25.3: Observations of Luminance, Contrast and Amplitude Resolution of Displays

Helge Seetzen[1], Hiroe Li, Linton Ye, Wolfgang Heidrich, Lorne Whitehead
University of British Columbia, Vancouver, BC, Canada
Greg Ward
[1]BrightSide Technologies Inc., Vancouver, BC, Canada

Abstract

Luminance, contrast ratio and amplitude resolution are rapidly growing display specifications. Through a series of human factor studies we have developed simple guidelines for these specifications including viewer preference for luminance, optimal contrast ratio and amplitude resolution under realistic conditions.

1. Introduction

In the past, conventional displays have been largely limited to a dynamic range similar to paper under office lighting conditions – approximately two or three orders of magnitude starting at a grayish black and finishing in the hundreds of cd/m^2. This paradigm of hundreds-to-one contrast ratio, limited luminance and an amplitude resolution in the hundreds of steps is shifting today. Novel display technologies are emerging with the potential of much higher contrast and brightness. Moreover, even existing display technology is being pushed to the limit with a strong increase in display performance. This trend proceeds unevenly with contrast ratio rising faster than luminance, and amplitude resolution remaining largely static. As a result, many display designs make sub-optimal use of the device capabilities.

This paper presents a series of human factor studies that aim to provide a basic framework of luminance, contrast ratio and amplitude resolution and their interaction. The use of a High Dynamic Range (HDR) display [1] as the imaging tool for the study, allows a large enough range for each variable to encompass all current and most near-future display technologies. The results of the study can be used to make design decisions for future displays as well as more realistic comparisons of existing devices.

2. Background

The Video Electronics Standards Association (VESA) and International Organization for Standardization (ISO) provide common guidelines for display specification including luminance and contrast ratio. Peak luminance is generally easy to measure and reported relatively accurately by the industry. Contrast ratio is significantly more challenging. A proper contrast ratio measurement requires a precise black level reading, which is generally low enough that even small variations in ambient illumination or measurement technique can have a major effect. This undesired but real uncertainty in the metric has led to a rapidly escalating competition of specifications in the display industry, where specification and actual results are rapidly diverging (e.g. Plasma displays with a specified contrast of 4,000:1 are often measured at less than 100:1 actual contrast) [See the Oct./Nov. 2003 PC Malta e-zine article by Tom Mainelli entitled "Tested contrast ratios rarely conform to vendors' specs" at www.pcworldmalta.com.]. There is ongoing and very active debate about new specifications either based on revisions of the measurement techniques found in VESA or new metrics.

3. Display Characteristics

Front-view displays are generally specified by two categories of characteristics: mechanical and perceptual. Mechanical characteristics such as size, aspect ratio and spatial resolution are usually easily quantifiable and comparable. While there might be some debate whether for example a large screen size is worth the increased display cost, it is quite clear that bigger is better in the absence of other difference.

Perceptual characteristics on the other hand by definition depend on some aspect of our perceptual apparatus, a mechanism that is in turn not fully understood. For the purpose of this paper we will focus on these characteristics, their relevance in terms of viewer preference and the interplay between different characteristics.

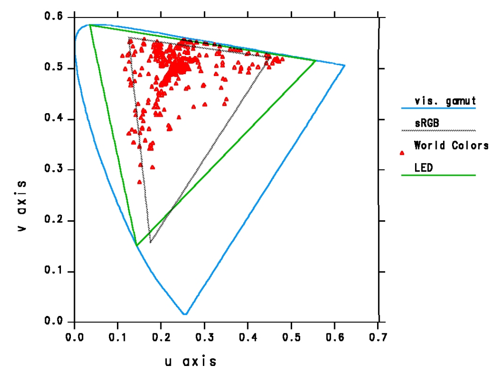

Figure 1: Color space comparison with real world colors. Sample data obtained from spectra of real world materials listed in appendix of Andrew Glassner's "Principles of Digital Image Synthesis," vol 2.

The most common perceptual characteristics of a display are its peak luminance, contrast ratio, amplitude resolution, temporal resolution, and color. For newer display technologies, color is arguably the closest to having reached the requirements of our visual system. Newly emerging wide color gamut displays have broadened the range of presentable colors considerably, but a comparison with real world chromaticity values shows that even the more limited sRGB gamut does a fairly good job of portraying the environment around us [Figure 1]. Whether in existing displays or emerging wide gamut solutions, the color (chromaticity) capabilities of displays will soon encompass our perceptual requirements.

A similar argument can be made for temporal resolution. Liquid Crystal Displays (LCD) have long left the 16ms response time barrier behind, Digital Light Projection (DLP) chips are getting faster and Cathode Ray Tubes were fast enough all along.

ISSN0006-0966X/06/3702-1229-$1.00+.00 © 2006 SID

Figure 2: Partial selection of images used in the first and second study

While there remain issues such as motion blur for LCD and color break-up for DLP, we have seen progress on all these fronts. More importantly, the gaps to be closed (if any) are relatively narrow and certainly not fundamental obstacles for any of the display technologies involved. One might debate whether today's display technologies offer adequate temporal resolution, but there is little doubt that tomorrow's will.

The story is quite different for the remaining three characteristics. The peak luminance, contrast ratio and amplitude resolution of most displays are significantly lower than the capabilities of the viewer. For the purpose of this paper, we will focus on these three characteristics, and explore their impact on viewer preference.

Peak luminance is the easiest characteristic to measure and describes the highest luminance value attainable by the display.

Contrast ratio, by the very definition of the word, measures the ratio between a bright and dark section of an image. A common way to measure contrast is to use a pre-defined black and white pattern such as the ANSI checkerboard. The highest *CR* is achieved in the absence of any ambient light or reflection of screen light back towards the display.

Amplitude resolution describes the number of distinct steps of luminance that can be portrayed by a display – for digital devices generally in terms of bit depth. Virtually all displays use a gamma response curve (often a gamma with a shallow start in the black levels) to distribute those available steps in an optimized fashion for our visual system.

4. Preference Studies

To investigate the relationship between peak luminance (hereinafter often abbreviated *PL*), contrast ratio (hereinafter often abbreviated *CR*) and amplitude resolution (hereinafter often abbreviated *AR*), and their impact on viewer preference, we have conducted three studies addressing the impact of each of these on viewer preference.

4.1 Experimental Constraints

Our visual system has been the object of study for decades and some of the fundamental processes of vision and perception are still not fully understood. We know that our capabilities regarding luminance and contrast perception vary greatly depending on the ambient luminance level, by which we mean the average luminance of the region around the display, which results in light that does not originate within the display entering the eye of the viewer. Even in the photopic range, ambient luminance levels and adaptation time continue to limit our abilities. Detailed descriptions of adaptation models can be found in the psycho-physics literature but for the purpose of this paper we can limit the ambient environment to one commonly encountered in home entertainment. CIE recommendations for illuminance levels in living room environments are 50-120 Lux, depending on the task. In this study we use an ambient illuminance of 100 Lux and a modestly reflective environment (test room with diffuse medium grey walls and no specular surfaces), corresponding to an average ambient luminance of about 20 cd/m^2.

4.2 Experimental Design

In order to investigate visual preference under the ambient conditions described above we have conducted three studies. Each study addresses one of the display characteristics of interest and in each case the results of the previous study are taken into account to provide secondary confirmation. The participant pool varied during the study with 38 participants for the first, 40 for the second and 12 for the third study, all between 18 to 35 years old. Approximately 1/3 of the participants were female for each study and all had normal or corrected to normal vision including color.

All three studies were conducted on 18" HDR Displays. These displays use a conventional LCD in front of a dynamically adjustable matrix of light emitting diodes (LED) which allows for a much higher dynamic range than the LCD alone. Participants sat approximately 1m away from the screen.

All participants were given approximately 10 minutes to adjust to the ambient environment prior to the study and had a chance to see an introductory series of random images spanning the dynamic range of the study. This gave them an overview of the study and helped to normalize the semantic scales used in the first study.

4.2.1 Luminance & Contrast Preference

The first study aims to establish a general overview of the impact of peak luminance and contrast ratio on viewer preference. For this purpose we selected four basic representative images, for each of which we produced sixteen test images by permuting 4 variations of peak luminance levels (1,600cd/m^2, 1,200cd/m^2, 800cd/m^2 and 400cd/m$^{2)}$, with 4 variations of contrast level. Contrast adjustments were done around the center point of the encoding range. Since most images have an average luminance below the center point of their encoding range, this means that a lower *CR* usually increases average image luminance even if *PL* stays constant. The image data was distributed in the luminance range between a value given by the *PL* divided by the *CR* (i.e. the "black" level) and the *PL* using a standard 2.5 gamma commonly used in television displays.

Each participant was exposed to two identical-looking 18" displays as described above. A randomly selected image from the set above was shown on both displays. On one of the displays the image was rendered normally with the appropriate variations in the LED backlight matrix. On the reference displays the LCD image was the same but the backlight was uniformly lit for a peak luminance of 400cd/m^2. The rendering algorithm for the HDR Display [2] was further constrained to make the above *PL* and *CR* adjustments entirely on the LED matrix so that the reference image actually remained unchanged for all 16 combinations per scene. This set-up ensured that color, spatial information, screen reflectance and so forth are approximately identical between the two displays. When the image appeared on each screen the participant was asked to rank the varying image in comparison to the reference image on a semantic scale employing four bi-polar adjectives: *bright – dim, deep – flat, pleasant – unpleasant, realistic – unrealistic*. A central mark on the scale indicates no perceived difference between the two displays.

In this fashion each participant went through the images in all combinations. One of the four images was repeated with each combination to estimate learning and other such effects occurring during the study.

Since the *PL* of the HDR Display used in the study is limited to a little over 1,600cd/m^2 we constructed a static display for a secondary higher *PL* study. Both LCDs where replaced by a calibrated stack of transparencies displaying the same series of LCD images. By replacing the LCD with transparencies the transmission of the system improved dramatically and *PL* levels of over 8,000cd/m^2 were achieved. The remainder of the second study is the same and we found that the results of the two studies align well in the area of luminance intersection (top *PL* of the first study and bottom *PL* of the second are both 1,600cd/m^2).

4.2.2 Contrast Preference

The first study provides a general overview of the *PL* & *CR* preference space. The rough division of *PL* and *CR* into only four choices each was necessary to maintain a manageable number of test images per participant. The second study focused on the relationship between *PL* and *CR*.

In a similar setup to the first study, each participant was exposed to a random image selected from a larger sample of 20 representative images. Each image was displayed at a randomly selected *PL* level and *CR*. Using the UP and DOWN keys of the keyboard the participant could adjust the *CR* of the image until it was most pleasing. The *CR* adjustment was designed to maintain the same average image luminance at all *CR* levels by adjusting contrast around the average point of the image data. Once the preferred *CR* setting was reached, the participant confirmed the selection and a new image was displayed. Each participant was shown all 20 images under 4 different random *PL* levels in this fashion.

4.2.3 Amplitude Resolution Preference

The previous studies provide a framework for *PL* and *CR* preference but assume that amplitude resolution of the displays is high enough at all *PL* and *CR* levels. Since the test HDR display offers a full 16-bit depth with an effective spatial dither due to the analog nature of the LED point spread function in the backlight, we are confident that this assumption is warranted in this case. However, ordinary display technologies do not have such a high amplitude resolution and so it was also desirable to investigate the minimum *AR* threshold per *PL* level. In other words, for a fixed *PL*, we investigated how many distinct luminance levels are necessary to present a visually smooth image, and how many linear bits are needed to present these levels.

These Just-Noticeable-Difference (JND) levels have been studied extensively by the psychophysics community with the two main models coming from Barten [3] and Ferwerda [4]. The Barten model is used mostly in image critical applications such as medical imaging and predicts approximately 1,000 JND over a luminance range of 0.05cd/m^2 to 4,000cd/m^2. Ferwerda's study suggests a much smaller number of JND (approximately 250 in the same range). This is not a case where there is a simple right or wrong, but there seems to be general agreement that for ordinary display applications, Ferwerda's estimate is on the low side, probably because his studies used a pulsing target and such a transient stimulus leads to higher perception thresholds.

Regardless of the discrepancy between the two studies, both apply to abstract test environments rather than TV screens in a typical living room environment. Since the objective of this study is to establish viewer preference under such common conditions, we carried out a modified version of the conventional contrast sensitivity or JND studies in such a setting.

Figure 3: Example target with ambient ring image (target/background contrast is greatly exaggerated)

Using a single HDR Display each participant was shown a 3 degree target (1m viewing distance) within a uniform background luminance. Adjusting the UP and DOWN key allowed the participant to adjust the luminance of the target until it was barely distinguishable from the background. Once the target was visible the participant would press ENTER and the background luminance would be set to the current target luminance. In this fashion the participant traversed the entire luminance range of the HDR display in JND steps.

An initial pilot study with 6 subjects indicated that the luminance steps at the extreme low end of the range of the HDR Display are larger than a single JND. This portion of the range (approximately below 1cd/m^2) has consequently been ignored in subsequent tests. The pilot also replaced the more conventional square transparent grating target with other geometric shapes to counter false positives resulting from a repetitive pattern. (This was a concern because the study took 30-45 minutes per participant.). The results of the pilot approximately matched the predictions of the Barten model - within 15% of predicted JND number and with a similar distribution for all participants. We are therefore reasonably confident that the target shape change and our experimental protocol are acceptable.

In order to represent common display viewing conditions the main study added a ring of image content at a distance of more than 6 degree from the target. By adjusting the average luminance level of the outer ring the study simulates the impact of surrounding image content on the area of interest. The average surrounding luminance (ASL) level remained constant for 4 participants and then changed so that for the total 12 subjects 3 ring images were used. The ring images were taken from representative television images adjusted for an ASL of 1,200cd/m^2, 800cd/m^2 and 400cd/m^2. The images were blurred strongly to avoid distracting spatial frequency content.. Figure 3 shows an example of such a setup with an outer image ring, a constant background area and finally a geometric target.

4.3 Experimental Results

The first study covers a large multi-variable space. First, all results are corrected for individual participant variation using the duplicate images inserted in each series (see Section 4.2.1). Next, the results on the semantic scales are linearized using the *Bright-Dim* scale as a guideline for the other three scales. A pilot study showed that the *Bright-Dim* scale very accurately matches the base three logarithm of the average luminance of each image (This comes as no surprise since in the given range of luminance this is a fair approximation of brightness perception). We therefore adjusted for non-linearity in the other scales by assuming consistency in the non-linearity of all scales. Finally, the three remaining scales were averaged into a single Perceived Image Quality (*PIQ*) scale. The results of all three scales were very similar and combining them in this fashion greatly simplifies presentation of the results. The second and third study are single variable designs and can therefore be used directly.

4.3.1 Luminance & Contrast Preference

Figure 4 shows the results of the first study including the high luminance data from the static display test (*PL* values above 1600cd/m^2). With the exception of two participants the results of the study were fairly consistent with at most 14% variation between the rankings of individual participants. The two outlying participants had generally far higher scores throughout and reached the top of the semantic scales prematurely at low *CR* and *PL* values. Since this saturated their results to the peak *PIQ* value for almost all cases we have discarded their results.

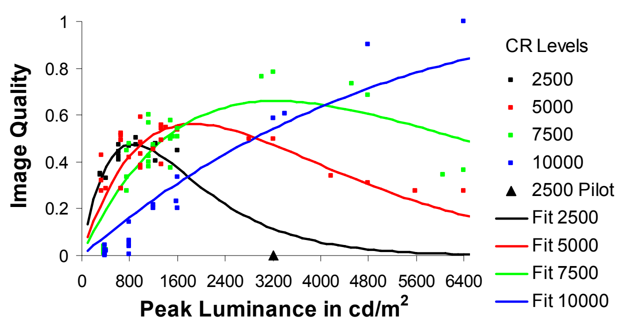

Figure 4: Viewer preference as a function of peak luminance at different contrast ratio levels (note that the *PL* 3,200cd/m^2 data point in the *CR* 2,500 series came from a pilot study)

In considering Figure 4, it appears that for each *CR* level, perceived image quality increases with *PL* value up to a maximum after which it decreases. In other words for each value of *PL*, there is an optimum value of *CR*. At low *PL* the best result is obtained with the lowest *CR* used in the study (2500). As *PL* rises this *CR* is insufficient to maintain the highest Perceived Image Quality and higher *CR* provide better results. This relationship holds not only for rising *PL* but also implies that high *CR* are not optimal for lower *PL* settings. To explain this effect we need to consider the image creation process in Section 4.2.1. The test images were scaled into the dynamic range between *PL* and *PL*/*CR* (the black level for this image) using a gamma of 2.5. At high *CR* and low *PL* this means that the bulk of the image content is shifted towards the dark region of the image – often to the point that shadow information is completely lost to the viewer. It might therefore be possible to gain some benefit from a higher *CR* at lower *PL* by adjusting the grey level distribution.

We have used a simple model to describe *PIQ* as a function of *PL* for different *CR* levels, as depicted in Equation (1):

$$PIQ(PL) = a \cdot PL \cdot e^{-b \cdot PL} \qquad (1)$$

A least-squares fit, adjusting the free parameters *a* and *b,* yields a reasonable fit with the data. For this fit, the optimal value of *PL* (PL_{opt}) and the corresponding maximal value of *PIQ* (PIQ_{max}) are given by Equation (2):

$$PL_{opt} = \frac{1}{b} \qquad PIQ_{\max} = e^{-1}\frac{a}{b} \qquad (2)$$

These values are shown for the four different values of *CR* in Table 1.

CR Level	2500	5000	7500	10000
PL_{opt}	856	1881	3256	12419
PIQ_{max}	0.4756	0.5621	0.6602	1

Table 1: Fit parameters for Figure 4

4.3.2 Contrast Preference

With the general relationship between *PL* and *CR* indicated above, we can use the results of the second study to establish the optimal *CR* range for each *PL* level. The results have been combined into bins of 100 *CR* levels each and averaged over each bin, as shown in Figure 5. The standard deviation is modest for most bins though there are a small number of outliers for most bins as one would expect for perception data. Equation (3) relates the optimal contrast ratio to peak luminance based on a least-squares fit. The function also agrees with the two high *PL* results from static display data in the first study (The two highest points in Figure 5 correspond to PIQ_{max} for the 7,500 and 10,000 *CR* lines).

$$OCR(PL) \cong 2862\ln(PL) - 16283 \qquad (3)$$

In view of the supporting data, it would be inappropriate to suggested that this value for optimal contrast function represents a sharply defined optimum; rather, it provides only a general guideline for the optimum contrast to achieved the optimal Perceived Image Quality for a given peak luminance. Figure 6 shows the results of the first study filtered by the optimal contrast function such that only results within 10% or 30% of the optimal contrast for that *PL* are considered. The resulting relationship between PL and PIQ with appropriate CR is logarithmic and well defined up to approximately 1,600cd/m2.

We also carried out a pilot study, with a much smaller number of images, for PL values up to 12,000cd/m2. Interestingly, it appeared that subjective ratings of image quality declined for PL values above 6,000-7,000cd/m2, regardless of contrast, suggesting that yet another effect is at work in this range. Likely, the problem in this range is simply discomfort glare given the modest selected ambient luminance level.

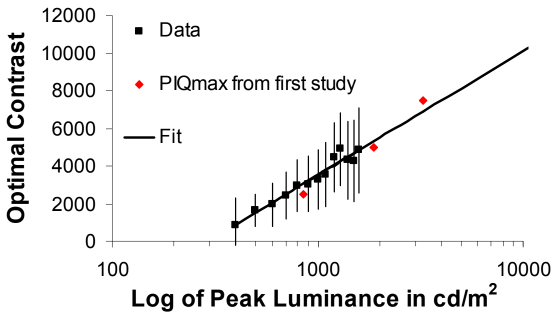

Figure 5: Log relationship between optimal contrast and peak luminance

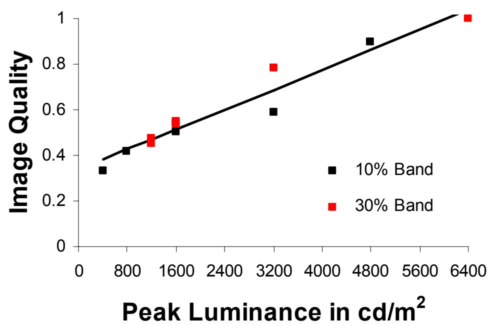

Figure 6: Viewer preference data filtered for optimal contrast

4.3.3 Amplitude Resolution Preference

Both sections above describe the relationship between *PL*, *CR* and *PIQ* under the given ambient conditions. As we have seen in Section 4.3.1 AR and grey level distribution play a critical role in this relationship. The *AR* of the HDR Display is fortunately far beyond the requirements of the Barten model above 1 cd/m^2 and it is therefore fair to assume that the results of both studies are unaffected by *AR* limitations. Yet this is largely not true for more conventional displays which are usually limited to 8-bit or 10-bit.

Figure 7 shows the results of the third study for each of the three ASL levels (see Section 4.2.3). The fourth (black) line is the Barten model which is the equivalent of an ASL of 0c/dm^2. Table 2 summarizes JNDs found within the range of the experiment.

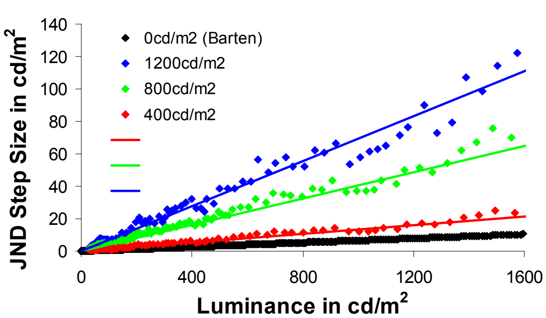

Figure 7: JND step size per luminance level under different average ring luminance conditions

The ASL impact is clearly significant in reducing the number of distinguishable steps. At the same time the perception model remains the same and the perception threshold stays approximately constant over the entire luminance range of the study. With the approximate ASL of specific application these values can then be used to estimate the required *AR* to remain below threshold. For television images for example ASL is usually between 20-30% of peak luminance.

Ring	0 (Barten)	400	800	1,200
JND	845	579	296	114
Threshold	0.007	0.013	0.041	0.069

Table 2: JND steps and thresholds per ring image luminance

5. Future Work

The consistency of results in all three studies is fairly high so future research will likely be directed towards expansion of the results rather than refinement. An obvious area of expansion is to map the high *PL* region of the study in more detail to remove some of the uncertainty in this area.

More fundamental additions to the framework would be an investigation of temporal and color gamut considerations and their impact on Perceived Image Quality. Such an expansion is possible with HDR Displays using rapid modulation of the LED backlight and RGB LED backlights respectively.

6. Conclusion

We have conducted a series of studies to provide simple design guidelines for displays. Specifically, the first study established the relationship between peak luminance of a display and viewer preference. We would suggest that common value estimation models could now be used to turn such viewer preference into commercially relevant factors which could then be compared to increased material cost for higher luminance designs. The second study has established preferred contrast ratio levels for specific peak luminance levels. Finally, we have provided some observations of optimal amplitude resolution requirements for displays as a function of peak luminance.

Combined, these results provide a guideline for display engineers to establish first the peak luminance of a display design, by balancing viewer preference against device cost, followed by appropriate choices of contrast ratio and amplitude resolution.

7. References

[1] H. Seetzen, L. A. Whitehead, G. Ward, “A High Dynamic Range Display System Using Low and High Resolution Modulators”, Proc. of the 2003 SID Symposium, May 2003.

[2] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. A. Whitehead, M. Trentacoste, A. Gosh, A. Vorozcovs, “High Dynamic Range Display Systems”, ACM Trans. Graph. 23, 3 (Aug. 2004), 760-768.

[3] P. Barten, “Physical model for the contrast sensitivity of the human eye”. Proc. SPIE, vol. 1666, 57-72, 1992.

[4] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, D. P. Greenberg, “A model of visual adaptation for realistic image synthesis”. Proc. of SIGGRAPH '96, 249-258, 1996.