Vision Research 90 (2013) 15–24

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Psychophysical contrast calibration ☆

Long To, Russell L. Woods, Robert B. Goldstein, Eli Peli *

Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Article history:
Received 11 September 2012
Received in revised form 13 February 2013
12 April 2013
Available online 30 April 2013

Keywords:
LCD
CRT
Luminance
Linearization
Display calibration
Contrast

ABSTRACT

Electronic displays and computer systems offer numerous advantages for clinical vision testing. Laboratory and clinical measurements of various functions and in particular of (letter) contrast sensitivity require accurately calibrated display contrast. In the laboratory this is achieved using expensive light meters. We developed and evaluated a novel method that uses only psychophysical responses of a person with normal vision to calibrate the luminance contrast of displays for experimental and clinical applications. Our method combines psychophysical techniques (1) for detection (and thus elimination or reduction) of display saturating non-linearities; (2) for luminance (gamma function) estimation and linearization without use of a photometer; and (3) to measure without a photometer the luminance ratios of the display's three color channels that are used in a bit-stealing procedure to expand the luminance resolution of the display. Using a photometer we verified that the calibration achieved with this procedure is accurate for both LCD and CRT displays enabling testing of letter contrast sensitivity to 0.5%. Our visual calibration procedure enables clinical, internet and home implementation and calibration verification of electronic contrast testing.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Visual psychophysical laboratory studies are usually conducted using electronic displays. In the clinic, electronic displays have been replacing the paper wall chart and optical projector tests of visual acuity and contrast sensitivity (CS) measurements starting with the 1980s introduction of the B-VAT system (Mentor O&O, Norwood, MA) (Williams et al., 1980). Electronic clinical test systems are in widespread use today (e.g. TestChart 2000 (Thomson Software Solutions, UK), Metrovision (Metrovision, France), Smart-System20/20 (M&S Technologies, Skokie, IL) and CST1800 (Stereo Optical Co, Chicago, IL)). Following the development of the basic electronic visual acuity chart many other clinical tests were incorporated into these systems including letter and grating CS in the B-VAT II-SG (Corwin, Carlson, & Berger, 1989) followed by a battery of binocular vision tests (Waltuck, McKnight, & Peli, 1991) that included distance stereoacuity testing (Rutstein & Corliss, 2000; Wong, Woods, & Peli, 2002). Many personal-computer based clinical vision test systems are now marketed either as integrated systems or as software packages to be used with existing computers and displays. In addition to the use in clinics, there has been a growing trend for remote visual testing using home computers (Dagnelie et al., 2003, 2008), smart phones, tablets (Dorr et al., submitted for publication), and over the Internet (Dagnelie, Zorge, & McDonald, 2000; Lavin, Silverstein, & Zhang, 1999). In-home testing has potential benefits in reducing costs, increasing convenience, recruitment of subjects for studies, monitoring of patients, and the ability to collect data frequently. However, home testing presents more challenges to standardization, display characterization and calibration.

The growing popularity of clinical letter CS testing using paper charts (e.g., Pelli–Robson chart (Pelli, Robson, & Wilkins, 1988), Reagan chart (Regan, 1988), and the Mars charts (Arditi, 2005; Dougherty, Flom, & Bullimore, 2005)) lead to the incorporation of letter CS testing in most clinic electronic vision test systems. While testing of visual acuity, stereo-acuity and other binocular functions is not very sensitive to chart or display luminance calibration, the testing of (letter) CS requires accurate luminance calibration of the display and, in most cases, higher luminance resolution than available with typical 8-bit displays and graphic cards. The enhanced luminance resolution is required to enable presentation of contrast levels near and below the human threshold for detection. A luminance calibration system with enhanced luminance resolution was provided with the early B-VAT II-SG that measured both letter CS and detection thresholds of sinusoidal gratings using only 6 bits of native luminance resolution. That system required a manual adjustment of display "brightness" to specific luminance values as measured with a photometer, as well as a flicker minimization (visual psychophysical) method to match the mean luminance of

---

☆ Previously presented: To L, Woods RL, Peli E. Visual calibration of displays for accurate contrast reproduction. SID International Symposium, May 31–June 5, 2009.

* Corresponding author. Address: Schepens Eye Research Institute, 20 Staniford Street, Boston, MA 02114, USA. Fax: +1 617 912 0112.

E-mail address: eli_peli@meei.harvard.edu (E. Peli).

0042-6989/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2013.04.011

gratings in the two hardware-modified domains of the expanded dynamic range. The difficulty associated with such calibration is further exemplified by the contemporary TestChart 2000 that recommends a proprietary light meter for calibration that can be either bought or rented from the manufacturer. A number of commercially available lab systems, such as the Cambridge Research Systems ViSaGe (Cambridge Research Systems Ltd., UK), come equipped with a photometer to facilitate a system calibration. Thayaparan, Crossland, and Rubin (2007) compared the TestChart 2000 to the Pelli–Robson and Mars charts and found that the coefficient of repeatability was 0.18 for the Pelli–Robson chart, 0.12 for the Mars chart, but only 0.24 log units for TestChart 2000. In addition, they found that the TestChart 2000 did not agree well with the Pelli–Robson chart which they attributed to the performance of LCD monitors at low contrast levels. They did not make any explicit statements as to which of these was the most accurate.

Most psychophysical studies involving electronic displays and manipulation of electronic images require accurate calibration of the display so that the luminance characteristics of the displayed images are known. Usually this is done by linearizing the relationship between the digital pixel representation and the luminance of the display (Brainard, 1989; Brainard, Pelli, & Robson, 2002). Historically, such studies were conducted using CRT displays and accurate and expanded luminance resolution was possible by combining the three color outputs of the graphic cards through a resistors net (video attenuator) to expand the luminance resolution of monochrome CRTs (Dakin et al., 2011; Falkenberg, Rubin, & Bex, 2007; Li et al., 2003; Niebergall, Huang, & Martinez-Trujillo, 2010; Pelli & Zhang, 1991; Watson et al., 1986). Calibration and linearization of such systems requires photometric measurement of the display voltage to luminance relations (the gamma function) followed by photometric verification of the successful calibration (Swift, Panish, & Hippensteel, 1997).

A linear luminance to digital image relationship is also required for many studies that can be safely conducted within the limited 8-bit display range (Haun, Woods, & Peli, 2012; Vera-Díaz, Woods, & Peli, 2010; Webster, Georgeson, & Webster, 2002). The same is true for most studies of image processing and image quality. If calibrations are not performed the impact of the display's non-linear voltage (pixel-level) to luminance gamma function may drastically affect the content of the displayed images (Peli, 1992a).

The quantization of luminance levels in electronic displays is particularly problematic at low luminance levels, where a change from one pixel value to the next pixel value produces a change in luminance that is a large fraction of the prior luminance. Thus, producing fine gradations of low contrasts on dark backgrounds is difficult or impossible (this limitation affects printed charts similarly). Therefore, paper charts and computer-based contrast sensitivity tests use gray letters on bright backgrounds. Note that the need to linearize the display may result in reduction of the dynamic range, as most linearization methods result in fewer available gray levels thus reducing the available dynamic range and reducing the luminance resolution below the original 8-bit depth. The resulting limited luminance resolution (about 6 bits) is insufficient to challenge human contrast sensitivity even at the bright end of the luminance range. The contrast generated with pixel values of 254 and 255 as the low and high luminances is easily detected by a normally-sighted observer, as the accelerating gamma function produces a ratio between these luminances that is higher than the pixel-value ratio suggests. The problem is even worse when we attempt to generate sinusoidal or Gabor patches since one has to operate near the middle of the display luminance range where every gray level step represents a higher fraction of the mean luminance or a larger change in contrast and where it may be necessary to generate a sinusoidal variation near this luminance over a spatial extent of at least 6 pixels (Pelli & Zhang, 1991; Woods, Nugent, & Peli, 2002).

CRT displays are rapidly disappearing from the consumer markets and are being replaced by LCD monitors. LCDs have the advantages of higher luminance, a larger color gamut (Sharma, 2002), and larger screen sizes. Offsetting these advantages are the disadvantages of more complex luminance response functions that may result in larger calibration errors (Sharma, 2002), the inability to use voltage-based luminance resolution expanders and strong sensitivity to viewing angle. If electronic displays are to be used clinically it is now necessary to be able to calibrate LCD screens.

We present a psychophysical display calibration procedure that enables (1) detection and elimination of display saturating non-linearity; (2) luminance calibration (linearization); and (3) measurements of luminance ratios of the three color channels (used in the color bit-stealing technique for luminance resolution expansion (Tyler, 1997a)), all *without use of a photometer*. This calibration approach can facilitate letter CS and other testing in the clinic, over the internet and at home.

2. Display saturating non-linearity detection and elimination

Electronic displays frequently have a saturating non-linearity at the bright end of the luminance range or a cut-off at the dark end. In a display with a saturating non-linearity, the luminance curve levels off prior to the digital input reaching the minimal or maximal RGB values. This saturating non-linearity reduces the number of unique grayscale shades displayable and further complicates the calibration process. This is particularly true in calibration procedures that fit a gamma function. The region of saturating non-linearity (high luminance) occurs where we most often test the limits of the contrast sensitivity of the visual system. A saturating non-linearity may occur in individual color channels (Fig. 1A). Though the calibration method in Colombo and Derrington (2001) accounted for saturating non-linearity, it did not include a procedure to detect whether saturating non-linearity occurred or a method to reduce or eliminate it. It is preferable to ensure that the display is not saturated before initiating a calibration process, as the saturation also limits the available dynamic range.

We used the pattern shown in Fig. 1B to visually detect saturating non-linearity at maximum luminance. The background consisted of four rectangular regions (gray and individual primary colors), each near its maximum level. Each bar had 8 square patches, arranged in decreasing order of luminance.[1] If all 8 patches in each bar were visible, there was no saturating non-linearity and the procedure continued to the next step. If any of the brighter patches were invisible, the observer adjusted the physical or software settings on the display, including brightness, contrast, and color profile until the patches with lowest-contrast/brightest-luminance (right most) became just visible. This procedure simultaneously ensured that there was no saturating non-linearity in any of the color channels.

The same procedure was repeated for low luminances, to control for cut-off, using a similar stimulus prepared for that range. At that end, the dimmest square patches would be indiscriminable if there was cut-off. At the end of the process, all test patches had to be visible simultaneously at both the high and low end luminances of the display. The display settings that achieve that were then locked (if such locking was provided by the display) and recorded for future experiments. The cutoff at the low end is often

---

[1] In a pilot experiment, we determined the best increments (on our displays) for the saturation test bars as follows: For the bright background: for grayscale, pixel-value increment = 2 (e.g. the squares were 253, 251, 249, etc.). For the color patches, the increments were green = 3, red = 4, blue = 5. For the dark background: grayscale increment = 3, all colors increment = 5.

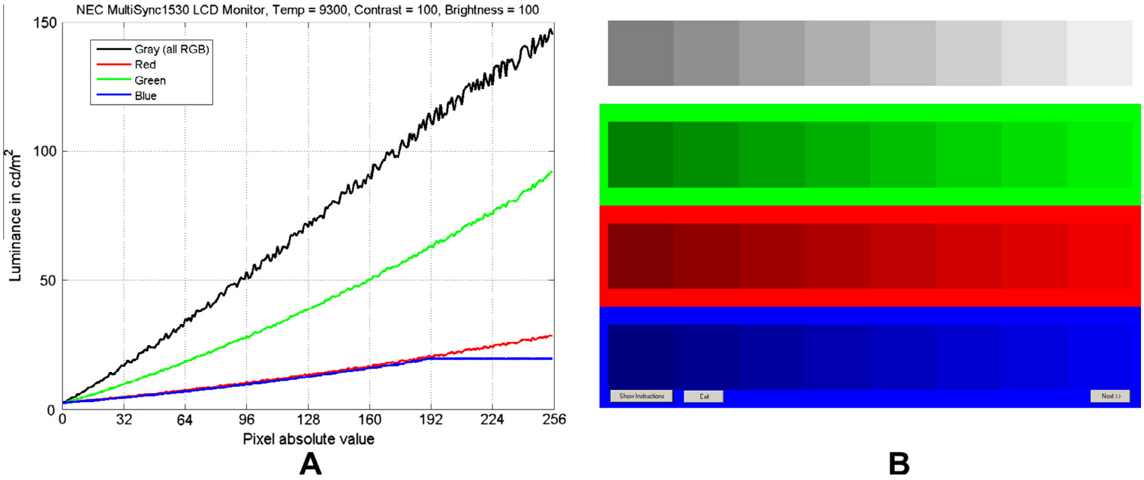


Fig. 1. (A) Luminance output of a LCD monitor where only one channel (blue) was saturated. The grayscale luminance (black) appears to be most "noisy" in the region of saturated-blue, but did not saturate itself. The data consists of one measurement at each pixel value for each color. The noise in the gray signal is photometer measurement noise and is the reason that we programmed the photometer-based procedure to take 10 samples at each RGB level. This figure is meant to illustrate saturation non-linearity and these data were not used to estimate gamma. (B) Pattern for detecting and removing saturating non-linearity at high pixel values. Square patches of decreasing luminance against the bright background to detect saturation in grayscale and individual color channels. To remove saturating non-linearity, an observer adjusted the manual controls of the display device until all eight patches in each zone were visible, and the rightmost patch was just visible against the background. A similar stimulus was used for cut-off testing at low pixel values.

only visible in gamma measurement curves if plotted on a log-luminance scale (unlike Fig. 1).

3. Luminance linearization

3.1. Contrast in the relative luminance domain

For onscreen presentation of an achromatic stimulus such as a letter, where the background luminance L_{bg} is higher than the letter (foreground) luminance L_{fg}, the contrast may be calculated by the Weber contrast:

$$C = \frac{L_{bg} - L_{fg}}{L_{bg}} = 1 - \frac{L_{fg}}{L_{bg}}. \tag{1}$$

Thus, the contrast is calculated from the ratio of the foreground to background. To reproduce any contrast on a given display, it is possible to characterize that display from luminances that are known relatively (i.e., proportionally) to one another. As also noted by Mulligan (2009), our visual calibration is possible since knowledge of absolute luminance (e.g. cd/m^2) is not required to reproduce a given contrast level. This works very well for most situations, but as described in Section 3.4, it does not work as well for low luminance backgrounds.

3.2. Visual estimation of display Gamma function

A gamma (γ) power model is often used to characterize the relationship between the RGB input levels and the luminance of a CRT display (Pelli & Zhang, 1991; Watson et al., 1986). Typically the light output of the display is measured with a photometer at different input levels, and then the data is fit to the model to obtain the gamma function parameter(s). The function is then inverted to provide the calibration needed to linearize the display luminance.

Besides photometer-based approaches, visual methods to estimate a gamma curve have been proposed that generally have asked the observer to equalize two luminance patches (Colombo & Derrington, 2001; Kay & Brandenberg, 2007; Peli, 1992a) or by nulling apparent motion (Mulligan, 2009). Colombo and Derrington (2001) tested both side-by-side and flicker minimization settings, but found the side-by-side configuration to be easier and quicker for subjects to complete. The Kay and Brandenberg (2007) solution was implemented in a software product (SuperCal, http://www.bergdesign.com) for Macintosh computers. Another company, Applied Vision Research and Consulting (Yang, 2013), developed an online calibrator, DisplayCal, which provides a rough estimate of the gamma value using a visual matching method.

On a CRT display, the native relationship between emitted luminance and input digital value (voltage) is monotonic but nonlinear. This nonlinearity may be approximated by a power function of exponent γ. We model the output relative luminance, $R(y)$ as follows:

$$R(y) = \left(\frac{y}{y_{\max}}\right)^{\gamma} (R_{\max} - R_{\min}) + R_{\min}, \tag{2}$$

where y is the 8-bit gray pixel value of the bitmap on the display, $y_{\max}$ is the maximum gray value used, γ is the display-dependent exponent, and $R_{\min}$ and $R_{\max}$ are the minimum and maximum luminance values (*following saturation correction*). In a relative luminance space, where $R_{\min} = 0$ and $R_{\max} = 1$, this becomes

$$R(y) = \left(\frac{y}{y_{\max}}\right)^{\gamma}. \tag{3}$$

This model can easily be adapted for estimation for both physical and relative luminance. Although in this paper we do not compare different gamma models, a recent review of other gamma models can be found in Besuijen (2007).

The model in Eq. (3) is characterized by γ that can be estimated as follows (Peli, 1992a). We collected n sample pairs of (y_i, R_i), $i = 1, \ldots, n$ by a series of pair-wise luminance matching tasks, when the observer was asked to match the gray level of a known relative luminance. The stimulus comprised two horizontally abutting squares (Fig. 2). The square patches were presented on a white background to maintain a display environment similar to a letter CS test, our test environment of interest. One 128-pixel square reference patch (3.4 cm on one display) was constructed from alternating horizontal lines, of two known (preset) relative luminance values. The observer was sufficiently far away from the screen that the alternating lines pattern was not visible and the reference patch therefore appeared to have blended into a uniform luminance. We did not use a checkerboard pattern because horizontal

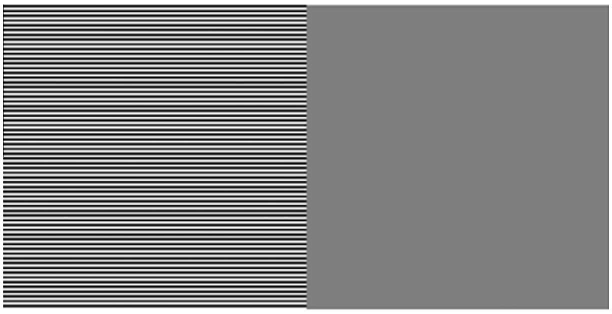


Fig. 2. Abutting square patches for the gray level matching task. The reference patch (left) has alternating lines of two preset luminance values. The calibration patch (right, in this example) was solidly filled with a single gray level and its brightness was adjusted by the observer, until the perceived brightness was as close as possible to the reference patch. Note that printed or displayed versions of this figure may be distorted due to sub-sampling of the alternating lines of the reference patch.

lines reduce inter-pixel independence on a raster-scanning CRT (Colombo & Derrington, 2001). In addition, using single lines allows the calibration to be conducted at a shorter distance. The other square, the calibration patch, was set uniformly to a single gray value, and the observer adjusted its luminance to visually match the reference patch. When a match is achieved the border between the two patches may no longer be visible and the two squares may appear to merge. At that point the calibration patch luminance is exactly half way between the luminances of the two levels represented by the alternating lines of the reference patch. The procedure for recursively generating the luminance matching patches is given in Step 2 of the online supplementary materials.

Gamma was estimated by minimizing the sum-of-squared-errors (SSE) in Eq. (4) using an optimization method, such as Gauss–Newton (Press et al., 1992).

$$\varepsilon(\gamma) = \sum((y_i/y_{\max})^{\gamma} - R_i)^2, \tag{4}$$

where (y_i,R_i) are pairs of matching pixel gray level and relative luminance levels obtained through the visual task.

3.3. Results of luminance estimation

To verify the results of our psychophysical method, we performed photometer-based (Minolta LS-100, Tokyo, Japan) calibration of a ViewSonic G810 CRT. Pairs of (y_i,L_i) were collected at 18 gray levels on a white background, where for each gray level $0 \leqslant y_i \leqslant y_{\max}$, L_i was the corresponding luminance (cd/m^2). Luminance levels were measured at the center of the screen using a square patch of the same size used in the psychophysical measurement.

Our psychophysical method used 7 matches. The photometer samples were taken in 15-gray-level intervals between 0 and $y_{\max}$ (18 samples). As seen in Fig. 3, both methods produced very similar gamma curves, the difference between the γ values was about 0.1%. The main difference between the curves is a non-zero minimum luminance on the photometric data. See Section 3.4 for a discussion of the effects of non-zero minimum luminance on contrast.

Three experienced observers and four initially naïve observers repeated the gamma estimation on an LCD monitor 10 times each (except one observer who did 6) over a period spanning 3 months. We analyzed the relative gamma, the psychophysically-estimated gamma divided by the gamma obtained with a photometer. There were no differences between subjects in relative gamma (ANOVA, $F_{6,57} = 0.16$, $p = 0.99$) or variability (Levene, $F_{6,58} = 0.60$, $p = 0.73$).

3.4. Effects of non-zero minimum luminance on contrast

As described above, the psychophysical method to estimate gamma uses a relative luminance range between 0.0 and 1.0. This definition of the relative luminance implies zero luminance for a black screen (when R = G = B = 0). In practice, because of reflected ambient light even in a dark room, backlight leakage (for LCD), and phosphor persistence (CRT), there is a positive luminance even when test pixels are set to zero (known as "black level"). Black levels are much lower with plasma, DLP and, particularly, OLED displays. In our experiments, we measured black levels of about 3–5 cd/m^2 when the displays were at such state. This "residual" luminance results in a difference between the contrast calculated from a relative luminance model, as applied in our method, and the contrast calculated from a model accounting for the absolute minimum luminance. For dark (foreground) on light (background) stimuli (as in a Pelli–Robson chart), the error in log-contrast is a function of the minimum and maximum luminances and the background luminance. For example, if the display's luminance range is from 5 to 100 cd/m^2 (as for our CRT), and the background is 100 cd/m^2, the error will be about 0.02 log units, while if the background luminance is 25 cd/m^2, the error will be about 1.0 log units. If the minimum luminance is 2 cd/m^2, those errors would be about 0.01 and 0.04 log units respectively, and, if the maximum luminance is 200 cd/m^2 (as for our LCD), those errors would be about 0.004 and 0.02 log units respectively. As can be seen in Section 6.2, for a bright background (near maximum luminance), those errors are negligible, being smaller than the measurement noise in those validations. These calculations also hold for the Michelson contrast definition. It is possible to reduce or eliminate these errors if the ratio of the minimum luminance to the luminance range is known or estimated. We did not implement this correction, as the errors were sufficiently small to ignore in our applications.

4. Color matching and bit-stealing for luminance resolution expansion

For a letter displayed on a white background of an 8 bit display with R = G = B, there are few possible displayable contrasts near the visible contrast threshold. Software based techniques to increase the luminance resolution include: spatial dithering – halftoning (Mulligan & Ahumada, 1992; Pappas & Neuhoff, 1992; Peli, 1992b; Ulichney, 1988), temporal dithering (Dorr et al., submitted for publication; Mulligan, 1993) and color dithering (bit-stealing: Tyler, 1997b). Because halftoning trades resolution for gray-scale and temporal dithering may result in visible speckling, we chose to implement bit-stealing, where a small, usually subthreshold, difference in hue is the only cost of the expansion.

Bit-stealing uses unequal levels of R, G, B to produce pseudo-gray luminance values that are inserted between the 256 values of luminance available with R = G = B. To compute the intermediate luminance, one needs to obtain the relative luminances of the primary colors. The ratio of the relative luminance were used to calculate (δ_R, δ_G, δ_B), which are combinations of increments of 0, 1, or 2 of each color gray level to be added to the three channels to alter the luminance. A more complete treatment is given in Step 4 of the online supplementary "How-To" guide. The luminance ratios of color pairs are device-specific, may also change with different display settings, and may vary between observers under some circumstances. Tyler suggested that such a ratio can be measured psychophysically using either a flicker test between pairs of colors, or a minimum distinct border match between adjacent color patches. We found with both approaches, that it was difficult even for an experienced observer to make the required judgments.

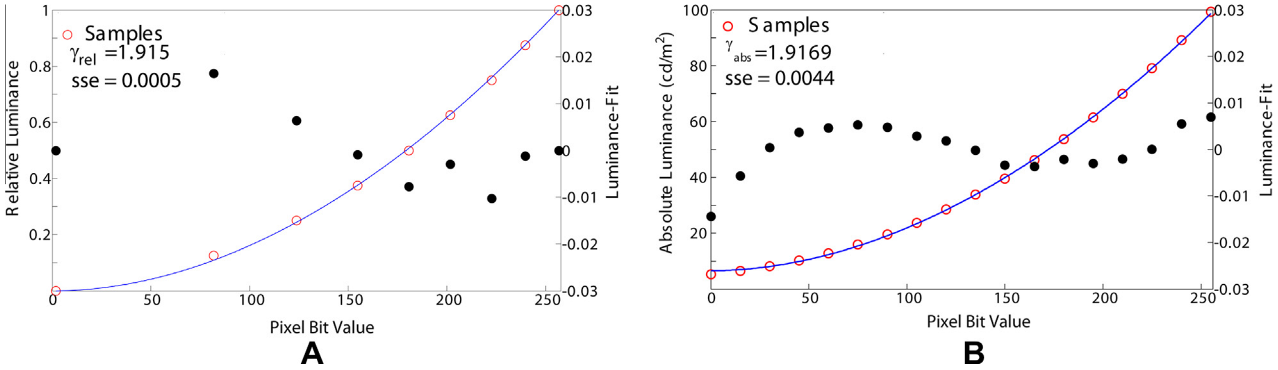


Fig. 3. Comparison between CRT gamma values estimated using the psychophysical method (A) and from a photometer-based measurement (B) (left axis scale). Residuals are shown as filled black circles and relate to the right axis scale. Note the non-zero luminance measured with photometer at gray value = 0.

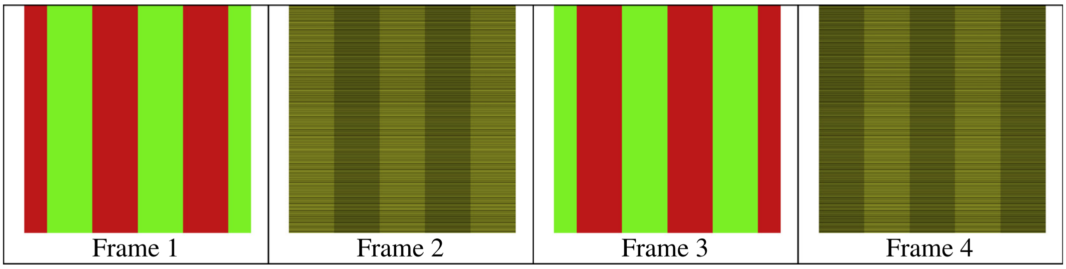


Fig. 4. The four-frame sequence used in the green/red equi-luminance matching task. In frames 1 and 3, the red bar remains constant and the green bar is adjusted according to the observer response. The sequence shown here with the green brighter than red will result in image motion to the left. Note that all bars in this figure are uniform (i.e. one color). On some displays and printers, the 2nd and 4th frames may show sampling/aliasing artifacts.

Therefore, we implemented an approach that we found to be easy for untrained observers.

4.1. Color luminance ratios measurement

To estimate the luminance ratios we implemented, at the suggestion of Jeff Mulligan (Personal communication, 2007), a motion illusion procedure (Anstis & Cavanagh, 1983; Mulligan, 2009). This technique has been used in several diverse studies including testing luminance contrast with IOLs (Pierre et al., 2007), where they used the method of adjustment until flicker, rather than motion, was perceived. We had tried this method, but found it difficult and thus switched to a forced-choice staircase. We modified the Anstis and Cavanagh technique slightly.[2] A sequence of four frames, arranged in the temporal order shown in Fig, 4, was played in a loop with a temporal rate of 5 frames per second. In frames 1 and 3 red and green bars alternated and in frames 2 and 4 bright and dark yellow bars alternated.

The sequence of frames creates a motion illusion of the vertical bars moving either to the left or right. A green bar, being brighter than the red bar, would cause the green bar at frame 1 to appear to "move" to the closer brighter yellow bar on frame 2, then onto the closer green bar at frame 3. This creates the illusion of the grating moving to the left. Likewise, a green bar darker than the red bar induces a rightward motion. When the green and red bars appear to have equal brightness, there is no apparent motion, just flickering bars. At each presentation, the observer reports in a forced-choice procedure whether the bars appear to be drifting left or right. Thus there is no need for a nulling of the motion to be perceived.

From the measured color ratios, we then generated the LUT to produce intermediate values of luminance (see Step 4 of the supplementary materials).

4.2. Results for color matching

Color ratios may vary between individuals based on individual differences in sensitivity to the primary colors of the display. Three experienced observers and four initially naïve observers with normal color vision repeated the luminance ratio estimation on an LCD, 10 times each (except one subject who did 6) over a period spanning 3 months (Fig. 5). We analyzed the relative color ratios; the psychophysically estimated color ratio divided by the ratio obtained using the photometer. There were differences between observers for the green/red ratio (ANOVA, $F_{6,56} = 25.2$, $p < 0.0001$) and for the red–blue ratio ($F_{6,56} = 113$, $p < 0.0001$). One subject was more variable than the others for green/red ratio (Levene, $F_{6,57} = 9.57$, $p < 0.0001$). For the red/blue ratio, the naïve subjects were less variable than the experienced subjects ($F_{1,62} = 9.61$, $p = 0.003$). Over the limited age range of these observers (22–49 y), there was a trend for older subjects to have a higher red/blue ratio, consistent with age related changes in the media but it was not statistically significant.

A summary of the color ratios of 6 LCDs, 6 CRTs, 2 HDTVs and 2 DLP projectors measured with a photometer are shown in Table 1. From that table, we set the hypothetical ranges for two luminance ratios. This was done by setting $\max(G/R) = \max(G)/\min(R)$; $\min(G/R) = \min(G)/\max(R)$; and similarly for R/B. Doing this we got the ranges: $G/R \in (1.5, 6.5)$ and $R/B \in (0.8, 5.0)$ that were inclu-

---

[2] In their method, the green bar luminance remained fixed whereas we fixed the red luminance. Since the green channel in most displays is brighter than the red channel at the same input pixel value, fixing the green channel carries the risk that the luminance at that pixel value is higher than the maximum luminance available for the red channel, whereas the luminance of any red pixel value will be within the range of the green channel. The same argument can be applied for luminance matching between red and blue (i.e. it is preferable to fix the channel that is expected to have the lower maximum luminance).

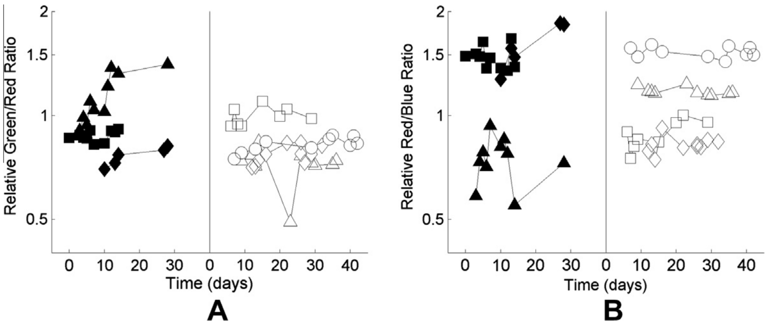


Fig. 5. Relative green/red ratio (A) and relative red/blue ratio (B) for an LCD obtained from 3 experienced observers (filled markers) and 4 initially naïve observers (open symbols) measured repeatedly over a period of weeks. The green–red ratio of that LCD monitor, measured with the photometer, was 2.43, and the red/blue ratio was 2.78.

Table 1
The distribution of values of the ratio of each color (R, G, B) to the total luminance for 16 displays.

Color	Median	Min	Max
Red	0.23	0.12	0.26
Green	0.67	0.64	0.79
Blue	0.10	0.08	0.14

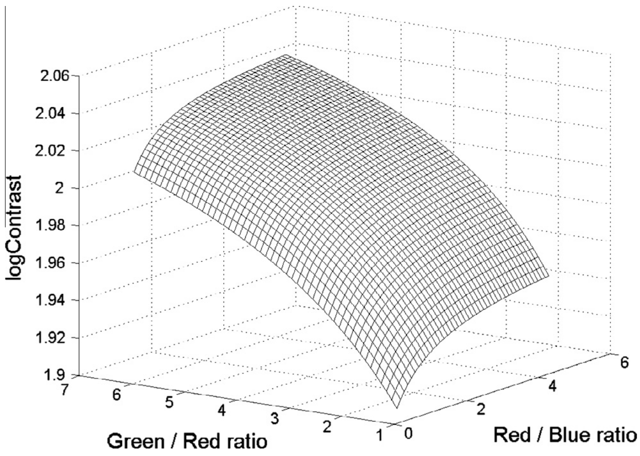


Fig. 6. Model of the variability of log contrast values with a range of color ratios. The output log-contrast was modeled with the ratios $G/R \in (1.5, 6.5)$ and $R/B \in (0.8, 5.0)$. An initial lookup table was generated using two fixed ratios $G/R = 3.5$ and $R/B = 2.0$. From the lookup table, RGB values ($R = 252$, $G = 253$, $B = 252$) corresponding to an intended log-contrast of 2.0 (1%) were extracted, and then used to calculate the contrast at each set of hypothetical color ratios in the above ranges. The log-contrast (on the z-axis) varied between 1.90 and 2.04.

sive of the subjective ratios measured by the subjects. Based on (R, G, B) values to produce a contrast of 2.0 log units (1%), extracted from a look up table generated using fixed ratios $G/R = 3.5$ and $R/B = 2.0$, we plotted (Fig. 6) the expected contrast when the color ratios varied within the above ranges. The range of contrast obtained was from 1.90 to 2.04 log units, equivalent to about 3 letters on the Pelli–Robson and Mars paper charts.

5. Liquid crystal display (LCD) versus cathode-ray tube (CRT)

CRTs have been replaced with LCD technology in most applications. The relationship between the voltage in an LCD pixel and the light intensity is an s-shaped curve that is nearly linear for the large region between the foot and shoulder of the s-curve (James Larimer, Personal communication, 2011). This difference from the CRT gamma function is controlled in most LCDs by electronically creating a desired display gamma function, thus providing backward compatibility with digital image content that was created for CRTs.

Several issues can affect contrast accuracy when displaying a stimulus on commercially available LCDs.

5.1. Gamma correction on LCD

We photometrically measured and fitted gamma functions to measurements of a CRT (ViewSonic G810) and a LCD (NEC MultiSync2090uxi). The residuals of the fits for both displays were of the same magnitude even though the LCD maximum luminance (200 cd/m^2) was twice that of the CRT (100 cd/m^2).

For commercial LCDs, the luminance output has likely been adjusted electronically to resemble the native gamma function of a CRT. Gamma correction is usually provided in the setup menu controls of many modern LCDs. While it is possible to set gamma to various values within the range specified by the manufacturer, we chose to select the display default value, as we expect the display to be optimized for this mode.

5.2. Effects of LCD top brightness on contrasts

For an LCD, there is usually a discontinuity in the light levels emitted between the 254 and 255 pixel values. At 255, the voltage to the LC cells that regulate the backlight transmittance is eliminated, allowing maximum transmittance. The difference between that light level and the level transmitted for the 254 level is not well regulated and can vary widely from other one-level transitions. Thus, a fitted gamma model may not properly represent low contrast stimuli with the background level set to 255 on an LCD. A simple solution is to change the maximum background luminance used on LCD to the well regulated 254.

5.3. LCD screen directionality

Despite recent advances to reduce the directional sensitivity along one dimension inherent in LCD technology (Badano et al., 2003; Krupinski et al., 2004), screen directionality remains a concern to be addressed. While early displays had this increased sensitivity set along the horizontal dimension, current displays usually are manufactured to have the directional sensitivity to be higher along the vertical dimension of the display. This effect is particularly crucial when using the display from a short distance such as in touch screen applications, in which case, different parts of the screen may be viewed from a sufficiently different angle to affect the imaging. To limit the impact of this effect in such an application we used a chin rest to ensure the angles and distances remained constant, and lowered the LCD on its base and tilted the LCD screen up by 18° so that the subjects' eyes were perpendicular to the center of the screen. This also made it easier and more comfortable for older subjects to see through any bifocal or multifocal near vision segment of their glasses.

Normally, we calibrated with the viewer or the photometer perpendicular to the center of the display. When we calibrated, psychophysically and photometrically, with our NEC MultiSync LCD display tilted 18° to the direction of the viewer or the photometer, we found no discernible difference in the calibrations compared to those done perpendicularly. Despite this lack of difference on that LCD, the importance of doing the calibration at the same angle as the contrast measurement cannot be over-emphasized. Care must be taken so that when moving sufficiently far away from the screen so that the alternating lines pattern is invisible, that the operator's eyes remain perpendicular to the center of the screen.

6. Verification

6.1. Validation measurement procedure

To validate our visual calibration, we compared contrasts produced with the visual calibrations to the photometrically measured foreground and background luminance ratios. Because photometer measurements are affected by many factors, such as display fluctuations, ambient or reflected light, and meter inaccuracy, a single measurement is inherently noisy. For a white background of 200 cd/m^2 and a contrast of 2.0 log units (1%), the foreground luminance has to be 198 cd/m^2. For the next lower nominal contrast value at 2.1 log units (0.79%), the expected foreground luminance has to be198.4 cd/m^2 (a difference of only 0.2%). Our luminance meter, the Minolta LS-100,[3] has a specified inaccuracy of ±0.2%. This could place the distinction between two nominal luminance values (0.4 cd/m^2) within the margin of errors limiting our ability to validate the results. To alleviate this, we measured, in random order, the background luminance and foreground luminances for 25 nominal values of contrasts, ranging from 0.0 to 2.5 log units in increments of 0.1 log units, each ten times. See Step 5 of the online supplementary materials for a more complete treatment of the procedure.

6.2. Results of verification procedure

Fig. 7 shows the contrasts obtained with photometer-derived and psychophysically-derived calibrations for a CRT and a LCD, for the range 1.8–2.4 log units. Those lower contrasts are more difficult to create, and only obtained through bit-stealing. For the higher contrasts (<1.8 log units), the measured contrasts were generally indistinguishable from the intended contrasts. The psychophysical calibration contrasts were very similar to those obtained using photometric calibration for both displays (ANOVA, $F_{1,264} = 0.09$, $p = 0.77$). For both calibration methods, the measured contrasts are more variable with the CRT than with the LCD (Levene, $F_{1,1068} = 607$, $p < 0.0001$), while for each display, the two calibration methods had the same variability (Levene, $F_{6,528} < 2.07$, $p > 0.15$). The source of this greater variability with the CRT is not known to us, and may not have been described before. A limitation of this calibration verification (and all others of which we are aware) is that the foreground and background are measured at different times (in the same location). This suggests that the CRT has larger variability of luminance over time than the LCD. It is possible that the actual instantaneous contrasts with the CRT were less variable than we measured, since temporal variations in luminance would affect all intended luminances at that time such that the contrasts would be maintained (even though the luminance was fluctuating).

7. Discussion

It is inevitable that many vision tests in clinics, for routine care and for clinical trials, will transition to electronic displays (for now, these are likely to be LCD rather than CRT, DLP, OLED or plasma). Paper-based charts are subject to problems (Crossland, 2004; Dougherty, Flom, & Bullimore, 2005), particularly effects of dirt, creasing, fading and difficulties obtaining and maintaining good illumination. It is also expected that CS testing will be more widespread and proper CS testing requires accurate calibration of the display system. Display systems are more vulnerable to miscalibration than paper charts as their parameters may be modified intentionally or otherwise. Some calibration problems mostly affect measurements of absolute thresholds and have little consequence for laboratory studies in which responses are compared across different conditions (e.g. Garcia-Perez et al., 2011). However, such miscalibrations are problematic in clinical studies when an individual's responses are compared to normative data or across clinics. Such miscalibrations of absolute contrast also affect large multi-laboratory studies, and were reported to occur in the Modelfest project (see Ahumada & Scharff, 2007). Difficulties in calibrating CS testing on a display were reported in a paper where the contrast levels used could only be specified to be monotonic (Chetrit et al., 2009). Such limited calibration does not enable comparisons across papers or even across locations or displays within a single study.

However, some of the problems we addressed here, such as unaccounted-for display saturating non-linearity or non-monotonic expanded gray scale, may affect any studies, as they can result in improper representation of some contrast levels across a single experiment. Thus, an appropriate calibration procedure is essential for successful implementation of these systems in the clinics and even more so in remote home testing. Evidence in the literature shows that improper calibration is not rare even in highly-equipped laboratories and must be endemic in clinics where the photometric equipment is usually not available to perform or test for appropriate calibration.

We developed and validated a visual calibration system that does not require a photometer and can be easily performed by a normally-sighted person with no prior psychophysics experience. While components of our system have been mentioned in the literature, and some have been implemented, to our knowledge this is the first example of combining all the necessary components in one system and of validating the effect by photometric measurement and comparison to photometric calibrations. Furthermore, most prior work was conducted with CRTs while we have expanded the applications to LCDs and addressed specific characteristics and limitations of LCDs. A previous study using a CRT (Colombo & Derrington, 2001), reported achieving consistent performance for contrasts of 4% and higher, while our systems performance was excellent down to contrasts of 0.5% (log contrast = -2.3) for both CRT and LCD monitors.

The visual calibration method has advantages and limitations. Some of these limitations are shared with photometric calibrations and some are specific to the visual calibration. The visual calibration is highly sensitive to display saturating non-linearity, as a monotonically-increasing gamma function is assumed. With photometric calibration a correction lookup table may be implemented without any model simply by inverting the measurement results. Sufficient elimination of the saturating non-linearity in some displays may be difficult. We noted, when evaluating the 16 different displays, that more expensive displays provided better and easier

---

[3] This is a fairly expensive photometer, costing about $3500.

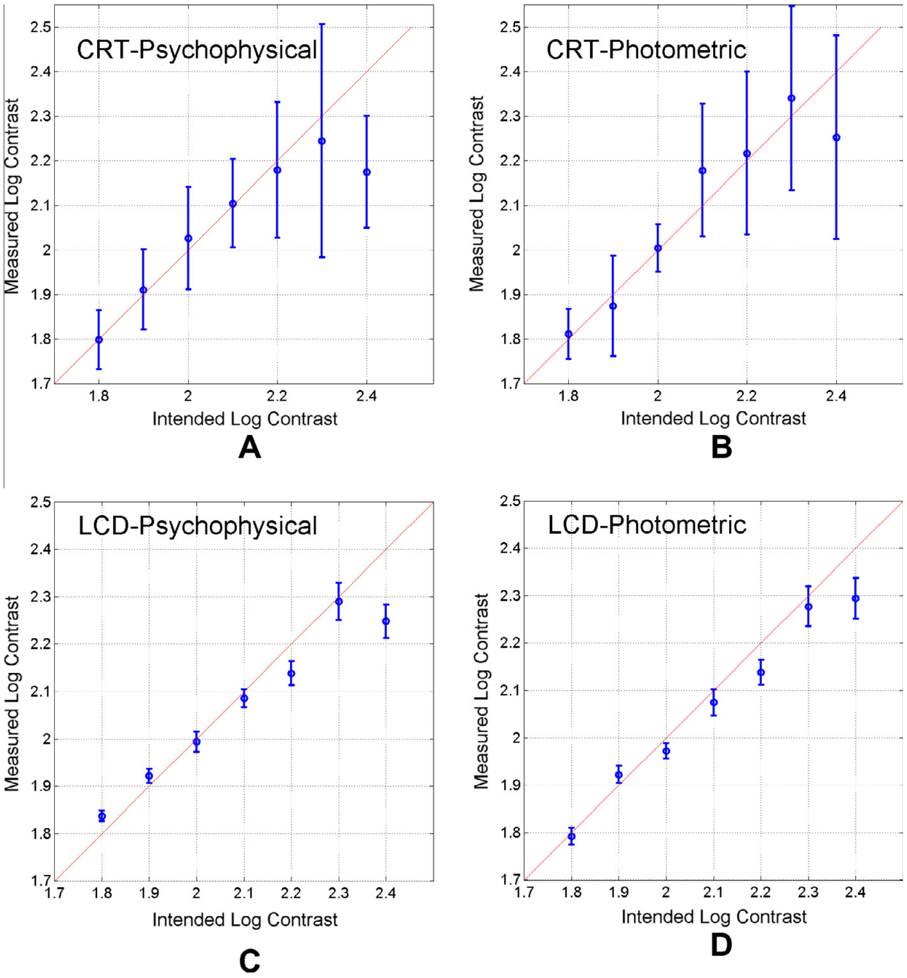


Fig. 7. Comparison of contrasts achieved using the psychophysical method against its intended values for the lowest contrast values (1.8–2.4 log units) of a CRT (A and B) and an LCD (C and D). For contrast below 1.8 log units (not shown), and for all conditions, the intended versus measured values fell exactly on the 45° line. Note that the error bars (standard deviations) for the LCD are smaller than the CRT, which suggests that low contrast stimuli presented on an LCD may be more stable than on CRT.

control of the parameters that are needed to reduce or eliminate saturating non-linearity. Meaningful display calibration must take into consideration room ambient light, scattering of light from regions outside the measurement patch, and even light reflected from the clothing of the observer. Many inexpensive photometric calibration methods, that attach a photocell to the display surface, do not account for these factors. Visual calibration naturally incorporates all of these aspects. In order to take full advantage of these benefits, it is preferable that the calibration is conducted under the same lighting condition and observation distance as used in the experimental session whenever possible.

The color ratio needed for bit-stealing may be affected by the calibrator (Fig. 5), color vision deficiency and age (yellowing of the crystalline lens). This needs to be addressed for both photometric and visual calibration. With visual calibration, using a calibrator who is from the expected subject population will naturally and directly adjust for these effects. The effect of color ratio is of interest only if its magnitude is meaningful. For an intended contrast of 2.0 log units, variation of the color ratio among devices and normally-sighted observers can result in a contrast of 1.90 to 2.04 log units. This range of 0.14 log units corresponds to about 3 letters in the Mars or Pelli–Robson charts. These errors are of the same magnitude as the coefficient of repeatability reported for these charts (Thayaparan, Crossland, & Rubin, 2007). That study (Thayaparan, Crossland, & Rubin, 2007) found worse repeatability for the TestChart 2000, a commercial system that uses bit-stealing but assumes color ratios of 1.0 in all cases (David Thomson, Personal communication, 2008). Under this assumption, the luminance output could be non-monotonic and would produce questionable results at low contrasts where the effect of bit-stealing is crucial. Thus, measuring the color ratios rather than using a generic value will eliminate a small, but systematic source of error in the measurements.

There are a number of limitations of our technique that also affect the photometric calibration technique. The bit-stealing technique which works well for general images, may be affected by the hue difference particularly for an application like our letter CS (Woods et al., in preparation) where we render large uniform regions against a background that is also large and uniform. When this happens, the stimulus and the background are each specified by a single entry in the look-up table and thus detection may be accomplished by the combination of luminance contrast and color contrast. It has been shown that slight color differences can affect luminance contrast threshold (Gur & Akri, 1990). This problem may be limited by modifying both the background and letter values by selecting entries from the look up table that are close in ratio to the intended contrast but are also closer in hue. Another solution may be achieved by dithering the luminance contrast slightly using the color bit-stealing across a narrow range for both regions thus trading the hue difference for a slight luminance noise (Tyler, 1997b).

Brainard (1989) and (Brainard, Pelli, & Robson, 2002) noted that the use of displays in psychophysics experiments implicitly relies on four assumptions: (1) phosphor constancy – that the relative spectral power spectra of the light emitted does not vary with the intensity of stimulation; (2) phosphor independence – that the emitted intensity of a phosphor is determined by the input value and is independent of the other two phosphors; (3) spatial independence – that the display's output at a location depends only on the input values for that location; and (4) single scale factor – that the relative intensities of the phosphors do not vary by location. Although that described CRTs, the treatments of how these assumptions affect the desired luminance is valid also for LCDs.

We have found that letter-CS (absolute values) and repeatability, measured using a computer-based test with CRTs and LCDs that were visually calibrated, were comparable to Pelli–Robson and Mars charts (Woods et al., in preparation). Our visual calibration was validated with a letter-CS test, consisting of gray letters on a white background, in mind. There are many other applications for which this technique may be appropriate, but would require additional validation. However, the measurement of letter-CS, because it operates at the limits of the human visual system and deals with minute differences in contrast is extremely demanding and thus we expect other applications of this technique to pose no difficulty.

Future technologies such as OLED and plasma may replace the LCD and they have one distinct advantage of black – zero pixel value-actually being black.

8. Conclusions

We have brought together several psychophysical techniques to develop a simple, easily deployed, display calibration technique. The procedure is usable for both CRTs and LCDs and has been validated for both. Although there are limitations in its general laboratory use, the availability of this calibration technique would enable CS measurements that can be done in the home, over the Internet, or in clinics at remote locations. We will make our software available upon request at no charge to non-profit institutions and with the proper execution of a material transfer agreement specifying rules for citation and prohibiting further distribution. The software will be supplied "as is" with no assurance of continuing support.

Acknowledgments

Supported in part by a grant from Johnson & Johnson Vision Care, Inc. and NIH Grants EY05957 and EY19100.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2013.04.011.

References

Ahumada, A., & Scharff, L. (2007). Lines and dipoles are efficiently detected (abstract). *Journal of Vision, 7*(9), 337.

Anstis, S. M., & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In J. D. Mollon & L. T. Sharpe (Eds.), *Color vision: Psychophysics and physiology* (pp. 66–77). London: Academic Press.

Arditi, A. (2005). Improving the design of the letter contrast sensitivity test. *Investigative Ophthalmology and Visual Science, 46*(6), 2225–2229.

Badano, A., Gallas, B. D., Myers, K. J., & Burgess, A. E. (2003). Effect of viewing angle on visual detection in liquid crystal displays. In *Proceedings of SPIE, medical imaging 2003: Visualization, image-guided procedures, and display, 5020, Bellingham, WA* (pp. 474–483).

Besuijen, J. (2007). Visual gamma measurement and methods to compare gamma models. *Journal of the Society for Information Display, 15*(8), 611–623.

Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *Color Research and Application, 14*(1), 23–34.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. Hornak (Ed.), *Encyclopedia of imaging science and technology* (pp. 172–188). Hoboken, NJ: John Wiley & Sons, Inc.

Chetrit, S., Gaudet, M., Wittich, W., Bailey, I. L., & Overbury, O. (2009). A comparative study of the efficiency of chart versus computer-generated contrast sensitivity testing in glaucoma patients and controls. *Canadian Journal of Optometry, 71*(3), 34–41.

Colombo, E., & Derrington, A. (2001). Visual calibration of CRT monitors. *Displays, 22*, 87–95.

Corwin, T. R., Carlson, N. B., & Berger, E. (1989). Contrast sensitivity norms for the Mentor B-VAT II-SG video acuity tester. *Optometry and Vision Science, 66*(12), 864–870.

Crossland, M. D. (2004). The role of contrast sensitivity measurement in patients with low vision. *Optometry in Practice, 5*(3), 105–114.

Dagnelie, G., Kramer, K. M., Seifert, G., Yang, L., & Havey, G. (2008). Bringing outcome measurement to the patient: Design of a calibration system for PC-based vision testing (poster presentation). In *Proceedings of the 9th international conference on low vision – Vision 2008, Montreal, QC*.

Dagnelie, G., Yang, L., Bahrami, H., Stone, J., & Melia, M. (2003). Vision tests for the home PC: Test validation and results from a lutein supplementation trial. *Journal of Vision, 3*(12), 57.

Dagnelie, G., Zorge, I. S., & McDonald, T. M. (2000). Lutein improves visual function in some patients with retinal degeneration: A pilot study via the Internet. *Optometry, 7*(13), 147–164.

Dakin, S. C., Greenwood, J. A., Carlson, T. A., & Bex, P. J. (2011). Crowding is tuned for perceived (not physical) location. *Journal of Vision, 11*(9), 1–13 (article no. 2).

Dorr, M., Lesmes, L., Zhong-Lin, L., & Bex, P. J. (submitted for publication). Rapid and reliable assessment of the contrast sensitivity function on an iPad.

Dougherty, B. E., Flom, R. E., & Bullimore, M. A. (2005). An evaluation of the Mars letter contrast sensitivity test. *Optometry and Vision Science, 82*(11), 970–975.

Falkenberg, H. K., Rubin, G. S., & Bex, P. J. (2007). Acuity, crowding, reading and fixation stability. *Vision Research, 47*(1), 126–135.

Garcia-Perez, M. A., Alcalá-Quintana, R., Woods, R. L., & Peli, E. (2011). Psychometric functions for detection and discrimination with and without flankers. *Attention, Perception, and Psychophysics, 73*(3), 829–853.

Gur, M., & Akri, V. (1990). Human contrast sensitivity is enhanced by color (abstract). *Investigative Ophthalmology and Visual Science, 31*(Suppl. 4), 264.

Haun, A. M., Woods, R. L., & Peli, E. (2012). Electronic magnification and perceived contrast of video. *Journal of the Society for Information Display, 20*(11), 616–623.

Kay, R. L., & Brandenberg, C. B. (2007). Nonlinearities of a display device by adaptive bisection with continuous user refinement. United States patent 7,304,482, December 4.

Krupinski, E. A., Johnson, J., Roehrig, H., Nafziger, J., Fan, J., & Lubin, J. (2004). Use of a human visual system model to predict observer performance with CRT vs LCD display of images. *Journal of Digital Imaging, 17*(4), 258–263.

Lavin, Y., Silverstein, A. D., & Zhang, X. (1999). Visual experiment on the Web. In *Proceedings of SPIE, human vision and electronic imaging IV, 3644, Bellingham, WA*.

Li, X., Lu, Z.-L., Xu, P., Jin, J., & Zhou, Y. (2003). Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. *Journal of Neuroscience Methods, 130*(1), 9–18.

Mulligan, J. (1993). Methods for spatiotemporal dithering. *The SID International Symposium Digest of Technical Papers, 24*, 155–158.

Mulligan, J. B. (2009). Presentation of calibrated images over the web. In *Proceedings of SPIE-IS&T electronic imaging, 7240, Bellingham, WA* (pp. 1–10).

Mulligan, J. B., & Ahumada, A. J., Jr. (1992). Principled halftoning based on human vision models. In *Proceedings of human vision, visual processing and digital display III, 1666, Bellingham, WA* (pp. 109–121).

Niebergall, R., Huang, L., & Martinez-Trujillo, J. C. (2010). Similar perceptual costs for dividing attention between retina- and space-centered targets in humans. *Journal of Vision, 10*(12), 1–14 (article no. 4).

Pappas, T. N., & Neuhoff, D. L. (1992). Least-squares model-based halftoning. In *Proceedings of SPIE, human vision, visual processing, and digital display III, 1666, Bellingham, WA* (pp. 109–121).

Peli, E. (1992a). Display nonlinearity in digital image processing for visual communications. *Optical Engineering, 31*(11), 2374–2382.

Peli, E. (1992b). Halftone imaging apparatus and method. United States patent 5,109,282, April 28.

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences, 2*(3), 187–199.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31*(7–8), 1337–1350.

Pierre, A., Wittich, W., Faubert, J., & Overbury, O. (2007). Luminance contrast with clear and yellow-tinted intraocular lenses. *Journal of Cataract and Refractive Surgery, 33*(7), 1248–1252.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing*. Cambridge: Cambridge University Press.

Regan, D. (1988). Low-contrast letter charts and sinewave grating tests in ophthalmological and neurological disorders. *Clinical Vision Sciences, 2*(3), 235–250.

Rutstein, R. P., & Corliss, D. A. (2000). BVAT distance vs. near stereopsis screening of strabismus, strabismic amblyopia and refractive amblyopia; A prospective study of 68 patients. *Binocular Vision and Strabismus Quarterly, 15*(3), 229–236.

Sharma, G. (2002). Comparative evaluation of color characterization and gamut of LCDs versus CRTs. In *Proceedings of SPIE, color imaging: device-independent color, color hardcopy, and applications VII, 4663, San Jose, CA* (pp. 177–186).

Swift, D., Panish, S., & Hippensteel, B. (1997). The use of the VisionWorks™ in visual psychophysical research. *Spatial Vision, 10*(4), 471–477.

Thayaparan, K., Crossland, M. D., & Rubin, G. S. (2007). Clinical assessment of two new contrast sensitivity charts. *British Journal of Ophthalmology, 91*(6), 749–752.

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

Tyler, C. W. (1997b). Why we need to pay attention to psychometric function slopes. 1997 OSA technical digest series: Vision science and its applications, 1, SuD2-1/240-243, Washington, DC.

Ulichney, R. A. (1988). Dithering with blue noise. *Proceedings of the IEEE, 76*(1), 56–79.

Vera-Díaz, F. A., Woods, R. L., & Peli, E. (2010). Shape and individual variability of the blur adaptation curve. *Vision Research, 50*(15), 1452–1461.

Waltuck, M. H., McKnight, R. N., & Peli, E. (1991). Visual function tester with binocular vision testing. United States patent 5,026,151, June 25.

Watson, A. B., Nielsen, K. R. K., Poirson, A., Fitzhugh, A., Bilson, A., Nguyen, K., et al. (1986). Use of a raster frame buffer in vision research. *Behavior Research Methods, Instruments and Computers, 18*(6), 587–594.

Webster, M. A., Georgeson, M. A., & Webster, S. M. (2002). Neural adjustments to image blur. *Nature Neuroscience, 5*(9), 839–840.

Williams, R. E., Decker, T. A., Kurtzman, C., & Kuether, C. L. (1980). Apparatus for generating and displaying visual acuity targets. United States patent 4,239,351, December 16.

Wong, B. P., Woods, R. L., & Peli, E. (2002). Stereoacuity at distance and near. *Optometry and Vision Science, 79*(12), 771–778.

Woods, R., Fullerton, M., Goldstein, R., McNiff, S. A., McIvor, T., & Peli, E. (in preparation). Search for low contrast letters as a measurement of contrast sensitivity.

Woods, R. L., Nugent, A. K., & Peli, E. (2002). Lateral interactions: Size does matter. *Vision Research, 42*(6), 733–745.

Yang, J. -B. (2013). Display calibration. <http://www.visionrc.com/Tweak/DisplayCal.aspx> Accessed 04.02.13.

Supplementary Material for

Psychophysical Contrast Calibration

A step-by-step guide for calibrating your display

without using a photometer

By Long To, Russell L Woods, Robert B Goldstein, Eli Peli

Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA

If you implement these procedures, please cite our accompanying paper in Vision Research.

Overview

The luminance calibration described here requires only responses from an operator. No instruments, such as luminance meters (photometers), are required. These procedures are implemented in software written for Windows computers by the authors (*LuminanceCalibration* program) which is available upon request at no charge to non-profit institution researchers and upon execution of a material transfer agreement specifying rules for citation and prohibiting further distribution. The software will be supplied "as is" with no assurance of continuing support. This guide will help the reader to implement their own software or modify our software if so desired. The physical instantiation of the calibration is a lookup table (LUT) with approximately 10 times the number of entries for a 24-bit display[1]. Each entry in the LUT is a set of RGB values that will produce a designated relative luminance that can be used to obtain a desired contrast.

The calibration procedure comprises five steps, two of which are threshold estimates in which the operator views and make judgments on a series of stimuli. First, the operator checks for display or signal nonlinear luminance saturation and cut-off, makes adjustments to the display (if necessary), and then records the display settings (*Fading Patches* task). Second, the operator adjusts the apparent brightness of a series of patches to match the brightness of an adjacent patch (*Matching Gray Patches* task). This step produces data for estimating the gamma (γ) of the display system. Third, the operator indicates the apparent direction of motion (right or left) of a series of alternating gratings (*Direction of Motion* task). This step produces a measurement of the luminance ratios of the primary colors of the display. The fourth step uses the estimated gamma and color ratios to build a LUT using a bit-stealing technique. With the appropriate software (either user-written or supplied by us), enough information is then available to generate relative luminance values, and hence contrasts, at desired levels on that display

---

[1] See below for why it is not exactly 2560 (10 times 256)

device. This LUT is then used by other software (in step five) to generate stimuli with desired contrasts for an experiment or clinical tests.

Periodic verification of the luminance calibration and of display settings is highly recommended. As discussed in the paper, separate calibrations are required for different displays and may be required for different viewers (e.g., people with color deficiency or different age, as aging affects spectral transmittance of the ocular media, especially the color of the crystalline lens).

Calibration preparation

A number of set up conditions should be taken into account prior to calibration to ensure the validity and accuracy of the results. Several of these conditions are not unique to calibration, but are good practice for any vision psychophysics experiment. When possible, these conditions are checked by our *LuminanceCalibration* software. At other times, however, compliance is dependent upon the prudence of the operator.

1. Physical location of equipment: Place the display device at one fixed location during calibration and subsequent tests. If the display device is moved, run the calibration again. For CRT displays, if nearby electronics are moved or change status (e.g. on versus off), the calibration should be conducted again.
2. Viewing position of operator: To reduce the effect of viewing angle (particularly important for LCD screens), the eye of the operator should be perpendicular to the center of the screen during the calibration process and during the experimental trials. For the *Matching Gray Patches* task, the operator must sit sufficiently far from the display device so that the stripes are not visible[2]. If the keyboard cable is too short, use an extension cable or a wireless keyboard to keep it within comfortable reach from the operating position. *If the experiment is conducted from a different distance, the equality of the viewing angle should be carefully controlled to match the calibration and the experiment conditions.*
3. Lighting conditions in the room: Use the same room lighting conditions during the calibration process as will be used for your use of the display (e.g. collecting data). Usually, this will be low lighting *(we recommend that room lights be turned off*). If any light is needed, care should be taken to avoid direct reflections from the surface of the display that may be visible to the viewer.
4. Screen resolution: This must be the same during calibration as will be used for your study. Preferably, this will be the "native" resolution (LCD, DLP) or the highest resolution that allows 24-bit or higher color depth (CRT).
5. Color depth: Set the color depth of the screen to 24-bit depth or higher. The *LuminanceCalibration* software (available from the authors) will not run with lower color depths, as the resolution is insufficient for measurements of human contrast sensitivity. The software will run at both 24- and 32-bit (the latter with an alpha channel for transparency).
6. Degauss the display device (if appropriate, CRT only).
7. Using the display control panel (menu), adjust the "color temperature" setting of the display for all color channels to their maximum level (100%). Contrast and Brightness levels will be adjusted as part of a saturating non-linearity elimination step below.

---

[2] Assuming a pixel size of 0.25mm, the lines should not be visible to a person with 20/10 visual acuity when seated 1.8m from the display.

8. The calibrating operator must have normal color vision, or the same color vision as the person to be tested. Age matching the calibrator to the study participants may be appropriate.

Step 1. Visual check for color and brightness saturating non-linearity

One key assumption of our method is a monotonically increasing relationship between the on-screen pixel value (when R=G=B, grey level) and the emitted luminance. However, on some display devices, especially LCDs, a saturating non-linearity may cause the maximum luminance to be reached at a pixel value below the maximum pixel value (y_{max}). In addition, at the low end of the luminance range, it is possible that the output remains unchanged until the input pixel value passes some non-zero value (cut-off). Saturating non-linearities may occur in one or all three component colors. Saturating non-linearity and cut-off reduce the luminance resolution and prevent a proper psychophysical luminance calibration using our approach that assumes the gamma function to extend over the full range of pixel values for all colors on the display device. Consequently, it is not possible to get accurate control of contrasts.

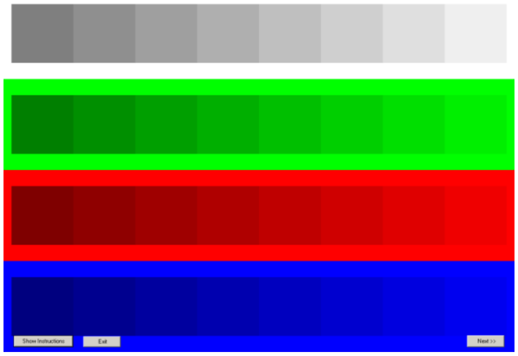


Figure S1. Demonstration stimulus for detecting and removing saturating non-linearities at high pixel values (this image has larger steps than the actual image so that eight patches in all rows will be visible on most displays). Square patches of decreasing luminance (from right to left) against the background to detect saturating non-linearity in grayscale and individual color channels. To remove saturating non-linearity, an observer adjusts the Contrast and Brightness manual controls of the display device until all eight patches in each zone are visible, and the rightmost patch is just visible against the background. A similar stimulus is used for Cut-Off testing at low pixel values.

Using a pattern such as the one shown in Fig S1, visual inspection by the operator can detect if the gray or color channels reach a saturation level before the pixel value reaching the maximum of 255 (or 254 on an LCD, since the 255 level is not well controlled on many LCDs). If any of the channels suffers from a saturating non-linearity, then the operator will need to adjust the display (using the contrast and luminance adjustments of the display menu controls) until the saturating non-linearity is eliminated for all 4 rows of patches. The brightest patch should be close to the maximum brightness of the surrounding frame. After these adjustments, if there is still a residual saturating non-linearity, then this display cannot be used with this psychophysical calibration method[3]. An illustration of a saturating non-linearity in only one color channel is provided in the paper. Three patterns, one of which is shown in Figure S1, are required. The first, the *Demonstration* pattern, is an example with larger increments between the successive patches. The second, *Saturation Test*, and the third, *Cut-Off Test*, patterns

---

[3] Unless you can determine the range over which it is monotonically increasing (e.g. grey level from 5 to 250), and restrict the range for calibration and use accordingly. In that example, grey level = 5 would be assigned a relative luminance = 0 and grey level = 250 would be assigned a relative luminance = 1.

are used for the saturating non-linearity check at high and low pixel values ranges, respectively. Each Fading Patches display is divided vertically into four color rows. From the top, these are white (grey), green, red and blue. There are eight patches in each row.

The details of the pixel values along each row for the *Saturation Test* (high pixel value) pattern are as follows: The background of each row should be of uniform luminance at the maximum of 255 for the color of that row (or 254 for LCDs). Along each row are 8 successive square patches of the same size, with diminishing contrasts from left to right against the strip background. The step increment, in pixel value, between adjacent patches, as well as between the last (right-most) patch and the background, varies from one row to another, but is otherwise constant along each row. The pixel values that we found empirically to work on the displays that we used in our study are listed below. Construction of the Cut-Off Test (low pixel value) pattern is comparable, except dark. The pixel values that we found to work are shown below.

On a bright background (for saturating non-linearity)

Patch No.	1	2	3	4	5	6	7	8	Background
Gray	247	248	249	250	251	252	253	254	255
Green	239	241	243	245	247	249	251	253	255
Red	231	234	237	240	243	246	249	252	255
Blue	223	227	231	235	239	243	247	251	255

* These values are for a CRT. For a LCD all values should be decreased by 1.

On a dark background (for cut-off)

Patch No.	1	2	3	4	5	6	7	8	Background
Gray	24	21	18	15	12	9	6	3	0
All colors	40	35	30	25	20	15	10	5	0

With the appropriate display settings, the eighth (right-most) patch in each row of all two test displays should be just detectable. When just detectable, all that may be seen of a right-most patch are the edges of the patch. It is important to set the last patches on the right to be just visible and to have the maximal range of luminances available.

If all eight patches are visible in all rows on both test displays, there is no saturating non-linearity or cut-off. Record the display settings and proceed to the next calibration step (**Matching Gray Patches**). If some of the patches are not visible, adjust the display settings until the last patches become just visible. To do this, start by first lowering the Brightness setting, then, if necessary, the Contrast setting. If that is not sufficient, try the Brightness and Contrast settings in combination. Once you have the best adjustment of the Brightness and Contrast settings, if one or more color sections (rows) of a test display remain incomplete (i.e. not all eight patches visible in that row), adjust the "color temperature" setting for that row. For example, if there are only seven patches visible in the blue row, adjust the Blue setting to make the eighth patch visible. Adjust the display's settings until all eight patches are visible in each color row. Note that following adjustment of one end of the luminance range the other end has to be rechecked to verify that it is still free of saturating non-linearity or cut-off effects (i.e. fixing one can upset the other).

Without satisfactory settings (all 64 patches visible: 32 on the *Saturation Test* pattern and 32 on the *Cut-Off Test* pattern), saturating non-linearities or cut-off may be present at either end of the luminance range, making it difficult to parametrically characterize the luminance output (fit a γ value). For our application (contrasts on light backgrounds), cut-off at the low end is not as much of a problem as a saturating non-linearity at the high end of the luminance range. Once a satisfactory condition is achieved, record the settings of the display and use them for all subsequent tests that use the generated LUT. These display settings are a crucial aspect of the psychophysical luminance calibration and must be used when using the relative luminance LUT (e.g. testing contrast sensitivity).

Step 2. Visual gamma calibration

This calibration step determines the grey-scale gamma (γ) function of the display (used to characterize the relationship between the RGB input levels and the luminance of a display). The operator's task is to match the brightness of two square patches, positioned side by side, at the center of the screen (Figure S2). One patch is solidly filled with pixels that all have a single pixel (grey) value (luminance); the other patch is made up of alternate horizontal lines of different pixel values (luminance). At this stage, all three color channels have the same pixel value. The task is repeated over seven pairs of abutting square patches. The procedure to select pixel values for each reference patch is described below.

For the initial reference patch, the alternating lines are set to $(R_1 = 0, y_1 = 0)$ and ($R_2 = 1$, $y_2 = y_{max}$). Subsequent reference patches are created by recursive partitioning of the prior displayable ranges. The procedure is summarized as follows: Create a reference patch with relative 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, where $0 = y1 < y3 < y2 \leq 255$ and that $R(y_3) = R_3 = ½$.

1. Create the next reference patch of relative luminance $R_4 = ¼$ 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.

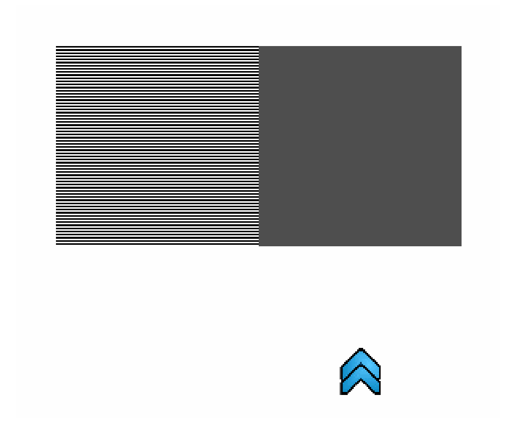


Figure S2. Matching of side-by-side gray patches. In this figure, the adjustable patch is on the right and filled using one gray level. Change the brightness of this patch, indicated by the arrow, until it looks the same as the other square patch. The onscreen location of the adjustable patch can be alternated between left and right in successive matches. Due to limitations in the reproduction process, the alternating dark and light horizontal lines in the patch on the left may not appear as a simple alternating pattern. For example, it may appear that a lower frequency gradient is present.

2. Similarly, create the next reference patch of relative luminance R_5 = ¾ by alternating

   $y = y_2$ (R_2 = 1) and $y = y_3$ (R_3 = ½). Luminance matching then yields y_5.

3. Continue recursive partitioning in this manner to obtain the values for

   (R_6= ⅛ , y_6), (R_7= ⅜ , y_7), (R_8= ⅝ , y_8), and (R_9= ⅞ , y_6)

In our software, the positions of the reference and calibration patches are alternated in space after each match to reduce side bias, local adaptation, and memory effects. The observer uses the keyboard to adjust luminance of the calibration patch (in our software, this is marked underneath with blue arrowhead on the screen: Figure S2).

The operator must sit sufficiently far from the screen so that the alternating horizontal line pattern is not visible, and, therefore, both patches look as if solidly filled. The program must allow the operator to adjust the solid-filled patch so that it matches the brightness of the alternating-lines patch (for example, using the *up* and *down* keyboard arrow keys). Once a match is found (i.e. they look the same and the edge between them seems to disappear), the matching value is recorded. After all the matches are found, our software estimates gamma by minimizing the sum-of-squares error, given by equation (4) of the paper. For reference purposes, the Matlab code to estimate this parameter gamma is included below. Note that the underlying optimization is either Gauss-Newton or Levenberg-Marquardt (LM), depending on the version of Matlab; LM is used in our integrated calibration software.

```
%% example pair of matching gray values and relative luminance
%% first  column: the gray values
%% second column: matching luminance values
      data = [185 1/2;...
              134 1/4;...
              223 3/4;...
               96 1/8;...
              162 3/8;...
              205 5/8;...
              239 7/8];

      options = statset('Display','iter');
      %% calling method to estimate gamma
 gamma = nlinfit(data, zeros(length(data),1), @objfunc, [1], options);
```

```
%% save this function to a file named objfunc.m
%% this function returns the residual error for each estimated value of gamma.
%% NOTE before running: set ymax to 254 for LCD and 255 for CRT
      function sse = objfunc(gamma, dataPair);
      sse=(dataPair(:,1)./ymax).^gamma - dataPair(:,2);
```

In our software, if the residual error from above fitting is higher than 0.001, the matching task is repeated to obtain new samples[4]. A typical value of γ is 2.2 for CRTs[5]. LCDs and DLPs often have software to replicate the typical CRT γ.

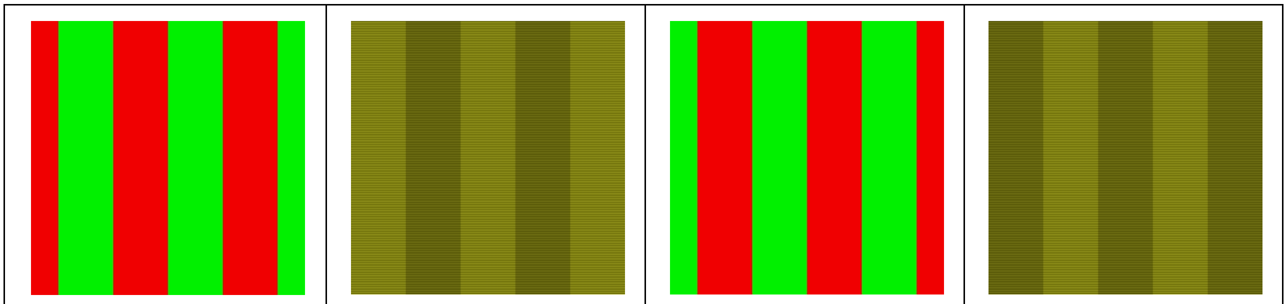
---

[4] Empirical observations indicated that, for values greater than 0.001, the operator probably made an error. For values less than 0.001, the fit is reasonable.

Step 3. Visual estimation of luminance ratio amongst primary colors

This calibration step measures the relative luminance of colors. It relies on a motion illusion of vertical bars that appear to be moving either to the left or right (discussed in the paper). It involves two tests that differ only in the pairs of colors that are tested. The first test has green and red vertical bars (to measure the relative luminance of red to green) and the second test has red and blue vertical bars (to measure the relative luminance of red to blue). Depending on the relative brightness of the bars, the bars will appear to move to the left or to the right. The operator's task is to indicate the direction of motion after each brief presentation. The appearance of the bars in the four frames of the sequence is shown in figures S3 and S4.

The stimulus is composed of four square frames, presented successively in a loop. In our implementation, each frame is a square image with size of 320x320 pixels, made up of five vertical bars of 64-pixels width. The composition of these frames is described below.



Frame 1 Frame 2 Frame 3 Frame 4

Figure S3. The 4-frame sequence used in the Green versus Red equi-luminance matching task. In frames 1 and 3, the red bar luminance remains constant and the green bar luminance is adjusted according to the observer's response. The green bar can be adjusted to appear either brighter or darker than the red bar. In frames 2 and 4, the bars appear to be dark and bright yellow. These bars are actually composed of alternating red and green horizontal lines (see figure S4 for a further description of the bars in frames 2 and 4). The values of the red horizontal lines are constants, but the green horizontal lines are adjusted according to the current value of green in frames 1 and 3 (i.e. multiplying by 17/16 and 15/16 respectively). The sequence shown here with the green brighter than red will result in perceived image motion to the left. Note that the alternating horizontal lines in frames 2 and 4 may result in sampling aliasing artifacts on some displays or in print.

Frame no. 1: Leftmost bar is red and remains fixed throughout the experiment at 240 (i.e. R=240, G=B=0). Next bar is green, is variable (depending on observer's prior response). There are two interleaving staircases, where the initial values for the upper staircase are (G=240, R=B=0; brighter than the red at that value) and for the lower staircase are (G=64, R=B=0; darker than the red at 240).

---

[5] http://www.w3.org/Graphics/Color/sRGB.html

Frame no. 2: Leftmost bar is composed of horizontal single-pixel rows of 64 pixels alternating between green and red. Both the red and green lines in this bar are brighter than their respective levels in frame 1 (set at 17/16 times the value from frame 1).

The next darker bar is also composed of alternating red/green bars, set at 15/16 their respective levels from frame 1. This bar therefore appears darker than the first bar. See figure S4 for an illustration.

Frame no. 3: same as frame 1, but of opposite phase (i.e. location of red and green bars are swapped).

Frame no. 4: same as frame 2, but of opposite phase.

The stimulus composes these four frames, playing in the above order in a continuous loop at the frame rate of 5 FPS (200ms each frame). We programmed the presentation using DirectX where all frames are drawn at runtime.

Figure S4. A close-up section of the bright and dark bars from frames 2 and 4. Each bar contains single rows of pixels with alternating color values. When viewed from afar, the alternating line pattern is invisible and blends into a shade of yellow, as seen in figure S3.

Each stimulus sequence is presented for the maximum of two seconds, but stopped upon getting a response. When played, this stimulus gives the appearance of motion, the direction of which depends on the relative luminance of the bars. If the green bar (in frames 1 and 3) is brighter than the red, the perceived motion is to the left; if the red is brighter, the perceived motion is to the right. The operator indicates (by using arrow keys on the keyboard) the perceived direction of the moving pattern. Estimation for each color ratio (green-red, and red-blue) is obtained using a staircase procedure illustrated in Figure S5.

After presentation of a stimulus, the observer reports whether the bars appear to be moving left or right. For the example shown in Fig. S3, if the motion is to the right, the green value is decreased in the next stimulus in the staircase; if the motion is to the left, the green value is increased. We implemented the following simple adaptive method for determining the match, but other psychophysical methods could be employed. We used two interleaved simple 1-up-1-down varying step-size staircases. The initial step size is 32 gray levels and reduces by half when the responses, i.e. the perceived motion directions, differ on two consecutive presentations (a reversal). The lower limit for the step size is set at 4 gray levels. The algorithm terminates when there are at least three reversals at the smallest step size in each staircase (6 reversals across both staircases). The color ratio is calculated as the average of these last 6 reversals.

Following the color ratio estimation (with green and red bars), a second estimation (with red and blue bars) should be conducted similarly. This procedure is the main reason for our recommendation to use an observer with normal color vision and age matched to the experimental subjects when performing the calibration.

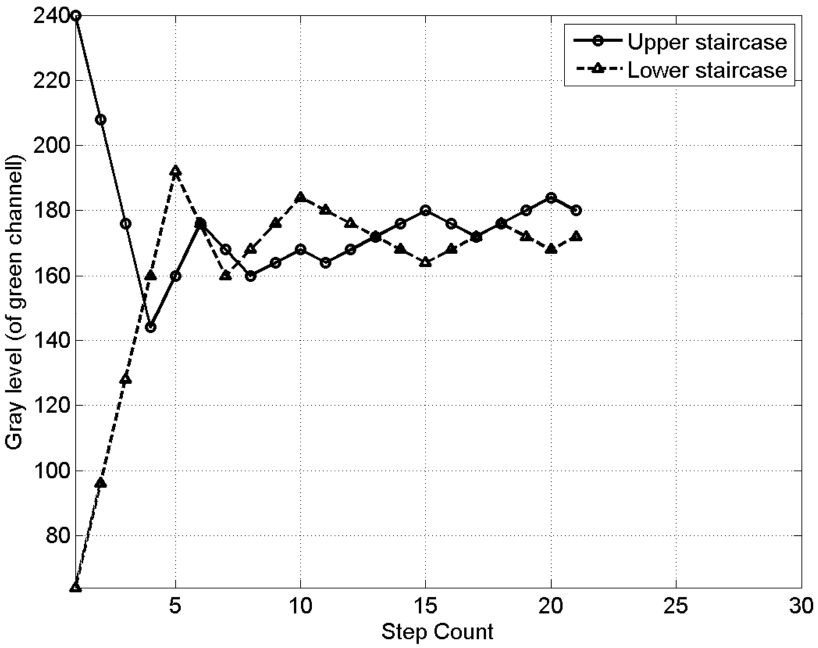


Figure S5. Example of two interleaved staircases during the color equi-luminance matching task. Every time the perceived motion changes, the staircase direction is reversed and the step size magnitude is reduced by half in the next stimulus until the minimal value is achieved after 3 reversals in each staircase. The ratio is calculated using $X_{G/R}$ = (Red/matchedGreen)$^{\gamma}$ where Red=240 and matchedGreen is the average of the last 3 reversals in each staircase (a total of 6 values).

Steps (2) and (3) are independent, so step (3) can be performed before step (2), if preferred. However, a value of gamma (from step 2) is still required to calculate the color ratios after completing step (3).

A simple test can be implemented to verify that the measured ratios are correct. This consists of creating an image consisting of a series of stripes (e.g. 50 pixels wide) of each of the bit-stealing increments (such as in the tables described earlier) in the expected luminance order based on the color ratios. If the color ratios are correct, then one should see (with difficulty) the stripes gradually increase in luminance and they should be monotonic, with no reversals if all is working properly.

Step 4. Lookup Table for relative luminance

The calibration procedures in steps (1) through (3) provide the information necessary to construct the bit-stealing lookup table (LUT) that will enable a program to obtain the correct pixel value settings to produce desired relative luminance values used to create a desired contrast. Instructions for constructing this LUT are given below. If one were to use only gray levels (R=G=B), only 256 entries (and thus values of luminance) would be available, and that is insufficient luminance resolution for many tasks. To generate the additional luminance values necessary for measuring contrast thresholds, intermediate luminance values must be obtained and generated using the bit-stealing technique. Thus, the LUT will consist of two types of entries:

those generated solely from γ (linearity correction), and those generated based on γ as well as the estimated color ratios (bit-stealing to increase luminance resolution). To increase the luminance resolution by a factor of ten, we had 9 additional luminance values between two consecutive grayscale entries. Each intermediate "bit-stealing" entry is given by $R = R_i + \Delta_R$, (equation S.2) where R is an intermediate desired luminance, R_i is the nearest grayscale entry, and the set of Δ_R values is calculated using equation S.3 (subject to the constraints given in equations S.6 and S.7). These constraints ensure that the bit-stealing luminance increment does not go beyond the next grayscale LUT entry.

To create the LUT using the estimated γ parameter and color ratios (which have been estimated through the motion direction psychophysical tasks), we use the following procedure. 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 to produce this luminance. The entries in the table are divided into two types: grayscale entries where luminance is dependent only on γ, and bit-stealing entries where luminance depends on the color ratios.

Grayscale entries have the following format:

$$R = (y / y_{\max})^{\gamma}, \; y_R = y_G = y_B = y, \tag{S.1}$$

where y is an 8-bit integer for each gray level and $y_{\max}$ is the gray value corresponding to the maximum used luminance (y_max = 255 for CRTs, 254 for LCDs. See Section 5.2 of main paper for details).

Across 16 displays (CRT, LCD, and DLP), we found that the median values of the luminances of R, G and B relative to the total luminance were about 0.2, 0.7, 0.1 (table 1 of the paper). On that basis, we chose nine color bit-stealing RGB combinations that provided approximately equal increments between subsequent grayscales. If the display of interest has a different set of relative luminances, other sets may be preferable to those described next.

For a bit-stealing entry, with the *RGB* increment $(\delta_R, \delta_G, \delta_B)$ selected from the set { (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) }, the relative luminance is calculated as

$$R = \left(\frac{y}{y_{\max}} \right)^{\gamma} + \Delta_R \tag{S.2}$$

where y is gray value of the preceding grayscale entry, and Δ_R is the intermediate luminance increment.

$$\Delta_R = (\delta_R p_R + \delta_G p_G + \delta_B p_B)\left(\left(\frac{y+1}{y_{\max}}\right)^{\gamma} - \left(\frac{y}{y_{\max}}\right)^{\gamma}\right), \quad \text{(S.3)}$$

where, (p_R, p_G, p_B) are the relative luminance contributions of each primary color and derived from the color ratios $X_{G/R}$ (green/red) and $X_{R/B}$ (red/blue) as follows:

$$p_R = \frac{X_{R/B}}{1 + X_{R/B} + X_{G/R}X_{R/B}}; p_G = \frac{X_{G/R}X_{R/B}}{1 + X_{R/B} + X_{G/R}X_{R/B}}; p_B = \frac{1}{1 + X_{R/B} + X_{G/R}X_{R/B}} \quad \text{(S.4)}$$

The RGB values corresponding to the above bit-stealing entry are:

$$y_R = y + \delta_R; y_G = y + \delta_G; y_B = y + \delta_B \quad \text{(S.5)}$$

To ensure that the bit-stealing luminance increment does not go beyond the next grayscale LUT entry, the following two constraints are verified for each entry in the LUT:

$$(\delta_R p_R + \delta_G p_G + \delta_B p_B) < 1 \quad \text{(S6)}$$

$$\left(y + \max(\delta_R p_R + \delta_G p_G + \delta_B p_B)\right) < y_{\max} \quad \text{(S.7)}$$

The total number of entries in the LUT may vary depending on the color ratios. In our application, we planned the table to have 9 additional luminance values between two consecutive grayscale entries, but that actual number of entries can vary depending on the intended application. This set $\{\delta_R, \delta_G, \delta_B\}$ of nine entries reflects every value in 0.1 increments between 0.1 and 0.9 and is based on an approximation of R:G:B=2:7:1, which is adequate to derive these integral bit-stealing entries. The test described at the end of step (3) should be used to confirm that the color ratios and chosen set are valid. While the number of grayscale entries is fixed, the number of bit-stealing entries may vary because of constraints in (S.6) and (S.7).

Step 5. How to use the Lookup Table

This procedure shows how to use the LUT to obtain background luminance values and RGB pixel value settings to produce a desired contrast. For a desired contrast C and background level, *R_{bg},* use equation S.8 to calculate the foreground luminance needed (*R_{fg}*). Since the LUT entries are discrete, it will be unlikely that exact entries for R_{bg} and R_{fg} will be found. Therefore, the LUT entries with a relative luminance closest to R_{bg} and R_{fg} should be used.

The relative foreground luminance R_{fg} needed to generate the intended contrast C on a background of maximum relative luminance, $R_{bg} = 1.0$. The value of R_{fg} is given by:

$$C = \frac{R_{bg} - R_{fg}}{R_{bg}} = 1 - R_{fg} \Rightarrow R_{fg} = 1 - C \tag{S.8}$$

Look up the RGB values from the table corresponding to the relative foreground luminance R_{fg}. Select the table entry with a relative luminance closest to R_{fg}.

Verification of Lookup Table data

This section is for readers who have access to a photometer and wish to use it to verify the validity of the lookup table generated by the above psychophysical procedure.

To compare the actual contrast (when the above RGB levels are used for the stimulus foreground) to the intended contrast, we took the following steps:

1. Use a photometer to measure the emitted luminance of a uniform patch on the screen with the above RGB values.
2. Calculate the actual contrast of this patch on a maximum-luminance background R_{bg}^{abs}.

   The superscript abs is for photometrical luminance measurement in cd/m^2.

$$C_{measured} = \left(\frac{R_{bg}^{abs} - R_{fg}^{abs}}{R_{bg}^{abs}} \right) \tag{S.9}$$

3. Plot measured contrast against intended contrast on a log scale as done in Fig 7 of the paper.