Using the human eye to characterize displays

Jennifer Gille[a] and James Larimer[b]
[a]Raytheon ITSS at NASA Ames, Moffett Field, CA
[b]NASA Ames Research Center, Moffett Field, CA

ABSTRACT

Monitor characterization has taken on new importance for non-professional users, who are not usually equipped to make photometric measurements. Our purpose was to examine some of the visual judgments used in characterization schemes that have been proposed for web users. We studied adjusting brightness to set the black level, banding effects due to digitization, and gamma estimation in the light and in the dark, and a color-matching task in the light, on a desktop CRT and a laptop LCD. Observers demonstrated the sensitivity of the visual system for comparative judgments in black-level adjustment, banding visibility, and gamma estimation. The results of the color-matching task were ambiguous. In the brightness adjustment task, the action of the adjustment was not as presumed; however, perceptual judgments were as expected under the actual conditions. When the gamma estimates of observers were compared to photometric measurements, problems with the definition of gamma were identified. Information about absolute light levels that would be important for characterizing a display, given the shortcomings of gamma in measuring apparent contrast, are not measurable by eye alone. The LCD was not studied as extensively as the CRT because of viewing-angle problems, and its transfer function did not follow a power law, rendering gamma estimation meaningless.

1. INTRODUCTION

As we move towards an "all digital" lifestyle, image appearance management across platforms has become increasingly important for non-professional users. Many web sites offer advice on monitor characterization, and new software products have been developed to allow color and tone scale characterization and adjustment for non-professionals[1,2]. Amateur digital photography and retail sales on the web especially need this kind of appearance management. It has been suggested that psychophysical experiments could be carried out over the web, if sufficient attention is given to monitor characterization by the potential observer[3]. The purpose of this paper is to examine some of the kinds of visual judgments currently in use for display characterization, especially on the web, and to compare those judgments to photometric measurements, for a desktop CRT and a laptop AMLCD, in the dark and in normal room lighting.

Various schemes have been proposed for monitor characterization by eye. It is suggested that monitor adjustments such as brightness and contrast be set to optimum levels by comparing nearly identical black or white patches. The tone scale function, parameterized by gamma (the exponent in the putative power function relationship between digital count input and luminance output of a display), is established by comparing the brightness of grey patches to black and white patterns. Colors are controlled by matching colored screen patches to external color swatches. These system characterization schemes need verification. In that process, the human visual system, photometric measurement, ambient lighting, display technology, and the model of display performance all interact.

The visual system is especially good at relative judgments but poor at absolute judgments. For instance, a 1% difference in luminance can be detected in a side-by-side visual judgment[4], yet finding an acceptable light level for house plants requires a simple rule such as "does best in a south-facing window". Sensitivity varies across individuals and over time, as do judgment criteria and levels of motivation for precision and accuracy. Therefore, simple, quick comparisons are needed for perceptual judgments.

Photometers are good at absolute measurements, and their judgment criteria and motivation levels do not vary. Because of this objectivity, they are often considered to be superior to perceptual judgments. However, photometers do not contain the kinds of noise-reducing circuitry that the visual system presumably does, and in

comparison judgments, the visual system can be considerably more reliable. Also, photometers need to be used correctly, for instance placed sufficiently far from the display and correctly focused.

Ambient lighting for computer monitor use is often eliminated in professional settings, but seldom in non-professional ones. It can be any combination of fluorescent, incandescent, and daylight, and is often considerably brighter than living-room TV-viewing levels. The addition of ambient light, to the extent that it is reflected by a display and thus represents a constant addition to the luminance output, will affect contrast[5]. The overall physical contrast ratio will be reduced (for example, a system that in the dark ranges from 0.8 cd/m^2 to 80 cd/m^2, and with a 5 cd/m^2 added ambient ranges from 5.8 cd/m^2 to 85 cd/m^2, will see its contrast ratio change from 100:1 to 15:1.) This reduction in overall physical contrast is accompanied by the visual effect that detail in the dark portions of the image can no longer be discerned; i.e. the image also has less apparent contrast.

Display technology issues include considerations like the following. LCDs, relying on polarization for light control, can be improperly measured by a photometer that is sensitive to polarization. On CRTs, the center of the screen is typically brighter than the edges. Viewing angle differences are quite strong for most LCDs. Displays may have a long warm-up period. Ambient light will reflect off of CRT phosphors in significant amounts. Characterization procedures need to obviate such problems.

Finally, the problem of display characterization is best understood as a problem of parameter estimation. The estimated parameters are always, if unconsciously, predicated on a model of display performance. The model must be correct; if it is not, the parameter estimates are meaningless. Likewise, if the estimation procedure is unreliable, the estimates must be regarded with suspicion.

2. BACKGROUND ON GAMMA

It is widely assumed among display users that the relationship between digital count (DC) input and luminance (L) output for a display is a power function,

$$L/L_{max} = (DC/DC_{max})^\gamma \hspace{3cm} (1)$$

where DC_{max} and L_{max} are maximum digital count and luminance values, respectively, and γ (gamma) is the power. The only parameter in this relationship is gamma. It is assumed that gamma is a fixed, fundamental, and perceptually important characteristic of a display.

The visual effect of manipulating gamma on a display is to change the apparent contrast of images. An image that looks right on a display with a lower gamma will lose detail in dark areas when presented on a display with a higher gamma. In the reverse situation, the grey tones are raised, resulting in a washed-out appearance.

Historically, for displays, gamma was measured as a property of the analog relationship of light input to signal output in the TV camera (set by NTSC standard to Signal = L$^{0.45}$), and of the relationship of applied cathode gate voltage to luminance output of a CRT (set by the physics of the electron guns and phosphors to the continuous power-law relationship L = voltage$^{2.5}$). The concatenation of these two functions implies a system gamma of 1.125 (gamma = (0.45)¥(2.5) = 1.125) for the relative relationship between natural scene values and screen values for broadcast television. This relationship applies to all television standards, NTSC, PAL, & SECAM. An image captured by a TV camera and sent to a CRT will not be a "linear" ($L_{out} = k¥L_{in}$) reflection of the original. The rationale for a system gamma of 1.125 instead of 1.0 has been described as making the picture look right in the dim TV viewing environment of the studio or living room by slightly pushing down the midtone greys and thus enhancing apparent contrast. A TV screen, of course, actually does have less overall contrast (smaller contrast ratio) than a real scene can have. A gamma of 1.125 relating L_{out} to L_{in} acts as a contrast-enhancing filter on most scenes because of its concave-upward shape as a transfer function[13].

In the analog domain of broadcast television the power relationship worked well, but in the desktop monitor environment the relationship is between digital counts (DC) and screen luminances (L), so there is an additional quantization of the signal that is unlike the standard television signal. The relationship between DC and L will reflect the transfer function used to digitally encode the image, software manipulations of the image, monitor

system LUTs and D-to-A converters, monitor settings, and factors inherent to the display technology used to manufacture the monitor. In other words, all bets are off: the relationship $L = f(DC)$ is determined by choices made by the manufacturers to differentiate their products and data processing manipulations and conventions, and not the physics of traditional CRT-based television standards. Also, computer monitors today are used in environments that are typically very different from the reduced ambient of the living room[10,7].

For the digital desktop monitor, as opposed to broadcast TV, the choice of transfer function to relate digital counts to luminances is actually a choice about the optimal assignment of bits to luminance regions. Still, the apparent contrast of an image is a potent perceptual factor in judging image reproduction quality. If DC-to-Luminance transfer functions vary across platforms, apparent contrast will vary, and image quality will be affected.

We reported[14] in 1998, as part of a study of color on the web, on the visibility of differences in images reproduced with different gammas defined as above in Equation 1. We found that images reproduced with a gamma of 1.9 were just distinguishable from images with a gamma of 1.8, and similarly 2.1 just distinguishable from 2.2. The addition of (simulated) ambient light at various levels did not change this sensitivity, although the images showed a loss of detail in the dark portions of the image, as if the gamma were changed, made higher, by the addition of the ambient. It is important to note that in the 1998 study, all image calculations were performed in luminance and rendered on a carefully characterized display using a lookup table that made no assumptions about the transfer function of the rendering device.

3. METHODS

Since the purpose of the present study was to examine the visual judgments of the general population of computer users in normal viewing environments, some internal validity (experimental control) was sacrificed for external validity (generalizability). Observers were chosen for variability in age, sex, color sensitivity, knowledge of vision science, and motivation for careful judgments. Photometric measurements were made by the observers themselves, with some exceptions. The ambient light had both a fluorescent lighting component and a variable daylight (through tinted film on glass) component. All the observers were daily users of computers at work or at school. The same procedures were followed for each. The CRT was a Sony Trinitron Multiscan300sf on an Apple G3 set to 1024x768 pixels and millions of colors, and the AMLCD was on an Apple G3 laptop set to 1024x768 pixels and millions of colors. All eight observers were tested on the CRT both in the light and in the dark, but only the first three were tested on the LCD because of the variability produced by small changes in viewing angle, as discovered during the observations[6].

Observers were first tested using the Farnsworth-Munsell 100 Hues Test for color sensitivity. Performance varied from superior through average to poor, and one observer was identified as a deuteranope. The performance of the deuteranopic observer, however, fell within the range established by the other observers on all the experimental tasks, including color matching of a blue.

Second, observers were asked to set the brightness and contrast levels of the display "to optimum values." These controls are understood to have the effect of raising and lowering the Luminance vs. Digital Count (DC) transfer function as a whole and stretching the high end up or down, respectively. The "dark light" level is also changed by these adjustments, and can be far above the minimum possible value (typically less than one nit.) The overall "gain" can be changed, and the overall maximum luminance can be changed under the user's control[7]. A common (although not necessarily the best) procedure is to display a very dark grey on a black background, and ask the observer to adjust brightness until the grey is just discernable from the black[8]. Next, a very light grey is shown on a white background, and contrast is adjusted until the grey is just discernable from the white. This may be done iteratively until an optimum setting for brightness and contrast is achieved. In the present case, a large black (DC = 0) rectangle with eight smaller rectangles (DC = 4, 8, 12, 16, 20, 24, 28, 32) was shown for the adjustment, and the judgment made between the darkest (DC = 4) small rectangle and the background. Similarly, a large (DC = 255) and smaller (DC = 251, 247, 243, 239, 235, 231, 227, 223) rectangle image was used for the white adjustment.

On the CRT, the observers were easily able to adjust the brightness knob for the black level as required. The contrast knob, however, had no useful effect on the judgment of the white. Contrast was simply set to maximum, and the very light grey was easily distinguished from white. On the LCD, only a brightness adjustment

was possible, and its action was to dim or brighten the backlight. In practice, it had no useful effect on the adjustment of the black nor white levels for any of the observers, and was simply set to maximum.

Third, observers viewed a 256-step ramp that went from black (DC = 0) to white (DC = 255) in 9-pixel-wide steps. They were asked to report any banding (visible edges), where it occurred, how visible it was, and how wide the bands were. On the CRT, narrow bands show differences between adjacent digital count levels; broader bands appear when some step differences are significantly larger than others, possibly due to D-to-A conversion implementation. On the LCD, there is no standard for producing all 256 greyscale levels, and thus we did not evaluate banding on this device[9].

Fourth, observers made perceptual estimates of gamma using a common brightness-matching procedure. Gamma is the exponent in the putative power function relationship between DC input and luminance output of a display. A DC image reproduced on a lower gamma display will look washed out, and reproduced on a higher gamma display will look muddy, compared to reproduction on the intended display. It is widely assumed that gamma is a fixed characteristic of a display and that much of the important information needed to characterize a display is communicated by identifying the value of gamma[10,11].

The brightness-matching procedure involved comparing grey squares at varying DCs to a background composed of 50% white and 50% black pixels, and deciding which grey square "disappeared into" (matched) the background best. The logic of the procedure is that the luminance of the black and white background should be half the maximum, so that by inserting a relative luminance of (.5) into the power relationship with the relative DC (i.e. DC / 255) identified as matching the background, one can solve for the exponent, gamma. In this study, observers made their judgments against four different backgrounds: alternating black and white horizontal lines, alternating black and white vertical lines, a 2x2 pixel checkerboard (the smallest size) and a 4x4 pixel checkerboard (i.e. the 2x2 checkerboard doubled.)

Four spatial patterns were tested because the gamma characterization procedures proposed on different web sites use different patterns, but more importantly because of the effects of the adjacent-pixel non-linearity for CRTs: horizontal black-and-white lines are typically brighter than vertical ones on a CRT[12]. Horizontal lines or line segments are widely, but not exclusively, suggested for the gamma estimation procedure.

Fifth, observers were shown the four background patterns used in the gamma estimation, and asked to rank order them for brightness.

Sixth, observers were asked to make a color match between a colored paper and a colored square on the screen. The match was made only in the light, of course, and care was taken that the colored paper and the monitor screen were at the same angle relative to the room lighting. The colored paper used was Sky Blue on the Macbeth Color Checker, and Adobe Photoshop was used to alter the color of the screen square to make the match. The color checker was propped against the monitor screen so that it occluded part of the colored screen region; this meant that the black border of the color checker intervened between the two areas being matched. In addition, the remainder of the screen was not put to black; thus, adaptation could be influenced by both the room lights and the monitor white point. The Photoshop alterations to the screen square were made at first by the observer until a crude match was achieved, and then the experimenter used the software in response to the observer's description of the differences between the screen and paper. Observers were told that they were to match color and brightness, but not to expect that the two squares would look identical, because of texture and other differences.

Seventh and finally, the observer made several photometric measurements using a Minolta CS-100 Colorimeter. The "luminance" and color coordinates for the Sky Blue paper, and the luminance and color coordinates for its screen match were measured. White through black were measured at DC = {255, 251, 192, 128, 64, 32, 4, 0}. The monitor was then turned off, and the black screen was measured.

4. RESULTS

Brightness Setting On the CRT, observers set the brightness control to a higher value in the light than they did in the dark, as shown in Figure 1. If the brightness knob is an offset that displaces the entire transfer function up

or down, this is the opposite result from what is expected. When the room lights are on, the absolute difference for DC = 0 vs. DC = 4 will not have changed, yet the higher background luminance of the screen due to the ambient should require a larger increment for discrimination, by Weber's Law. If the displacement model is correct, setting the brightness level higher should only intensify this effect, by raising the background further without changing the increment. We conclude that the action of the brightness knob is not an offset. The photometric measurements at these low light levels are not completely reliable, and thus cannot confirm the conclusion that the displacement model is incorrect for this monitor.
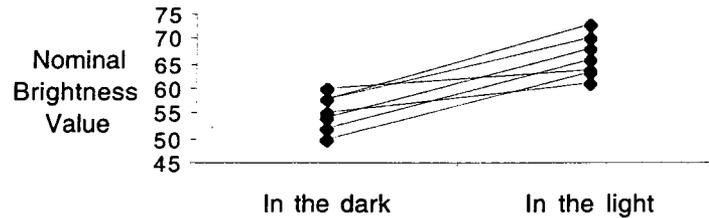

Figure 1. CRT brightness settings, in the dark and in the light. The brightness adjustment is manipulated and the nominal brightness value is read from an on-screen display. A higher number corresponds to a higher brightness setting. Results for the eight observers are plotted individually.

Banding On the CRT, banding was generally more visible in the dark than in the light, as expected, because small increments should be more visible on a smaller pedestal. In the light, the banding was most obvious in the mid-tones. In the dark, banding was generally visible everywhere, most obvious in the dark greys and least obvious in the light greys.
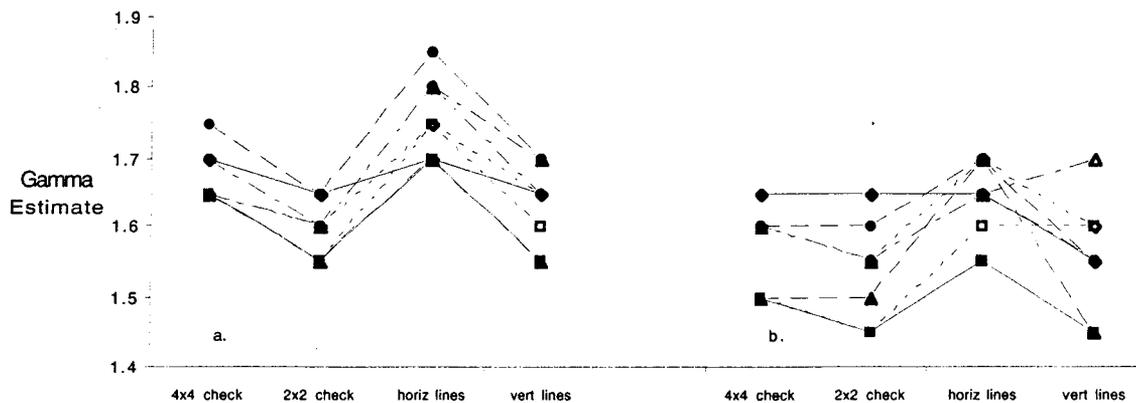

Figure 2. Perceptual gamma estimates by background pattern for eight observers on the CRT, measured in the dark (a) and in the light (b)

CRT Gamma: Perceptual Measurements in the Dark and in the Light Figure 2 shows the results of the perceptual judgments of all observers, for each background, on the CRT, in the dark (a) and in the light (b). In the dark, for all observers, the judged gamma is highest for the horizontal-line background and lowest for the vertical lines and 2x2 checkerboard, as predicted by the adjacent-pixel non-linearity for CRTs. In the light, this pattern is less clear, although the horizontal-line background still produces the highest gamma estimates. This finding was supported by the rank orderings of the brightness of the background patterns; observers consistently ranked the

horizontal-line background as the brightest, and the vertical lines and 2x2 checkerboard as the darkest. Note that the variation across backgrounds, in the dark vs. light, and among observers, is greater than the 0.1 unit threshold found for 1.8 and 2.2 in the Gille et al. study; that is, these are probably not negligible differences.

Figure 3a shows all 32 pairs of observer gamma judgments (it seems like fewer because there were many duplicate judgments across observers.) Comparing the perceptual gamma estimates in the dark and in the light, we see that the estimates are lower in the light, not only inconsistent with the presumed loss of detail in the shadows for images in the light compared to in the dark, but also inconsistent with the idea that gamma is a fixed characteristic of a given display. If the effect of the ambient is purely additive, and the transfer function in the dark is a true power function, then the perceptual estimate of gamma in the light should be the same as the estimate in the dark.

Figure 3a.
Perceptual Gamma Estimates, 32 pairs
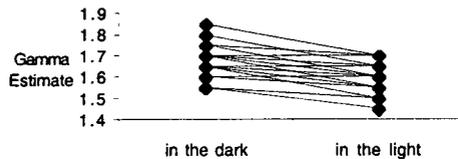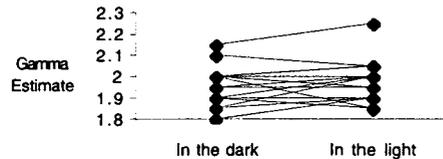Brightness setting adjusted

Figure 3b.
Perceptual Gamma Estimates, 24 pairs
Brightness setting held constant

However, there were two changes made to the display between dark and light conditions: not only was ambient added or not, but also the brightness control was altered by the observer. To disambiguate these effects, two more measurements were made, holding the CRT brightness setting constant.

The first was a physical measurement of average background luminance for the four patterned backgrounds, i.e., 4x4 and 2x2 checkerboards and vertical and horizontal lines, in the light and in the dark, compared to the average of the maximum (DC = 255) and minimum (DC = 0) display luminances. The results are shown in Figure 4. The horizontal-line background was closest to the average of black and white. (Note that the average is not the same as half the maximum.) Simple arithmetic indicates that this is expected if the transfer function from DC to L is a power function plus a constant increment equal to the minimum black luminance .
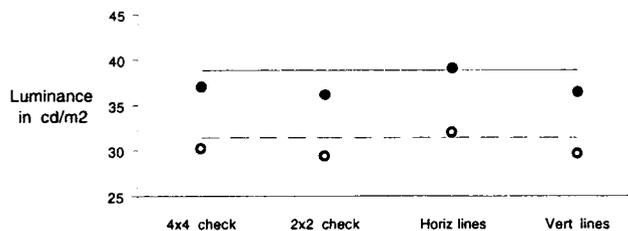

Figure 4. On the CRT, the luminance of four background patterns (circles) compared to the average of black and white (lines), in the dark (open circles, dashed line) and in the light (closed circles, solid line).

The second set of measurements was gamma estimates across backgrounds, in the dark and in the light, with the CRT brightness setting held constant, for 6 observers, some of whom had participated in the basic study and some of whom were new observers. The results are shown in Figure 3b. In this case, where the screen brightness adjustment was not manipulated, gamma estimates in the dark and in the light were the same (mean estimates of 1.96 and 1.97, respectively). The changes in perceptual gamma estimates must have been due to the brightness adjustment changes, and this is inconsistent with the assumption that the brightness adjustment simply shifts the

transfer function up and down. Note that the wide variability in estimates is due not only to individual differences, but also to the background differences.

CRT Gamma: Models of CRT Response Measured in the Dark   There are three models that have been applied to characterize CRT monitor response measured in the dark (i.e. no ambient). The first model, the "good enough" model in Equation 1, characterizes the monitor to within measurement error as a power function:

$$L/L_{max} = (DC/DC_{max})^\gamma$$

The second model, the "dark light" model, recognizes that the monitor output with zero-value input is in fact never zero, but some dark light value (DL), which presumably simply shifts the power function upwards in luminance. We saw above that this model might be appropriate for CRTs with added ambient light:

$$L/L_{max} = (DC/DC_{max})^\gamma + DL \qquad (2)$$

The third model, which we will call the "three parameter" model, is based on the widely-accepted CRT characterization by Berns, Motta and Gorzynski[15], which includes gain, offset, and power (gamma) parameters:

$$L/L_{max} = (Gain \yen DC/DC_{max} + Offset)^\gamma > 0 \qquad (3)$$

and $L/L_{max} = 0$ otherwise.

This model is based on their "understanding of generic signal processing of computer-controlled displays and of CRT metrology", and the characterization has been incorporated into the sRGB[16] specifications. Specifically, the sRGB standard for the DC to L transfer function is defined as:

$$L/L_{max} = ((DC/255 + 0.055) / 1.055)^{2.4} \qquad (4)$$

The function in Equation 4 is virtually identical[17] to the "good enough" power function with gamma equal to 2.2, a point that we will return to later.

$$L/L_{max} = (DC/255)^{2.2} \qquad (5)$$

Notice that these three models have one, two, and three parameters, respectively. This means that model fits to photometric data will have to be evaluated not only in light of the adequacy of model, but also in the degrees of freedom available to improve fit[18]. Also, the second and third models cannot be described as an instance of the general linear model (even using a log transformation), and the two or three parameters must be estimated using numerical methods. We used the fmins() function in Matlab, which uses the Nelder-Mead simplex (direct search) method, and least-squares estimation in our fitting procedures. The fmins() function requires initialization values (start values) for the parameters in the fitting procedure. The multivariate structure relating parameter values to the measure of fit can be thought of as a surface whose lowest point is being sought. This surface need not be monotonic, and the search procedure may find a local minimum instead of the desired global one. Common practice is to run the fitting procedure several times with different initialization values, hopefully avoiding misidentifying local minima as the global minimum. We have found that the least-squares surfaces for the third model, especially, can have many local minima, casting doubt on the estimation process, or at least making it not very practical.

As an alternative and non-equivalent procedure, a straight line could be fitted to only the straight-line portion of the log-log plots, as in the definition of film gamma[19]. This is an attractive procedure for many reasons, but of course the straight-line portion must be carefully chosen, because in the log-log space, errors at the low end are magnified in importance compared to errors at the high end. In this study, we used the log-log plots to identify regions of the transfer function that can be well fit by a power function, and then we restricted the fitting procedure to that region.

a. in the dark   b. in the light

Gamma Estimate

4 —
3.5 —
3 —
2.5 —
2 —
1.5 —
1 —

perceptual, horiz lines   "good enough"   "dark light"   "three param"   perceptual, horiz lines   "good enough"   "dark light"   "three param"
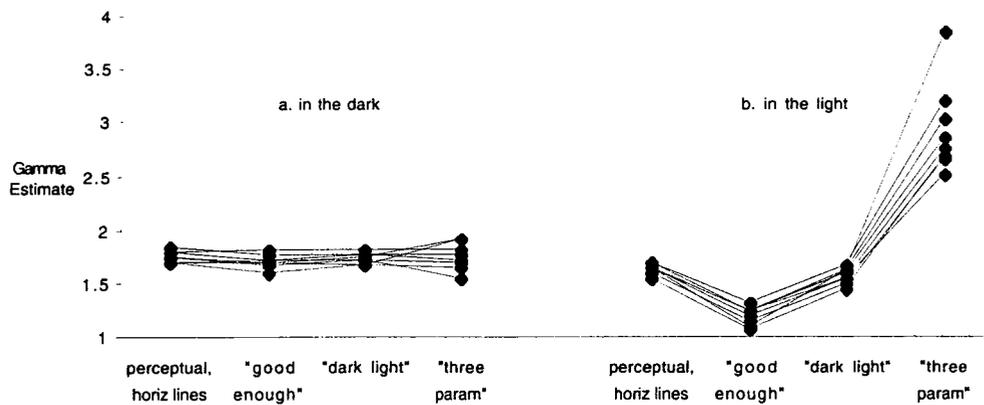
Figure 5. Gamma estimates for the CRT transfer function, from perceptual judgments using horizontal lines and from fits of three models to photometric data, measured in the dark (a) and in the light (b), for eight observers.

<u>CRT Gamma: Estimates Derived from Photometric Measurements in the Dark</u> Figure 5a shows the results of gamma estimation using the photometric data from each observer for each of the three models compared to the perceptual judgments from the horizontal—line background, in the dark. The figure is scaled for direct comparison to the data in Figure 5b, but generally the estimates are rather stable and differ across models by no more than .2 units. Remember that individual observers adjusted the monitor to different brightness settings, so the estimates based on photometric data are expected to vary across observers as the perceptual estimates did.

To further analyse the shape of the DC-to-Luminance transfer function in the dark, careful measurements were made at a given brightness setting for DC = {0, 15, 31, 47, 63, 79, 95, 111, 127, 143, 159, 175, 191, 207, 223, 239, and 255}, and the measurements shown in Figure 6a along with the three best-fitting instances of each model. The gamma estimates were consistently 2.1. The perceptual judgment was simulated by interpolating the DC value closest to the average of black and white, again obtaining 2.1. In Figure 6b, the log-log plot of Luminance vs. DC shows deviation from linearity only at DC = 0 (where the photometric measurements are noise-limited and therefore unreliable); the slope of the line is 2.1.
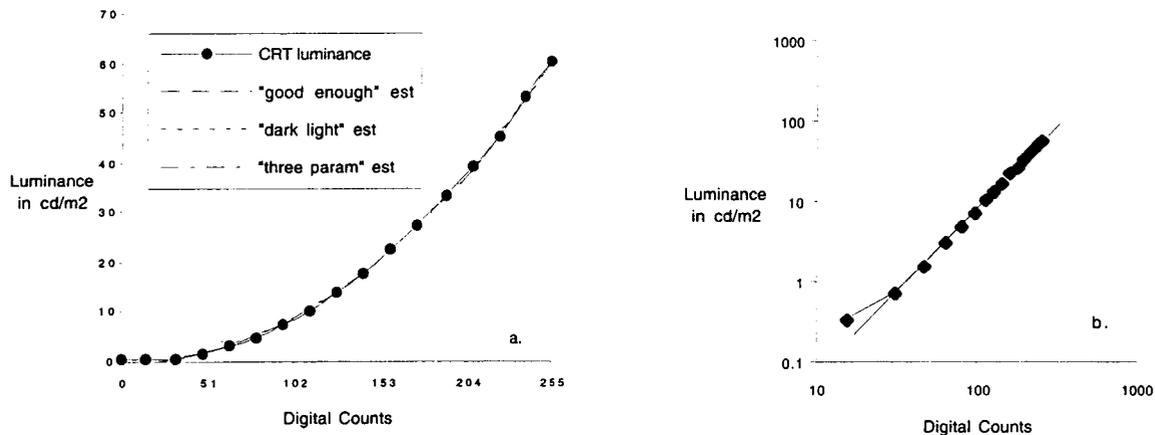

Figure 6. (a) CRT transfer function, measured in the dark, with fits from three models.
(b) Log-log plot of the same transfer function; line fitted by eye.

CRT Gamma: Estimates Derived from Photometric Measurements in the Light Figure 5b again shows the results of gamma estimation using the photometric data from each observer for each of the three models compared to the perceptual judgments from the horizontal—line background, this time in the light. The perceptual estimates and the estimates from the dark light model are about equal. The estimates from the "good enough" model are lower than the perceptual estimates, because they are forced through zero and are left to fit the upper part as well as possible, with the result that the curve is less bowed than it would be otherwise. The estimates from the "three parameter" model are quite variable, suggesting an unwanted instability of estimation.

Once again, to further analyse the shape of the DC-to-Luminance transfer function in the light, careful and more extensive measurements were made at the given brightness setting, and the results shown in Figure 7a along with the three best-fitting instances of each model. None of the models can fit the in-the-light data as well as it did the in-the-dark data. The associated gamma estimates are 1.7, 2.2, and 5.7 (!). Working backwards to predict the perceptual gamma that would be found for these data, we obtain 2.1, as expected from the measurements in the dark.

The "good enough" model provides the poorest fit. The "three parameter" model fits best, but produces an unlikely estimate for gamma. The dark light model gives the best combination of fit and reasonable gamma estimate. Three different models, all with gamma as a parameter, produce three significantly different estimates on the same real data. Which is the "real" gamma[7]?
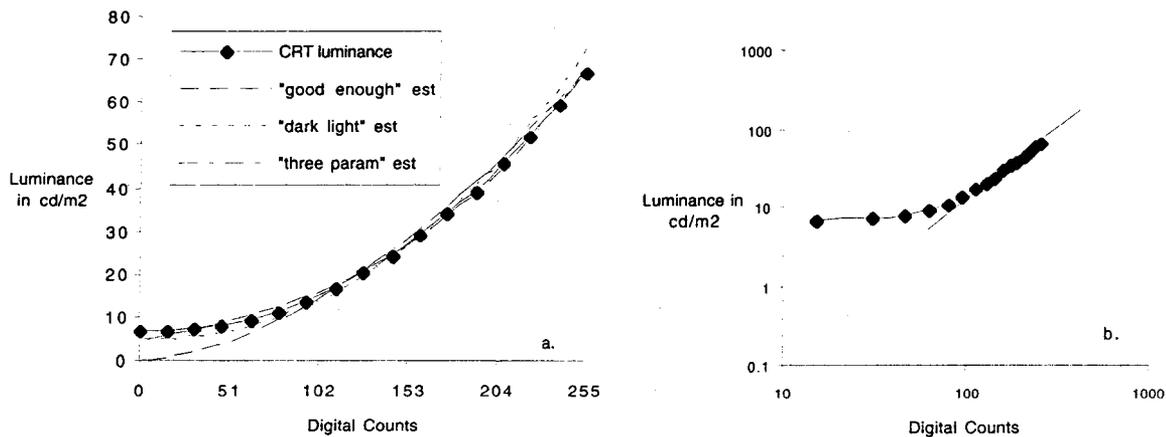

Figure 7. (a) CRT transfer function, measured in the light, with fits from three models.
(b) Log-log plot of the same transfer function; line fitted by eye.

Figure 7b shows the log-log plot of Luminance vs. DC. The deviations from linearity are apparent at the low end, where the offset due to ambient swamps the CRT output. Gamma estimates were calculated for the ten upper points of the data (linear portion of the log-log plot) for each of the three models, and the results shown in Figure 8. The gamma estimates for the first two models are 1.77 and 1.73, and are well-constrained. The fitting of the third model was extremely sensitive to start values in the numerical analysis; excellent fit was obtained, for instance, with a gamma of —99. An outcome of 1.75 is reported for the three parameter model only for the trivial reason that the value was closest to the others. All three of these fits, however, fail to detect, in the truncated data set, the offset that is in the complete data set, and have intercepts close to zero. The slope of the linear portion of the log-log plot was also 1.77.

From the measurements in the dark and the offset measured at DC = 0 in the light, we could hypothesize that the true function is the dark light model with gamma = 2.1 and offset = 6.84. This function is also plotted on Figure 8 ("perceptual" est), and is the only one that provides a good fit to the data. Thus, the best empirical fit to these data would use the "perceptual" gamma estimate in the model that is a power function plus an offset, of the form of Equation 2.
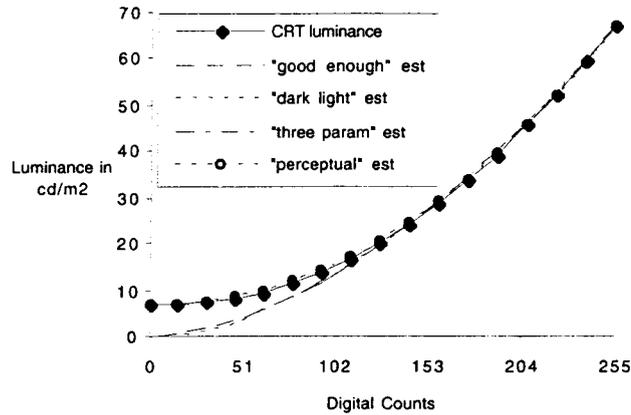
Figure 8. CRT Transfer function measured in the light, with fits to the truncated data set from four models.

LCD Gamma: Perceptual Measurements in the Dark and in the Light   The gamma estimates for the three observers in the light and in the dark are plotted in Figure 9. The wide variability among observers is attributed to viewing angle differences[20]. Gamma estimates were for the most part the same across backgrounds, consistent with the absence of the adjacent-pixel non-linearity of a CRT. Gamma estimates were the same in the light and in the dark, consistent with not having changed the brightness adjustment.
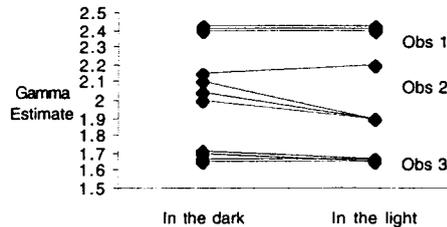
Figure 9. Perceptual gamma estimates, LCD, in the dark and in the light; 12 pairs with slight offsets for visibility.

LCD Gamma: Estimates Derived from Photometric Measurements   The photometric measurements made by the observers in the light and in the dark, and the estimates of gamma derived from them, were quite variable. The photometer was handheld and the LCD used has a strong angular dependence. Newer liquid crystal displays are beginning to appear in the marketplace that are much more isotropic than the display used here, and viewing angle dependence is not an insurmountable problem for liquid crystal displays. However, because of the noise in the observers' measurements, those data are not reported here. Nevertheless, the form of the transfer function, which is presumed to emulate the power function of a CRT, is of interest[20].

More extensive measurements of the LCD transfer function at a fixed viewing angle were made in the dark. That function and the best fits by the three models is plotted in Figure 10a. The most striking feature here is that the actual transfer function cannot be a power function, because of the decelerating portion at the high end. The gamma estimates of 2.0, 2.0 and 1.5 from the three models are essentially meaningless because of the poor fits. The presumed perceptual gamma estimate would be 1.9, and the fit from that estimate is also plotted in Figure 10a.

The log-log plot of the LCD transfer function is shown in Figure 10b. The most linear portion is the middle 11 points, eliminating the bottom two and top four. When the three models are fit using only these points, the results are gamma estimates of 2.8, 2.7, and 2.8, and good fits up to about DC = 210, then sharply diverging to a predicted maximum of about 170 cd/m^2. The slope of the linear portion of the log-log plot was 3.0.
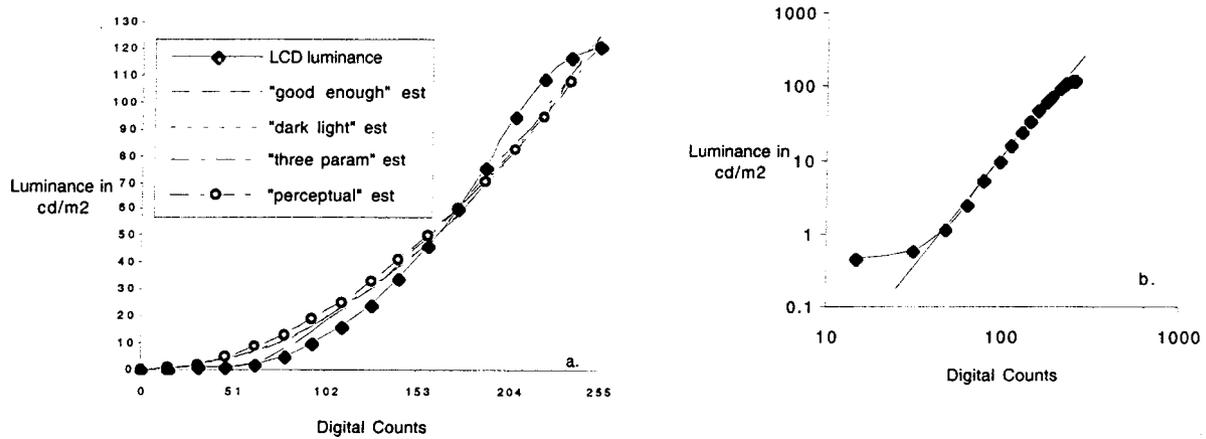

Figure 10. (a) LCD transfer function, measured in the dark, with fits from four models.
(b) Log-log plot of the same transfer function; line fitted by eye.

The LCD and CRT displays differed in many ways. The LCD compared to the CRT under the conditions of this study had a higher maximum white luminance (120 vs. 60 cd/m^2), a higher overall contrast ratio in the dark (300:1 vs. 200:1), and a different effect of ambient (addition of 5 cd/m^2 vs. 6 cd/m^2, i.e. 4% vs. 10% of maximum display luminance). A discussion of why these and other differences might lead to the choice of a different gamma or transfer function model is beyond the scope of this paper.
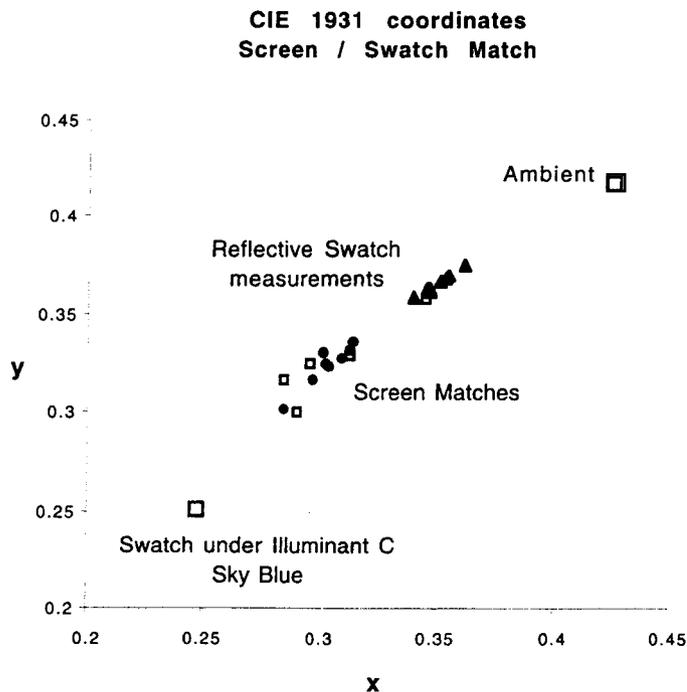

Figure 11. Results of the color-matching task on the CRT, in CIE 1931 coordinates.

CRT Color Match Figure 11 shows the results of the screen-to-colored-paper color match for the eight observers. The measured CIE 1931 color coordinates of the ambient light and the calibration values for the colored paper under Illuminant C are labeled and shown as large open squares. The measured values for the colored paper are labeled and shown as filled triangles, and the measured screen matches are labeled and shown as filled circles. Also plotted, as small open squares, but not labeled, are: screen white in the dark ~(.28, .32), screen white in the light ~(.30, .32), D50 ~ (.35, .36), D65 ~(.31, .33), and 9300 deg K ~(.29, .30).

When shown the Macbeth Color Checker in the light but away from the monitor screen, observers described the Sky Blue paper as "blue" or "grey-blue", and the grey sequence as "grey". Against the monitor screen, the Sky Blue paper was described as "battleship grey" or "grey", and papers in the grey sequence as "tan" or "taupe". Later in the experimental sequence, when observers made photometric measurements of screen black through the colorimeter, which limits field of view considerably, several exclaimed, "that's brown!" That is, the observers exhibited color constancy ordinarily, the screen contributed to adaptation, and the ambient light was a strong colorant.

The results show that observers exhibited neither perfect color constancy (matches do not fall at the Sky Blue locus), nor the ability to make a perfect color match (matches and actual color coordinates are well separated.) Observers found the task very difficult, and often complained that the color of the paper seemed to change as they made their matches. Similar results were found for the LCD[21,22,23].

5. CONCLUSIONS

Monitor adjustments, when available, are important, but their actions and their effects on transfer functions may not be simple transforms. If settings are changed, the monitor should be characterized again.

On the CRT, **banding** was visible, and varied from DC region to DC region. This argues that the transfer function was not perceptually uniform, and that perhaps 256 levels of grey without dither are not sufficient for a perfectly smooth ramp on the display used. Banding was less apparent in the light than in the dark, as expected from the reduced contrast.

The concept of **gamma** is not well defined, yet as a possible measure of apparent contrast, it is perceptually salient, and therefore should have a consistent physical definition that relates to the perceptual variable. In this study, we looked at several problems relating to the use of gamma to characterize displays.

First, different common models of display transfer functions can have very different values for gamma and yet describe virtually identical functions. The published example of this is the sRGB standard[16,10,17], where the standard transfer function, using the Berns et al. model with a gamma of 2.4, is indistinguishable from the "good enough" model with a gamma of 2.2 (Equations 4 and 5.)

With added ambient, these effects (estimates differing with different models) were pronounced. (Note that the sRGB standard does not account for the levels of added ambient found in this study[10], although they are typical for the workplace.) At various points in this study, we used six different methods to estimate gamma. One was perceptual, matching a grey patch to a black and white background. The other five were all based on photometric data. The three power-law models, "good enough", "dark light", and "three parameter", were fit to the data by numerical analysis. Log-log plots of the data were examined for linear regions (i.e. regions that might obey a power law), and the models applied to the restricted regions. Finally, the perceptual judgment was simulated by interpolating the DC value closest to the average of black and white. These methods did not consistently agree, and with added ambient produced very different results. Again, which model/method gives the "real" value?

Second, within the numerical analyses, the three-parameter model was over-parameterized, and could produce very different estimates that all fit the data about the same. The numerous local minima required numerous start values, and the process was not practical, although it is based on the most well-founded of the three models.

Third, although a power function was a good characterization for our CRT transfer function, it was not at all for the LCD. Since an LCD *could* emulate a CRT, one can imagine that the best assignment of bits to luminance regions (best transfer function) may vary among technologies, or that the power function that is native to the CRT is not actually optimum in some sense.

In any event, the most consistent description of the transfer function and estimate of gamma for our data was as follows, and was not surprising. For a reasonably adjusted CRT, in the dark, a simple power function fits the photometric measurements well. The estimates of gamma from the simple model and perceptual matching of grey patches to horizontal lines (lines in the scan direction) are congruent. For a reasonably adjusted CRT, in the light, a simple power function plus additive offset fits the data well, and again the gamma estimates from that model and perceptual matching are congruent. Within this framework, there will still be variation among individuals in their judgments.

Using this "most consistent" definition, we found that gamma was the same in the dark and in the light, although it changed with changing monitor adjustments. This meets the assumption that gamma is a fixed property of a display, at least for a given adjustment. However, it fails to capture the perceptual variable of interest that makes gamma important: apparent contrast. As ambient light is added, apparent contrast is reduced. Could manipulating gamma compensate for this effect, as was done with broadcast TV for adaptation effects? What parameter does capture the concept of apparent contrast? Should we be looking for a transfer function other than the one native to CRTs?

Finally, our **color-matching** task highlighted what is already known: matching real-world objects to screen images is a very complex business. The color matching has to be done in the light for the reflective object, which means that the illuminant has an effect not only on the color coordinates for the object, but also on the screen white point. Color constancy and chromatic adaptation, here due to the two light sources of screen and ambient, are strong, competing attributes of perception.

6. DISCUSSION

How good is the human eye at characterizing displays? Quite good, for the tasks that it is commonly given. All observers were able to adjust monitor brightness appropriately and such that the low end did not bottom out, in both the dark and in the light. However, the effect of the brightness adjustment was not an additive offset, as supposed. The perceptual estimate for gamma, on a CRT, assuming the two-parameter transfer-function model (power function plus additive offset), and using horizontal lines to generate the half-luminance (actually average luminance), was simple, quick and accurate both in the dark and in the light. In some sense, it was superior to photometric measurements.

However, other important transfer-function variables, notably the offset and maximum luminance, were not measured by the eye. The eye simply can't make those kinds of absolute judgments alone[24], and yet these are important variables affecting apparent contrast, a salient factor in image appearance, and the same factor that the value of gamma affects. Ambient light, which is not under the control of the display system, is the major contributor to the offset.

On our LCD, the perceptual estimate of gamma was meaningless, both because the transfer function decelerated at the high end and because of the viewing angle dependence of the display we used in the experiment. Future liquid crystal displays are expected to have better viewing angle performance and therefore may be amenable to this methodology in the future. We did not attempt to relate the gamma of the power function that characterized the transfer function in the region from DC = 0 to DC = 210 to apparent contrast.

The visual check for banding in a luminance ramp, while somewhat casual, did show the sensitivity of the eye for this type of task. The 256 levels of the CRT were not enough for a smooth percept, and even non-uniformities in relative step size were discernable. Predictably, banding was less apparent in the light.

What we haven't shown in this study is that a power-law transfer function is perceptually uniform[25], nor that there is an optimum value for gamma[26]. In fact, the confounded effects of ambient and gamma on apparent

contrast suggest that there probably is no single best value for gamma, despite its historical convenience for analog CRT systems. In early broadcast TV, gamma correction solved a technology problem; then the incomplete gamma correction addressed a perception problem in a simple and effective way, but one that relied for its success on the use of a single technology in rather uniform settings. The inherent CRT transfer function was found by happy accident to be adequate for the assignment of digital counts to luminance regions on a CRT, but again this relies on a single technology and particular viewing conditions. The real goal of transfer-function manipulations has always been to improve image appearance, largely by controlling apparent contrast. It would seem unnecessarily self-limiting to identify a clever solution or serendipity with fundamental perceptual principles, in a time of emerging display technologies.

Finally, the color matching task that we chose was not simple, not quick, and not conclusive.

ACKNOWLEDGMENTS

REFERENCES

1. MacDonald, Lindsay W., (2000) Assessment of monitor calibration for Internet imaging. *Internet Imaging*, Giordano B. Beretta, Raimondo Schettini, Editors, *Proceedings of SPIE Vol. 3964*, 162-167. Contains a detailed introduction to the problems of monitor characterization.

2. Hunt, Robert W. G., (2000) How to shop on the web without seeing red. *IS&T/SID Eighth Color Imaging Conference: Color Science, Systems, and Applications*, 2-7.

3. Lavin, Yingmei, Silverstein, A. and Zhang, X. (1999) Visual experiment on the web, in *Human Vision and Electronic Imaging IV*, Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors, *Proceedings of SPIE Vol. 3644*, 278-289. Discusses many of the monitor characterization issues dealt with in the present study. Of particular interest is the more extensive perceptual gamma estimation procedure, where the transfer function is measured at several points, rather than just one.

4. Wandell, Brian A. (1995) *Foundations of Vision* (front endpaper). Sinauer Associates, Inc., Sunderland, Massachusetts. See: http://white.stanford.edu/html/numbers/numbers.html

5. What is here called "ambient" is termed "ambient flare" or "veiling glare" when applied to CRTs. See: Holm, Jack (2000) Photographic and Imaging Manufacturers Association, Inc. PIMA 7667 Working Draft 2.0. Photography — Electronic Still Picture Imaging — Extended sRGB color encoding — e-sRGB. http://www.pima.net/standards/it10/PIMA7667/PIMA7667_WD2.PDF

6. Today s dominant liquid crystal technologies require careful and precise control of viewing angle to produce consistent results of photometric properties. These controls are rarely obtained in the work environment, so procedures that require this level of precision are not useful in that setting.

7. Poynton, Charles (1998) The rehabilitation of gamma, in *Human Vision and Electronic Imaging III*, Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors, *Proceedings of SPIE Vol. 3299*, 232-249. See also http://www.inforamp.net/~poynton/notes/brightness_and_contrast/index.html

8. This works because it is then guaranteed that the function is not "bottomed out", at least not at or after DC=4 (or whatever small value is used). The brightness control on a CRT with a viewable edge to the rastered region of the screen can be manipulated to make this edge (between the rastered region set to black and the unrastered region) invisible; however, this setting will not produce the maximum peak brightness. See: http://www.aim-dtp.net/aim/calibration/blackpoint/crt_brightness_and_contrast.htm

9.   On the LCD, there is no standard for producing all 256 greyscale levels. Some devices employ a 6 bit DAC for the 6 most significant bits and use spatial dither to generate the 2 least significant bits; others use an 8-bit DAC, and still others use hybrid systems that combine voltage rails and resistor networks to directly control the fields delivered to the liquid crystal pixel. Because of this variety of driving schemes for liquid crystal displays, banding was not evaluated on our LCD device.

10.   Stokes, Michael, Anderson, M., Chandrasekar, S. and Motta, R., (1996). A Standard Default Color Space for the Internet — sRGBhttp://www.w3.org/Graphics/Color/sRGB

11.   Stone, Maureen, Newman, T., and Callow, M. (1997) It may be Gamma to you, but it's Greek to me. http://www.parc.xerox.com/red/members/stone/vrml-cfwg/gamma/default.html

12.   Horizontally adjacent pixels can affect each other's brightness on a CRT, because the horizontal bandwidth of the CRT depends upon the properties of the video amplifier. CRTs with an inadequate amplifier are unable to produce a 100% modulation of pixel brightness in horizontally adjacent pixels when the monitor is set to its highest spatial resolution. The vertical bandwidth is less critical because the temporal frequency response of the video amplifier need only be good enough to modulate at line times which are between two and three orders of magnitude less than the horizontal video amplifier bandwidth requirement. In particular, alternating white and black pixels within one row of the raster scan will produce less intensity variation and therefore darker whites than the same spatial variation oriented vertically. For this reason, white and black alternating columns of pixels are generally darker on average than alternating black and white rows. Therefore, using a relative luminance of (.5) in the power function would be incorrect when vertical bar patterns are used with this procedure. See: Klein, Stanley A., Hu, Q.J., and Carney, T. (1996) The adjacent pixel nonlinearity: Problems and solutions. *Vision Research, 36,* 3167-3181.

13.   Schade, Otto H. Sr. (1975) *Image Quality - A Comparison of Photographic and Television Systems.* RCA Laboratories, Princeton, NJ.

14.   Gille, Jennifer, Luszcz, J. and Larimer, J. (1998) Error diffusion using the web-safe colors: how good is it across platforms? in *Human Vision and Electronic Imaging III,* Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors, *Proceedings of SPIE Vol. 3299,* 368-377.

15.   Berns, Roy S., Motta, R. J. and Gorzynski, M. E., (1993). CRT colorimetry, part I: theory and practice. *Color Res. Appl., 18,* 299-314.

16.   http://www.srgb.com/sRGBstandard.pdf; IEC/3WD 61966-2.1: Colour measurement and management in multimedia systems and equipment — Part 2.1: Default RGB colour space — sRGB

17.   http://www.srgb.com/sRGBGammaCalculation.pdf

18.   The third model is reduced to a two-parameter model if it is "pinned" at the maximum luminance (at 255, $L/L_{max} = 1$ and $DC/255 = 1$, so that Gain = (1 — Offset)), but we did not use that procedure here. The third model can also be extended from a two- or three-parameter model to a three- or four-parameter model by an additive offset, as shown and recommended in the following references (18a, 18b). Since the conclusions we draw related to the perceptual judgments would not be altered by including yet another model with three or four parameters to fit, we have opted for showing the results of indiscriminately applying the third model.

18a. Roy S. Berns and Naoya Katoh (1997) CIE Expert Symposium '97 Colour Standards for Imaging Technology 21-22 November 1997. The Digital to Radiometric transfer function for computer-controlled CRT displays. Two slightly different presentations online: http://www.cis.rit.edu/people/faculty/berns/PDFs/cie_CRT.pdf http://www.cis.rit.edu/people/faculty/berns/PDFs/cie_pers.pdf

18b.   Roy S. Berns (1997) Response to IEC CD 1966-3 Colour Measurement and Management in Multimedia Systems and Equipment, Part 3: Equipment using cathode ray tubes. http://w3.hike.te.chiba-u.ac.jp/IEC/100/PT61966/makuhari/berns_re.pdf

19.   *RCA Electro-Optics Handbook*, (1974). RCA Technical Series EOH-11.

20.   Wright, Steven L., Ho, K., and Lien, A., (2000) Status of TFTLCD color and metrology. *IS&T/SID Eighth Color Imaging Conference: Color Science, Systems, and Applications*, 301-304.

21.   Braun, Karen M., and Fairchild, M. D., (1996) Psychophysical generation of matching images for cross-media color reproduction. *IS&T/SID Fourth Color Imaging Conference: Color Science, Systems, and Applications*, 214-219. Contains analyses of competing influences in mixed adaptation.

22.   Henley, Sharron A., and Fairchild, M. D., (2000) Quantifying mixed adaptation in cross-media color reproduction. *IS&T/SID Eighth Color Imaging Conference: Color Science, Systems, and Applications*, 305-310. Contains analyses of competing influences in mixed adaptation.

23.   Brainard, David, and Ishigami, K., (1995) Factors influencing the apearance of CRT colors. *IS&T/SID Third Color Imaging Conference: Color Science, Systems, and Applications*, 62-66.

24.   This discussion is not completely accurate. Although we haven't shown it here, what is of greatest importance is not the absolute light levels, but the overall contrast ratio, white to black. One can imagine a method of judging overall contrast ratio by eye in the dark, using a ruled neutral-density wedge that covers a white (DC = 255) region of the screen and abuts a black (DC = 0) region so that a judgment of equal brightness can be made. An implementation in the light would be much more difficult.

25.   The perceptual uniformity of the CRT transfer function is a notion that has permeated the literature but that warrants examination; see Poynton (op. cit.), Stokes (op. cit.) and Stone (op. cit.).

26.   Similarly for the idea that gamma = 2.2 is perceptually optimum; see Stokes (op. cit.), and Stone (op. cit.).