17.3: Visual Calibration of Displays for Accurate Contrast Reproduction

Long To, Russell L Woods, Eli Peli
Schepens Eye Research Institute, Harvard Medical School, Boston, MA

Abstract
We developed a visual calibration method (free of instruments) to characterize color display response, and demonstrate its capability to produce wide dynamic range achromatic display accurately as needed for measuring contrast sensitivity. We performed calibration and verification on both CRT and LCD monitors, and report differences between them.

1. Introduction
Vision science experiments are conducted using on-screen stimuli, where the stimulus contrast often plays an important role in the experimental outcome. Similarly, medical and other critical display applications require a high level of precision in luminance reproduction. The ability to accurately control the contrast is often an important requirement [1], [2]

On a generic computer system the display nonlinearity and the limited bit-depth of the graphic card poses constraints to both the range and accuracy of displayable contrasts. While there are commercial systems specifically designed for vision science (e.g. Cambridge Research Systems, Rochester, Kent, UK), they are often expensive and have a proprietary software interface (where the calibration information might not always be accessible to other applications).

We developed a visual calibration method which comprises a series of psychophysical tasks to be completed by an operator with normal vision without using any measurement instruments. The method is applicable to most general-purpose displays. The calibration procedure results in a lookup table, indexed by relative luminance, which is used to set luminance levels for a gray-scale stimulus accurately and over a wide dynamic range. The calibration is designed to be general-purpose, and at the time of our implementation it was targeted at a computerized letter contrast sensitivity test. Such an application measures observer's contrast threshold by presenting letters of variable contrasts on a bright background, and hence it is very important to generate precise contrast particularly at low levels near the threshold. The standard 8-bit dynamic range is insufficient and it is even further reduced following linearization of the display. Thus, a method to expand the dynamic range is required.

The paper is organized in the following sections. Section 2 describes the visual method for gamma estimation. Section 3 shows how to obtain the ratio of luminances between primary colors (*RGB*), also by means of a psychophysical task. Section 4 describes generating the lookup table with bit-stealing to increase the displayable contrast resolution. Section 5 contains verification results, comparing the contrasts calculated from directly-measured luminance against the intended contrasts. Section 6 discusses some important issues that we noted when calibrating LCD displays.

2. Estimation of display nonlinearity
We define the relative luminance (R) domain as the operating range between the maximum and minimum utilized luminances of a monitor. On an 8-bit grayscale display, the maximum and minimum are usually achieved when all three *RGB* channels are set to 255 (white) and 0 (black), respectively. There were differences in accurately calibrating and displaying contrasts on a CRT compared to a similar display on an LCD (section 6).

On a conventional CRT display, the relationship between emitted luminance and input digital (voltage) value is monotonic but often nonlinear, which can be modeled and estimated by different methods [3]. We model this relationship as a power function of exponent γ (gamma) by the following approximation:

$$R(y) = (y/y_{max})^{\gamma}, \qquad (1)$$

where y is the 8-bit gray value of the bitmap on display, y_{max} is the maximum gray value (usually 255), $R(y)\in[0,1]$ is the corresponding relative luminance, and γ is the device-dependent exponent (typically between 1.8 and 2.2 on a CRT).

The selection of the above model over other forms of nonlinearity estimation [3] is motivated by the requirement to have the model traversing both the relative and physical luminance ranges, so that the results of a visual calibration can be compared and verified objectively with photometer measurements. We do not compare different gamma models in this paper.

Estimation of γ in equation (1) was performed as follows. We collected n sample pairs of (R_i, y_i), i=1…n by a series of pair-wise luminance matching tasks, where the observer found a gray level match for a known relative luminance. In this psychophysical procedure, the stimulus comprised two horizontally abutting squares, presented on a white background, as illustrated in Figure 1. One square patch (128-pixel height by 128-pixel width), which we call the reference, was constructed from alternating horizontal lines, of two known relative luminance values. From the observation distance during the calibration, the alternating lines were not visible and the reference patch therefore appeared to have a uniform luminance. The other square, the calibration patch, was uniformly set to a single y, and that y was controlled by the observer. The matching task required an observer to adjust the y of the calibration patch to match the apparent brightness of the reference patch. When the match was achieved, the border between the two patches became invisible and the two appeared to merge into a single rectangle. This task was easy to perform and had good repeatability (95% confidence limits of γ for eight observers were 1 to 3%).

For visual calibration of gamma, Colombo and Derrington [4] had observers match the luminance of a uniform patch to reference patches with varying spatial duty cycles (that could be seen by the observer) and Besujien [3] had observers match a uniform patch to a reference patches that temporally-flickered or comprised spatially-alternating-lines, at only one pairing of grey levels. We implemented the approach proposed by Peli [5], wherein a uniform patch was matched to a series of reference patches comprising alternating lines set initially to ($R_1 = 0, y_1 = 0$) and ($R_2 = 1, y_2 = 255$). Subsequent reference patches were created by recursive partitioning of the luminance range, described as follows:

ISSN/009-0966X/09/3901-0216-$1.00 © 2009 SID

1. Create a reference patch with luminance of $R_3 = ½$ by alternating lines of $y = y_1$ ($R_1 = 0$) and $y = y_2$ ($R_2 = 1$). The matching task is then performed to yield y_3 such that $R(y_3) = R_3 = ½$.
2. Create the next reference patch of $R_4 = ¼$ by alternating $y = y_1$ ($R_1 = 0$) and $y = y_3$ ($R_3 = ½$). Luminance matching of two square patches then yields y_4.
3. Similarly, create the next reference patch of $R_5 = ¾$ by alternating $y = y_2$ ($R_2 = 1$) and $y = y_3$ ($R_3 = ½$). Luminance matching then yields y_5.
4. Continue recursive partitioning until enough samples have been collected.

This task was programmed, so the observer used the keyboard to change the calibration luminance and to record matches. Figure 1 shows a typical stimulus presentation. The positions of the reference and calibration patches were interchanged randomly from trial to trial to avoid side bias and memory effects. Since the alternating lines were not visible at the viewing distance (about 1.5m), an arrow head indicated to the observer which of the two patches was under control.

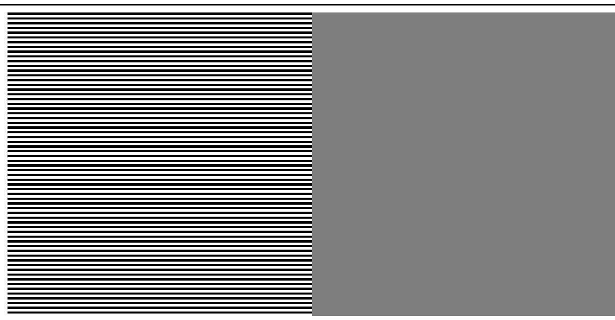


Figure 1. Abutting square patches for the gray level matching task. The reference patch (left) has alternating lines of two preset luminance values (if it looks different then that it is due to aliasing in the production of the image on the display used). The calibration patch (right) is solidly filled with a single gray level and its brightness was adjusted by the observer, until the perceived brightness was identical to the reference patch.

The γ was estimated from the collected data by minimizing the sum-of-squared-errors (SSE) in equation (2) using the Gauss-Newton optimization method [6]. The initial value of γ in the minimization was set at 1.8, which is typical for a CRT device.

$$\varepsilon(\gamma) = \Sigma \left((y_i / y_{max})^{\gamma} - R_i \right)^2 \qquad (2)$$

3. Increasing the dynamic range

For the achromatic stimuli we typically used, adhering to the grayscale (y) rule of $R=G=B$ yields 256 possible shades of gray on a 3×8-bit color display. When a dark letter is presented on a bright background, 255 contrast levels can be obtained this way. For testing the contrast threshold of a person with normal vision, almost all of these available contrasts levels are well above the threshold level, making them redundant in such a test. On the other hand, there would be few displayable contrasts within the vicinity of the true threshold (about 1%). This limited contrast resolution often results in an unacceptable quantization error.

To increase the number of displayable contrasts, especially near the threshold, it is necessary to produce more than 256 shades of gray. Using a generic video card and CRT display, such improvements in bit-depth can be achieved with an analog system of resistors [1]. This is not possible however with digitally driven LCD displays. We have applied bit stealing, a strategy proposed by Tyler [7]. By allowing a small variation amongst the RGB channels, for example setting $R = G = B + 1$, we gain extra luminance levels with minimum change in achromaticity of the stimulus (see equation (6) for step increments), expanding the number of available luminances by about ten-fold [7].

Previous implementations of bit-stealing have required photometric measurement of the luminance of each color as proposed by Tyler [7]. Since the gamma remains similar for each color, the luminance of each color channel is scaled by its maximum luminance. Thus, it is only necessary to determine the relative luminance of the three colors. These color ratios are device-specific and may change with monitor settings.

Color-ratio estimation

The color ratios could be estimated from a minimum-distinct-border task (as used for gamma estimation) or a heterochromatic-flicker task [7]. While relatively easy to implement, we found that even experienced observers had difficulty making the judgments for those tasks. Instead, we adapted a method of equi-luminance calibration based on a motion illusion [8] to estimate the color ratios. Each stimulus was a looping playback of four frames, arranged in the order shown in Figure 2, presented at a temporal rate of 4 frames per second. The temporal arrangement creates a motion illusion of the vertical bars appearing as moving either to the left or right. A green bar brighter than the red bar would cause the green column at frame 1 to appear to “move” to the brighter yellow column on frame 2, then onto the green column at frame 3. This creates the illusion of the grating moving to the right. Likewise, a darker green bar induces an apparent leftward motion. For our color-ratio estimation, the bright red bar was fixed at 240 units (in a 8-bit scale), and the green bar was adjusted according to observer responses. This setup is slightly different from Anstis and Cavanagh [8], where the green bar luminance remained constant, as green is usually brighter than red at the same input level, keeping green fixed at a high level may set its luminance outside the displayable range for the red channel, whereas any red value would be well within the luminance range of the green channel. The same logic was applied for the matching luminance across red and blue channels.

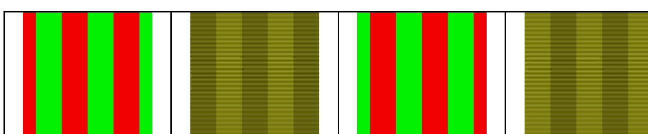


Figure 2. Individual frames for the Green-Red equi-luminance matching task. In frames 1 and 3, the Red bar remained constant at 240 and the green bar was adjusted according to the observer response. In frames 2 and 4, the brighter yellow bars comprised alternating red and green lines, set at 17/16 their value in the red/green frame. Respective lines in the darker bars were set at 15/16 their value in the red/green frame.

The transition between right and left movement occurred at the level of green that resulted in equal apparent brightness of the red and green bars. That point of equi-luminance was found using a 2-alternative, forced-choice staircase procedure with adaptive step sizes. Two interleaving staircases were run with the fixed color (red) channel set to 240 units. For the variable (green) channel, the upper staircase gray-level started at 240 units and the lower staircase started at 64 units. At each presentation, the observer reported whether the vertical bars appeared to be drifting left or right (with the keyboard). The variable color channel was decreased or increased depending on the reported motion direction. The red-blue color ratio was determined in the same way, with the blue channel fixed at 240 and the red (variable) channel had starting values of 240 and 128 units.

The magnitude of the step was controlled by an adaptive procedure, which started at 16 units, and was reduced as the test progressed. The step size was reduced by half when the responses differed on two consecutive stimuli. The lower limit for the step size was set at 4 units. The algorithm terminated when there were at least three reversals at the smallest step size in each staircase. The equi-luminance value of the variable channel was calculated as the average of the last six reversals across two staircases.

4. Luminance lookup table with bit-stealing

A luminance lookup table (LUT) is an index of relative luminance levels and corresponding input *RGB* values. To display a given contrast, an application sends a query to the table to retrieve the RGB values for the relative luminance required to achieve this contrast. For our application, we assume that all presentations are made on a fixed background of higher luminance (L_B), and the foreground luminance (L_F) is directly deduced using the following relationship:

$$\text{Contrast} = 1 - (L_F / L_B) \quad (3)$$

We created the LUT using parameter γ and color ratios, estimated by the above psychophysical tasks. Each entry in the table has four elements, R, y_R, y_G and y_B, with R being the relative luminance from 0 to 1, and the other three being *RGB* values representing this luminance. To facilitate the table description, the entries are grouped into two categories. Level-1 entries are dependent only on γ, and level-2 are bit-stealing entries, which were derived using both γ and color ratios. It is not necessary to make this distinction between entries in a normal usage of the table.

Level-1 entries have the following format:

$$R = (y / y_{max})^{\gamma}, \; y_R = y_G = y_B = y, \quad (4)$$

where y is an 8-bit integer for each gray level. As seen in the formula, level-1 entries are derived directly from γ. Level-2 entries, on the other hand, correspond to intermediate luminances between consecutive entries from level 1, and are formulated as follows

$$R = (y / y_{max})^{\gamma} + \Delta_R, \quad y_R = y + \delta_R, \; y_G = y + \delta_G, \; y_B = y + \delta_B, \quad (5)$$

where y is the 8-bit gray level,

$0 < \Delta_R < [\, ((y+1)/ y_{max})^{\gamma} - (y / y_{max})^{\gamma}]$. The bit-stealing *RGB* entries are generated with the assumption that green emits the brightest luminance, red the second and blue the least (as found on all displays tested).

$$(\delta_R, \delta_G, \delta_B) \in \{ (0,0,1), (1,0,0), (1,0,1), (1,0,2), (2,0,1), (2,0,2), (0,1,0), (0,1,1), (1,1,0) \} \quad (6)$$

subject to

$$(\delta_R p_R + \delta_G p_G + \delta_B p_B) < 1 \quad (7)$$

and

$$(y + \max(\delta_R, \delta_G, \delta_B)) \le y_{max} \quad (8)$$

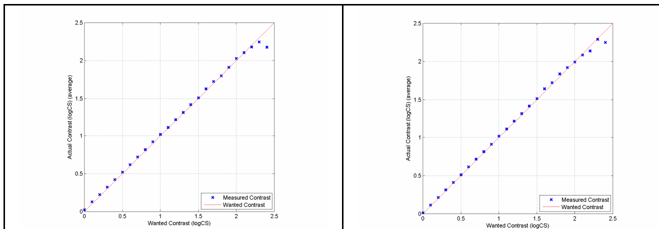
where p_R, p_G and p_B are the relative luminance contributions of each primary color in an achromatic patch, all directly calculated from the color ratios. In addition, the relative luminance increment Δ_R is calculated as:

$$\Delta_R = (\delta_R p_R + \delta_G p_G + \delta_B p_B)(((y+1)/ y_{max})^{\gamma} - (y / y_{max})^{\gamma}) \quad (9)$$

The total number of entries in the lookup table may vary as it depends on the estimated color ratios. While the number of level-1 entries are fixed, the level-2 entries are subject to the constraints in (7) and (8), which may exclude some value sets (δ_R, δ_G, δ_B). These constraints are imposed to ensure that any referenced luminance corresponds to a unique and displayable set of RGB values. The levels shown in (6) were chosen to give approximately equal increments between the level-1 luminances, based on the relative R: G: B luminances of about 0.3: 0.6: 0.1, which was typical of the monitors that we measured.

5. Verification of displayed contrasts

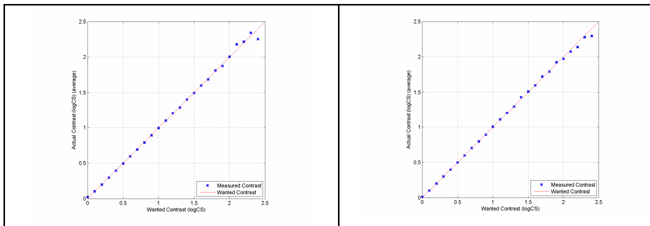
As described in section 4, the lookup table contains *RGB* correspondence for an array of luminance values. We performed a validation test by using the lookup table obtained with the visual calibration to generate a series of stimuli at different contrast levels, and compared the results to the contrast found with a photometer (Minolta LS-100, Japan) and equation (3). Measurements were made in logarithmic increments of 0.1 for stimuli of nominal log(Contrasts) ranging from 0 (100% contrast) to 2.4 (0.4% contrast). Ten photometric measurements were taken of luminance at each gray level for the corresponding foreground and background. We then calculate the average at each measurement point and substituted the parameters into equation (3) to compute the measured contrast. The results are presented in Figure 3, which demonstrates, for both CRT and LCD displays, that the measured contrasts were almost identical to its lookup-table value up to a about log(Contrast) of 2.3 (0.4% contrast)



CRT LCD

Figure 3. Comparison of contrasts achieved using the visual calibration method against its nominal values.

To further demonstrate the quality of the visually-calibrated lookup table, we performed a separate experiment where the lookup table was generated using photometer measurements of both gamma and color ratios. We then ran the same validation test on the photometer-generated lookup table, with the results in Figure 4. This demonstrates that the quality of the visually-calibrated lookup table was as good as the photometer-based procedure.



CRT LCD

Figure 4. Comparison of contrasts achieved using a photometer-based calibration method against its nominal values.

6. Notes about calibrating a LCD display

CRTs have been used widely in vision-science experiments, and their characteristics are well known. However, as the newer LCD technology becomes more popular for general applications, CRTs are also becoming difficult to replace. We noted some differences between LCD and CRT displays that affected the visual calibration. Such differences have not often been described.

(A) The luminance output of a LCD has likely been *corrected* internally to resemble the native performance of a CRT. This gamma correction is provided in the majority of modern LCDs and can usually be specified from a display menu. We discovered that, at least on some LCD displays, there is an *uncorrected* gamma setting and the accuracy of the contrasts produced from the lookup table is worse when the specified gamma differs from the *uncorrected* gamma of the LCD. When possible, the native gamma of the LCD is estimated by setting the gamma to *uncorrected*, measuring the gamma in this *uncorrected* gamma state, then applying the measured gamma and then conducting the calibration.

(B) Luminance calibration is best when there is no color correction (e.g. no “color temperature”) or other software control of RGB (e.g. “auto brightness”, “economy mode”).

(C) The relationship between brightness and physical device state of a LCD is the reverse of that for a CRT. For a CRT, the maximum brightness is achieved with maximum DAC voltage. For a LCD, the maximum brightness is achieved when all the liquid crystals are turned off (“random” alignment), as this transmits the most backlight. The transition from this “off” state (y = 255) to an active state (y = 254) reduces the luminance output. In the LCD displays that we examined, the change in luminance when this small voltage is applied results in a larger reduction in luminance than the next increment (i.e. y = 254 to 253). This produces a step in the gamma curve (i.e. not smooth). Consequently, when the gamma curve is fit, it fails to fit well at high RGB values, due to the step. This causes a substantial problem when very low-contrast stimuli are generated on a maximum-brightness background (our application). The problem was overcome by simply setting y_{max} = 254 (i.e. not using y = 255), resulting in a significant improvement in the validation test.

(D) Any display device may suffer from saturation, which may be caused by settings of the device parameters (e.g. certain color temperature settings), or due to a manufacturing defect. Saturation makes it difficult to characterize the monitor and to display contrast levels properly, as the assumption of a smoothly increasing gamma function is not met. Saturation may occur in only one color channel. We developed a set of test screens to detect saturation. An operator can use these test screens to determine quickly if the display is subject to saturation, and try to correct it before calibration.

7. Conclusion

Visual calibration of both CRT and LCD displays to obtain a wide dynamic range of achromatic gray levels is practical and simple to implement and apply. With the increased use of LCD displays and the increasing sophistication of their electronic driving schemes, much has to be learned to safely apply these displays (and other emerging display technologies) in critical applications such as visual science and medical displays. This has been a first step in our attempt to investigate such use and within the range of our application domain we believe that we have found it to be achievable.

8. Acknowledgments

Jeff Mulligan suggested the Anstis and Cavanagh [8] motion illusion as an easier task to estimate the color ratios. Supported in part by NIH grants EY12890 and EY05957 and a grant from Johnson & Johnson Vision Care.

9. References

[1] D.G. Pelli and L. Zhang, "Accurate control of contrast on microcomputer displays," *Vision Research*, 31, 1337-1350 (1991).

[2] A.B. Watson, K.R.K. Nielsen, A. Poirson, A. Fitzhugh, A. Bilson, K. Nguyen and A.J. Ahumada, Jr., "Use of a raster framebuffer in vision research," *Behavior Research Methods, Instruments, and Computers*, 18/6, 587-594 (1986).

[3] J. Besuijen, "Visual gamma measurement and methods to compare gamma models," *Journal of the SID*, 15/8, 611-623 (2007).

[4] E. Colombo and A. Derrington, "Visual calibration of CRT monitors," *Displays*, 22, 87-95 (2001).

[5] E. Peli, "Display nonlinearity in digital image processing for visual communications," *Optical Engineering*, 31/11, 2374-2382 (1992).

[6] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1992)

[7] C.W. Tyler, "Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis," *Spatial Vision*, 10/4, 369-377 (1997).

[8] S.M. Anstis and P. Cavanagh, A minimum motion technique for judging equiluminance, in J.D. Mollon and L.T. Sharpe, Editors *Colour vision: Psychophysics and physiology.*, 66-77 (Academic Press, London, 1983)