https://doi.org/10.2352/ISSN.2169-2629.2018.26.336

Assessing Color Discernibility in HDR Imaging using Adaptation Hulls

Timo Kunkel, Robert Wanat, Jaclyn Pytlarz, Elizabeth Pieri, Robin Atkins and Scott Daly
Dolby Laboratories, Inc.; Sunnyvale, CA

Abstract

Objectively predicting the discernibility of color differences is a common requirement when assessing the performance of a display device and is the domain of color difference metrics. Metrics commonly used for this task assess the color discriminability of two stimuli based directly or indirectly on the viewing environment with a known, constant adaptation luminance level L_A.

These metrics were originally derived for color assessment of reflective and transmissive media as well as low dynamic range displays, where L_A can both be maintained and estimated with reasonable certainty for any likely stimulus pair. With High Dynamic Range (HDR) and Wide Color Gamut (WCG) displays, using Steady State metrics is becoming increasingly challenging when assessing the discernibility of two similar stimuli over the display's full range of capability. This is especially true if spatially or temporally varying HDR content is displayed, causing the Human Visual System (HVS) to frequently and unpredictably change L_A, and with that visual sensitivity.

To overcome these challenges, we present the concept of an "Adaptation Hull" color difference metric. Rather than using a specified adaptation luminance that is in most cases substantially different than the stimuli under test, an Adaptation Hull metric instead considers an optimal adaptation state where the HVS has the highest sensitivity to color differences.

Introduction

The mass-market adoption of High Dynamic Range (HDR) and Wide Color Gamut (WCG) technologies has led to display devices that are capable of rendering brighter highlights, deeper blacks and more intense colors compared to traditional displays. Due to this development, the demand to assess the rendering capabilities of elements in HDR imaging ecosystems has also significantly increased over the previous years, for example to guide content mapping as well as display design, calibration and comparison.

One important aspect when determining the performance of HDR imaging ecosystems is the objective assessment of color discernibility or color difference, for which several metrics have been proposed and standardized. Generally, these metrics are based on color models that assume a static, known or well-defined viewing environment. Viewing a static stimulus under a static viewing environment keeps the HVS in a balanced simultaneously perceivable dynamic range or Steady State [1]. Therefore, we refer to those metrics and their underlying color models as "Steady State".

It can be argued that the reason why the reproducible color range (intensity and chromaticity) could have been modelled sufficiently while assuming a single state of adaptation is because typical reflective, transmissive and SDR display color targets are often assumed to not significantly invoke many of the human visual system's (HVS) local or temporal adaptation processes. With emissive HDR and WCG imaging systems this assumption no longer holds true as the content shown on HDR-capable displays is more likely to cause significant changes in adaptation state of the HVS [2, 3, 4]. For example, recent work shows that pupil area, one of the constituent mechanisms responsible for luminance adaptation, could vary within the ratio of 3:1 during HDR content viewing [5]. Moreover, increasing screens sizes cause the image to occupy more of the visual field which in turn causes the display luminance to become the dominant factor driving the HVS adaptation.

In the following, we discuss how viewing HDR and WCG content causes larger changes in adaptation and why this is significant for determining discernibility of colors. Given that the state of adaptation cannot be feasibly established during content playback, we argue that a sensible choice for evaluating color differences is to instead take the conservative approach and assume that the adaptation is at its most sensitive state. As a consequence, we propose the establishment of a new class of color difference metrics called "Adaptation Hull" metrics. Using this approach improves efficiency when assessing the rendering capabilities of an HDR/WCG display by separating display specific properties from viewing environment- and content-dependent aspects of Steady State models that are not required or unknown.

Previous Work

Objective color difference metrics have been a long-time research subject. In 1976, the CIE standardized its first ΔE metric which was developed alongside the CIELAB color space and is today known under the name ΔE_{76} [6]. This metric was further improved in the revision of the standard now known as ΔE_{94} [7]. The latest revision of the standard, ΔE_{2000} [8], is the industry standard in color difference estimation. In all three standards, the input data is represented by two CIELAB triplets of the colors being compared. Because of the use of CIELAB to represent the input, these models require a known white point and a controlled adaptation. The error metric is calculated as the Euclidean distance between the colors, either in CIELAB space or in the transformed LCh space (lightness, chroma and hue). In the later revisions, the distance along each axis is additionally scaled by the functions that compensate for hue nonlinearity of the CIELAB space, add the crispening effect (highest HVS sensitivity around adaptation point) and model the effect of chroma.

Another CIE standard that can be used to predict color differences is a color appearance model known as CIECAM. Originally released as CIECAM97s, it was later revised as CIECAM02 [9]. Even though the main goal of the model is predicting absolute and relative color correlates such as lightness, colorfulness, hue, etc., it can be converted into a color space with uniform color differences. There exist several such conversions, the most popular of which is the CIECAM-UCS [10]. CIECAM-based models require information about the contents of the visual field, including the relative CIEXYZ coordinates of the patches compared and those of the white point, luminance of the background, absolute adaptation luminance and the viewing conditions, including the degree of adaptation. The model is intended to be used only with single patches and does not take spatial information about the stimuli into account.

This limitation of not considering spatial information was addressed by Kuang et al. in iCAM [11], an image color appearance metric which incorporates local adaptation by using lowpass-filtered image luminance as an approximation of adapting luminance. The model uses the IPT color space [12] to improve hue linearity while retaining the color correlate calculation formulas from CIECAM02. Because of this, the model still belongs to the Steady State family because the knowledge about the adaptation state is necessary for its calculation, even though it is implicitly calculated based on the contents of the image.

For completeness, we should mention CMC [13] which is also based on the CIE LCh color space. Although CMC is a tolerance instead of a difference metric, it serves a similar purpose. Tolerances predict color errors, but with the distinct goal of determining when color errors become objectionable.

The aforementioned color difference metrics have found widespread use in the paint, material and textile industries. They are also widely used to assess the accuracy and rendering performance of stimuli shown on emissive display devices. Using Steady State metrics for this task has been acceptable because until now most displays have offered a fairly limited dynamic range and color gamut. Nevertheless, there has been an interest in approaches that are not based on Steady State principles.

Properties of adaptation hulls can already be attributed to earlier models such as DICOM [14], developed in 1992 for use in interchange of achromatic medical images. The part of the standard related to storing of images uses an encoding scheme which does not expect adapting luminance when viewing the content, instead assuming the conservative full adaptation to the stimulus. This allows for a more efficient quantization of the luminance space, up to the limits of visual detection. To predict the detection threshold, the contrast sensitivity function (CSF) by Barten [15] is used, with a constant stimulus spatial frequency of 4 cyc/deg, which is close to the highest sensitivity of the HVS across the majority of the photopic range.

A similar model was developed under the name Perceptual Quantizer (PQ) [16] and standardized by SMPTE as ST.2084 [17]. PQ makes an even more conservative prediction than DICOM on detection thresholds by using the spatial frequency at which the contrast sensitivity is the highest at each luminance. As a result, it can be used to encode the range between 0.0 and 10,000 cd/m^2 where, when using 12-bit codewords, the luminance differences between consequent code words never exceed the detection threshold.

An early approach on extending PQ to also include a chromatic component was proposed by combining it with CIECAM02-UCS [18]. This is achieved by defining the CIECAM02 white-point, L_A and background Y_b to be relative to the stimulus luminance.

More recently, PQ became the basis of the IC_TC_P color space [19], which is calculated by applying the PQ nonlinearity to LMS cone responses assuming D_{65} adaptation and then converting the result into three channels, chromatic C_T and C_P and achromatic I. The conversion matrix was optimized to maximize hue linearity using the Ebner and Fairchild dataset [12] that is the basis of the IPT color space. Starting from data from MacAdams' experiments [20], a new color difference metric called ΔIC_TC_P was proposed with the goal of predicting near-threshold color differences. The metric does not predict color differences in specific conditions, instead offering insight into the possibility of perceiving differences under the highest visual system sensitivity, when fully adapted to the compared colors.

In the following sections, we will explain the need for such a color difference metric, evaluate its performance and offer insights into typical use cases. There has been a recent experiment comparing the accuracy of ΔIC_TC_P against other color difference metrics conducted by Safdar et al. [21]. Because their assumed use case reflected a Steady State scenario for which ΔIC_TC_P is not intended, ΔIC_TC_P performed worse than the other tested metrics.

Behavior of the HVS when Displaying HDR Content

Due to the HVS' adaptive processes, color differences that may not be discernible under a static state of adaptation can become visible if adaptation is variable and can manifest, for example, as contouring or color inaccuracies. At the same time, the range and rate of adaptation change is difficult to predict as it would not only have to include the time course of adaptation but also predict fixations and saccadic motion of the eyes (in the following, we relate the term ' stimulus' to what the HVS perceives for the duration of a fixation, independent of the complexity of the imagery content). One potential but complex option could be to apply existing Steady State models exhaustively over all potential states of adaptation from the lower to the upper boundary of the display in order to identify the discernibility of two stimuli.

To avoid such complexity, a new metrical approach is desirable that factors in unpredictable adaptive processes that hinder the use of Steady State metrics. In order to support the concept behind the proposed new metrical approach, we first discuss several aspects that facilitate adaptive processes in the context of consuming typical image and video content on HDR/WCG displays.

Spatial Impact on Adaptation

Perceptual color models factor in the spatial properties of the HVS. For example, CIECAM02 describes the area covered by a stimulus to be 2°, which in turn is extended to 10° by the background of the stimulus [9]. Everything beyond 10° visual angle is the stimulus surround which is intended to comprise the ambient illumination. Conceptually, the surround attributes the vision cues that motivate adaptation.

When applying color models in context of an SDR display, the background and surround parameters have to be defined as well, which is typically not straight-forward. It is therefore alluring to resort to a simplification when providing those parameters to the color models: the background consists of the active screen area while the surround is defined as everything beyond such as room illumination.

With large HDR/WCG display devices, this problem becomes more complex. HDR displays and TVs in general are constantly increasing in size, with 55" and 65" displays being common and even 75" or larger becoming available to average consumers. At the same time, the viewer's distance to the display is typically not changing e.g. due to the way their living room is set up. Therefore, when viewing content, a modern display subtends an increasing viewing angle of the viewers' visual field, for example from 37° and 65° when viewing 3 and 1.5 picture height away, respectively (in accordance with UHD TV design). Further, content such as video imagery is not necessarily well defined in size.

With such viewing set-ups, the color model's 10° visual angle covering the background area might easily fall inside the viewing area of the display and with that, the displayed content can impact the surround and consequently the state of adaptation. This concept is illustrated in Figure 1.

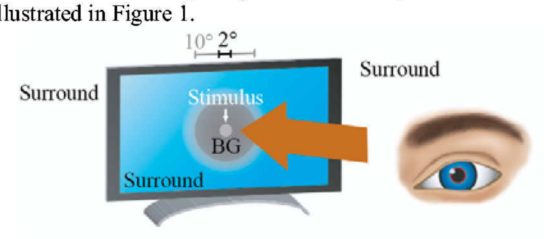

Figure 1: The area of the background (BG) and surround as defined by CIECAM02 in context of a large HDR/WCG display. The visual area of images often extends beyond the 10° visual angle for the background specified in CIECAM02. Therefore, the displayed content has a direct impact on the surround and with that potential changes in adaptation.

Further, it is well known that visual glare and light scattering in the optical media reduces perceivable contrast (e.g. [22]). Therefore, it has been argued that very dark areas requiring dark adaptation do not occur even on HDR displays [23]. However, that work did not consider a wide range of image statistics or video content. Nevertheless, if spatial properties such as fixations of the eyes to different areas of the displayed content are considered, the distribution function of glare in a scene has also to be factored in. In the latter scenario, there will be areas where glare is pronounced, reducing perceivable dynamic range. However, it is also realistic that glare might not have a significant impact even though a bright stimulus is present on the screen. For example, if a viewer fixates on an area sufficiently far from the high-luminance area, veiling glare will trail off to insignificant values beyond 16° visual angle [24, 25]. This is feasible with today's TVs and viewing behavior and needs to be accounted for when predicting discernibility of colors.

Temporal Impact on Adaptation

Typical content displayed on HDR/WCG displays is not limited to still images and usually contains a temporal component that is not only defined by the progression of actual imagery but also influenced by creative storytelling. For example, a dark scene can directly follow a bright one, and a scene with higher dynamic range can follow one with a lower dynamic range, causing the HVS' adaptation to change. Example scenes where this would be the case are given in Figure 2. There, the histograms are calculated from the source HDR images, illustrating their respective dynamic range (DR) brackets and central tendencies. It is visible that over time, the DR covered by all three scenes, spans more than the 3.6 Orders of Magnitude (OoM) identified for the HVS' Steady State [2].

In this context, the time-course of adaptive processes has also to be considered. Some of those processes happen faster than others when watching content. Even though full adaptation to low scotopic luminance can take tens of minutes [26], Stokkermans and Heynderickx [27] have shown that at least partial dark adaptation such as from 25 to 0.001 cd/m^2 is carried out in the order of seconds, which is a realistic assumption for scene length.

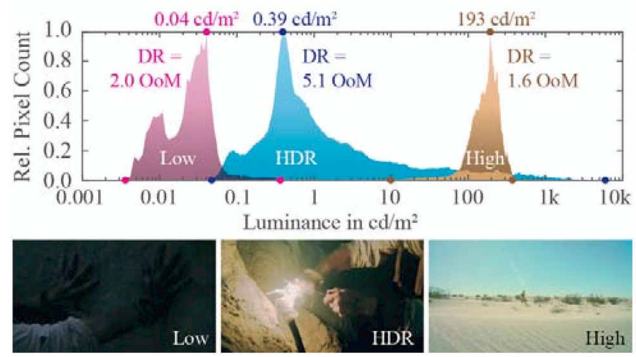

Figure 2: Typical luminance distribution of an HDR image as well as a very dark and bright scene. All these ranges can be displayed on today's HDR/WCG displays. When shown in sequence, e.g. as elements of artistic storytelling, the scenes will be long enough for the HVS to adapt to each luminance range in order to optimize sensitivity. Images from 'One-Way-Ticket' (Dolby). The images are tone-mapped for print.

Deriving an Adaptation Hull Metric

We have established that the HVS' state of adaptation cannot be predicted easily when watching content on emissive HDR/WCG displays because adaptive processes can change the sensitivity of the HVS in an unpredictable manner. Therefore, the discernibility of two stimuli shown on such displays will also differ as a function of adapting luminance. For example, at some adaptation luminance levels those stimuli are discernible while at others they are not.

Figure 3 illustrates, on a simplified level, how to establish an adaptation hull from a Steady State color volume. A Steady State color volume can be illustrated as a double cone that has its largest diameter where the contrast sensitivity is the highest for the prevailing environment (here a function of the state of adaptation $L_{A,1}$). This double cone converges to white on the top and to black at the bottom due to non-linear (sigmoidal) response compression [28]. Due to this, the sensitivity to distinguish color differences can be considered highest where the cone has the largest diameter. This is where a 1 JND 'thick' disc can be extracted (1) as contrast sensitivity would be lower at any other intensity under $L_{A,1}$ such as depicted by (2). Now, the adapting luminance can be changed to $L_{A,2}$ (3) and the procedure (1) is carried out again. Because of the shift of adaptation, there are colors shifted in and out of the discernibility region when moving from a Steady State based on $L_{A,1}$ to $L_{A,2}$ (4). The extraction of ' discs' can now be repeated for a range of adapting luminances, here denoted $L_{A,2}$ to $L_{A,5}$ (5) forming a stack of the highest sensitivities at their given L_A. If carried out at each JND step along the intensity axis, we end up with an adaptation hull which is limited by the display boundaries, here denoted $L_{A,min}$ to $L_{A,max}$ (6). This cylindrical representation is a simplification as the sensitivity of the HVS to chromatic stimuli is reduced when lowering L_A, ultimately leading to achromatic stimuli in the scotopic range. Also, the actual capabilities of a real display would have an impact on the shape of the ' cylinder'.

It is important to note that due to this process of just considering the highest sensitivity at each adaptation luminance, the concept of lightness as it is described in the color appearance literature [9] is

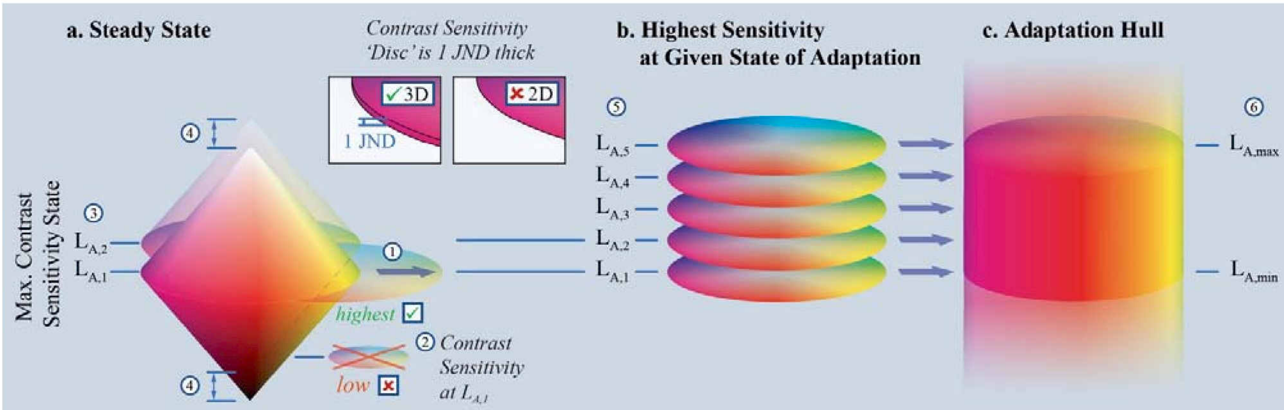

Figure 3: Establishing an Adaptation Hull Model from a Steady State Model. Explanation in text. Please note that this illustration only serves to explain the concept.

not modelled. The model predicts if the differences between colors are discernible while making no predictions on the appearance of such differences. Table 1 summarizes the conceptual differences between Steady State and adaptation hull approaches.

Table 1: Comparison of Features between Steady State and Adaptation Hull Metrics

	Steady State	Adaptation Hull
Designed to predict discernibility of two colors based on:	Their color appearance under a single state of adaptation	Highest sensitivity of HVS of any plausible state of adaptation
Adaptation Level has to be known?	Yes	No
Efficiency to identify highest sensitivity throughout the PQ Range?	Metric would have to be calculated for each JND step of the desired range	Metric is run once
Metric can predict lightness?	Yes	No (not required for task)

In summary, when the color differences being evaluated are in the region of maximum discernibility for a Steady State space the results are similar to the adaptation-hull space which is independent of the state of adaptation. However, when the color differences are in a region of lower discrimination in the Steady State space, the results diverge, and the Steady State space is likely to under-represent the discernibility in comparison to the adaptation hull.

Accuracy of an Adaptation Hull Metric based on IC_TC_P

After we have introduced the concept and application field of adaptation hull metrics, we are interested in how an actual adaptation hull metric based on IC_TC_P compares against a Steady State one. We have chosen CIEΔE_{2000} (called ΔE_{2000} for simplicity) for this task because of its widespread adoption and its common use in context of our application with HDR/WCG displays.

Pieri & Pytlarz [29] created a color difference dataset covering the adaptation luminance levels of 1000, 25 and 0.1 cd/m^2 (all luminance levels are reported in photopic luminance only) and compared them against the aforementioned metrics. These results are shown on the right side of the plot in Figure 4. Discernibility errors between the dataset and the ΔE_{2000} predictions are based on a D_{65} adapting white point, with adapting luminance equal to the luminance of the test pairs. We see that the error is significantly higher with the Steady State ΔE_{2000} metric for the 0.1 cd/m^2 case. ΔIC_TC_P, which is based on an adaptation hull shows a low discrepancy between metric and dataset for all three adaptation levels, supporting the applicability of the adaptation hull concept.

We also analyzed the effect of adaption on this data set using the ΔE_{2000} metric. The results are shown in Table 2. We used the STRESS metric [30] for objective analysis.

Table 2: Comparison of STRESS Performance with Different Assumed Adaptation States

	ΔIC_TC_P	ΔE_{2000}		
Topic	Hull	Adapt to Luminance	Constant $203\frac{cd}{m^2}$ D_{65}	Adapt to Color and Luminance
All	44.5	78.7	70.2	84.1
Color	43.1	80.8	60.5	82.0
Neutral	42.5	64.7	67.3	63.4
$0.1\frac{cd}{m^2}$	56.8	66.4	74.9	75.6
$25\frac{cd}{m^2}$	36.0	51.0	44.2	87.2
$1000\frac{cd}{m^2}$	32.2	55.9	63.0	88.9

In the STRESS metric, a lower value means better performance. We see that no matter the tested adaption state or range of colors, the adaptation hull metric, ΔIC_TC_P performs better. We also see that the best performing adaption state of ΔE_{2000} is different depending on the colors being analyzed. The constant adaptation state introduced by ITU-R BT.2408-1 [31] of 203 cd/m^2 at D_{65} performed

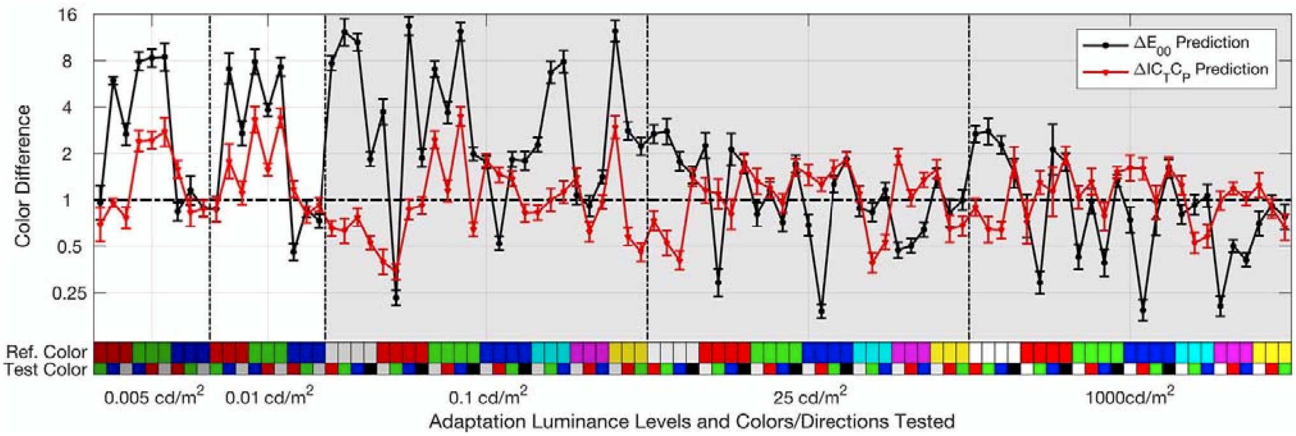

Figure 4: Predicted color differences for measured detection thresholds from the Pieri and Pytlarz experiment [29] (in grey area on the right starting at 0.1 cd/m²) with the results of the experiment described in this paper on the left (0.01 cd/m² and below). The error bars represent standard deviation of the mean calculated using bootstrapping over 1000 iterations. The vertical dash-dotted lines separate adaptation luminance levels.

well for the saturated colors and the 25 cd/m^2 patches, whereas adapting to the color and luminance level performed better for neutrals. Then again, adapting to only the luminance (keeping color fixed at D_{65}) performed the best for low and high luminance patches. So, even in this experiment with a well-controlled adaption state, it is difficult to determine the proper adaptation parameters to get the best performance from ΔE_{2000}.

Color Difference Experiment at Mesopic Luminance

To determine whether ΔE_{2000} continues the trend of over predicting color differences at mesopic luminances below 0.1 cd/m^2, we conducted an experiment at even lower adapting luminance levels than the ones reported by Pieri & Pytlarz. We consider this a pilot study for a larger experiment in the future where we intend to measure color perception at or around the absolute threshold of color vision.

To achieve the necessary stability at low luminance, we used a Dolby PRM-4220 reference monitor. Because the native black level of a PRM is 0.005 cd/m^2, we used two stacked sheets of Rosco E-Colour #211 ND 0.9 filters, lowering the peak luminance from 600 cd/m^2 to 8.2 cd/m^2. This allowed us to test two new luminance levels, 0.005 cd/m^2 and 0.01 cd/m^2, as well as repeat the experiment at 0.1 cd/m^2 as a sanity check. We chose those luminances following an earlier session aimed at determining at what luminance the authors could no longer discern colors. The native primaries of the PRM were shifted by the filters to the values shown in Table 3. The display was calibrated with the filters attached using a PhotoResearch PR-740 spectrophotometer before carrying out the experiment.

Table 3: Chromaticities and peak luminance of the primaries and white point of PRM4220 display used in the experiment

Color	x	y	Y
Red	0.6718	0.3243	1.554
Green	0.2705	0.6827	5.8924
Blue	0.1492	0.0529	0.7567
White	0.2773	0.2958	8.203

We initially attempted to ensure equiluminance of the test stimuli at mesopic levels by using the CIE 191:2010 standard for mesopic luminance [32]. However, the tested colors were more saturated than the street lighting for which CIE 191:2010 was designed. As a result, the red and blue primaries had, respectively, lower and higher ratios of scotopic luminance to photopic luminance than allowed by the model. Even after a 30-minute adaptation period, the difference between the apparent brightness of these two colors when matched in mesopic luminance was highly conspicuous.

We decided to take a different approach than the previous experiment and rather than allowing the observer to fully adapt to the patch luminance, we instead showed an adaptive field at 8 cd/m^2 peak luminance across the entire display between each set of comparisons. The sets were composed of 30 stimuli comparisons and took on average about 135 seconds for each participant to complete. We designed the experiment this way with the intention of preventing significant rhodopsin regeneration and minimizing the influence of rod activation during the comparisons. Previous research has shown that only a small amount of rhodopsin can be regenerated during a period of a single set of comparisons [26]. This experiment setup allows for readaptation to photopic levels, which produces results that are more applicable to predicting color differences in creative media. This is because even on HDR displays, scenes that are fully in the mesopic range are not long enough for complete dark adaptation.

In the experiment, prior to viewing the test stimulus, each observer was shown the 8 cd/m^2 adaptation stimulus for 30 seconds. The test stimulus consisted of a noise image with 1/f spatial frequency composition and a mean luminance equal to the chosen adaptation level in photopic luminance. The noise pattern had the same color as the reference stimulus which matched the chromaticity of one of the primaries (see Table 3). We followed a 4-alternative-forced-choice design in which four uniform squares were shown in the middle of the screen in a 2-by-2 pattern. The side of each patch subtended 2 degrees of visual angle with a 0.5-degree gap between the squares.

During each comparison, one randomly chosen patch was mixed with some amount of the test color, either one of the other primaries or white. The goal of the observer was to choose which patch was different from the others by pressing a corresponding button on a

HORI Fighting Stick Mini 4 controller. We decided not to limit the time the stimulus was displayed on the screen. After each comparison, only the noise pattern was shown for two seconds before the onset of another comparison stimulus. We used the QUEST procedure [33] to provide suggestions for color mixes and to fit a psychometric function to the results. After 30 comparisons, the final QUEST estimate of a 62.5% correct choice threshold was used as the result of the comparison set. The observer was then shown a bright adaptive field at 8 cd/m^2 for 15 seconds before starting another set of comparisons, switching either test color, reference color or adapting luminance.

Five observers, two male and three female, completed the experiment, all with normal or corrected-to-normal vision. All of the observers have extensive experience in color imaging, three of whom are authors of this paper. The age of the observers ranged from 24 to 36. Due to the strenuous nature of the experiment, which had to be conducted in a dark room with very low luminance stimuli, the task was split into two sessions, with comparisons at 0.005 cd/m^2 and 0.1 cd/m^2 conducted one day and comparisons at 0.01 cd/m^2 conducted another day.

The results from each observer were averaged and are shown on the left side of Figure 4, alongside the results from the previous experiment. These results show ΔIC_TC_P and ΔE_{2000} differences calculated between the test and reference color pairs and then averaged across observers. The error bars represent standard deviation of the mean calculated using the bootstrap method with 1000 iterations. For ΔE_{2000}, a reference white was not provided as part of the stimulus so we assumed the white point to be 1/0.18 (approx. 5.55) times higher than the adapting luminance to simulate the highest sensitivity of the HVS. In reality, inferring the white point luminance in a scene that does not contain pixels that appear white (such as our stimuli) is a much more difficult task than inferring adaptation. This is another advantage of the ΔIC_TC_P metric because it can be used accurately with less information about the scene.

The results at 0.1 cd/m^2, measured only as a sanity check, follow the results from the previous experiment closely and thus are omitted from this plot for the sake of simplicity. The same trend that was apparent from the results of the previous experiment can also be seen at low luminance, where the performance of ΔE_{2000} suffers due to the lower adapting luminance levels. This is because ΔE_{2000} was based on datasets measured mostly at photopic levels. Compared to ΔE_{2000}, ΔIC_TC_P provides a better estimate of chromatic JNDs in mesopic conditions, on average. As a result, we suggest the use of ΔIC_TC_P for color difference prediction, especially when mesopic vision is involved.

Linearity Above Threshold

While ΔIC_TC_P was designed with near-threshold color differences in mind, it does not mean that the model is incapable of predicting suprathreshold differences. The term ' suprathreshold' refers to the extended range of color differences with delta values larger than the in the previous section identified threshold at 1.0. In the context of color difference metrics, it is desirable that such values are linear, i.e. doubling the error value should make the subjective error magnitude twice as high. Please note that this linearity is only meaningful for the extent of the verification dataset and not for arbitrarily large values.

To test the suprathreshold performance of ΔIC_TC_P, we compared its predictions with experimental results from Witt [34] which were measured for reflective patches in a controlled environment with known adapting luminance. The data represents subjective magnitude estimation of color differences between painted color patches. The data is represented in CIE 1964 XYZ 10° observer space which we transformed to 1931 2° observer by applying a diagonal matrix which is the ratio between the D_{65} white point in both spaces. This is only an approximation of the correct conversion, but a more accurate matrix would require the knowledge of the paints used by Witt in the experiment.

As mentioned earlier, for the data to be useful above the detection threshold, it needs to be linear with perceivable differences. To model whether this is the case, we performed linear regression for each of the color directions to fit a straight line between the experimental results and predictions of ΔIC_TC_P for the same colors. We omitted one of the patches in the dataset, number 8 in the 'Yellow' set, because we suspect that the luminance of that yellow patch was misreported (as it is half that of the other patches), although we have not been able to confirm this. All the remaining patches were used in the comparison, which is shown in Figure 5. To help us establish if there exists a certain point at which ΔIC_TC_P is no longer linear with subjective assessments, the results are presented as r^2 and RMSE of the linear regression of ΔIC_TC_P to the experimental results. A perfect color difference metric would have an r^2 very close to 1 and a low RMSE, up to the limits of the noise in the experimental data. If a model stopped being linear with the experimental data above a certain error level, its r^2 would drop and RMSE would increase.

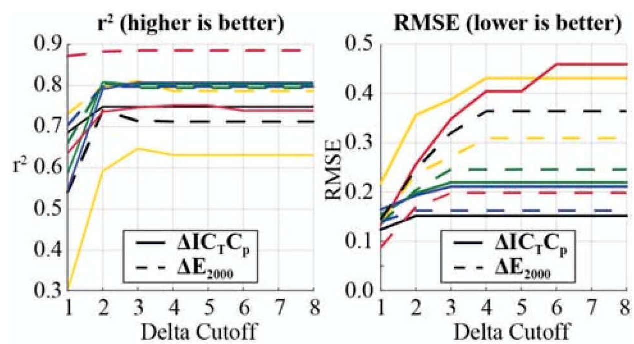

Figure 5: Linearity above threshold for both ΔE_{2000} and ΔIC_TC_P. The line colors represent the color quadrants and achromatic colors of the dataset [34].

We plotted the r^2 and RMSE values for both ΔE_{2000} and ΔIC_TC_P as a function of a maximum error value considered. At delta cutoff 8, we present errors across all the samples because ΔIC_TC_P error never exceeded that value. At error cutoff 7, we ignored all samples whose ΔIC_TC_P is above 7 and so on, down to a cutoff of 1, indicating only samples at or below detection threshold. A sudden dip in the r^2 and an increase in RMSE above a certain value would indicate that the model is no longer linear with the experimental data and the errors above this value are no longer perceptually meaningful. We performed the same calculations for ΔE_{2000} to provide a comparison. As can be seen, both ΔIC_TC_P and ΔE_{2000} remain relevant across the whole range tested in the experiment. While this is unsurprising for ΔE_{2000}, which was fitted to this dataset and is known to be useful for suprathreshold color errors, ΔIC_TC_P was never intended for such large differences. Nevertheless, it still performs well.

The linear regression resulted in different slopes of the line depending on the color tested for both color difference metrics. This means that even though it is possible to compare the errors along a single hue direction, comparing the error between different hues or between the chromatic and grey patches will not necessarily produce a meaningful result.

Conclusions

Steady State metrics have many benefits for predicting if two stimuli on reflective targets, textiles, as well as emissive SDR displays, can be distinguished under the same, known viewing environment.

However, in order to predict the rendering requirements for HDR/WCG display systems where the state of adaptation of an observer is not known, we have introduced the concept of an "Adaptation Hull" metric class using ΔIC_TC_P as an example. This new approach increases efficiency, shows low errors predicting existing datasets and provides linearity above thresholds while at the same time not requiring the prior knowledge of the state of adaptation.

Acknowledgement

We would like to thank Jack Holm for his useful comments and all of the participants in our experiment.

References

[1] H. Davson, Physiology of the Eye, 5th ed. Pergamon Press, 1990.

[2] T. Kunkel and E. Reinhard, "A reassessment of the simultaneous dynamic range of the human visual system," Proc. 7th Symp. on Applied Perception in Graphics and Visualization, pp. 17-24, 2010.

[3] S. Daly, T. Kunkel, X. Sun, S. Farrell, and P. Crum, "Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance," Proceedings vol. 8651, Human Vision and Electronic Imaging XVIII, 2013.

[4] S. Daly, T. Kunkel, X. Sun, S. Farrell, and P Crum, "41.1: Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays," SID Display Week, pp. 563-566, 2013.

[5] S. Daly, E. Gitterman, and G. Mulliken, "Pupillometry of HDR viewing," IS&T HVEI Electronic Imaging, 2018.

[6] CIE Publication DS 014-4.3/E:2007: Colorimetry - Part 4: CIE 1976 L*a*b* Colour Space (Report). Draft Standard. 2007.

[7] CIE Publication 119:1995, Industrial Colour Difference Evaluation. CIE, Vienna, Austria,1995.

[8] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl., vol. 26, no. 4, pp. 340–350, 2001.

[9] M. D. Fairchild, Color Appearance Models, Wiley-IS&T Series in Imaging Science and Technology, 2nd edition, 2005.

[10] M. R. Luo, G. Cui and C. J. Li, "Uniform Color Space based on CIECAM02," Color Res. & Appl., vol. 31, no. 4, pp. 320-330, 2006.

[11] J. Kuang, G. M. Johnson and M. D. Fairchild, "iCAM06: A refined image appearance model for HDR image rendering," Journal of Visual Communication and Image Representation, vol. 18, no. 5, pp. 406-414, 2007.

[12] F. Ebner and M. Fairchild, "Development and testing of a colour space (IPT) with improved hue uniformity," Proceedings of the 6th Color Imaging Conference: Color Science, Systems, and Applications, pp. 8–13, 1998.

[13] F. J. J. Clarke, R. McDonald, B. Rigg, "Modification to the JPC79 Colour–difference Formula," Journal of the Society of Dyers and Colourists, vol. 100, pp. 128-132, 1984.

[14] W. D. Bidgood Jr. and S. C. Horii, "Introduction to the ACR-NEMA DICOM standard," Radiographics, vol. 12, no. 2, pp. 345-355, 1992.

[15] P. G. J. Barten, "Contrast Sensitivity of the Human Eye and its Effects on Image Quality," SPIE Optical Engineering Press, 1996.

[16] S. Miller, M. Nezamabadi and S. Daly, "Perceptual Signal Coding for More Efficient Usage of Bit Codes," SMPTE Motion Im. J., vol. 122, no. 4, 2013.

[17] SMPTE Standard ST 2084:2014 – High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays.

[18] T. Kunkel, G. Ward, B. Lee and S. Daly, "HDR and Wide Gamut Appearance-based Color Encoding and its Quantification," 30th IEEE Picture Coding Symposium, San Jose, CA, 2013, pp. 357-360.

[19] Dolby Labs, Inc.: What is ICTCP – Introduction. White paper, version 7.1, United States, 2016

[20] D. L. MacAdam "Visual sensitivities to color differences in daylight," J. Opt. Soc. Am. A, vol. 32, no.5, pp. 247-273, 1942.

[21] M. Safdar, G. Cui, Y. J. Kim and M. R. Luo, "Perceptually uniform color space for image signals including high dynamic range and wide gamut," Opt. Express 25, pp. 15131-15151, 2017.

[22] G. Spencer, P. Shirley, K. Zimmerman and D. P. Greenberg, "Physically-based glare effects for digital images," in SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, ACM, pp. 325–334, 1995.

[23] J. McCann and A. Rizzi, "Retinal HDR images: Intraocular glare and object size," 16th Color and Imaging Conference Final Program and Proceedings, pp. 16-21, 2008.

[24] D. Maes, T. Vyvey, J. Looy, B. Vandevelde and G. Stojmenovik, "The Perceived Intra-Frame Dynamic Range in a Cinema Environment," PQS 5th ISCA/DEGA Workshop on Perceptual Quality of Systems, pp. 1-18, 2016.

[25] M. Murdoch and I. Heynderinckx, "Veiling glare and perceived black in high dynamic range displays," J. Opt. Soc. Am. A, vol. 29, pp. 559-566, 2012.

[26] T. D. Lamb, D. Trevor and E. N. Pugh, "Phototransduction, dark adaptation, and rhodopsin regeneration the proctor lecture," Investigative Ophthalmology & Visual Science, vol. 47, no. 12, pp. 5138-5152, 2006.

[27] M. G. M. Stokkermans and I. E. J. Heynderickx, "Temporal dark adaptation to spatially complex backgrounds: effect of an additional light source," J. Opt. Soc. Am. A, vol. 31, pp. 1485-1494, 2014.

[28] J. M. Valeton and D. van Norren, "Light adaptation of primate cones: an analysis based on extracellular data," Vision Research, vol. 23, no. 12, pp. 1539-1547, 1983.

[29] E. Pieri and J. Pytlarz, "Hitting the Mark - A New Color Difference Metric for HDR and WCG Imagery," SMPTE Motion Imaging Journal, vol. 127, no. 3, pp. 18-25, 2018.

[30] P. A. García, R. Huertas, M. Melgosa and G. Cui, "Measurement of the relationship between perceived and computed color differences," Optical Society America, 24(7):1823-29, July 2007.

[31] Rep. ITU-R BT.2408-1. "Operational practices in HDR television production," p. 3, April 2018.

[32] Commission Internationale de l'Éclairage, "Recommended system for mesopic photometry based on visual performance." CIE Publication 191, 2010.

[33] A. B. Watson and D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Perception & psychophysics*, *33*(2), 113-120, 1983.

[34] K. Witt, "Geometric relations between scales of small colour differences," Color Res. & Appl., vol. 24, no. 2, pp. 78-92, 1999.

Author Biography

Timo Kunkel is a color and imaging researcher at Dolby Labs, Inc. His main areas of research are human perception-based color models, advanced display technologies and psychophysics. His current research efforts include the development of concepts for HDR and WCG display assessment and metrology. He received his PhD in Computer Science from the University of Bristol, UK and a MSc in Physical Geosciences from the University of Freiburg in Germany.

Robert Wanat graduated West Pomeranian University of Technology, Szczecin in 2011 with an MSc in Computer Science and Bangor University in 2015 with a PhD in Computer Graphics. He works as a Senior Color Scientist for Dolby Labs in Sunnyvale, CA. His research interests include visual perception modelling, low light color reproduction and effect of viewing conditions on image appearance.

Jaclyn Pytlarz is a Senior Engineer at Dolby Labs in Sunnyvale, CA where she has worked inside Dolby's Advanced Technology group since 2014. Her research includes vision science surrounding technologies for HDR and WCG displays. She holds a B.S. degree in Motion Picture Science from Rochester Institute of Technology and is currently pursuing her M.S. in Computational and Mathematical Engineering at Stanford University.

Elizabeth Pieri is an Engineer on the Imaging Applied Research Team at Dolby Labs, based in Sunnyvale, CA. Her research includes vision science as it relates to HDR and WCG display technologies, with a particular focus on color perception and display mapping. She received a B.S. degree in Motion Picture Science from Rochester Institute of Technology in 2016.

Robin Atkins' career in imaging science began while designing High Dynamic Range displays at Brightside Technologies. These displays revealed a fascinating host of new challenges in perception and color appearance – which he is now working to address as the manager of the Applied Vision Science Group at Dolby Labs. Robin's current focus is on designing algorithms for mapping high dynamic range and wide color gamut content to the multitude of endpoint displays that we use every day. He has degrees in Engineering Physics and Electrical Engineering.

Scott Daly is an applied vision scientist at Dolby Laboratories, with specialties in spatial, temporal, and chromatic vision. He has a BSEE from NCSU and an MS in bioengineering from the University of Utah. Past accomplishments include a team technical Emmy in 1990, and the Otto Schade award from SID in 2011. He is a member of IEEE, SID, and SMPTE.