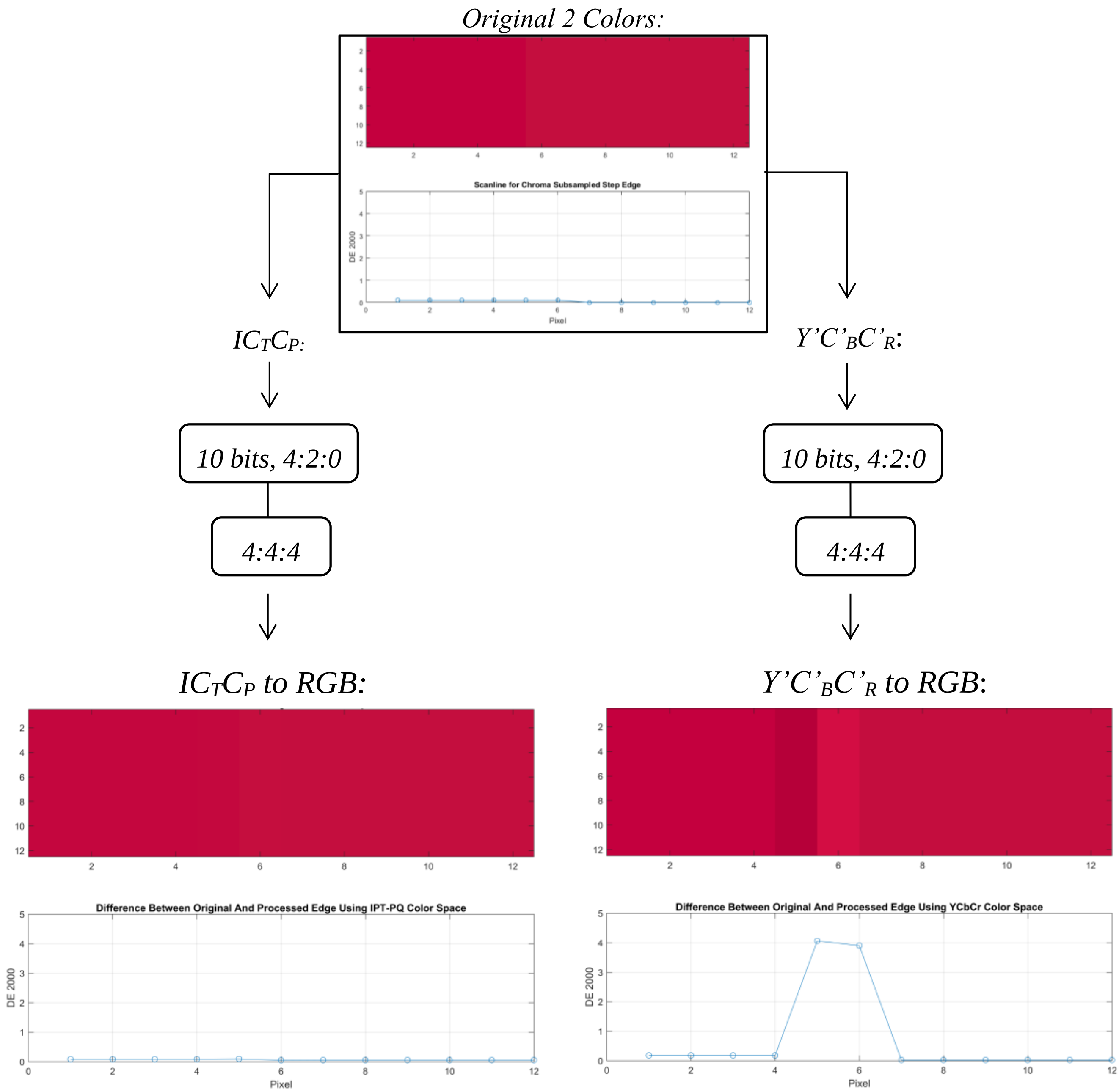
Constant luminance vs. constant intensity hue linearity comparison (PQ)



Colour sub-sampling: Figure 29 shows a practical example of a colour sub-sampling distortion due to NCL encoding. Two very similar colours with a Δ*E*2000 of 0.1 were sub-sampled to 10 bits 4:2:0 in $Y'C'_BC'_R$ and IC_TC_P and reconstructed. Due to the poor decorrelation between Y' of $Y'C'_BC'_R$ and luminance (Y), errors introduced in chroma during sub-sampling spread to the luminance and became more visible with a Δ*E*2000 of 4.0. Constant intensity IC_TC_P has a higher tolerance for chroma error and the colours remain indistinguishable with a Δ*E*2000 of 0.2.

FIGURE 29

Colour sub-sampling distortions based on correlation with luminance



9 Some considerations on the use of high dynamic range in TV image capture, mastering, distribution and presentation

This section focuses on the operational issues introduced by HDR imagery, offering a number of operational considerations on the desirable amount of image dynamic range, in relation to the processing that may be required along the various stages that television images go through, from image capture to production, postproduction, mastering, versioning, distribution and presentation of television programmes to the public at large.

The introduction of HDR imagery poses a number of housekeeping challenges, associated with the increased number of picture formats that will be in use.

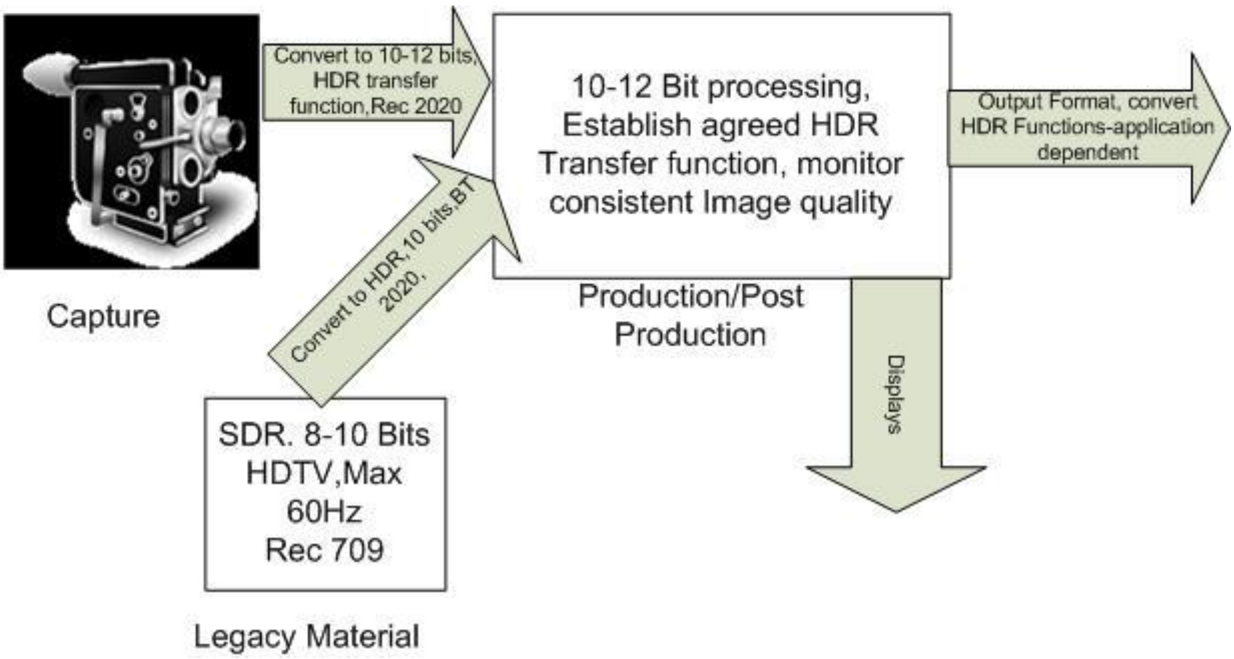
9.1 Television image capture, production, postproduction and mastering

During image capture and production, depending on the envisaged subsequent image postproduction, it may be desirable to capture television programme images at HDR even if they are not intended for distribution as HDR images. Since image capture and production at HDR provides an extended image postproduction headroom, programme directors may wish to exploit that headroom in order to achieve their creative intent.

As a generalization, there are two forms of image capture; live events, and content that will or could be subject to further signal processing.

Live capture typically uses image parameters that will be retained during the entire production process.

Content captured for subsequent post-processing and for multiple distribution channels is increasingly using image parameters that are not defined by standards organizations; in particular the image pixel depth, typically 16 bits, and the image transfer function(s) are left to a manufacturer's choice which in many cases is proprietary.



The above illustration is intended to show possible conversion points in the production process. The output from a capture camera for non-live transmission would typically be about 16 bits raw data (and could also be in some proprietary HDR format). In non-real-time file processing, the 16 bits could well be edited. However, if legacy material is to be added to the production, some conversion processes will need to be performed to create either the final HDR output or some intermediate image format. Suffice to say that interface limitations will need to be addressed if serial digital interfaces (SDI) interfaces are used.

The image dynamic range of the programme at the end of the postproduction process will normally be the one required for programme mastering and distribution. However, at this time it is not clear which various options will be chosen by programme producers. Clearly, consideration will need to be given to available emission bandwidth and to the environment in which end users will watch television.

Programmes most likely will be mastered at the image dynamic range required to meet the needs of the most demanding media targeted for their distribution.

Historically, during the production process image quality is monitored and checked for consistency throughout a given production; the introduction of HDR images complicates this process. The display devices used in production will most likely be watched under different viewing conditions to those found in the home or elsewhere; the peak brightness and other display settings will be different. In the absence of some strict guidelines and recommended operating practices, the intended creative intent of HDR programmes may fail to be adequately displayed to end users.

9.2 Television programme versioning

When a programme master has been completed, it is often necessary to generate various versions of it, depending on the requirements of the targeted distribution markets and media, each of which may have its own specific needs, e.g. in terms of language, subtitles, moral codes, etc.

Programme versioning also potentially includes the generation of versions at different image standards, e.g. at different image dynamic ranges or resolutions, depending on the targeted markets.

9.3 Television programme presentation

In broad terms, the mission of broadcasting is to provide information, education and entertainment to the public at large. The entertainment mission implies that programme presentation should be pleasurable: programmes should be presented to the public, conveying the creative intent of the programme director in an attractive way, to maximize the quality of viewers' experience, e.g. without causing viewers' annoyance or discomfort.

In this respect, an extended image dynamic range may play a role in improving viewers' satisfaction on condition that the programme director properly uses it and the end user watches it under appropriate viewing conditions, yet to be defined and recommended.

The statement above begs a question: how much image dynamic range is needed to optimize viewers' satisfaction? Obviously, too little would make little difference to viewers, and too much could well cause viewer's discomfort. It also begs the question of what should be done with inserting legacy material into an HDR production?

The limit to image dynamic range in the home depends mainly on two boundaries, namely

– the peak image brightness that the consumer display can provide without compromising other aspects of its performance such as colour fidelity, display life and resolution, which depend on the design of each display model, and
– the minimum image brightness at which consumers can still discern details on the screen of their display; this depends on the design of each display model and on the consumers' prevailing viewing environment, namely their "home" environment;

It also depends on other considerations such as pixel depth, available bandwidth, and real-time or non-real-time transfers of data etc., which are not directly related to home viewing.

It is important that a reference viewing environment has been defined in Recommendation ITU-R BT.2100; this requirement has not been part of previous image Recommendations.

How bright should peak brightness be, for how long should peak brightness appear on the screen, how much of the screen area should be allowed for peak brightness? Concerns have been expressed that there might be a risk of a video version of "loudness" until such time as a Recommendation on Operational Practices for HDR-TV is in place.

9.4 The typical home viewing environment

Concerning the consumer's television viewing environment, Recommendation ITU-R BT.2022 (08/2012) specifies "general viewing conditions for subjective assessment of quality of SDTV and

HDTV television pictures on flat panel displays". Section 1.2 of its Annex 1 specifies general viewing conditions in a home environment clarifying that, in its context, *"the home viewing environment is intended to provide a means to evaluate quality at the consumer side of the TV chain. General viewing conditions in § 1.2 reproduce a home environment. These parameters have been selected to define an environment slightly more critical than the typical home viewing situations".*

In view of the recent date of approval of Recommendation ITU-R BT.2022 (08/2012), the viewing environment specified therein can be assumed to still closely represent a typical home viewing environment, although it could be that some extensions may be required for consumers' HDR viewing in the home.

Recommendation ITU-R BT.2022 specifies the home viewing environment as being characterized by a value of 200 lux of incident light falling on the screen from the environment (e.g. from the windows during the day or the room lamps at night, as well as the light from the TV set itself, diffused back to it by the room walls and furniture). Of course, with an illumination of 200 lux on the TV screen from the environment, details in deep dark areas of the screen will appear to be somewhat washed out and the image dynamic range, as perceived by home viewers, will appear to be reduced with respect to its appearance when viewed in a dark environment.

Based on available indications, the reflectance of the screen of a modern consumer television set can be expected to be of the order of about 2% of the incident light falling on it, when the set is inactive and it uses a "black" screen with a semi-matte surface. This means that, in the consumer viewing environment specified in Recommendation ITU-R BT.2022, the luminance of the blackest image black would be of the order of 4 cd/m^2.

References

[1] L.B. Wolf. Relative brightness of specular and diffuse reflection. Optical Engineering, 33(1), pp. 285-293, Jan 1994.

[2] V. Interrante, H. Fuchs, and S. Pizer (1997), Conveying the 3D shape of transparent surfaces via texture. ICASE report No 97-27. NASA contractor report 201705.

[3] R. Dror, A. Willsky, and E.H. Adelson (2004), Statistical characterization of real world illumination. JOV 4, pp. 821-837.

[4] A. Blake and H. Bulthoff (1991), Shape from specularities: computation and psychophysics. Phil Trans from the Royal Society of London. Series B, 33, pp. 237-252.

[5] S. Hecht, C. Haig, A.M. Chase, The influence of light adaptation on subsequent dark adaptation of the eye, Journal of General Physiology, 20 (1937), pp. 831–850.

[6] M.G. M. Stokkermans and I. Heynderickx (2014),Temporal dark adaptation to spatially complex backgrounds: effect of an additional light source. JOSA A V31 #7, pp. 1485-1494.

[7] R. Mantiuk, S. Daly, and L. Kerofsky (2010), "The luminance of pure black: exploring the effect of surround" SPIE Electronic Imaging Conference, San Jose.

[8] Rempel, A., Heidrich, W., Li, H., and Mantiuk, R., Video viewing preferences for HDR displays under varying ambient illumination," [Proc. of the 6th Symposium on Applied Perception in Graphics and Visualization], 45{52, ACM (2009).

[9] S. Daly, T. Kunkel, X. Sun, S. Farrell, and P. Crum (2013), "Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance "SPIE Electronic Imaging, HVEI, Burlingame, CA, Feb 2013.

[10] R. De Valois and K. De Valois (1990), Spatial Vision, p26, Ch.2, Oxford Science Publications., Oxford University press.

[11] N. Salvaggio (2008), Basic photographic materials and processes, 3rd edition. Focal press.

[12] S. Miller, M. Nezamabadi, S. Daly, "Perceptual Signal Coding for More Efficient Usage of Bit Codes," SMPTE Mot. Imag. J., 122(4):52-59, May-June 2013.

[13] P.Heckbert, "Non-Overshooting Hermite Cubic Splines For Keyframe Interpolation". New York Institute of Technology. February 1985.

[14] F. Ebner, M. Fairchild, "Derivation and Modelling of Hue Uniformity". Rochester Institute of Technology. July 1998.

[15] "Encoding Color Difference Signals for High Dynamic Range and Wide Gamut Imagery", Froehlich *et al*., Color and Imaging Conference, Volume 2015, Number 1, October 2015, pp. 240-247(8).

[16] J. Pytlarz, R. Atkins, T. Lu, F. Pu, P. Yin, T. Chen, W. Husak, "Overview of IPT-PQ Color Space", Doc. m37266, October 2015, Geneva, Switzerland.