Visually Determining Gamma for Softcopy Display

R. Victor Klassen and Raja Bala, Xerox Innovation Group, Webster, New York, USA
Nathan Klassen, Monroe Community College, Rochester, New York, USA

Abstract
We propose estimating the gamma value for the blue channel of a softcopy display via a grey-balancing task. A "near-neutral" patch is displayed along with a neutral reference pattern made from black and white pixels. The red and green values for this near-neutral patch are derived from a separate calibration of these 2 channels. The user then adjusts the blue value until the near-neutral patch appears neutral with respect to the surrounding reference. Gamma for the blue channel is estimated from the selected value. Implementations include interactive adjustment with mouse and/or keyboard, or selecting from a fixed set of patches the one that appears most neutral. When compared with the standard technique, inter- and intra-observer variation in the settings for blue were substantially reduced.

Introduction
Soft proofing continues to gain importance especially in the graphic arts and production colour markets. We expect it to play an increasingly important role in distributed and remote colour management applications. To be useful, soft-proofing depends on a calibrated display. At the high end in the graphic arts market, users are willing to calibrate their displays using expensive colour measurement instruments. Further down market, users may use interactive visual calibration software which they may buy (from e.g. Adobe or Monaco Systems) or discover on a Macintosh system (on which it comes bundled). While visual techniques are less accurate than their measurement-based counterparts; they are relatively inexpensive, with sufficient quality for many applications.

An important colour characteristic of display devices is the 1-dimensional tone response of each of the R, G and B primaries. For CRTs and many LCDs, this tone response is described by a power-law relationship between input digital value and displayed luminance.[1] One simplified form of the power-law function is given as follows:

$$Y = \begin{cases} K\left(\dfrac{D - D_o}{D_{\max} - D_o}\right)^{\gamma} & if \quad D > D_o \\ 0 & if \quad D \le D_o \end{cases} \tag{1}$$

where D is the input digital count, Y is the fractional displayed luminance, D_0 is the black offset value below which there is no discernable luminance, D_{max} is the maximum digital input value, K is the gain, and γ is the exponent often referred to as "gamma". The tone response in Eqn (1) must be derived individually for each of R, G, B, thus resulting in potentially different values for K and gamma for the 3 channels.

As indicated earlier, the most accurate method of calibrating the tone response is to take radiometric measurements corresponding to multiple digital input values along the tonescale, and deriving the function parameters via some form of regression, data fitting or interpolation.[1] A simpler and less costly approach is to use visual tasks (i.e. no instrumentation) to estimate the tone response. The focus of this paper is on estimating gamma using purely visual tasks. In the following formulation, it is assumed that K=1, D_{max} = 255 (i.e. an 8 bit system), and the offset D_0 is obtained from a separate preceding visual task.*

Several techniques have been proposed for visually estimating gamma for displays.[2-6] Perhaps the most well-known approach is shown in Fig. 1. The task involves using the slider to adjust the digital value of the continuous-tone patch on the right until its luminance matches the average luminance of the halftone pattern on the left, generated using alternating on/off lines.

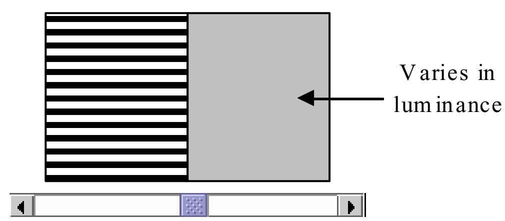

Figure 1. Gamma determination by 50% luminance matching.

The task can be executed on a greyscale R=G=B stimulus (as shown in Fig. 1), in which case it is assumed that the same gamma applies for all three channels. Alternatively, the task can be carried out separately for R, G, and B to derive potentially different gamma values for the 3 channels. The assumption is that the fractional luminance Y_{HT} of the halftone pattern is 50%, i.e. it is halfway between the luminances at full-off and full-on. The value of gamma is estimated from the digital value needed to match the 50% fractional luminance as follows:

$$Y_{HT} = 0.5 = \left(\frac{D_{select} - D_o}{255 - D_o}\right)^{\gamma}$$

$$\Rightarrow \quad \gamma = \frac{\log(0.5)}{\log\left((D_{select} - D_o) / (255 - D_o)\right)} \tag{2}$$

where D_{select} is the digital value selected in the visual luminance matching task.

The technique just described is in use within many commercially available display calibration tools, and, to the authors' knowledge, was first described by Cowan[2].

Visual tasks that assume the same gamma for the three channels use greyscale (R=G=B) images or patches, and are generally simple to execute. However, the equi-gamma assumption is often incorrect. The Photoshop 3.0 calibration tool attempts to correct for this assumption by having the users perform a grey-balance adjustment jointly with the 50% greyscale luminance matching task. This is an iterative procedure that could produce inconsistent results from observer to observer. Apple's display calibration tool provides similar functionality.

Since the power-law response is a channel-wise phenomenon, it makes more sense to estimate gamma separately for each of the three channels. The problem is that luminance judgments are very difficult to perform for the blue primary. Vision scientists (e.g. Wandell[7], p. 328) believe that the blue (short-wavelength) sensor response contributes little if any to the human visual system's luminance channel. The medium- and long-wavelength sensors also respond, but to a much lesser extent, to light generated by the blue phosphor of a CRT. Hence relatively large changes to the strength of the blue signal yield small changes in the visual (luminance) response. The resulting difficulties in the visual task produce large variances in the estimated gamma value for blue.

Perhaps more important is the finding that the short wavelength sensor response is not used in edge detection[8]. In the typical setup (as in Fig. 1), the user attempts to reduce the strength of the edge between the two halves of the field. It is easy to bring the edge strength down to threshold, as the long and medium sensors are scarcely stimulated by a field containing only pure blue and black.

On the other hand, we do use our short-wavelength cones for hue discrimination. As evidenced by the data that supports the various colour difference metrics such as ΔE94, ΔECMC, and CIEDE2000[9], short-wavelength cone contributions to colour differences are roughly the same as those of the other cones, at least in the neighbourhood of the neutral axis. This suggests that a grey-balancing task does not result in a disadvantage in the gamma estimate for blue relative to the other primaries. It does not say whether grey balancing is more or less sensitive than luminance matching. However, experience has shown that grey balance is an important indicator of perceived image quality, or the presence or absence of a colour cast. One might expect that the limits to our ability to adjust grey balance would be no worse than our limits in noticing colour casts in images.

Grey-Balance for Calibrating the Blue Channel

We propose a visual method of determining the gamma for the blue primary that is more consistent than the luminance matching task. It is based on the insight that accurate gamma estimation for blue is important not for luminance reproduction, but for proper colour-balance, most importantly grey-balance. Thus it makes sense to use grey-balancing, rather than luminance matching, as the criterion for selecting the blue gamma value. Our notion of neutral is somewhat affected by the white point of our adapting environment, so providing a reference white is helpful.

The idea then is to design a visual task to find a patch best representing neutral, given calibrated digital values for the red and green primaries that produce 50% fractional luminance. (The latter are obtained from any standard approach, e.g. the task shown in Fig. 1.) A large patch is displayed within a larger surround, which contains white and preferably a checkerboard or line pattern, to establish a reference for the neutral axis. This is shown in Fig. 2. The user adjusts a slider, causing only the digital input to the blue channel to change, while the red and green inputs are fixed at the calibrated 50% luminance level. This changes the hue of the patch in the middle, moving it along a line from yellowish to bluish. The user selects the value at which the patch appears most nearly neutral with respect to the surround. Effectively the task is to match the chromaticity of the patch with that of the halftone pattern (which by definition is the same as that of the display white). We then use the selected value to estimate gamma for the blue primary using Eqn (2).

Note that the proposed method of grey-balancing relies upon the so-called "chromaticity constancy" assumption which states that different levels of a pure primary produce the same x-y chromaticity coordinates. This assumption is usually valid for CRT displays, but is violated in some LCDs[10] (especially the low-cost versions found in laptop computers.) When the assumption is violated, the estimate of gamma obtained by single-primary luminance matching can be significantly different from that obtained by grey-balancing (the difference being systematic, and larger than inter-observer variations). In the next section we will discuss the implications of this.

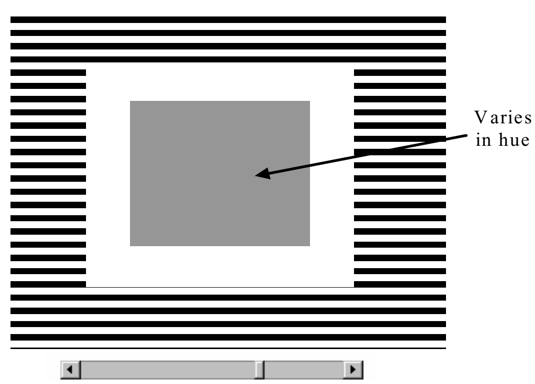

Figure 2. Blue gamma determination via grey-balancing. The user adjusts the slider until the patch in the middle appears neutral "grey" with respect to the surround.

Experimental Results

CRT Experiment

The proposed visual calibration technique was implemented as a Java applet and tested on a Gateway CRT display. In a separate visual experiment, it was determined that D_0=5. Five observers were then first asked to perform the 50% luminance matching task shown in Fig. 1 separately for each of the R, G, B channels. All but one observer performed the task twice, providing a total of 9 observations. Gamma estimates for R, G and B were then calculated using Eqn (2).

Each observer was then asked to perform the grey-balance task shown in Fig. 2 to determine a second gamma estimate for the blue channel. The R and G values were taken from the previous task. Eqn (2) was again used, where this time D_{select} is the blue digital value that produces the best grey-balance match with the surround.

In Table 1, statistics are compared for the nine gamma estimates for the blue channel from the luminance matching vs. grey-balancing tasks. The precision used in this implementation resulted in a quantization step of 0.04 for gamma values.

Table 1: Statistics for Blue Gamma Estimates from Standard Luminance Matching vs Proposed Grey-balancing

	Mean	Std. Dev s	Min	Max	Max single-observer range
Luminance matching	2.24	0.138	2.07	2.51	2.17-2.51
Grey-balancing	2.26	0.027	2.25	2.33	2.25-2.29

The results show that the average gamma estimates from the two approaches are the same (i.e. within quantization precision). The standard error of the mean is □/√N, where N is the number of observations. In this experiment, the standard errors were .046 and .009 for the luminance matching and grey-balance tasks respectively. This indicates that the grey-balance approach estimate (i.e the average of the observers' grey-balance-based estimates) lies within the standard error of the luminance matching approach (2.194-2.286), while the mean of the luminance matching is nearly within two standard errors of the mean of the grey balancing approach (2.242-2.278). However the proposed grey-balancing task produces substantially less variance than the standard luminance matching task. This is true both across observers, and across repetitions of the task by a single observer. Based on these (admittedly small) statistics, an Excel simulation indicated that an individual using the grey balancing approach would select a value within the range 2.189-2.333 95% of the time, whereas using the luminance matching approach, they would be outside this range over 70% of the time. An F test on the two variances indicates statistical significance at the 99.99% confidence level.

While it is instructive to examine the consistency of the gamma estimates, what is of ultimate interest is the image quality from the resulting correction. That is, we would like to see how variances in gamma estimates translate to differences in colour reproduction of images. To this end, two calibrated RGB images were gamma-corrected using the gamma estimates from Table 1. The gamma for R and G were chosen from a single observer's response to be 2.21 and 2.17 respectively. To examine the worst-case scenario, the B channel was corrected with the minimum and maximum gamma values for the luminance matching task (from Table 1 these are 2.07 and 2.51). Prints were made of images corrected using each of the two methods. To generate these prints, the gamma-corrected image was assumed to be in sRGB space (equivalently, the Gateway display was assumed to be an sRGB display). This image was mapped to CMYK using the printer characterization transform for a Xerox DocuColor12, and then printed to this device.

We observed significant differences between the two corrected images, indicating that variations in the gamma estimate for the B channel can indeed have a strong effect on the final reproduction. The same procedure was repeated for the grey-balancing task. The differences between the two extreme cases were very difficult to find. This shows that the consistency of the blue gamma estimate considerably affects the consistency of the resulting colour reproduction, and that the proposed approach produces far more consistent images than the standard approach.

An equivalent quantitative experiment was also performed. An 8×8×8 uniformly sampled RGB grid was generated. These RGB values are to be interpreted as raw device values driving the CRT. The R and G channels were raised to powers of 2.21 and 2.17 respectively. The B channel was raised to the minimum gamma value of 2.07 obtained from the luminance matching experiment. The result is a set of RGB values linearized in luminance according to the visual gamma estimates. These RGB values were converted to XYZ and then to CIELAB, assuming sRGB primaries and white point. A second set of CIELAB data was obtained using the same procedure, but assuming the maximum gamma of 2.51 from the luminance matching experiment. CIE ΔE differences were computed between the two data sets, and are shown in Table 2. Clearly the variations in observers' response to the visual task produce some significant ΔE errors. While one could analytically compute the maximum ΔE induced by a given change in blue gamma (differentiating L*a*b* wrt gamma), we believe this exercise to be at least as instructive.

The same calculation was performed using the minimum and maximum blue gamma estimates from the grey balancing approach. These are also included in Table 2. Clearly, the grey balancing approach results in far less intra- and inter-observer variation, thus offering a more consistent and robust approach to gamma estimation for the blue channel.

Table 2: Observer Variations from the Luminance Matching vs. Grey Balance Approach, Measured in CIE 1976 ΔE Units.

	CIELAB ΔE		
	Average	95^{th} Percentile	Maximum
Luminance Matching	5.41	11.2	15.2
Grey Balancing	0.98	2.03	2.80

All of the above analysis was based on data collected from five observers, all colleagues of the first two authors. Because the difference in variances was so large, we were convinced of the effect. In order to validate the results with a larger set of observers, the third author repeated the experiment, slightly modified, with 25 observers. The experiment differed only in the following respects: a different monitor was used under different (but consistent) lighting conditions, and the experimenter set the values for red and green to a predetermined value to improve consistency and reduce observer time. The pool of observers was primarily teenagers in the local home-schooling community and their parents. Twelve observers were female and thirteen were male. Table 3 summarizes the results.

Table 3: Results on 25 Observers (50 Observations)

	Average	Standard deviation
Luminance Matching	1.635	0.184
Grey Balancing	1.623	0.129

The difference in the standard deviations is substantially less striking, however an F test on the variances (whose ratio is 2.05) indicates that the difference is significant at the 99% confidence level.

LCD Experiment

Recall the earlier concern about the efficacy of the method in the case where chromaticity-constancy fails. To address this concern, the same visual tasks were performed on a laptop LCD found to violate chromaticity-constancy. The corrected electronic images obtained from both the luminance matching and the grey-balancing tasks were compared with calibrated prints viewed in a light booth. The general observations are:

- Consistency in observer responses in the grey-balance task is again superior to that in the luminance matching task
- The biggest differences in the images are seen near the neutral axis. The grey-balancing approach corrects input pixels with approximately equal R, G, B values to render with a chromaticity near that of the display white point. This is not the case with the luminance matching approach.
- In terms of overall quality, it was found that the grey-balance approach produces a closer match to the print than the luminance-matching approach in a few image regions. In no cases did the grey-balance approach produce a worse result.

Thus the proposed approach offers not only a significant advantage in consistency of results, but also a potential advantage in image quality for displays that do not conform to the chromaticity constancy assumption built into the standard CRT model.

Discussion

In addition to the basic approach described earlier, one can envision the following extensions and variants:

- Currently only the blue value is adjusted until a chromaticity match with the surrounding pattern is achieved. If this does not suffice (as might be the case for low cost LCDs), an additional control on e.g. the red channel may be necessary to achieve a satisfactory chromaticity match. This would then be used to estimate the gamma for both the blue and red channels.
- Especially for displays that violate chromaticity constancy, multiple levels could be matched. That is, in addition to matching the grey at 50%, lines combining that 50% grey with 100% white or black could be matched to grey patches in a subsequent step. This would provide values for a multi-parameter model, rather than the single-parameter gamma model.
- In addition to matching the chromaticity of the patch with that of the neutral surround, another control could be added to achieve a luminance match between the two stimuli. This could be implemented within the same or a separate panel.
- Instead of a slider-based adjustment, the user could select from a fixed set of patches the one closest to neutral (see Fig. 3). The patches would span the possible range of gamma values.

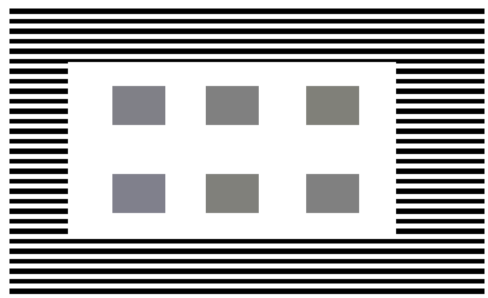

Figure 3. Possible display selection approach. The user selects the patch that appears most neutral with respect to the surround.

- A variant of the above approach is to select from a small set of patches the one closest to grey, and then a new set is presented with the selected patch at its centre, and a narrower range surrounding it. This is repeated until the desired level of precision is reached. For example, assuming monitors have gammas in the range 1.0 to 2.5, the first set might be produced with gamma values of 1.375, 1.75, 2.125. If the user selects 2.125, the next set would be 1.9375, 2.125, 2.3125. On each step, the set would represent a narrower range of gammas, until the desired precision is reached. The assumption in this approach is that if the user selects a given patch from a set of three equally spaced patches, then the “true” value is between the value for that patch plus half a space and the value minus half a space. This assumption can be relaxed by making the sets shrink more slowly.

Conclusions

It is relatively easy to obtain a visually based estimate for gamma for red and green using the typical approach of comparing an on-off halftone pattern to a mid-level continuous-tone patch. However, estimates of gamma for blue obtained in this way tend to have very poor precision. This is to be expected given the limited contribution of the blue channel to luminance judgments in the human observers. Human observers are very good at detecting small deviations in chromaticity from neutral, assuming a good neutral reference is available. By replacing a luminance-based judgment with a hue/chromaticity judgment we achieved a substantial increase in precision of visual estimates of gamma. This not only improves the precision of the value of a display calibration parameter; it also produces a significant improvement in the judged appearance of images displayed on a device calibrated using the visual approach.

References

* One example of such a task involves displaying a foreground pattern with varying grey levels against a perfect black (D=0) background, and seeking a visual match between foreground and background.

1. R. S. Berns, R. J. Motta, M. E. Gorzynski, "CRT Colorimetry. Part 1: Theory and practice", *Color Research & Appln*, Vol. 18(5), pp. 299-314, Oct. 1993.
2. W. B. Cowan, "An inexpensive scheme for calibration of a colour monitor in terms of CIE standard coordinates", *Comp. Graphics*, Vol. 17, No. 3, pp. 315-321.
3. J. Gille, J. Larimer, "Using the human eye to characterize displays", SPIE Vol. 4299, pp. 439-454, 2001.
4. J. Gille, L. Arend, J. Larimer, "Display characterization by eye: contrast ratio and discrimination throughout the grayscale", SPIE Vol. 5292, pp. 218-233, 2004.
5. G. J. Braun, "Visual display characterization using flicker photometry techniques", SPIE Vol. 5007, pp. 199-209, 2003.
6. M. Ueda et al., "Display characteristic determining device", U.S. Patent 5,923,315, 1999.
7. B. Wandell, *Foundations of Vision* Sinaur Associates, Sunderland MA, 1995.
8. B.W. Tansley, R.M. Boynton, "Chromatic border perception: the role of red- and green-sensitive cones", *Vision Research*, 18pp. 683-697, 1978
9. G. Sharma Ed., *Digital Color Imaging Handbook*, Chapter 1, pp. 28-40, CRC, 2003.
10. G. Marcu, K. Chen, "Gray tracking correction for TFT-LCDs", *Proc. IS&T/SID's 10th Color Imaging Conference*, pp. 272-276, 2002.

Author Biography

R. Victor Klassen received a B.Sc. in Physics and a Ph.D. in Computer Graphics from the University of Waterloo, then moved to Xerox, where he is now a Principal Scientist. His current interests include physical limitations of color measurement and reproduction, and limitations of the visual system to discriminate colours. A member of ACM SIGGRAPH, IEEE-CS, and IS&T, he has served on the program committees of SIGGRAPH, Graphics Interface, and RIDT..