ITUPublications
Recommendations

International Telecommunication Union
Radiocommunication Sector

Recommendation ITU-R BT.2163-0

(11/2023)

BT Series: Broadcasting service (television)

Objective measurement algorithm for evaluation of the brightness of high dynamic range television

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 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 Recommendations

(Also available online at https://www.itu.int/publ/R-REC/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
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

***Note**: This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2023

RECOMMENDATION ITU-R BT.2163-0

Objective measurement algorithm for evaluation of the brightness of high dynamic range television[1,2]

(Question ITU-R 142-3/6)

(2023)

Scope

This Recommendation specifies a measurement algorithm for the purpose of determining the Image Level, based on mean image luminance, which may be useful for assessing the brightness of individual images. Further metrics, based on the Image Level, are Temporal Image Level and Image Level Response which may be useful in modelling the response of the human visual system to a sequence of images.

Keywords

Video measurement, brightness, television production, high dynamic range, HDR, television, HDR-TV, international programme exchange

The ITU Radiocommunication Assembly,

considering

a) that high dynamic range television (HDR-TV) offers an extremely wide dynamic range of image luminance levels;

b) that large jumps in brightness can be surprising to viewers;

c) that monitoring of image brightness facilitates understanding of the viewer experience;

d) that the state of human visual system adaptation is affected by the temporal sequence of images;

e) that Recommendation ITU-R BT.2100 specifies two forms of HDR-TV, Perceptual Quantization (PQ) and Hybrid Log-Gamma (HLG), and that production in both formats is expected;

f) that a nominal peak luminance of 1 000 cd/m^2 is commonly used for HLG reference displays;

g) that reference surround environment conditions are specified in Table 3 of Recommendation ITU-R BT.2100;

h) that existing video measurement methods employed in production do not provide a numeric value representing subjective image brightness;

i) that, to facilitate the evaluation of subjective video brightness of programmes, standardized measurement methods are useful,

recognizing

that use of such algorithms will supplement, rather than replace, conventional waveform monitoring,

---

1 This Recommendation is not a replacement for the guidance contained in Recommendation ITU-R BT.1702 which is for the protection of the vulnerable section of the viewing population who have photosensitive epilepsy, and who are therefore prone to seizures triggered by flashing lights, including certain types of flashing television images.

2 The algorithms may need to be revised after testing on a wider variety of moving images and when tested in conjunction with a metering specification that is currently under development.

recommends

1 that when a measure of subjective brightness of a high dynamic range image, produced as per Recommendation ITU-R BT.2100, is required to facilitate programme production and exchange, the Image Level algorithm specified in Section 1 may be used;

2 that the Temporal Image Level measurement algorithm, specified in § 2, may be used to assess the contribution of an image sequence to viewer adaptation;

3 that the approximation to the response of the eye, Image Level Response, specified in § 3, may be useful for evaluating the significance of changes in brightness within a programme,

further recommends

1 that this Recommendation should not be used as a tool in brightness regulation, impose operator restrictions or limit content creation;

2 that due to limited verification of algorithms with moving images, the ITU-R may wish to consider early updates and improvements to this Recommendation.

1 Image Level measurement algorithm

This section specifies the Image Level (IL) measurement algorithm, which may be useful for assessing overall image brightness. The IL algorithm is the basis of the other metrics specified in this Recommendation in §§ 2 and 3.

The input to the IL measurement algorithm is a non-linear Recommendation ITU-R BT.2100 HLG or PQ signal, *R'G'B'*, in the range [0:1].

The algorithm consists of four stages:

1 conversion to linear display light;

2 calculation of the luminance component;

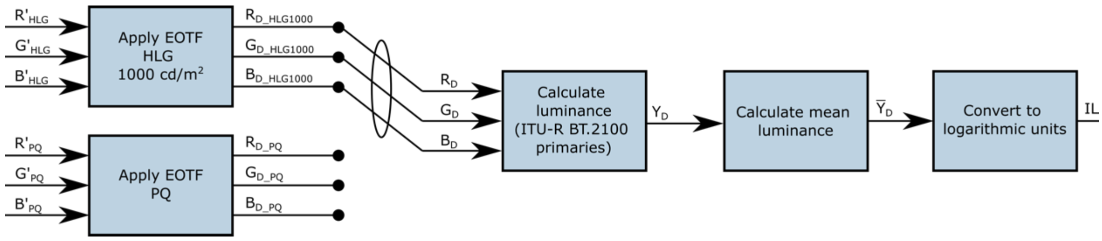
3 calculation of the mean display luminance;

4 conversion to logarithmic units.

Figure 1 shows a block diagram of the algorithm.

FIGURE 1

Block diagram showing the Image Level measurement algorithm



1.1 Conversion to linear display light

For HLG signals $R'G'B'_{HLG}$, the HLG EOTF specified in Recommendation ITU-R BT.2100, Table 5 is applied, using a value of 1.2 for γ as specified in Note 5f to Table 5 of Recommendation ITU-R BT.2100 for a display with a nominal peak luminance of 1 000 cd/m^2. The resulting signal is display linear light $R_D G_D B_{D\_HLG}$ in the range [0:1 000] cd/m^2.

For PQ signals $R'G'B'_{PQ}$, the PQ EOTF specified in Recommendation ITU-R BT.2100, Table 4 is applied. The resulting signal is display linear light $R_D G_D B_{D\_PQ}$ in the range [0:10 000] cd/m^2.

Hereafter the component linear display light signal is referred to as $R_D G_D B_D$, irrespective of whether the signal originated as HLG or PQ.

1.2 Calculation of the luminance component

The linear displayed luminance signal Y_D is calculated using the equation

$$Y_D = 0.2627R_D + 0.6780G_D + 0.0593B_D$$

An approximate method is described in Annex 1.

1.3 Calculation of the mean display luminance

The mean display luminance $\bar{Y}_D$ for the video frame, in units of cd/m^2, is then calculated.

$$\bar{Y}_D = \frac{1}{H \times V} \sum_{h=0}^{H-1} \sum_{v=0}^{V-1} Y_D(h, v)$$

where:

$Y_D(h,v)$: linear displayed luminance of the pixel value at horizontal position h, vertical position v

H : number of horizontal pixels

V : number of vertical pixels.

1.4 Conversion to logarithmic units

The mean display luminance in cd/m^2 is converted to base 2 logarithmic units, with a normalization value of 1 cd/m^2.

NOTE – A mean display luminance of 1 cd/m^2 produces an IL value of zero.

$$IL = \log_2 \frac{\bar{Y}_D}{1}$$

Background research for this algorithm is supplied in Annex 2.

2 Temporal Image Level algorithm

Viewing a sequence of images affects the adaptation state of a human observer and therefore the perceived impact of the Image Level. The Temporal Image Level (TIL)[3] may be useful for assessing image brightness over time. The TIL is calculated as follows:

$$TIL(t) = \begin{cases} IL(t) & \text{if } t = 0 \\ TIL(t-1)\left(1 - \frac{1}{\tau+1}\right) + IL(t)\left(\frac{1}{\tau+1}\right) & \text{if } t > 0 \end{cases}$$

where t is the frame number starting with 0, and τ is the characteristic time of decay, a constant which is set to:

$$p(t) = \begin{cases} 1 & \text{if } t = 0 \\ IL(t) - TIL(t-1) & \text{if } t > 0 \end{cases}$$

[3] The TIL algorithm has only been verified under limited conditions with mostly static image content such as those described in Annex 4. The TIL algorithm may need to be revised after testing on a wider variety of moving images.

$$\tau = \begin{cases} 22\dfrac{f}{24} & \text{if } p(t) \geq 0 \\ 800\dfrac{f}{24} & \text{if } p(t) < 0 \end{cases}$$

where:

f : framerate (in Hz) of the video.

The derivation of the functional form of TIL is shown in Annex 3, and the time of decay value choices are described in Annex 4.

3 Image Level Response algorithm

The Image Level Response (ILR) is used to model the response of the human visual system to the luminance of an image. It may be useful for assessing the instantaneous level of viewer annoyance arising from changes in Image Level, for example at a shot-change. It does not model the decay in subjective annoyance following a change in brightness. However, it may provide a useful input to metering devices that, when combined with other measures specified in this document, could model that decay. It is dependent on both the luminance of the image and the level of adaptation of the eye. In a Recommendation ITU-R BT.2100 reference environment, this response can be approximated using the IL and TIL quantities as follows:

$$ILR = \frac{(2^{IL})^{n_c}}{(2^{IL})^{n_c} + (2^{TIL})^{n_c}}$$

where:

ILR : approximate response of the human visual system

n_c : constant equal to 0.57.

Annex 1
(informative)

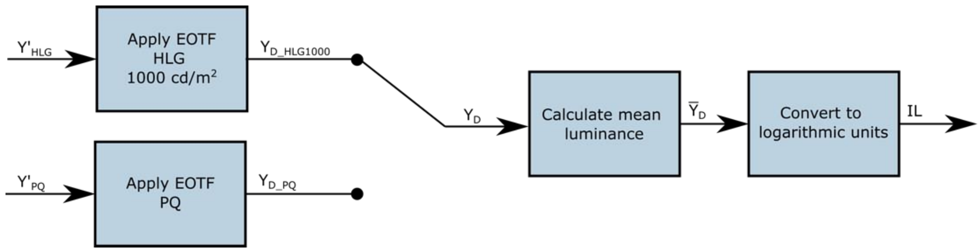
Approximation of the display luminance calculation for simplified hardware implementation

This Annex describes an approximate method of calculating the display luminance component, Y_D, that simplifies the implementation of the IL algorithm described in § 1. Where hardware resources are restricted, this method could be implemented in devices that are used for a visual aid only. Saturated colours might give rise to lower measurement values using this approximation. Devices using this method would not be suitable for numerical analysis of the IL values, and hence would not need to log IL values.

Figure 2 shows a block diagram of the approximate algorithm. Its input is a non-linear HLG or PQ luma signal, Y', in the range [0:1]. If the signal is IC_TC_P, the I signal can be used in place of Y' in the diagram below.

FIGURE 2

Block diagram showing the approximate IL measurement algorithm



Conversion to approximate[4] linear display light

For HLG signals Y'_{HLG}, the HLG EOTF specified in Recommendation ITU-R BT.2100, Table 5 is applied directly on Y'_{HLG}, using a value of 1.2 for γ as specified in Note 5f to Table 5 of Recommendation ITU-R BT.2100 for a display with a nominal peak luminance of 1 000 cd/m^2. The resulting signal approximates display linear light Y_{D_HLG} in the range [0:1 000] cd/m^2.

For PQ signals Y'_{PQ}, the PQ EOTF specified in Recommendation ITU-R BT.2100, Table 4 is applied directly on Y'_{PQ}. The resulting signal approximates display linear light Y_{D_PQ} in the range [0:10 000] cd/m^2.

The approximate linear display light signal is referred to as Y_D, irrespective of whether the signal originated as HLG or PQ.

The approximate value of Y_D is then used in the remaining calculations described in §§ 1.2 and 1.3.

Annex 2
(informative)

Objective metrics for brightness measurement in high dynamic range television

This Annex describes background research that has informed development of the Image Level algorithm described in § 1. After an overview of related work, a subjective test is described that established values to use as brightness data for a set of test images. These values are then used to test a number of potential objective metrics for measuring image brightness.

Related work

There is an extensive body of work on brightness perception and adaptation, and this is described in BBC White Paper 341 [1].

Experiment to establish brightness values for test images

The term *brightness* denotes “the extent to which an area appears to exhibit light” ([2], p. 69). It is distinct from *lightness*, which relates to the apparent reflectance of an object, regardless of how it is lit ([2], p. 70). Brightness is a subjective quantity that cannot be measured directly, so a subjective test methodology to create a set of ground truth brightness measurements is developed. The intention

---

4 Note that to calculate the display luminance accurately, the EOTFs in Recommendation ITU-R BT.2100 are applied on *R'G'B'* signal components, not on the luma component as described here.

is to create an objective brightness metric that matches the subjective results as closely as possible, and is based on the displayed luminance values.

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.

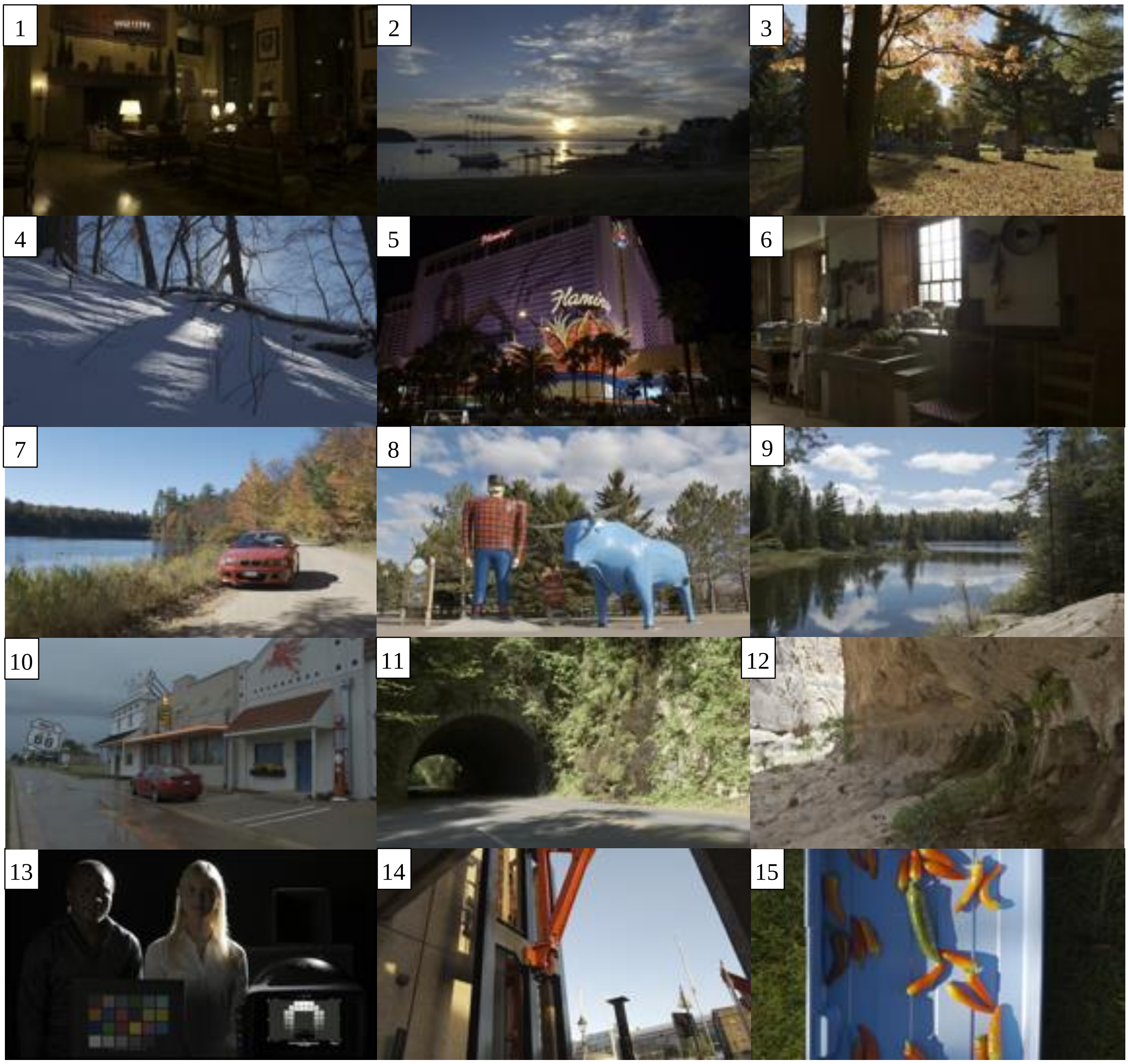
The test image and grey slate were shown on a SIM2 HDR47E display using its calibrated LogLUV mode. The slate levels ranged from 0 to 4 000 cd/m^2, with 400 steps following an exponential function such that the step size was 3.9×10^{-10} cd/m^2 at black and 50 cd/m^2 at the top end. Like all LCD screens, 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. The actual luminance of the screen for input slate levels at intervals of 10 cd/m^2 were recorded 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/m^2. The lights were positioned behind the screen, directed towards 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 Fig. 3.

FIGURE 3

Test room set-up

FIGURE 4

Images used for the tests



Note to Fig. 4: Images 1 to 12[5] are from [3], image 13 from [4][6] and images 14 and 15 were created by BBC R&D.

The LogLUV input to the SIM2 display does not have a brightness control, so an ordinary PLUGE signal cannot be used to calibrate the black level. Hence a set of specially-generated test signals that included sub- and super-black 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 offset was found to be 0.005 cd/m^2. This is lower than would be expected if the lights were positioned in front rather than behind the screen.

---

5 These images are reproduced by permission of Mark Fairchild.

6 This image is reproduced by permission of the Stuttgart Media University.

Fifteen images are used in this study and shown in Fig. 4. The first 12 images were taken from Mark Fairchild's HDR Photographic Survey [3], and these were supplemented by one image (number 13) from the Stuttgart Media University [4] and two images created by BBC R&D (numbers 14 and 15). Since the dynamic range of the raw images was greater than that expected for use in HDR television, the images are scaled to look aesthetically pleasing (as judged by a small number of expert viewers) with a smaller dynamic range. This is equivalent to adjusting the camera iris. The test set included bright and dark images, and several images with regions of both light and shade. The images were converted to Recommendations ITU-R BT.2100/BT.2020 colour primaries and shown at a resolution of 1 920×1 080 pixels to match the maximum resolution of the display.

Each image is displayed at four peak display luminance levels, 500, 1 000, 2 000 and 4 000 cd/m^2, using a scene-referred approach to luminance scaling. A gamma function appropriate to the display peak luminance was applied, following Recommendation ITU-R BT.2100, Note 5f to Table 5. 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 scene-referred systems.

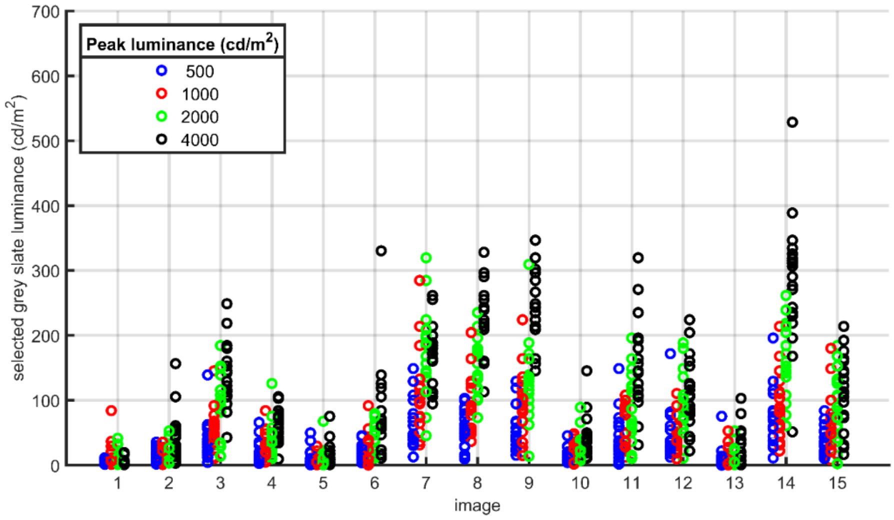
Subjects were seated at a distance of 1.9 m from the display, which corresponds to 3.2 times the screen height. Each subject was screened for normal visual acuity before the test, then given written instructions. Two training images were provided, and three "dummy" images were included at the start of 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.

Results of experiment to establish brightness values for test images

Figure 5 shows the individual responses of all 20 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 Fig. 4) 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 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.

FIGURE 5

Selected grey slate luminance levels for individual test subjects, according to the image and peak display luminance. Each circle corresponds to one subject's rating of that image



Objective metrics

For the purposes of this study, models that relate the displayed pixel luminance values to the overall perceived brightness values collected from the subjective tests are developed. The model will eventually need to operate on signal values rather than displayed light levels if it 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 $Y'C_b'C_r'$ images with 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 Recommendation ITU-R BT.2100, 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 Recommendations ITU-R BT.2100/BT.2020 colour equations.

The following models produce a numerical value for the overall brightness from the displayed pixel luminance values of our test images. The displayed luminance is defined 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 $M = 1\,080$ and $N = 1\,920$.

1) Mean display luminance

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

2) Mean log10 display luminance

$\frac{1}{MN}\sum_i \sum_j \log_{10}(Y_D(i,j))$. Following Fechner [5].

3) Mean PQ inverse EOTF of display luminance

$\frac{1}{MN}\sum_i \sum_j \mathrm{EOTF}^{-1}(Y_D(i,j))$, with EOTF^{-1} as defined in Recommendation ITU-R BT.2100 Table 4.

4) Mean display luminance raised to a power

$\frac{1}{MN}\sum_i \sum_j (Y_D(i,j))^p$. Following Stevens [6]. Here, values are tested from $p = 0.2$ to $p = 1$.

5) Mean CIE 1976 lightness

$\frac{1}{MN}\sum_i\sum_j L^*(Y_D(i,j))$, with L^* as defined by the CIE [7], including the linear section at low luminance levels. The displayed light values corresponding to 75% signal level as the reference white are used, 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, 1 000, 2 000 and 4 000 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 [8]. In the form used here, it is assumed that the viewer is fixated at the centre of the screen, and approximate that all pixels subtend the same angle at the retina, omitting 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: calculate the average luminance of only those pixels in the central quarter of the screen.

8) Percentiles

The nth 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. Percentiles from P_{10} to P_{100} are tested.

9) Percentile ranges

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

Calculate the mean display luminance of only those values between P_{25} and P_{75}. The range P_{10} to P_{90} is also tested.

The mean selected grey levels for the 60 test images (15 images at four peak brightnesses, see Fig. 4) were used as ground truth brightness values to evaluate the models. Pearson's correlation coefficient and Spearman's rank correlation coefficient between each model and the ground truth are reported.

Correlation of objective metrics with subjective results

The Pearson correlation coefficients and Spearman rank correlation coefficients for all of the brightness metrics studied are shown in Table 1. The results are also presented graphically in Figs 6 and 7.

The simplest metric, the mean displayed luminance (metric 1), is a very good match, with a Pearson correlation coefficient of just under 0.96. Its performance is shown in Fig. 8, left plot. The various non-linear scaling methods in metrics 2-5 only cause the data to deviate from the straight line of metric 1, as shown for metric 5 with an exponent of 0.33 in the right-hand plot in Fig. 8, where a curved line would clearly be a better fit. The only exception is raising the pixel luminance values to a power of between 0.8 and 0.9 before averaging (see Fig. 7, left), which offers a very slight improvement over metric 1.

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) perform very poorly, but the higher percentiles correlate well with the subjective test results. This 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 correlation drops for percentiles above the 90th, which implies that very small, bright areas have a smaller effect on the overall brightness.

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. All of our test images contained some dark regions, so the percentile ranges in this case are not any more informative than the individual percentiles.

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

FIGURE 6

Correlation coefficients for all studied metrics
Parameters giving the highest correlation are used for metrics 4 and 8

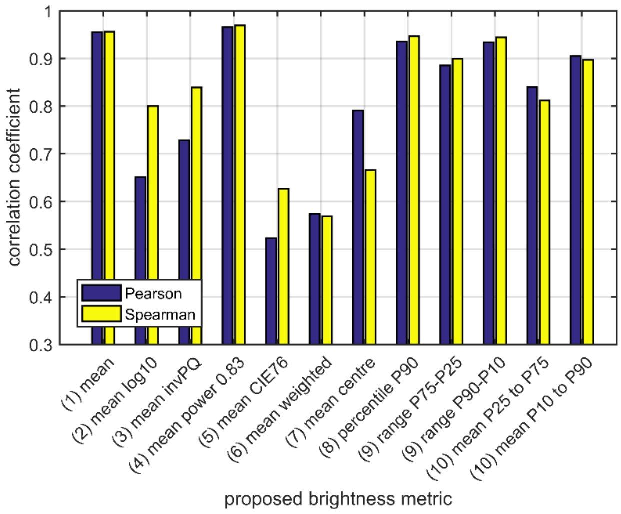


FIGURE 7

Correlation coefficients for metric4, the mean displayed pixel luminance values raised to a power, with exponents from 0.2 to 1 (left); and correlation coefficients for metric 8, percentiles 10 to 100 of the displayed pixel luminance values (right)

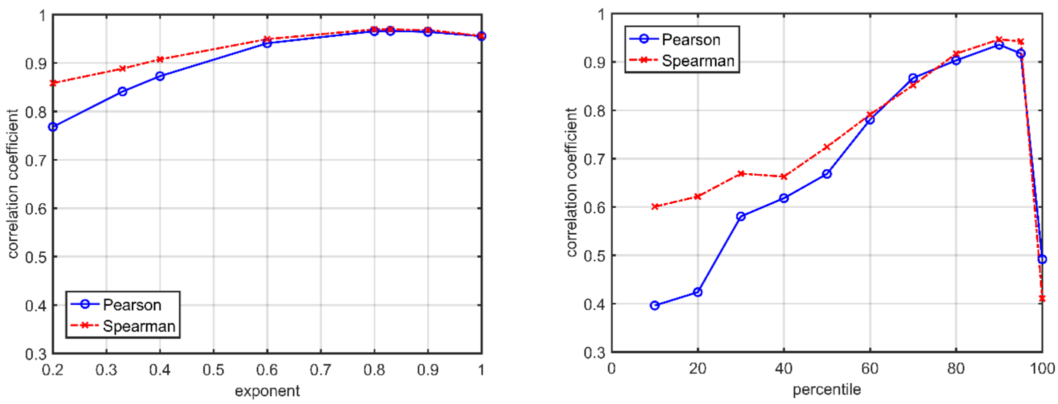


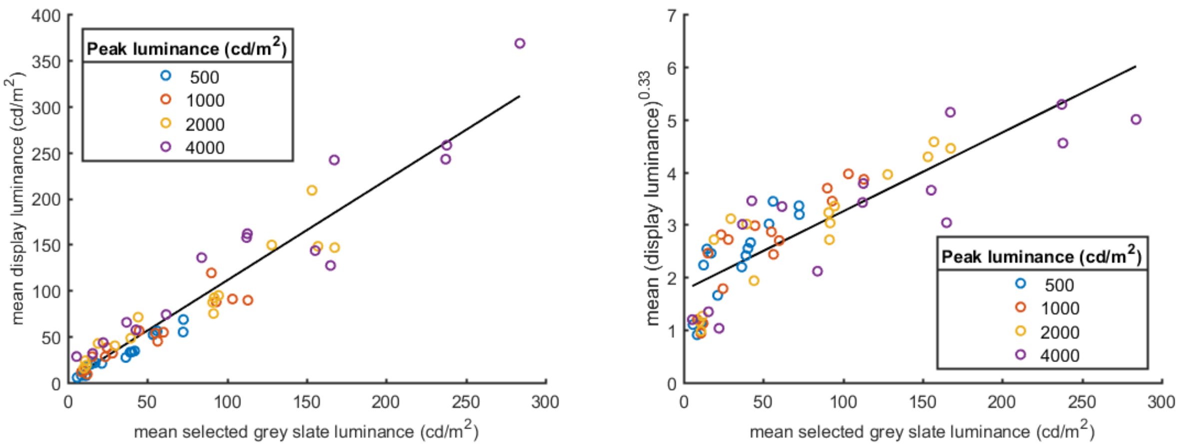
TABLE 1

Pearson and Spearman rank correlation coefficients for the studied brightness metrics. The best correlation in each section of the Table is highlighted in bold

Metric	Parameter value (if applicable)	Pearson correlation coefficient	Spearman rank correlation coefficient
1) Mean display luminance	**-**	**0.955378**	**0.955877**
2) Mean log10 display luminance	-	0.651608	0.800111
3) Mean PQ inverse EOTF of display luminance	-	0.728708	0.839344
4) Mean CIE 1976 Lightness	-	0.523273	0.626341
5) Mean display luminance raised to a power	0.2	0.767964	0.858405
	0.33	0.841025	0.888358
	0.4	0.872807	0.907530
	0.6	0.940758	0.949264
	0.8	0.965687	**0.969492**
	0.83	**0.966131**	0.969214
	0.9	0.964285	0.968158
6) Mean weighted displayed luminance	-	0.573798	0.568769
7) Mean of values in centre of screen	**-**	**0.791089**	**0.666296**
8) Percentiles	P_{10}	0.395950	0.600389
	P_{20}	0.423570	0.621506
	P_{30}	0.580292	0.668797
	P_{40}	0.617995	0.662573
	P_{50}	0.668203	0.724034
	P_{60}	0.780425	0.790886
	P_{70}	0.866544	0.851570
	P_{80}	0.903016	0.916921
	$\boldsymbol{P_{90}}$	**0.935327**	**0.946374**
	P_{95}	0.917317	0.942317
	P_{100}	0.491586	0.410593
9) Percentile ranges	$P_{75} - P_{25}$	0.885345	0.899194
	$\boldsymbol{P_{90} - P_{10}}$	**0.934097**	**0.944540**
10) Mean of values within a specified range	P_{25} to P_{75}	0.839611	0.811892
	$\boldsymbol{P_{10}}$ **to** $\boldsymbol{P_{90}}$	**0.905243**	**0.89686**

FIGURE 8

Metric 1, mean display luminance (left), and metric 5, mean display luminance raised to the power of 0.33 (right), each plotted against the mean selected grey slate luminance (used as a measure of the subjective brightness, see Annex 2 for details), for each test image



Conclusion

Ten classes of objective brightness metric that relate the displayed pixel luminance values to the overall image brightness are studied. Previous measurements of the perceived overall brightness of a set of HDR images determine the ground truth data for evaluation of the studied metrics.

The best performing metric raises the displayed pixel luminance values to a power of 0.83 before calculating the mean, but this was only marginally better than a simpler metric that finds the mean displayed pixel luminance values directly. The improvement offered by first raising the values to a power is very small, and likely to be within the limits of experimental uncertainty, so the simpler method is preferred for real-time applications. The high correlation of 0.96 suggests that this simple metric will be an effective basis for a brightness monitor.

References

[1] Noland, K.C., Pindoria M. and Cotton, A., *"Modelling Brightness Perception for High Dynamic Range Television"* Research & Development White Paper, 2019, pp. 1, 2. Available: http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP341.pdf

[2] Hunt, R. W. G., "*The Reproduction of Colour in Photography, Printing and Television*," 4th ed. Fountain Press, 1987.

[3] Fairchild, Mark, *"The HDR Photographic Survey,"* website, http://markfairchild.org/HDR.html

[4] Fröhlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A., and Brendel, H., *"Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays,"* Proc. SPIE 9023, Digital Photography X, 90230X, 7 March 2014. Images available online at https://www.hdm-stuttgart.de/vmlab/hdm-hdr-2014/

[5] Fechner, G. T., "*Elemente der Psychophysik*," Breitkopf und Härtel, 1860.

[6] Stevens, S. S., "*To Honour Fechner and Repeal His Law*," Science, vol. 133, no. 3446, pp. 80-86, 1961.

[7] Commission Internationale de l'Eclairage, "CIE 1976 L*u*v* colour space," 2014. [Online] Available: https://cie.co.at/eilv/159.

[8] Moon, P. and Spencer, D. E., "*The Visual Effect of Non-Uniform Surrounds*," J. Optical Soc. America, vol. 25, no. 3, pp. 233-248, 1945.

Annex 3
(informative)

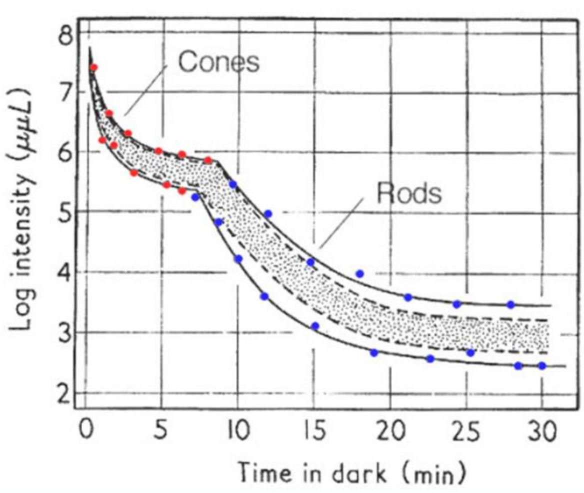
Background to the Temporal Image Level functional form

Introduction

Human vision adapts over time to the content that it sees. Such adaptation is in some sense dependent on the content that the observer has been viewing in the recent past. It has long been known that for a stimulus presented at a fixed luminance and for a fixed duration, the adaptation level of the observer is related to the product of the presented luminance and its duration (i.e. the total energy to which the observer was exposed) [1] [2] [3]. If after fully adapting to such a fixed luminance level, the stimulus is removed, then dark adaptation follows which takes around 30 minutes to fully take effect. The curve of dark adaptation as function of time is shown in Fig. 9.

FIGURE 9[7]

Dark adaptation curve



Note to Fig. 9: The shaded area represents 80% of the group of subjects. (Hecht and Mandelbaum's data from Pirenne M. H., Dark Adaptation and Night Vision. Chapter 5. In: Davson, H. (ed), The Eye, vol 2. London, Academic Press, 1962.)

It can be seen that rods and cones adapt along similar curves, but in a different light regime. In the fovea only cones exist, so the portion of the curve determined by the rods would be absent. As mentioned above, dark adaptation curves depend on the pre-adapting luminance. Further, the duration of the pre-adapting luminance has an effect on dark adaptation. Shorter durations of pre-adapting

[7] This Figure was taken from: https://webvision.med.utah.edu/book/part-viii-gabac-receptors/light-and-dark-adaptation/.

luminance result in faster adaptation. This suggests that exposure to luminance longer ago results in a smaller effect on the current state of adaptation.

It can be hypothesized that the current state of adaptation of an observer exposed to video content can be approximated by integrating the luminance of past video frames in a weighted manner, so that frames displayed longer ago are given a lower weight. The equivalent in terms of image processing would be to integrate each pixel location individually over a certain number of preceding frames. This integration, however, would be equivalent to applying a temporal low-pass filter to each pixel location. Thus, in principle it would be possible to determine the state of adaptation of the visual system of an observer exposed to video by convolving a low-pass filter with the video itself.

Such a convolution is computationally expensive, and additionally the effect can be computed in a much more efficient and biologically plausible manner. To this end, the response of neurons in the (human) brain can be well modelled by (generalized) leaky integrate-and-fire models. Neurons exhibit a relation between neuronal membrane currents at the input stage and membrane voltage at the output stage[8]. It is known that neurons leak some potential according to their membrane resistance, so that at time t the driving current $I(t)$ relates to the membrane voltage V_m as follows:

$$I(t) = \frac{V_m(t)}{R_m} + C_m \frac{dV_m(t)}{dt}$$

where:

R_m : membrane resistance

C_m : capacitance of the neuron.

This is in essence a leaky integrator[9]. It is possible to multiply by R_m, and introduce the membrane time constant $\tau_m = R_m C_m$ to yield[10]:

$$\tau_m \frac{dV_m(t)}{dt} = -V_m(t) + R_m I(t)$$

It is reasonable to assume that at time $t = 0$ the membrane voltage is at a certain constant value, i.e. $V_m(0) = V$, and that at any time after that the input vanishes, i.e. $I(t) = 0$ for $t > 0$. This is equivalent to a neuron beginning adaptation to the absence of input. For a photoreceptor, for example, this would be the case when dark adaptation begins, but it is noted that this process is not unique to photoreceptors. The resulting closed-form solution of the above equation is then:

$$V_m(t) = V\, e^{\frac{-t}{\tau_m}} \qquad \text{for } t > 0$$

It can be seen that this equation indeed qualitatively models the dark adaptation curves of Fig. 9. Note also that this equation is essentially equivalent to the model proposed by Crawford in 1947 [4] [5]. Leaky integration has been shown to be an appropriate model of the adaptive behaviour of neurons implicated in human vision.

For values of t approaching 0, the derivative of the above function tends to $\frac{-V}{\tau_m}$, so that the initial rate of change can be controlled through the parameter τ_m. To arrive at a computationally efficient formulation, the above differential equation is rewritten as follows:

$$\tau_m(V_m(t) - V_m(t-1)) = -V_m(t) + R_m I(t)$$

so that:

---

8 https://en.wikipedia.org/wiki/Biological_neuron_model#Leaky_integrate-and-fire.

9 https://en.wikipedia.org/wiki/Leaky_integrator.

10 https://neuronaldynamics.epfl.ch/online/Ch1.S3.html.

$$(\tau_m + 1)V_m(t) - \tau_m V_m(t-1) = R_m I(t)$$

which subsequently leads to:

$$V_m(t) = \frac{\tau_m}{\tau_m+1}\left(V_m(t-1) + \frac{I(t)}{C_m}\right)$$

The structure of this equation suggests that the output of a neuron V_m at time t is a function of the output of the neuron at time $t-1$, as well as the input I at time t.

For the purpose of implementing this model as a leaky integrator that can be applied to images or values derived from images, the membrane resistance R_m may be set to 1, so that:

$$V_m(t) = \frac{\tau_m}{\tau_m+1}\left(V_m(t-1) + \frac{I(t)}{\tau_m}\right)$$

To apply this model in a broadcast setting, a single adaptation level per frame is preferable, rather than a per-pixel adaptation level. This may be achieved by noting that the steady-state adaptation $L_a(t)$ of frame t may be approximated by some average luminance of a frame, such as for instance the image level (IL). The temporal state of adaptation $L_T(t)$ is then given by:

$$L_T(t) = \frac{\tau_m}{\tau_m+1}\left(L_T(t-1) + \frac{L_a(t)}{\tau_m}\right)$$

The effect of applying this method is that of a temporal low-pass filter, albeit without the computational complexity associated with such filter operations. It is therefore included as a measure of human visual adaptation to moving content.

The above equation can be rearranged to:

$$L_T(t) = L_T(t-1)\left(1 - \frac{1}{\tau_m + 1}\right) + L_a(t)\left(\frac{1}{\tau_m + 1}\right)$$

References

[1] F. A. Mote and A. J. Riopelle. The effect of varying the intensity and the duration of preexposure upon foveal dark adaptation in the human eye. J. comp. physiol. Psychol., pp. 46(1):49-55, 1953.

[2] Ledda, Patrick, Luis Paulo Santos, and Alan Chalmers. "A local model of eye adaptation for high dynamic range images." In Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa, pp. 151-160. ACM, 2004.

[3] Mote, F. A., and A. J. Riopelle. "The effect of varying the intensity and the duration of preexposure upon foveal dark adaptation in the human eye." The Journal of general physiology 34, No. 5 (1951): pp. 657-674.

[4] Crawford, B. H. "Visual adaptation in relation to brief conditioning stimuli". Proc. R. Soc. Lond. B 134, No. 875 (1947): pp. 283-302.

[5] Pianta, Michael J., and Michael Kalloniatis. "Characterisation of dark adaptation in human cone pathways: an application of the equivalent background hypothesis". The Journal of physiology 528, No. 3 (2000): pp. 591-608.

Annex 4
(informative)

Verification of the time constant in the temporal image level

Introduction

This Annex describes research performed to investigate the appropriate τ for the time of decay algorithm contained in § 2.

Subjective testing

A set of subjective tests was developed by Dolby Laboratories. The novelty in this experiment is to study subjective tolerance to brightness jumps over changes in time. Two versions of the experiment were conducted to estimate the visual responses to transitions from bright-to-dark-to-bright and those from dark-to-bright-to-dark. The former was done by, first, showing observers a bright image for 10 seconds, then switching to a dark image for a variable amount of time, and finally switching back to the bright image for 7 seconds. In a similar vein, the latter was done by first, showing observers a dark image for 10 seconds, then switching to a bright image for a variable amount of time, and finally switching back to the dark image for 7 seconds. Table 2 showcases the different durations of the middle image that were used in both versions of the experiment. Prior to the experiment, a series of Beta tests were run to determine the optimal time durations to use for each version.

TABLE 2

Durations of the adaptation time course implemented in the experiment

Experiment version	Duration of middle image seconds
Bright-to-dark-to-bright	0.5, 5, 10, 25, 40, 60, 90, 120, 150
Dark-to-bright-to-dark	0.5, 2, 5, 10, 15, 20, 30, 60

Following the transition of interest (from the middle image to the final image), subjects were asked to rate their experience. Table 3 showcases the scores and terminology affiliated with the impairment scale. The terms that indicate a degree of pain/annoyance each have two values to add gradation. Additionally, the two versions of the test use different language to reflect the observer feelings following the brightness junction.

TABLE 3

Impairment scale used to rank the junction for both versions of the experiment

Score(s)	Terminology bright-to-dark-to-bright	Terminology dark-to-bright-to-dark
1, 1.5	Painful	Annoying
2, 2.5	Notably Painful	Notably Annoying
3, 3.5	Slightly Painful	Slightly Annoying
4	Not Painful	Not Annoying

Subjective participants for this experiment come from a mixture of 17 expert and non-expert viewers. A Recommendation ITU-R BT.2100 viewing condition with a 5 cd/m^2 surround was used with a

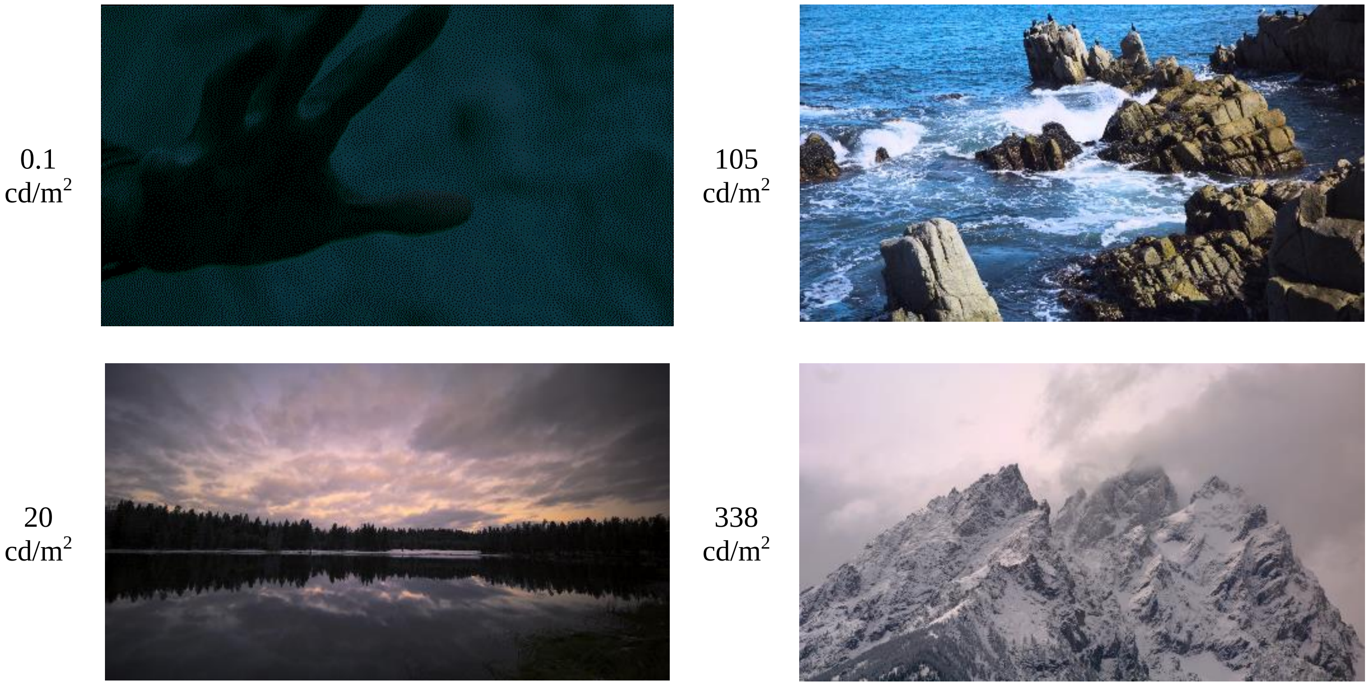
Christie 4K 6P laser projector system. The laser projector was calibrated to a peak luminance setting of 1 000 cd/m^2 and had a measured black level of 0.0004 cd/m^2. The images were situated such that they subtended a horizontal angle of 3.2 picture heights from the observer's viewing position. The image signal was sent over standard digital interface (SDI) after being encoded with the PQ electro optic transfer function.

Test images

A total of four images were used in this study (shown in Fig. 10). One image served as the "dark" image while the other three images served as different levels of the "bright" image. The testing procedure was repeated for each bright image. All images were tone mapped for display within the luminance range of the projector. The dark image was created by Dolby Laboratories, Inc. (Dolby) and the three bright images were licensed to Dolby by Spears & Munsil[11]. The images, along with their calculated mean display luminance values, are shown in Fig. 1. The mean display luminance values were calculated using the Image Level algorithm in § 1.

FIGURE 10

Test images at specified mean display luminance values on the 1 000 cd/m^2 projector



Test results

Within the two versions of the experiment, three different "bright" images were evaluated to test the transition to different mean display luminance values. This was done to show how the mean subjective score (MSS) changes with brightness jump magnitude on top of how it changes with the time course of adaptation.

Figure 11 shows the mean ratings and 95% confidence intervals for the three images across the nine adaptation time trials. An MSS of 4 represents a transition that does not introduce any eye pain or subjective annoyance, while scores at 2 or below suggest notable eye pain or annoyance.

[11] Spears & Munsil Ultra HD Benchmark (2023). https://www.biaslighting.com/products/spears-munsil-ultra-hd-benchmark-2023.

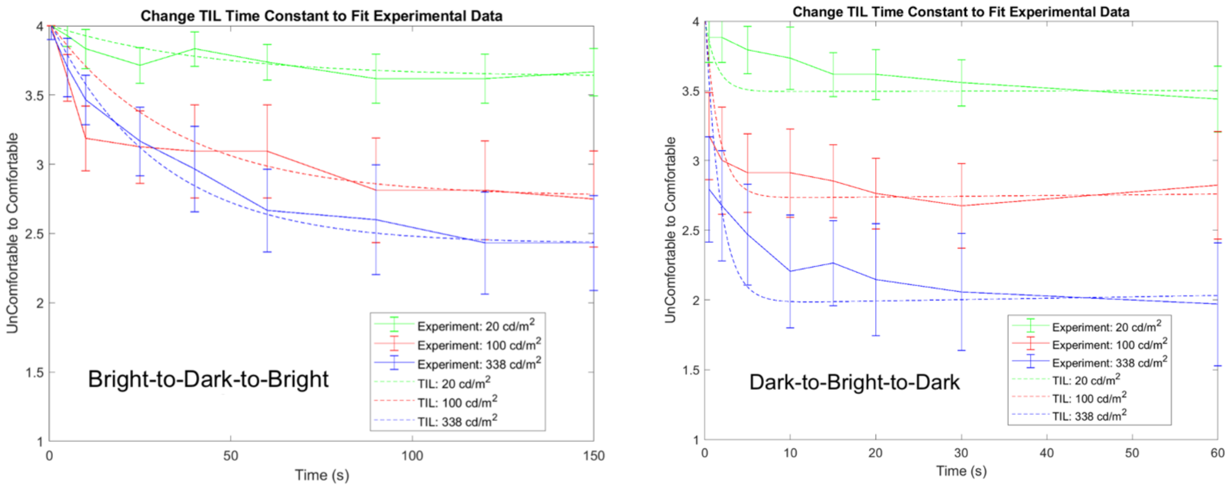
To fit an appropriate τ for the TIL equation it is necessary to relate adaptation time to MSS results. The ILR in § 3 is a closer approximation of subjective experience.

Hence, following conversion from mean display luminance to image level, the TIL equations were applied with various τ to estimate the level of adaptation to the dark image following the initial ten second exposure to the bright image and the different adaptation times. Once the TIL estimates were made for each combination of bright image and adaptation time, the ILR values were calculated. The response values were then scaled such that they reflect the MSS range spanned by their respective brightness values. This was done via the standard range normalization and scaling procedure.

These TIL-estimated MSS values with optimal τ are plotted with dashed lines alongside the experimental data with solid lines for the experiment in Fig. 11.

FIGURE 11

Mean subjective scores of brightness transitions for differing adaptation times compared to approximate generated mean subjective scores



The optimal τ values for this functional form are given below (within the confidence intervals). To keep adaptation times low, for § 2 the lower end of the optimal values is used.

$$\tau = \begin{cases} [22 \text{ to } 25]\,\frac{f}{24} & \text{if } p(t) \geq 0 \\ [800 \text{ to } 2\,000]\,\frac{f}{24} & \text{if } p(t) < 0 \end{cases}$$

The time decay more closely follows the trends of the experimental data for the bright-to-dark-to-bright version. There is a notable deviation for 100 cd/m^2 at 10 seconds. Future work investigating this outlier would be helpful. The time of decay constant does not follow the dark-to-bright-to-dark experimental data as cleanly as it does for the bright-to-dark-to-bright version. This largely comes from the increased influence the brightness junction magnitude brings in short durations. Observers noticed an immediate afterimage with even the shortest time duration of a bright image. The current functional form does not allow for an immediate drop followed by a slow roll-off. Further work investigating this phenomenon would be helpful.

____