



The World Wide Gamma

Nathan Moroney, Giordano Beretta

HP Laboratories
HPL-2010-129

Keyword(s):

display, electro-optical conversion function, lightness scaling, online experiments

Abstract:

Displays continue to evolve in numbers, specifications and applications, with many of these displays being used to access content via the World Wide Web, on telecommunications networks or using proprietary content services. This paper uses an experimental technique to estimate the "gamma" for displays on the World Wide Web. A lightness partitioning experiment was performed by 404 volunteers on the World Wide Web and the resulting data was analyzed to estimate a relationship between lightness and display values. While there are many uncontrolled variables and sources of uncertainty, robust non-parametric statistical analysis results in very low standard errors. The fitted function has an offset for the black point, but the remainder of the lightness versus display digital count data is well fit with a linear function. Overall the data is well fit with an offset of about -0.04 and a gamma of roughly 2.36.

External Posting Date: September 21, 2010 [Fulltext] Approved for External Publication
Internal Posting Date: September 21, 2010 [Fulltext]
To be presented at the 18th IS&T/SID Color Imaging Conference, November 10, 2010, San Antonio, TX

© Copyright The 18th IS&T/SID Color Imaging Conference, 2010

The World Wide “Gamma”

Nathan Moroney and Giordano Beretta; Hewlett-Packard Laboratories, Palo Alto, CA, USA

Abstract

Displays continue to evolve in numbers, specifications and applications, with many of these displays being used to access content via the World Wide Web, on telecommunications networks or using proprietary content services. This paper uses an experimental technique to estimate the “gamma” for displays on the World Wide Web. A lightness partitioning experiment was performed by 404 volunteers on the World Wide Web and the resulting data was analyzed to estimate a relationship between lightness and display values. While there are many uncontrolled variables and sources of uncertainty, robust non-parametric statistical analysis results in very low standard errors. The fitted function has an offset for the black point, but the remainder of the lightness versus display digital count data is well fit with a linear function. Overall the data is well fit with an offset of about -0.04 and a gamma of roughly 2.36.

Introduction

Displays vary in their size, aspect ratio, spatial and temporal resolution, video processing hardware and software, color modulation technology, and in many other features and specifications. In this paper we focus on the “gamma” or electro-optical conversion function. The history and nuances of the term “gamma” are long and convoluted.^{1,2} In this experiment the focus is on the relationship between perceived lightness and display values. Display primaries, white point and variation in the non-linearity across primaries are not considered in this paper.

Lightness partitioning was one of the experimental techniques used to derive the Munsell value scale.³ This task uses an array of variable or adjustable stimuli and an overall objective to create equal lightness steps between neighboring stimuli. Figure 1 shows a diagram of the mechanism used by Munsell, Sloan and Godlove in 1933 to conduct their lightness partitioning experiments, which they referred to as *value step experiments*. Observers adjusted individual sliding stripes to create their partitions. They reported close agreement between their value step experiments and the just-noticeable difference experiments. Therefore we consider the lightness partitioning technique to be a fundamental and reliable technique for lightness scaling.³⁻⁵

We choose to use the World Wide Web for our experimentation. This allows access to a much larger pool of potential participants⁶ and potentially a larger sampling of displays. It is simply not practical to measure large numbers of displays “in the wild” on an ongoing basis. With the assumption of some degree of consistency in lightness partitioning across observers, the central question is: can we use an experimental technique to infer the average display gamma? There are many potential sources of uncertainty and variability; such as display type, viewing angle, ambient illumination, veiling glare, display cleanliness, possible participant color deficiencies or anomalies, operating systems, browsers, color management systems, and

many other considerations. However, with a large enough population it might be possible to derive average estimates with a sufficiently low standard error to be useful for some applications. The goal is not an estimate of individual displays but of an overall gamma for the World Wide Web.

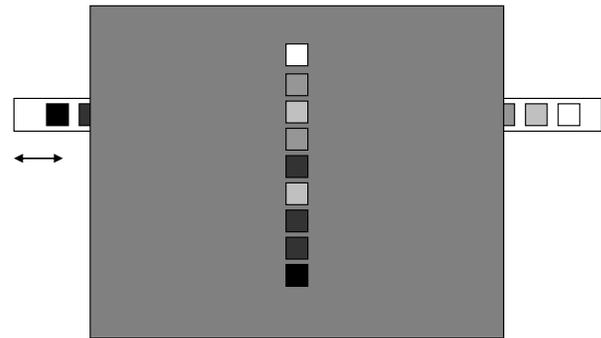


Figure 1. The physical device used by Munsell, Sloan and Godlove for lightness partitioning. A 26 by 35 cm matte 21.6 percent reflectance grey board had seven 1.2 cm square holes cut into it. A fixed white 88.3 percent reflectance anchor is shown on the top and a black 2.9 percent reflectance anchor is shown on the bottom. Seven sliding scales were then used to create equal lightness steps. Only one sliding scale is shown.

Previous related research⁷⁻¹⁵ includes gamma estimation tools, gamma sensitivity analysis and visual calibration techniques. These include matching of one or more patches to a reference spatial or temporally dithered patch. These approaches make implicit assumptions of the lightness values of the dithered anchors. This paper differs from this previous research in two significant ways. First we use the lightness partitioning task used by Munsell, Sloan and Godlove to assess the display non-linearity. This is a slightly more complex task but the end result is a direct estimate of the lightness steps and more data for model fitting. Second we deployed a web-based application to harvest the results from many participants to attempt an overall estimate of the World Wide “gamma”. Additional optional data, such as display type and participant information, are collected but they are not considered in this initial publication.

Experiment

Using the World Wide Web, 404 volunteers performed a lightness partitioning task using their display. The volunteers were recruited using blogs and email lists starting in May 2008 and continuing to the present. The specific task was, given black and white anchors, the participant was to create equal lightness steps for five intermediate patches. The black and white anchors were

the left and right of the five adjustable patches as shown in Figure 2.

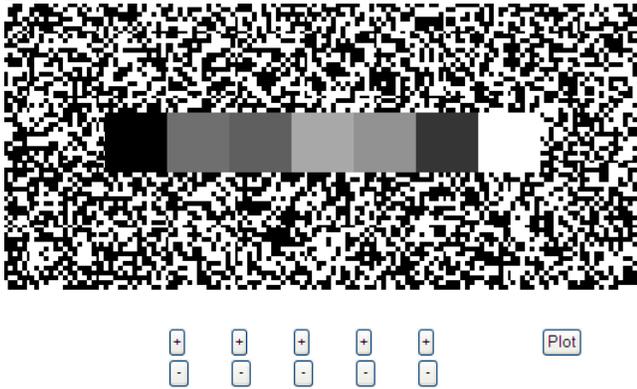


Figure 2. Screen shot of the lightness partitioning experiment with the black anchor shown on the left and the white anchor on the right. The plus and minus buttons below the intermediate five patches were used to make the corresponding lightness patch above the buttons lighter or darker. The 'plot' button was used to submit the data and graph the results.

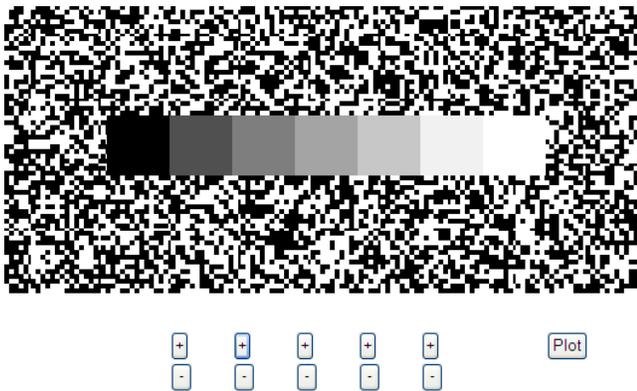
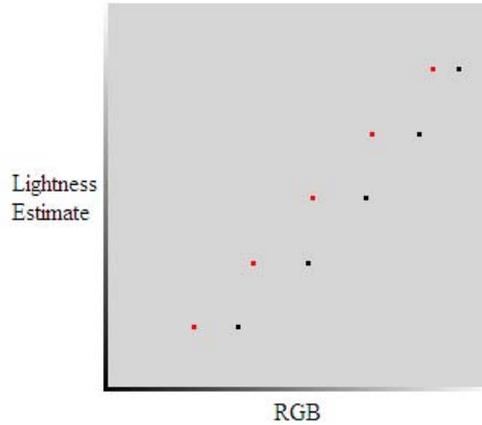


Figure 3. Screen shot of lightness partition after adjustments. This is an approximate representation of a single example partition.

The black anchor was a square with red, green and blue values of 0, 0, 0. The white anchor was a square with red, green and blue values of 255, 255, 255. The five intermediate patches were randomized to equal RGB values between 0 to 255 for each experiment. The patches were 50 by 50 pixels and for a 100 ppi display would be roughly 1 degree at a viewing distance of 33 cm. Unfortunately the size will vary based on the display resolution, but consistent representation of size is a known shortcoming of HTML. A coarse 4 by 4 pixel black and white dithered background was used to reduce the effect of crispening on lightness scaling.¹⁶ Crispening is the increased sensitivity to perceptual differences as a background color falls between two patch colors. A screen shot of an approximate lightness partitioning is shown in Figure 3.



Optional Feedback

Display: Ambient Illumination:
 Comments:

Figure 4. Individual results plotted for the current display, viewing conditions and observer, shown as black points, and the current global average, shown as red data points.

Once each observer completed their lightness partitioning their results were immediately shown relative to the current global average. At this point, additional optional data was collected from the participants. This is shown in Figure 4. The current global experimental average is shown as red points in the figure while the results for the current display and participant are shown as black data points. This rudimentary graph was implemented directly in the web-application and therefore lacks the sophistication of current graphing applications. In the case of Figure 4, the results are significantly different from the global average. The black points of the current experimental data are consistently below the red points of the current global average.

The submitted data was also screened for potentially disruptive participants. The 404 participants analyzed in this paper all passed a minimal screening test. Specifically the submitted data values were summed per participant. An observer whose summed value was greater than 950 or less than 450 were discarded from subsequent analysis. As a result data from 28 participants was discarded. The result is that the data is overall roughly monotonic and extreme submissions, such as all patches set to black or white, were eliminated.

Results

A visualization of all of the responses is shown as an image in Figure 5. This image is a "meta step ramp" in that each pixel for a given step is a single participant. That is each step in the ramp consists of the first 400 observer responses shown as a 20 by 20 square. To achieve a square aspect ratio for each lightness level, the last four participants are not shown but this omission is only used for the visualization in Figure 5. The image gives an indication of both variations in the data and an overall level per

square. This step ramp has effectively been dithered with real world deviance.

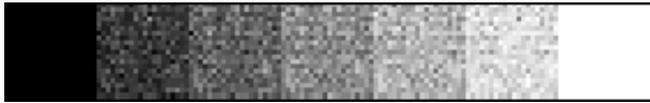


Figure 5. Visualization of the first 400 of the raw responses. Each individual square is a single participant's partition value. The left black and right white anchors are shown as a single value. The image shows deviations, outliers and yet overall lightness levels

The Anderson-Darling test for normality¹⁷ at the 5% level is only achieved for one of the five partitions. This suggests that a majority of the observer data does not follow the normal or Gaussian distribution. This can also be seen in the superimposed normal quantile-quantile plots¹⁸ shown in Figure 6. The experimental data is shown for the five partitions with a line through the corresponding first and third quartiles. For a normal distribution a majority of the data points would fall on the straight line. The deviations at the ends of the distributions support the use of robust non-parametric methods, such as the median, median absolute deviation¹⁹ and bootstrapping.²⁰

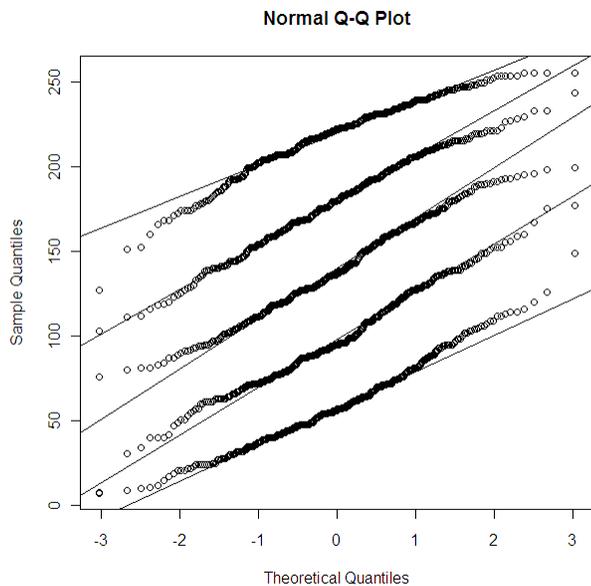


Figure 6. Superimposed quantile-quantile plots for the raw observer partitions. The data progresses from darkest to lightest partition from the bottom to the top of the figure. For each partition a line is shown through the first and third quartiles.

The median value per partition, shown in Figure 7, exhibits an offset for the black point. That is, the intercept of the fitted function relating the lightness partition and the display digital counts was negative. This can be interpreted as meaning that all digital count values below roughly 17 will, on average, appear black to participants. From other experimental analysis²¹ we know that on average a majority of observers will call this anchor the color term “black”. The remainder of the lightness scale is well fit

with the solid-line linear function. The maximum lightness point corresponds to the display white point and we know from other experimental analysis²¹ that on average a majority of observers will call this patch the color term “white”. Likewise the central partition point will on average be called “gray” or “grey” by a majority of observers. This suggests given hundreds of participants and displays - there is in fact on average a good match of display non-linearity and corresponding lightness scale.

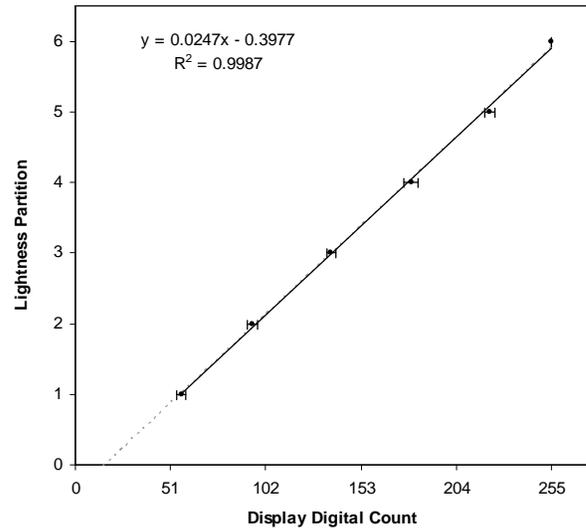


Figure 7. Median display digital count versus lightness partition for 404 experimental participants. The straight line is the fitted linear function, the dotted line is the extrapolation of the fit to the black point and the data is shown plotted with error bars corresponding to plus or minus one bootstrapped standard error.

The results shown in Figure 7 also show a fitted linear function as a solid line and an extrapolation to the black point as a dotted line. Error bars are also shown for the digital counts for each lightness partition. These error bars are plus or minus one bootstrapped standard error. For a normalized abscissa and ordinate the slope is 1.05. The results are also shown in tabular form in Table 1.

Table 1. Median, median absolute deviation (MAD) and one bootstrapped standard error (BSE) for the five lightness partition levels where partition one is the darkest and five the lightest.

Partition	1	2	3	4	5
Median	57	95	137	180	222
MAD	22	27	28	27	19
BSE	1.2	1.5	1.2	1.9	1.4

Discussion

There are many sources of uncertainty and variability for estimating the World Wide “Gamma” using crowd sourcing. However the initial results shown in Figure 7 are encouraging. The use of hundreds of displays and participants results in bootstrapped

standard errors that are less than plus or minus 2 digital counts. Achieving this level of precision in a laboratory setting is likely impractical. For example while a team of ten graduate students might be able to recruit and carry out this experiment on 40 volunteer observers each in parallel it would still require the sourcing of over 400 different displays. For the world wide “gamma” the resulting correlation for the linear fit is an r^2 value of 0.9987, which is quite good.

These results can be fit by specific monitor or display models. For example, a simplified gain-offset-gamma or GOG model²² of the form $((1-\text{offset})^X + \text{offset})^Y$ can be fit to the data assuming an initial L^* lightness partition. An exhaustive search to two decimal places for the minimum standard deviation of the differences results in a fitted offset of -0.04 and a gamma of 2.36. The simplified GOG model was used to reduce the complexity of the fitting process. The standard deviation of the differences was used as a measure that minimized differences without the possibility of simply minimizing the cancellation of positive and negative differences. Other fitting methods and criteria are also possible and can be performed using the published data. Given the relative decline in CRTs this fit is provided for comparison purposes only. However it is interesting to note that the estimated value for the overall gamma is not that far off from the sRGB²³ non-linearity and also a recent recommendation² by Poynton, but the offset would appear to differ. Note that these values are not optimal or best but simply the fit to the median experimental data collected from hundreds of displays and participants. The raw observer data has also been posted online.²⁴ We anticipate variations on the experiment to investigate the display non-linearity on a channel by channel basis.

References

- [1] C. Poynton, The rehabilitation of gamma, Human Vision and Electronic Imaging III, Proc. SPIE 3299, pp. 26-30 (1998).
- [2] C. Poynton, “Picture rendering, image state, and BT.709”, from <http://www.poynton.com/notes/PU-PR-IS/Poynton-PU-PR-IS.pdf> (accessed April 5, 2010).
- [3] A.E.O. Munsell, L.L. Sloan and I.H. Godlove, Neutral Value Scales. I. Munsell Neutral Value Scale, JOSA, 23, pp. 394-411 (1933).
- [4] Surajit Nundy and Dale Purves, A probabilistic explanation of brightness scaling, Proc. Nat. Acadm. Sci., 99(2), pp. 14482-14487 (2002).
- [5] P. Whittle, Brightness, discriminability and the ‘crispness effect’, Vision Res. 32(8), pp. 1493-1507, (1992).
- [6] Ulf-Dietrich Reips, Internet experiments: Methods, guidelines, metadata, Human Vision and Electronic Imaging XIV, Proc. SPIE 7240(1), 724008 (2009).
- [7] Bilissi, E.; Jacobson, R.E.; Attridge, G.G., Just noticeable gamma differences and acceptability of sRGB images displayed on a CRT monitor, The Imaging Science Journal, 56(4) pp. 189-200 (2008).
- [8] W. Han, J. Shi, P. He and L. Yun, Color reproduction from desktop display to projector based on visual matching, Chinese Optics Letters, 7(8), pp. 748-752 (2009).
- [9] P. He, J. Shi, X. Huang and Q. Li, Investigation on color shifts for different gamma of display system in CIECAM02-based uniform color space, Proc. IEEE Int. Conf. Comp. Sci. and Soft. Eng., pp. 30-33 (2008).
- [10] Ricardo Motta, Visual characterization of color CRTs, Proc SPIE 1909, pp. 212-221 (1993).
- [11] Snjezana Soltic and Andrew N. Chalmers, Modeling the effects of gamma on the colors displayed on cathode ray tube monitors, J. Elect. Imaging, 13(4), pp. 688-700 (2004).
- [12] J. Gille and J. Larimer, Using the Human Eye to Characterize Displays, Human Vision and Electronic Imaging VI, Proc. SPIE 4299, pp. 439-454 (2001).
- [13] J. Gille, N. Arend and J. Larimer, Display characterization by eye: contrast ratio and discrimination throughout the grayscale, Human Vision and Electronic Imaging IX, Proc. SPIE 5292, pp. 218-233 (2004).
- [14] Y. Lavin, A. Silverstein and X. Zhang, Visual Experiment on the Web, Human Vision and Electronic Imaging IV, Proc. SPIE 3644, pp. 278-289 (1999).
- [15] X. Zhang, Y. Lavin and D. A. Silverstein, Display Gamma is an Important Factor in Web Image Viewing, Human Vision and Electronic Imaging VI, Proc. SPIE 4299, pp. 455-462 (2001).
- [16] N. Moroney, Factors affecting lightness partitioning, Proc. SPIE, 4663, pp. 35-42 (2002).
- [17] Anderson-Darling test for normality, from <http://pbil.univ-lyon1.fr/library/nortest/html/ad.test.html> (accessed April 5, 2010).
- [18] NIST/SEMATECH e-Handbook of Statistical Methods, “Quantile-Quantile Plot” from <http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm> (accessed April 5, 2010).
- [19] C. Park and B.R. Cho, Development of Robust Design Under Contaminated and Non-normal Data, Quality Engineering, 15(3), pp. 463-469 (2003).
- [20] R Library: Introduction to bootstrapping. UCLA: Academic Technology Services, Statistical Consulting Group, from <http://www.ats.ucla.edu/stat/R/library/bootstrap.htm> (accessed April 5, 2010).
- [21] N. Moroney, Unconstrained web-based color naming experiment, Proc. SPIE 5008, pp. 36-46 (2003).
- [22] Roy S. Berns, Ricardo J. Motta and Mark E. Gorzynski, CRT colorimetry. Part 1: Theory and practice, Color Res App., 18(5), pp. 299-314 (1993).
- [23] International Electrotechnical Commission, Part 2-1: Colour Management – Default RGB Colour Space – sRGB, First Edition, IEC 61966-2-1 (1999).
- [24] <http://www.mostlycolor.ch/2010/10/world-wide-gamma-data-technical-report.html>

Author Biography

Nathan Moroney is with the Print Production Automation Lab at Hewlett-Packard. He has a Masters Degree in Color Science from the Munsell Color Science Laboratory of RIT and a Bachelors degree in color science from the Philadelphia University. Nathan is a fellow of the IS&T and was the technical chair for CIE Technical Committee 8-01, which developed the CIECAM02 color appearance model. He was the past general co-chair for the IS&T/SID Color Imaging Conference and is a contributor to mostlycolor.ch.

Giordano Beretta is with the Print Production Automation Lab at Hewlett-Packard. He did his graduate work in computational geometry at ETH, before joining Xerox PARC in 1984 to work on color reproduction. After a stint in strategic planning and becoming the technical advisor for Color at Canon, he joined HP, where he is a member of the Commercial Print Automation project. He is a Fellow of the IS&T and the SPIE, and recipient of the IS&T Bowman Award.