

Modelling Brightness Perception for High Dynamic Range Television

Katy C. Noland, Manish Pindoria and Andrew Cotton
BBC Research & Development
BBC Centre House,
56 Wood Lane, London, W12 7SB, UK.

Abstract—High dynamic range (HDR) television offers greater contrast and more immersive images than conventional television, and as such it is an important part of the overall ultra-high definition television package. Standardisation is almost complete, and the industry is already taking the first steps in HDR programme production. The extended dynamic range and brighter screens associated with HDR make sudden jumps in brightness possible. To ensure consistency between programmes and to avoid uncomfortable brightness shifts at programme junctions, some production guidelines are needed for HDR brightness, just as guidelines have been necessary for audio loudness.

In order to develop production guidelines for brightness, brightness perception must be understood. In this paper we report the results of subjective tests that measured the overall perceived brightness of a set of HDR images. We then propose ten classes of potential objective metric that relate the displayed pixel luminance levels to the overall subjective brightness level, and evaluate them using our test results as ground truth. The most effective metrics tested are the mean of the pixel luminances, the mean of the pixel luminances raised to the power of 0.82, and the 96th percentile of pixel luminances, all of which performed similarly well. The mean displayed pixel luminance is preferred, since it is the simplest to implement. The effectiveness of these simple objective metrics suggests that real-time brightness monitoring in production is a realistic goal.

Keywords—HDR; UHD; brightness; perception; vision; human visual system

I. INTRODUCTION

High Dynamic Range (HDR) television [1] increases the contrast range of television images, and so can exploit the capabilities of ever-brighter consumer displays to create a more natural and immersive viewing experience in the home. Standards for end-to-end television production and distribution are nearing completion, HDR televisions are available in the shops, and the first HDR programmes are being produced.

With the introduction of HDR television and brighter displays, there comes the potential for greater jumps in brightness between and within programmes. Whilst an increased dynamic range offers new creative opportunities, care needs to be taken that unwanted changes in brightness are not uncomfortable. Viewers should not need to adjust the contrast control on their television within a programme, at programme junctions, at breaks for advertisements, or when switching between channels. A need for production guidelines for brightness levels has therefore been identified, to prevent unwanted uncomfortable

brightness shifts. The guidelines are likely to resemble requirements for audio loudness levels [2][3][4] in many respects.

Before appropriate guidelines can be set, it is necessary to be able to measure brightness. The term brightness denotes “the extent to which an area appears to exhibit light” [5, p. 69]. It is distinct from the relative brightness, known as lightness, which relates to the apparent reflectance of an object, regardless of how it is lit [5, p. 70]. Brightness is a subjective quantity that cannot be measured directly, so subjective test methodology must be used.

In this paper we report subjective tests that measured the perceived brightness of a set of HDR images. We then use the results as ground truth values for development of an objective brightness metric. We propose ten classes of metric, many of which are based on existing knowledge of brightness perception, and we compare their performance in terms of correlation with the ground truth data. It is proposed that the best metric be used as a basis for brightness monitoring in production, and for setting production guidelines.

II. RELATED WORK

There is an extensive body of work on brightness perception and adaptation, from which we highlight here a few key results. We also describe the relevant existing guidelines for best practice television production.

A. Brightness Perception

Weber’s law, described by Fechner [6, p. 136], states that the just noticeable difference (JND) at a particular stimulus level is proportional to the absolute stimulus level, for all our senses, including vision. Extending this relationship to supra-threshold values implies a logarithmic relationship between absolute luminance and perceived brightness. However, it is known that the JND varies with the adaptation state of the eye [7] and, at low light levels, depending on whether primarily the rods or cones in the eye are being stimulated [8, p. 506]. The electro-optical transfer function (EOTF) defined for the Perceptual Quantizer (PQ) approach to HDR [1] uses JNDs that are more precisely specified for luminance levels suitable for HDR television, based on a model by Barten [9]. The inverse PQ EOTF could therefore also be interpreted as a mapping from luminance to brightness, if counting JNDs is considered appropriate as a measure of brightness difference.

Stevens strongly contested this use of JNDs [10]. He proposed instead a power law relationship between the luminance

and brightness, with the exponent varying according to the size of the source and the adaptation state of the eye [11]. A power of 0.33 was found for a 5 degree stimulus and dark-adapted eye, a value that is often quoted [12], [13]. Bartleson and Breneman modelled brightness perception of patches in non-uniform backgrounds [14], and derived a modified log relationship that has been shown to have the same form as Stevens' model when the parameters are chosen to account for viewing conditions [15]. The CIE 1976 lightness measure [16] [5, p. 116] is often used for measuring perceived relative brightness, but it is unlikely to be useful for measuring absolute brightness, since it requires luminance levels to be normalised to a reference white. Without the normalisation step, it is fundamentally a power relationship with an exponent of 0.33.

The studies discussed so far relate to the brightness of small, uniform stimuli, but this is not sufficient for a model of overall image brightness: for that we also need a method of combining the contribution of the different parts of the image. Bauer showed that Stevens' power relationship also holds for the estimated mean brightness of a group of test patches [17], so we might expect it also to hold for a combination of pixels. However, a viewer's estimate of the arithmetic mean brightness is still not quite what we are looking for. A more suitable measure would be the adaptation state of the eye induced by looking at the image, sometimes referred to as the *equivalent background brightness* [7, p.19] [18].

Bartleson used the luminance of perceived "middle gray" as the adaptation level [19]. Barlow [7, p. 1] suggested that average luminance is the driver for adaptation, which implies that a simple calculation of the mean of each pixel's contribution may be appropriate. Moon and Spencer [20] made use of the *Holladay principle*, that the adaptation caused by a non-uniform visual stimulus can be modelled as a uniform visual stimulus that produces the same adaptation state. They extended the principle to describe the equivalent adaptation luminance as an additive function of individual source luminances, weighted by a function of distance from the fixation point. However, for normal viewing, the viewer's eyes are free to move, so the Moon and Spencer weighting function is likely to be inappropriate in this case.

Adaptation is not instantaneous, and complete adaptation to darkness can take many minutes [21]. However, adaptation to a lesser difference in luminance is much faster, and can pass largely unnoticed [7, p. 1]. Brightness guidelines should aim to keep the adaptation level within the range of comfortable sudden changes.

B. Television Production Practice

In standard dynamic range (SDR) television, conventions that map specific scene luminance levels to appropriate signal levels are widely used, although they are not officially standardised. They include "diffuse white" or an ice hockey rink at around 90% signal level, "flesh tones" at around 50–70% signal level, and grass or an 18% reflectance card at 50% signal level. Conventional SDR displays are generally too dim to allow good quality pictures to be produced using only part of the signal range, so these best-practice conventions have developed to ensure that full use is made of the available dynamic range. This has the secondary effect of maintaining a



Fig. 1. Test room set-up, showing the display and lighting arrangement.

degree of brightness consistency between programmes, which means that large jumps in brightness do not usually occur. There has not therefore been a specific need for brightness guidelines for SDR.

For HDR, displays are generally brighter, so there can be more flexibility in the signal range used for individual scenes and hence there is a greater chance of brightness jumps occurring scene-to-scene. Some brightness guidelines are therefore required. Guidelines to ensure good quality pictures are also needed, which are likely to be more flexible than those for SDR. For Hybrid Log-Gamma HDR [1], a reference level for graphics of 75% signal level has already been agreed, which allows headroom for highlights above the graphics level. However, the focus of this paper is brightness measurement, independent of whether the pictures are artistically pleasing.

III. EXPERIMENT

We conducted subjective tests to find the perceived overall brightness of a set of HDR images. Test subjects were asked to adjust the brightness of a grey slate until it matched the perceived overall brightness of a test image. The luminance of the grey slate is known, and can be used as a numerical value that is representative of the brightness of the image. Subjects were able to switch freely between the test image and grey slate, and were able to take as much time as needed to give their answers. Although we did not specifically account for the time to adapt between the test image and grey slate, as the subject approached their answer the brightness difference would tend to zero, meaning no adaptation would be necessary.

The test image and full-screen grey slate were shown on a SIM2 HDR47E display using its calibrated LogLUV mode. The slate levels ranged from 0 to 4000 cd/m², with 400 steps following an exponential function such that the step size was 3.9×10^{-10} cd/m² at black and 50 cd/m² at the top end. Like most LCD screens, when displaying a full screen in a single colour the SIM2 is not able to accurately display the input brightness, especially at high luminance levels. We therefore measured the actual luminance of the screen for input slate levels at intervals of 10 cd/m², and mapped the intended grey slate values to these measured values (interpolated where necessary) before presenting the results. Two adjustable LED lights illuminated the wall behind the display such that the light reflected off the wall measured D65 white at 5 cd/m². The lights were positioned behind the screen, directed towards



Fig. 2. Images used for the tests. Images 1-12 are from [22], image 13 from [23] and images 14 and 15 were created by BBC R&D.

the wall, to minimise light falling directly on the screen. There was no other source of light in the room. The test set-up is shown in figure 1.

The LogLUV input to the SIM2 display does not have a “brightness” control, so an ordinary PLUGE signal [24] cannot be used to calibrate the black level in the normal way. Hence a set of special test signals that included sub- and super-black patches at a range of black levels was used to find the required offset, and this offset was added to the test images before display. The black level was found to be 0.005 cd/m^2 . This low level was achieved because the lights were positioned behind rather than in front of the screen, so reflections were minimal.

We used 15 images, shown in figure 2. The first 12 images were taken from Fairchild’s HDR Photographic Survey [22], and these were supplemented by one image (number 13) from Hochschule der Medien, Stuttgart [23] and 2 images created by BBC R&D (numbers 14–15). Since the dynamic range of the raw images was greater than that expected for use in HDR television, we scaled and clipped the images to look aesthetically pleasing (as judged by a small number of expert viewers, on the same SIM2 screen in dim lighting) with a smaller dynamic range. This is equivalent to adjusting the camera iris. It may have been possible to produce more subjectively pleasing images using a more sophisticated tone-mapping operator, but the pure scaling approach minimised the number of subjective judgements made when producing the test material. The test set included bright and dark images, and several images with regions of both highlights and shadows. The images were converted to LogLUV format for the SIM2 display, and shown at a resolution of 1920×1080 pixels to match its maximum resolution.

We showed each image at four peak display luminance levels, 500, 1000, 2000 and 4000 cd/m^2 . The signal range was scaled to the display luminance range, and a gamma function appropriate to the display peak luminance was applied, following ITU-R BT.2100 [1, note 5e, p. 7]. This is simply a method of increasing the range of brightnesses used in the test, and does not limit the applicability of the results to any particular approach to display adaptation.

Subjects were seated at a distance of 1.9 m from the display, which corresponds to 3.2 times the screen height. Each subject

was screened for normal visual acuity before the test, then given written instructions. Two training images were provided, and three “dummy” images were included at the start of the test. Results for the training and dummy images were discarded. The images were presented in a different random order for each subject, and care was taken that the same image (at a different brightness) never appeared twice consecutively. Twenty subjects completed the test. All were working at the BBC and so are likely to have some familiarity with subjective testing for images and video.

IV. PROPOSED METRICS

For the purposes of this study, we developed models that relate the displayed pixel luminance values to the overall perceived brightness. The model will eventually need to operate on signal values rather than displayed light levels if it is to be used for signal monitoring, but at this stage it is kept independent of signal format so that it can be applied to any HDR image.

The test images were stored as Hybrid Log-Gamma $Y'C'_bC'_r$ images with BT.2020/BT.2100 colour primaries [25][1] and 4:2:2 colour subsampling. After upsampling the colour difference components and converting to $R'G'B'$, the displayed luminance values were calculated according to ITU-R BT.2100 [1, table 5]. First the HLG opto-electric transfer function is removed to find the scene linear light signals $R_S G_S B_S$, then gamma and scaling are applied according to the peak luminance of the display to find the display colour components $R_D G_D B_D$. Finally displayed luminance values are calculated from the displayed colour components, using BT.2020/BT.2100 colour equations [25, table 4].

We implemented the following proposed models that produce a numerical value for the overall brightness from the displayed pixel luminance values of our test images. We define the displayed luminance for a particular pixel as $Y_D(i, j)$, where i and j are pixel indices with $i \in 0 : M - 1$ and $j \in 0 : N - 1$. For our images $M = 1080$ and $N = 1920$.

1) Mean display luminance

$\frac{1}{MN} \sum_i \sum_j Y_D(i, j)$. As a baseline metric we calculate the mean of all displayed pixel luminance values.

- 2) **Mean log2 display luminance**
 $\frac{1}{MN} \sum_i \sum_j \log_2(Y_D(i, j))$. Following Fechner [6].
- 3) **Mean PQ inverse EOTF of display luminance**
 $\frac{1}{MN} \sum_i \sum_j \text{EOTF}^{-1}(Y_D(i, j))$, with EOTF^{-1} as defined in ITU-R BT.2100 Table 4 [1].
- 4) **Mean display luminance raised to a power**
 $\frac{1}{MN} \sum_i \sum_j (Y_D(i, j))^p$. Following Stevens [11]. We test values from $p = 0.2$ to $p = 1$.
- 5) **Mean CIE 1976 lightness**
 $\frac{1}{MN} \sum_i \sum_j L^*(i, j)$, with L^* as defined by the CIE [16], including the linear section at low luminance levels. We use the displayed light values corresponding to 75% signal level as the reference white, as this has been defined as the reference level for graphics for Hybrid Log-Gamma HDR. These levels are 120, 203, 344 and 581 cd/m^2 for peak brightness levels of 500, 1000, 2000 and 4000 cd/m^2 respectively.
- 6) **Weighted mean display luminance**
 $\frac{1}{MN} \sum_i \sum_j Y_D(i, j) \cdot \cos \theta_{ij} / \theta_{ij}^2$, where θ_{ij} is the angle subtended at the eye between pixel (i, j) and the centre of the screen, with a minimum value of 0.75° . This metric follows Moon and Spencer [20, eqn. 3]. In the form used here, we assume that the viewer is fixated at the centre of the screen, we make the approximation that all pixels subtend the same angle at the retina, and we omit constants that would not affect correlation coefficients.
- 7) **Mean of values in centre of screen**
 $\frac{4}{MN} \sum_{i=M/4}^{3M/4} \sum_{j=N/4}^{3N/4} Y_D(i, j)$. A simplified version of Moon and Spencer's weighting, we calculate the average luminance of only those pixels in the central quarter of the screen.
- 8) **Percentiles**
The n th percentile, P_n , is the luminance level below which n percent of all pixel luminance levels fall. These relate to the distribution of display luminance levels. We test percentiles from P_{10} to P_{100} .
- 9) **Percentile ranges**
We calculate the interquartile range $P_{75} - P_{25}$, and the difference between the 90th and 10th percentile $P_{90} - P_{10}$. These describe the spread of the displayed luminance values.
- 10) **Mean of values within a specified range**
We calculate the mean display luminance of only those values between P_{25} and P_{75} . We also test the range P_{10} to P_{90} .

The mean selected grey levels for the 60 test images (15 images at 4 peak brightnesses) were used as ground truth brightness values to evaluate the models. The base-2 logarithm of the ground truth and of the output of each model was calculated, to improve the perceptual uniformity of the error space, then Pearson's correlation coefficient and Spearman's rank correlation coefficient were found.

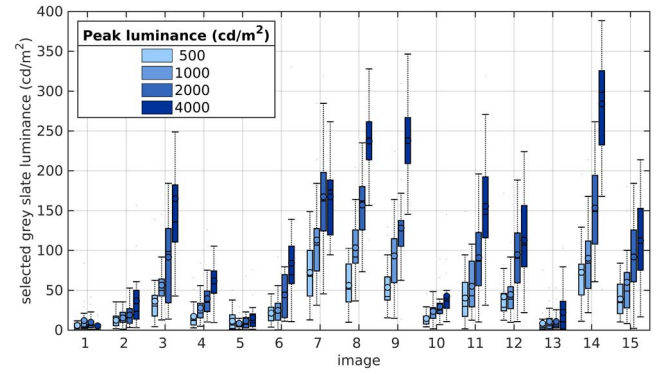


Fig. 3. Box plot showing the selected grey slate luminance levels for each image and peak display luminance. The boxes span the interquartile range, whiskers are at 1.5 times the interquartile range from the box limits, median values are shown with horizontal bars, and the mean values with circles.

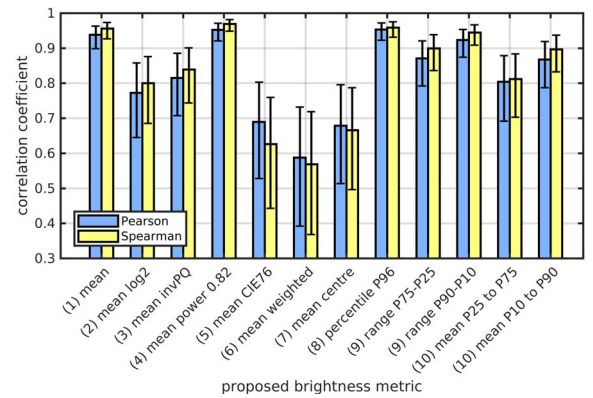


Fig. 4. Correlation coefficients for all proposed metrics (see section IV), with 95% confidence intervals (calculated using Fisher's z transform). The parameters giving the highest correlation are used for metrics 4 and 8.

V. RESULTS

Figure 3 shows a box plot of the selected grey slate luminance levels from the twenty subjects, for each image at each peak display luminance. It should be noted that the images might not all have different subjective brightness levels, even at different peak display luminance levels. For example, test image 1 (see figure 2) is mainly dark with a few strong highlights. When the display luminance increases, the greatest subjective difference is in the perceived brightness of the highlights, so the overall brightness of this image may not be affected by a change in display peak luminance if the level of a small area of highlights is not perceptually important. This is exactly the kind of effect we would like to investigate in order to develop an effective brightness metric. Nonetheless, the results show a general trend for images to appear brighter as the display peak luminance increases, and show that there is a spread of bright and dark images in the test set. Some of the results show a high variance, which could potentially be reduced using an alternative test methodology.

The Pearson correlation coefficients and Spearman rank correlation coefficients for all of the brightness metrics (see section IV) are shown in table I. The results are also presented graphically in figures 4, 5 and 6.

TABLE I. CORRELATION COEFFICIENTS BETWEEN THE BASE-2 LOGARITHM OF THE GROUND TRUTH VALUES AND THE BASE-2 LOGARITHM OF THE OUTPUT OF EACH BRIGHTNESS METRIC. THE PEAK CORRELATION IN EACH SECTION OF THE TABLE IS IN **BOLD**.

Metric	Parameter value (if applicable)	Pearson correlation coefficient	Spearman rank coefficient
1) Mean display luminance	—	0.94	0.96
2) Mean log2 display luminance	—	0.77	0.80
3) Mean PQ inverse EOTF of display luminance	—	0.82	0.84
4) Mean CIE 1976 Lightness	—	0.69	0.63
5) Mean display luminance raised to a power	0.2	0.84	0.86
	0.33	0.88	0.89
	0.4	0.89	0.91
	0.6	0.93	0.95
	0.8	0.95	0.97
	0.82	0.95	0.97
6) Mean weighted displayed luminance	—	0.59	0.57
7) Mean of values in centre of screen	—	0.68	0.67
8) Percentiles	P_{10}	0.65	0.60
	P_{20}	0.67	0.62
	P_{30}	0.71	0.67
	P_{40}	0.71	0.66
	P_{50}	0.75	0.72
	P_{60}	0.80	0.79
	P_{70}	0.84	0.85
	P_{80}	0.89	0.92
	P_{90}	0.92	0.95
	P_{96}	0.95	0.96
P_{97}	0.89	0.88	
P_{100}	0.39	0.41	
9) Percentile ranges	$P_{75} - P_{25}$	0.87	0.90
	$P_{90} - P_{10}$	0.92	0.94
10) Mean of values within a specified range	P_{25} to P_{75}	0.80	0.81
	P_{10} to P_{90}	0.87	0.90

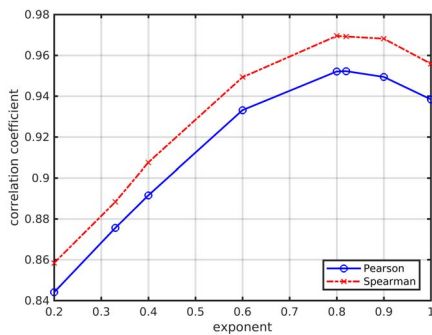


Fig. 5. Correlation coefficients for the mean displayed pixel luminance values raised to a power, with exponents from 0.2 to 1 (metric 4).

The simplest metric, the mean displayed luminance (metric 1), is a very good match, with a Pearson correlation coefficient of 0.94. Its performance is shown in figure 7. The various non-linear scaling methods in metrics 2–5 only cause the data to deviate from the straight line of figure 7, making a curved line a better fit. The only exception is raising the pixel luminance values to a power of around 0.8 to 0.9 before averaging (see figure 5), which offers a slight improvement over metric 1. The peak was found to be at a power of 0.82. However, the confidence intervals overlap extensively (see figure 4), so we cannot conclude that the difference is significant.

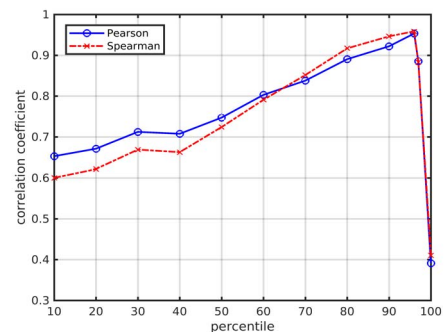


Fig. 6. Correlation coefficients for percentiles 10 to 100 of the displayed pixel luminance values (metric 8).

The mean weighted display luminance (metric 6), which relies on viewer fixation, performs poorly, and the mean of the values in the centre of the screen (metric 7) is also much less effective than the mean of all values. This shows that pixels near edge of the screen make an important contribution to the overall brightness.

The lower percentiles (metric 8) also perform poorly, but the higher percentiles correlate well with the subjective test results, with the 96th percentile giving the highest correlation overall (see figure 6). However, its confidence interval extensively overlaps that for metric 1. It is also likely to be strongly dependent on the images used for the test, and may

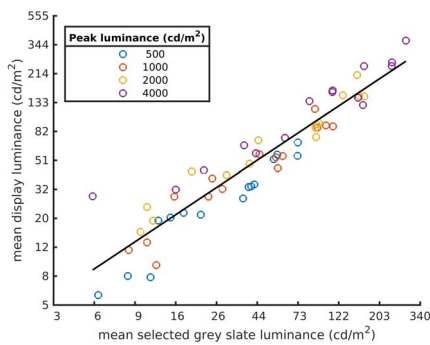


Fig. 7. Mean display luminance (metric 1) plotted against the mean selected grey slate luminance, for each test image. Axes use a log scale.

not perform so well if, for example, some artistic tone-mapping were applied. Metric 1 is likely to generalise better. The variation in performance for the different percentiles suggests that the luminance distribution in the darkest parts of the image is not an important factor in the perceived overall brightness, but that the level of the brightest parts is a significant driver.

The percentile ranges (metric 9) perform similarly to the percentile corresponding to the higher limit of the range, i.e. the correlation for $P_{90} - P_{10}$ is similar to that for P_{90} alone, and the correlation for $P_{75} - P_{25}$ lies between P_{70} and P_{80} . Both had reasonably good correlation with the subjective results.

The mean of values within a specified range (metric 10) also performs reasonably well. Again, the higher correlation comes from the wider range, i.e. the range that is most similar to metric 1.

VI. CONCLUSION

We have conducted subjective tests to determine the perceived overall brightness of a set of HDR images. We then used the results as ground truth data for evaluation of proposed objective brightness metrics that relate the displayed pixel luminance values to the overall image brightness.

The best performing metrics are the 96th percentile of the displayed pixel luminance values, the displayed pixel luminances raised to a power of 0.82 before calculating the mean, and the mean displayed pixel luminance values calculated directly. The performance differences between these three metrics are very small, and their confidence intervals overlap, so the simplest method—direct calculation of the mean—is preferred for real-time applications. The high correlation of 0.94 suggests that this simple metric will be an effective basis for a brightness monitor.

In order to extend this work to provide production guidelines for HDR video brightness, further investigation is required into tolerance of brightness shifts. It will also be necessary to relate the model to signal luminance levels.

A related challenge for the industry is to characterise “good quality” HDR content, that makes best use of the available dynamic range. Suitable signal levels for graphics and reflectance charts, and appropriate signal ranges for flesh tones, sports fields and other common content could be part of this, forming parallels with the existing conventions for SDR television.

ACKNOWLEDGMENT

We would like to thank Tim Borer, Simon Thompson, and all the test subjects.

REFERENCES

- [1] ITU-R, “Image parameter values for high dynamic range television for use in production and international programme exchange,” Recommendation ITU-R BT.2100-0, July 2016.
- [2] —, “Algorithms to measure audio programme loudness and true peak audio level,” Recommendation ITU-R BS.1770-4, October 2015.
- [3] EBU, “Loudness Normalisation and Permitted Maximum Level of Audio Signals,” EBU Recommendation R128, June 2014.
- [4] ATSC, “Techniques for Establishing and Maintaining Audio Loudness for Digital Television,” ATSC Recommended Practice A/85, March 2013.
- [5] R. W. G. Hunt, *The Reproduction of Colour in Photography, Printing and Television*, 4th ed. Fountain Press, 1987.
- [6] G. T. Fechner, *Elemente der Psychophysik*. Breitkopf und Haertel, 1860.
- [7] H. B. Barlow, *Dark and Light Adaptation: Psychophysics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1972, pp. 1–28.
- [8] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*. John Wiley & Sons, Inc., 1967.
- [9] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*. USA: SPIE Optical Engineering Press, 1999.
- [10] S. S. Stevens, “On the Psychophysical Law,” *The Psychological Review*, vol. 64, no. 3, pp. 153–181, 1957.
- [11] —, “To Honour Fechner and Repeal His Law,” *Science*, vol. 133, no. 3446, pp. 80–86, 1961.
- [12] C. Poynton and B. Funt, “Perceptual uniformity in digital image representation and display,” *Color Research & Application*, vol. 39, no. 1, pp. 6–15, 2014.
- [13] G. Ward and Anywhere Software, “High dynamic range image encodings,” 2006.
- [14] C. J. Bartleson and E. J. Breneman, “Brightness perception in complex fields,” *J. Opt. Soc. Am.*, vol. 57, no. 7, pp. 953–957, Jul 1967.
- [15] K. F. Choi, “Relationship Between Bartleson and Breneman’s Brightness vs. Luminance Equation and Stevens’ Power Law,” in *IS&T’s 47th Annual Conference/ICPS*, 1994, pp. 433–436.
- [16] Commission Internationale de l’Eclairage, “CIE 1976 L*u*v* colour space,” 2014. [Online]. Available: <http://eiv.cie.co.at/term/159>
- [17] B. Bauer, “Does Stevens’ Power Law For Brightness Extend to Perceptual Brightness Averaging?” *The Psychological Record*, vol. 59, pp. 171–186, 2009.
- [18] W. S. Stiles and B. H. Crawford, “Equivalent Adaptation Levels in Localized Retinal Areas,” *The Physical and Optical Societies Report, Joint Discussion on Vision*, pp. 194–211, 1932.
- [19] C. J. Bartleson, “Optimum image tone reproduction,” *Journal of the SMPTE*, vol. 84, no. 8, pp. 613–618, August 1975.
- [20] P. Moon and D. E. Spencer, “The Visual Effect of Non-Uniform Surrounds,” *J. Opt. Soc. Am.*, vol. 35, no. 3, pp. 233–248, 1945.
- [21] W. S. Stiles, “Adaptation, Chromatic Adaptation, Colour Transformation,” *Anales Real Soc. Espan. Fis. Quim.*, vol. 57, Series A, pp. 149–175, 1961.
- [22] M. Fairchild, “The HDR Photographic Survey,” 2006–07. [Online]. Available: <http://rit-mcsl.org/fairchild/HDR.html>
- [23] J. Fröhlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, “Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays,” in *Proc. SPIE*, vol. 9023, 2014. [Online]. Available: <https://hdr-2014.hdm-stuttgart.de>
- [24] ITU-R, “Specifications and alignment procedures for setting of brightness and contrast of displays,” Recommendation ITU-R BT.814-2, 2007.
- [25] —, “Parameter values for ultra-high definition television systems for production and international programme exchange,” Recommendation ITU-R BT.2020-2, October 2015.