

Perceptually optimal boundaries for wide gamut TVs

Justin Laird^a, Ingrid Heynderickx^{a,b}

Philips Research Europe, High Tech Campus 34, 5656AE, Eindhoven, The Netherlands;
Delft University of Technology, Mekelweg 4, 2628CD, Delft, The Netherlands

ABSTRACT

Displays are coming on the market that can produce a larger range of colors over EBU and this has led to much research on the topic of how to use the additional color gamut volume provided by the displays. Some research has focused on different methods to map colors from small to large gamuts, whereas this paper focuses on defining the required gamut boundaries for natural content. Two gamuts were created from the results of a psychophysical experiment that asked observers to choose their preferred image in terms of saturation. One gamut corresponded to the median value of their choices and the second gamut corresponded to the 90% value of their choices. The results indicated that even at the 90% value, the resulting gamut was smaller than that of the wide-gamut display for most hues and actually closer to EBU for some hues. These results are display independent, at least when considering modern displays that can reach luminance values above 250cd/m².

Keywords: Color appearance, color management, psychophysical experiment, wide color gamut display, gamut mapping, LCD TV

1. INTRODUCTION

The area of backlighting for displays has seen tremendous growth and development in recent years. Researchers in this area have used a wide array of technologies to make displays brighter^{1,2}, increase their range of producible colors³⁻⁵, as well as making them more efficient^{3,6}. While not a comprehensive list by any means, these highlight a few of the more popular aspects of new developments in display technology. Out of these three, the increase in brightness and producible colors has garnered a lot of attention and they both seemingly promise great things for the next generation of displays⁷. Indeed, there are indications that as displays become brighter they are better able to mimic scenes from the real world, given that the contrast ratio also increases⁸. The increase in the amount of colors, on the other hand, seems to come more as a technological push than a perceptual need. As such, there has been research carried out to determine the possibilities that exist, from a perceptual standpoint, for these displays over traditional displays⁹.

The range of colors producible by a device is referred to as a gamut and in this paper a display that can create more colors than the EBU¹⁰ (or Rec. 709 standard) is referred to as “wide color gamut” or simply “wide-gamut” display. One important aspect for wide-gamut displays is the fact that transmission and encoding practices still use traditional standards. Although there are new standards such as xvYCC¹¹, it is most common to encode and transmit image content according to the EBU standard and therefore most video that arrives at the television does not contain colors outside the EBU gamut. The question then is how to map colors from a smaller gamut into a larger one or in other words how to use the extra space made available at the display. Up to now manufacturers have relied on color mapping/enhancement algorithms to handle this issue and as such there has been some research carried out that dealt with new mapping methodologies for mapping between small and large color gamuts^{9,12-14}.

Previous studies into wide-gamut color mapping have revealed strengths and weaknesses of current mapping strategies from a perceptual standpoint. For example, one of the main drawbacks is the tendency for these algorithms to use the entire gamut that is available. For wide-gamut TVs coupled with higher luminance levels (above 250 cd/m²) it has led to very intense, bright colors that observers find displeasing. There are different approaches to this problem; however, three obvious possibilities would be to 1) decrease the overall display luminance, 2) tune color mapping algorithms for optimizing colors, or 3) redefine a gamut within the display gamut. Since it has been shown that increasing the brightness of displays while taking into account other factors like contrast ratio is beneficial for image quality⁸, decreasing the overall luminance level of displays would not be the best choice. Imposing stricter guidelines for tuning the algorithms in terms of how and where to map the colors is a viable option and has received a lot of attention though making a robust algorithm is very time consuming and even then may not give good results across all image content or

different wide gamut displays. The third approach is to define, independent of any algorithm, areas of color space that should be avoided in wide-gamut displays. This is a fundamental issue and from a perceptual standpoint is more interesting since in theory the results can be applied to any new, bright, wide-gamut television. This also has the added benefit of decreasing tuning time imposed on color mapping/enhancement algorithms and would allow optimal use of brighter displays (i.e. more freedom in enhancing colors without worrying about fluorescent colors).

Therefore, an experiment was conducted in order to define a perceptually optimal gamut for natural images. Boundaries were determined for specific locations within color space using natural scenes. These boundaries essentially create a “virtual” gamut in which the color mapping/enhancement algorithms operate. Since the majority of images shown on televisions are natural content, that was considered an important basis for creating the virtual gamut.

2. EXPERIMENTAL

2.1 Setup

The goal of this experiment was to obtain observers’ preference of saturation as it related to natural material. The results were then used to define absolute boundaries within color space that should not be crossed by a color mapping/enhancement algorithm. The images and results were defined in the CIELab¹⁵ color space as this is the most widely accepted perceptually uniform color space.

The experiment was performed on a well-characterized, wide color gamut 42” LC-TV. The resolution for this display was 1366x768 and it had a peak brightness of 300 cd/m². This display was previously characterized using known techniques to build a characterization model¹⁶ with results seen in Table 1. To test the model, 2060 RGB values were sent through the forward model to predict corresponding XYZ values. These same RGB values were shown and measured on the display, with the measured and predicted values compared in CIELab. A more accurate color difference formula, known as delta E 2000, (CIEDE2000 or simply ΔE_{00}), was used. This formula corrects for inaccuracies within CIELab and provides a more realistic value of perceived color difference¹⁵. As seen in Table 1, on average the model can predict colors within a ΔE_{00} of 1 (which is approximately 1 JND).

Table 1 - Results of characterization model for SQ using 2060 random RGB colors.

| | Average | Maximum | Minimum | Standard Deviation |
|-----------------|---------|---------|---------|--------------------|
| ΔE_{00} | 0.88 | 2.77 | 0.01 | 0.00 |

It is believed that the majority of content viewed on a TV is natural material; therefore two scenes containing natural material were chosen. However, the required gamut boundary should not be limited to sensitive memory colors, such as the blue of sky or the green of grass. Therefore, scenes with a limited amount of context, that was expected to be relatively insensitive to over-saturation, were selected. This resulted in scenes, shown in Figure 1, that were contextually simple, not related to memory colors, and nearly monochromatic. Although the images were simple there was enough context to be meaningful to observers. In other words, observers realized that the left image in Figure 1 was part of a car door and the right image was the outside wall of a house. The size of the images was approximately 800x600 pixels and thus smaller than the display resolution. The images were displayed at their native resolution surrounded by a grey background.

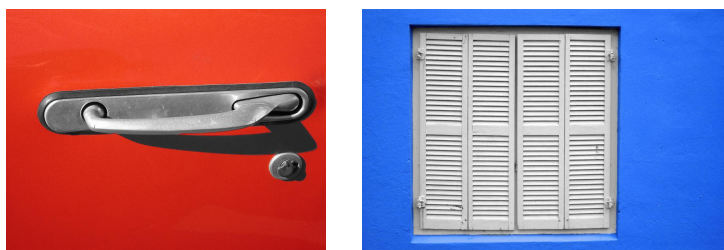


Figure 1 - Two scenes used in experiment that were nearly monochromatic (i.e. small variation in L^* , C^* and hue). The chromatic parts in the scenes were altered in hue and lightness.

The chromatic parts in the two scenes of Figure 1 were altered in hue, chroma, and lightness in order to sample different areas of color space. For the experiment, the observers adjusted chroma for a given hue and lightness until the image became unnatural. So, in practice chroma was gradually varied for 6 hues and 3 lightness levels per hue tested for each scene. With one exception that is discussed below, this resulted in 17 conditions per scene and a total of 34 conditions that each observer had to judge. The 6 hues were red, green, blue, cyan, magenta and yellow. The lightness levels varied according to hue but always in the order of brightest to darkest. The only exception was yellow, for which there were only 2 lightness levels.

2.2 Mapping images

To generate the images with various chroma levels the following procedure was used. Using the measurement data from the characterized display, the images were transformed to XYZ and then CIELab using standardized formulas¹⁷. Once in CIELab, the coordinates $L^*a^*b^*$ were converted to LCh, which corresponded to linear values Lightness (L^*) and chroma (C^*) and the polar value hue (h), all seen in Figure 2. The next step was to select all the pixels in the image, of which the hue had to be rotated. In particular, these were the pixels that had a chroma value above 5. For this experiment the hues, shown in Figure 3, were chosen to correspond to the red, green and blue primaries of the display and their corresponding secondary colors cyan, magenta and yellow.

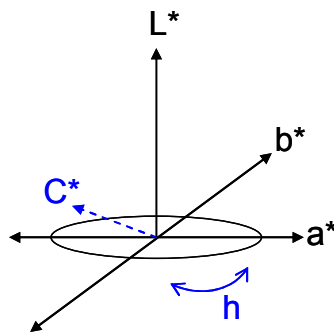


Figure 2 - The coordinates of CIELab: $L^*a^*b^*$ and the corresponding C^* (chroma) and polar coordinate, h (hue).

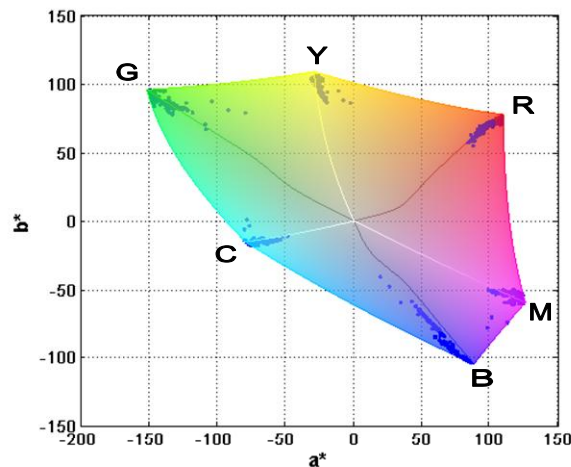


Figure 3 - Hues chosen for this experiment correspond to the primaries and secondaries of the characterized display. Each hue is shown in this figure in bold letters. The points at each hue are pixels groupings explained further in the paper.

After the chromatic pixels were shifted to the hue of the primaries, they were adjusted in lightness to match the lightness of the primary. Then a lightness value above and below the target was selected so that there were a total of three lightness values at each hue. Note that in each case the average value of lightness and hue of the chromatic areas in the scene were set. Since these were real images there were small variations in all three dimensions of color.

After the hue and lightness were set, the chromatic pixels were shifted to the neutral axis and from there vectors that emanated from the neutral axis to the gamut boundary of the display were made. These vectors corresponded to 46 equal steps in chroma while maintaining lightness and hue from the neutral axis to the gamut boundary. These were the values that the observers used in the experiment. Note that in this setup increasing chroma is analogous to increasing saturation since the lightness is being held constant.

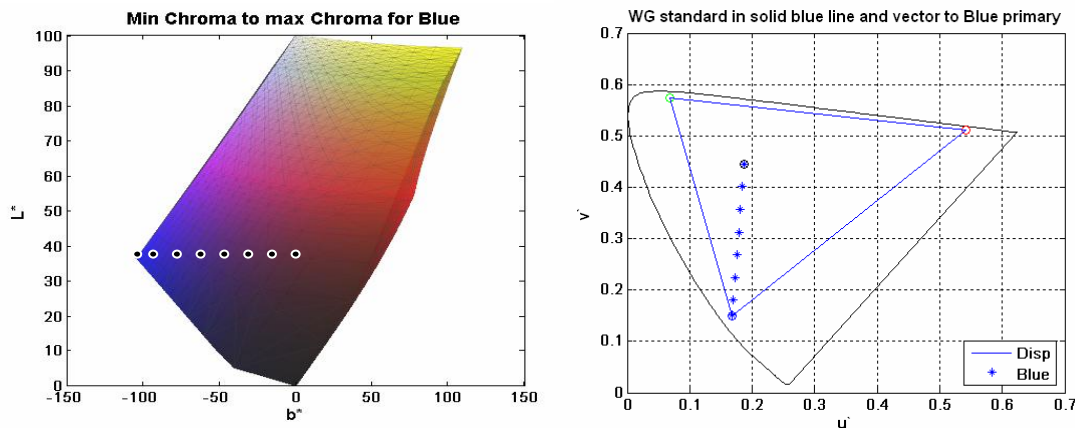


Figure 4 - Example mapping shown of color point in two visualizations of a gamut. Note that lightness and hue are maintained, while only chroma (saturation) is increased.

In Figure 4 an example mapping is shown within two gamut visualizations, a three dimensional volume on the left and a chromaticity diagram on the right. In each case the point furthest inside the gamut is a neutral color and then as the point is moved outwards, its chroma increases until it reaches the gamut boundary. In the chromaticity diagram, on the right, this is seen as increasing saturation from the whitepoint (circled) to the blue primary. This is the vector that observers traversed when they were making their choices. So there were essentially 34 of these vectors made, (17 per scene), where the end positions of the vectors in color space varied in hue and lightness. For natural images, instead of there being a single point at each position in the vector there were clusters of pixels where the average pixel in the cluster corresponded to each point in the vector. Figure 3 shows an example for one scene. The points correspond to the cluster of pixels at each hue for a single lightness level. The clusters are at the end of the vectors, corresponding to the gamut boundary.

2.3 Observers and experimental procedure

There were 32 color-normal observers in total, 62% were male and 38% female and the median age was 30 with a minimum and maximum age of 22 and 55 respectively.

Observers sat 3 meters from the display in a darkened room with dim lighting uniformly lit behind the TV. There was a single image viewed on the display with a background that corresponded to 20% of the luminance of the whitepoint. There were a total of 34 conditions (vectors) and for each condition there were 46 possible choices (corresponding to the 46 chroma positions along the vectors). Each observer randomly started somewhere along the vector and then manually progressed (forward or backward) through the vector. The task of each observer was to select for each condition their preferred saturation level. If they expressed uncertainty in making their choices they were instructed to progress forward through the vector until the image became too saturated and then to step backwards until they were satisfied. These instructions were given since for many observers there was a range of answers and not a single “preferred” image. So in effect, observers often made choices based on their upper level of preference. In the end, the results were a number between 1 and 46 for each of the 34 conditions, where 1 corresponded to the neutral axis and 46 to the gamut boundary.

3. RESULTS/DISCUSSION

3.1 Preferred saturation

The first step in the analysis was to perform an analysis of variance (or ANOVA) on the entire dataset with the factors lightness, hue, scene and observer. The “Prob>F” seen in Figure 5 is also referred to as the p-value, which signifies the probability that the null hypothesis is true, (i.e. whether the mean of all levels of a factor are equal). A small p-value indicates that the probability is high that the null hypothesis is not true, (i.e. that the mean of the levels within a factor are different). As seen in Figure 5, there was a strong effect of all the factors on the preferred saturation as all the p-values were <0.001.

| Analysis of Variance | | | | | |
|----------------------|-----------|------|----------|-------|--------|
| Source | Sum Sq. | d.f. | Mean Sq. | F | Prob>F |
| Lightness | 1321.56 | 2 | 660.78 | 15.44 | <0.001 |
| Hue | 10555.55 | 5 | 2111.11 | 49.34 | <0.001 |
| Scene | 3942.13 | 1 | 3942.13 | 92.13 | <0.001 |
| Observer | 60198.68 | 31 | 1941.89 | 45.38 | <0.001 |
| Error | 44842.78 | 1048 | 42.79 | | |
| Total | 120023.71 | 1087 | | | |

Figure 5 - Results of ANOVA for all results compiled together.

The effect of hue on preferred saturation indicates that the wider color gamut is not appreciated to the same extent in all hue directions. In practice, the hues blue and yellow corresponded to higher saturation levels. Since the actual lightness levels used in the experiment were different for different hues, it is logical that observers' had a different optimal saturation level at these different lightness levels. In general, observers preferred higher saturation values for lower lightness levels. As the preferred saturation level depended on the scene, the analysis was repeated for both scenes separately. Figure 6 shows the mean results of the two scenes when averaged over hue, lightness and observers. On the y-axis are the two scenes and on the x-axis are the choices (with the range cropped between 25-31 instead of the full range 1-46). Figure 6 clearly shows that observers allowed higher saturation for scene 1 (i.e the car door), with a mean choice of around 30, than for scene 2 (i.e. the wall) that had a mean choice around 26.

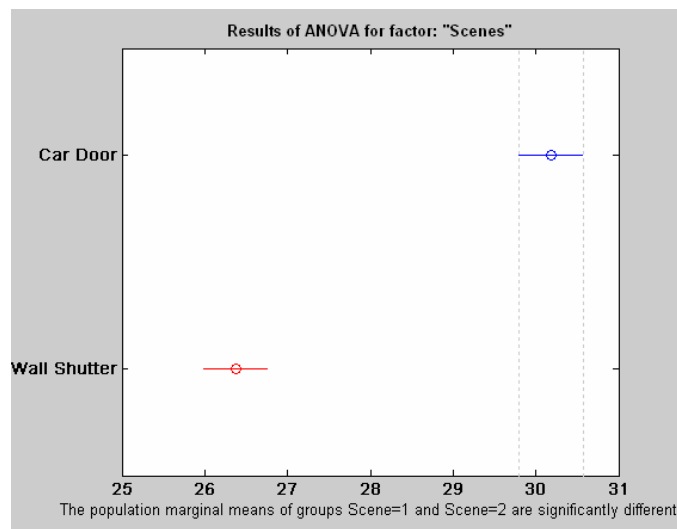


Figure 6 - Results graphically shown for the two scenes. Scene 1 is the car door and scene 2 is the wall with shutters. The X-axis is the mean value of observers' choices (1-46) when averaged over hue, lightness and all observers. The error bars are the 95% confidence interval about the mean.

It was hypothesized, prior to the experiment, that if there was an effect of scene, observers would allow higher saturation for the *wall* scene (scene 2) than for the *cardoor* scene (scene 1). The reasoning was that the *cardoor* scene contained less texture and therefore less variation in lightness. This resulted in a more condensed grouping of pixels and when

those pixels were moved towards the gamut boundary they quickly became very intense, and therefore, were expected to be displeasing to viewers. However, as shown in Figure 6, observers preferred higher saturation values for the *cardoor* than for the *wall*. Comments from observers after the experiment indicated that it probably was a cognitive effect where apparently observers had a more preconceived idea of what the wall of a house should look like than what a car door should look like or the “requirement” for a car door was less strict in the observers’ mind.

Because the results across both scenes cannot be grouped, the remaining analysis in this paper will focus on a single scene. Since the purpose of this experiment is to determine the optimal perceptual *boundaries*, the scene that has higher saturation in the result should be used as it represents the outer most boundary. From Figure 6 it is evident that this is the *cardoor* scene, and therefore the remaining analyses will use only this scene.

Since the ANOVA also indicated a significant effect of observer on the results, it was checked if any observers were outliers. The Tukey post-hoc test indicated one observer that fell into a group by herself and was significantly lower in her saturation preference than all other observers. Box plots also showed that this same observer was an outlier in multiple conditions. Given this outcome, that observer was eliminated from the dataset.

3.2 Establish boundaries

The resulting dataset now consists of 31 observations for the 17 conditions (6 hues, 3 lightness levels, except for yellow – which had 2 lightness levels) for the *cardoor* scene (scene 1). For each condition, the boundary was defined based on two selection criteria. The first criterion was the median (50%) of the histogram of all choices per condition. This represents approximately the average preferred saturation level of observers. The second criterion is the 90% value of the histogram of observers’ choices per condition. This point is interesting since it encompasses 90% of observers’ choices and thus represents the extreme boundary that a display manufacturer might want to use. If the 90% choice is used for determining gamut boundaries, however, more reliance on proper gamut mapping would be needed to prevent over saturation of colors/images for a considerable amount of viewers.

Figure 7 shows the histogram of observers’ choices for the red hue at the highest lightness level as an example. The values corresponding to the 50% and 90% criterion are added as an illustration. It shows that the majority of observers chose a value between 22 and 36, which is far below the maximum of 46 that corresponds to the gamut boundary of the wide-gamut display used in this experiment. This means that all observers prefer a red color that is less saturated than the current wide-gamut display primary when tested at an equivalent brightness level. Similar plots, though not shown, are made for each condition.

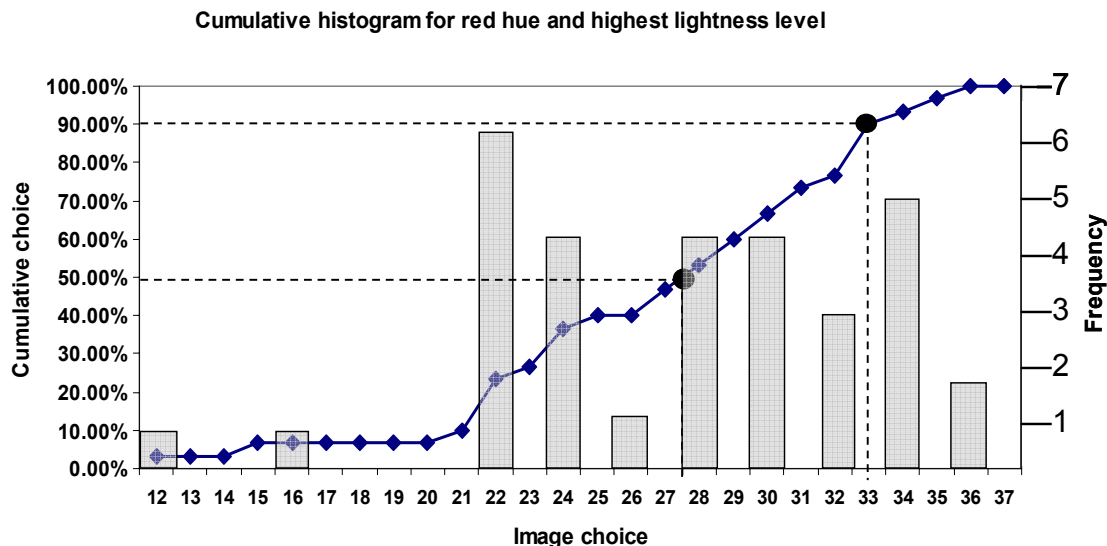


Figure 7 - Example histogram for red hue at the highest lightness level. The solid line is the cumulative histogram in terms of percentage of observers included. The 2 large circles are the 50% and 90% choice respectively.

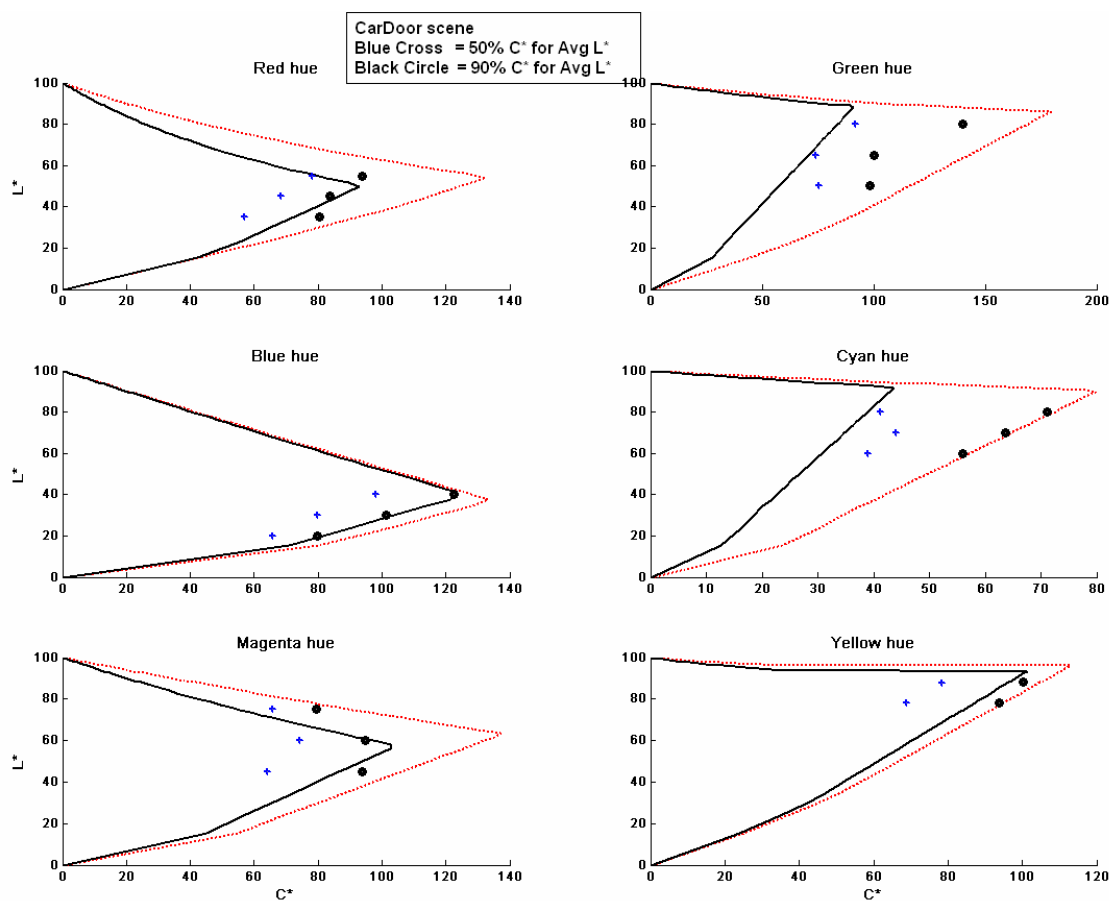


Figure 8 - Results of gamut boundary. The blue dots represent the gamut boundary when using the median observer choice and the black dots represent the boundary when using the 90% value for observers' choices. The black line is the EBU gamut and the red line represents the wide-gamut display.

Shown in Figure 8 are the results for each hue and lightness level for scene 1 using all 31 observers. There are six subplots corresponding to the six hues and within each subplot there are 6 markers (3 crosses and 3 dots). Yellow is the exception where there are 4 markers (2 crosses and 2 dots). The crosses represent the 50% value of observers' choice and the dots represent the 90% value. The solid line in each subplot represents the EBU gamut boundary and the dotted line represents the wide-gamut display boundary. CIELab L^* is represented on the y-axis and C^* (chroma) on the x-axis.

One of the first things to notice from Figure 8 is that for hues red, blue and magenta even 90% of observers are within or close to the EBU gamut. Only in cyan and yellow is the wide-gamut display boundary reached. In the yellow region it is seen that the majority of observers choose a boundary that is within the EBU gamut, where only the upper ~15% of observers choose a boundary outside of EBU. In the green region of color space, a wide-gamut display has some added value, but it is noted that 90% of the observers prefer a boundary that is halfway between EBU and the wide gamut. Given these results it would seem that the EBU primaries of red and blue are already saturated enough at the luminance levels of current LC-TVs, and only the green primary should be extended. If a multi-primary display is being created, then these results show that it would be perceptually more beneficial to extend in the cyan or yellow region rather than in magenta. From an engineering standpoint it should be noted that for a three-primary device a more saturated green primary than what is shown in Figure 8 may be needed in order to reach the perceptually optimal gamut in the cyan and yellow region.

Again, the idea for the experiment is to define a new gamut where more traditional mapping and/or enhancement algorithms can operate safely without resulting in oversaturated colors/images. If manufacturers used the 50% gamut

then the underlying assumption is that observers would prefer a display with less saturated primaries but higher luminance. This would be called a “tall-gamut” rather than a “wide-gamut” and would allow more generalized mapping techniques. However, if the 90% gamut is used then the resulting display gamut would approach that of current “wide-gamut” TVs and consequently, greater care will be needed to prevent intense oversaturated colors.

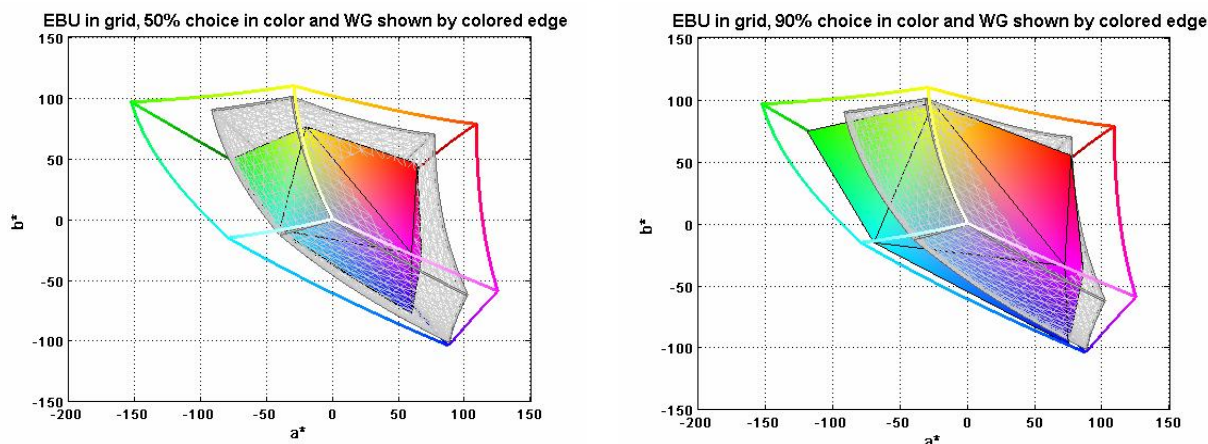


Figure 9 - Gamuts shown in CIELab color space. The outer colored line represents the wide-gamut display. The black/white mesh represents the EBU gamut and the inner colored gamut is the new virtual gamut. Left shows the gamut when using 50% value of observers' choice. Right shows the gamut when using the 90% choice.

Figure 9 shows how the virtual gamut might look when using either the 50% or the 90% value of observers' choice. These are the 3D volumes of the gamut within CIELab, though Figure 9 shows only the a^* -vs.- b^* plot (looking down the L^* axis). This figure reiterates that for an LC-TV having high luminance values the preferred yellow, red, magenta and blue primaries remain close to the EBU primaries and green and cyan in fact are expanded in saturation.

These boundaries found here exist only for natural images and it is not known if it can be extrapolated to all image content such as computer generated colors/images; for example the boundary in the green region of color space is not relevant for scenes including traffic lights but only to scenes containing natural green colors. It is believed that this is okay since observers do not have an internal requirement for how computer-generated content should look. Further verification would be needed to determine if these boundaries hold for other more complex natural and computer generated material.

It should be noted that luminance plays an important role in this outcome. The results shown in this paper are valid for current state-of-the-art televisions that usually have luminance values of 250 cd/m^2 or greater. For TVs with lower luminance values verification of these results is needed. This is analogous to the real world, where colors do not typically reach very high saturation levels but can reach high luminance levels. Table 2 gives the boundary point having the highest luminance per hue (each hue has 3 lightness levels and only the highest per hue is shown). Since lightness is held constant as chroma increases, the luminance is the same for both the 50% and 90% choices. The importance of Table 2 is that the saturation of the chosen boundaries is valid for displays having comparable brightness levels – for displays which are dimmer the results can be drastically different.

Table 2 - Luminance (in cd/m^2) of hues from the *cardoor* scene tested in this experiment.

| | Luminance |
|---|-----------|
| R | 68 |
| G | 168 |
| B | 33 |
| C | 168 |
| M | 143 |
| Y | 213 |

4. CONCLUSION

In this experiment we have tested observers' preferred saturation for natural images shown on a wide-gamut display and converted these results into gamut boundary points. The results show that the majority of observers prefer a gamut that is in size closer to EBU for some hues than a typical wide-gamut display. These results seem to imply that the perceptually optimal display should have red and blue primaries closer to that of EBU but with a more saturated green primary. Another option of course would be to have a multi-primary display with red, blue and magenta primaries close to that of EBU but more saturated green, cyan and possibly yellow primaries.

However, this conclusion does not mean that wide-gamut displays are not desirable and should not be pursued and developed further. Wide-gamut displays offer many opportunities to color mapping/enhancement designers over traditional displays. If a wide-gamut display were developed and the virtual gamut implemented into that particular wide-gamut display then there could be special cases where mapping only parts of an image into the wide-gamut region of a display would be nice (such as neon lights, fireworks, etc...). Another point is that in order to reach the boundary points in cyan and green region found in this research, the primaries would need to be more saturated than what was shown and certainly more saturated than EBU.

The results of this experiment are valid even when considering wide gamut transmission standards, such as xvYCC, because the maximum saturation of the natural image transmitted would probably not be as saturated as what can be obtained with LEDs or laser backlights.

The boundaries determined in this experiment are relevant for displays having a peak luminance value similar to that of current LC-TVs. It is not known whether similar results would be obtained for displays having lower luminance values.

These results, though interesting, are not yet valid for all content as only two simple, natural scenes were tested. The results cannot be extrapolated to all images and the next step is to validate these results using more complex, natural scenes. Furthermore, there is no reason to believe that the choices observers made would extend to computer generated scenes. This is indeed an interesting issue to explore since more and more content is computer generated.

These results do however show that designers of wide-gamut displays should carefully consider how color mapping algorithms behave and that it is possible that creating extremely large color gamuts are not the best option for natural content.

REFERENCES

1. Hiroaki S. Wide-color-gamut monitors: LED-backlighting LCD and new phosphor CRT. Proc. SPIE 2004. p 151-160.
2. Roth S, Ben-David I, Ben-Chorin M, Eliav D, Ben-David O. Wide Gamut, High Brightness Multiple Primaries Single Panel Projection Displays. SID 2003. p 118-121.
3. Erno H. A. Langendijk OB, Frank Budzelaar, Frank Vossen. Dynamic Wide-Color-Gamut RGBW Display. SID 2007. p 1458-1461.
4. Brennessholtz MS, McClain SC, Roth S, Malka D. A Single Panel LCOS Engine with a Rotating Drum and a Wide Color Gamut. SID 2005. p 1814-1817.
5. Roth S, Caldwell W. Four Primary Color Projection Display SID 2005. p 1818-1821.
6. de Greef P, Hulze HG. Adaptive Dimming and Boosting Backlight for LCD-TV Systems. SID 2007. p 1332-1335.
7. de Vaan AJSM. Competing Display Technologies for the best Image Performance. Color Imaging Conference 2006. p 274-279.
8. Seetzen H, Heidrich W, Stuerzlinger W, Ward G, Whitehead L, Trentacoste M, Ghosh A, Vorozcovs A. High Dynamic Range Display Systems. ACM Transactions on Graphics 2004;23(3):760-768.
9. Muijs R, Laird J, Kuang J, Swinkels S. Subjective Evaluation of Gamut Extension Methods for Wide-Gamut Displays IDW 2006.
10. Parameter values for the HDTV standards for production and international programme exchange ITU-R Recommendation BT.709-5:2002. (2002)
11. Multimedia systems and equipment - Colour measurement and management - Part 2-4: Colour management - Extended-gamut YCC colour space for video applications - xvYCC. IEC 61966-2-4. (2006)

12. Kang B, Morovic J, Luo MR, Cho M. Gamut Compression and Extension Algorithms Based on Observer Experimental Data. ETRI Journal 2003;Volume 25(Issue 3):156-170.
13. Kotera H, Mita T, Hung-Shing C, Saito R. Image-dependent Gamut Compression and Extension. 2001. p 288-292.
14. Ok HwL, Seong D.; Choe, Won H.; Park, Du, S.; Kim, Chang Y.;. Color processing for multi-primary display devices. IEEE International Conference on Image Processing 2005. p 980-983.
15. Berns R. Billmeyer and Saltzman's Principles of Color Technology. New York: John Wiley & Sons; 2000.
16. Day EA, Taplin L, Berns R. Colorimetric characterization of a computer-controlled liquid crystal display. Color Research & Application 2004;Volume 29(Issue 5):365-373.
17. Wyszecki G, Stiles WS. Color Science: Concepts and Methods, Quantitative Data and Formulae. New York: John Wiley & Sons; 1982.