Vision Res. Vol. 31, No. 7/8, pp. 1337–1350, 1991
Printed in Great Britain.

0042-6989/91 $3.00 + 0.00

ACCURATE CONTROL OF CONTRAST ON MICROCOMPUTER DISPLAYS

DENIS G. PELLI* and LAN ZHANG

Institute for Sensory Research, Syracuse University, Syracuse, NY 13244-5290, U.S.A.

(Received 7 February 1990; in revised form 4 September 1990)

Abstract—Off-the-shelf microcomputers can now display arbitrary 8-bit images, but accurate control of these images requires dealing with several undesirable properties of real digital to analog converters (DACs) and analog video monitors. The limitations of DACs and video monitors are presented in the form of a model for their calibration and use in vision experiments. Low contrasts can be accurately rendered by summing a small accurate a.c. signal and a large less-accurate d.c. signal (Watson et al., 1986; *Behavior Research, Method, Instrument and Computers, 18,* 587–594). Exploiting that idea, this note presents an easy-to-build passive resistor network, a video attenuator, that combines the outputs of three 8-bit DACs to render low contrasts with 12-bit accuracy at the display. Measurements confirm the 12-bit accuracy.

Video attenuator   Contrast   Monitor calibration   Microcomputer graphics

INTRODUCTION

All but the cheapest microcomputers now include color graphics. These computers are appealing for vision research. They are cheap. Good compilers are available, making them easy to program. Stimulus generation is almost entirely digital, up until the Digital to Analog Converters (DACs) that drive the analog monitor. Digital image synthesis is flexible, powerful and easy to understand and work with. This paper is concerned with the limitations of the analog part, as the DAC and monitor have many undesirable properties that must be understood and dealt with if the monitor is to display the image we want.

Computer graphics are typically implemented by a video card driving a color video monitor. The video card is a digital framestore representing each pixel of the final image by a small number, usually of 1–8 bits.† The monitor is a swept-raster cathode ray tube. To implement color, each pixel's number is transformed, by reference to a lookup table, to three numbers that drive three 8-bit DACs. (There is little reason to buy less-than-8-bit DACs since they cost nearly as much. More-than-8-bit DACs are only available on two very expensive systems, Adage and Pixar.) The three DACs drive separate guns in the color monitor. The "red", "green" and "blue" guns shoot electrons at red, green and blue phosphors on the face of the cathode ray tube. The glow of the phosphors produces a high-resolution color image.

The framestore approach to making images is easy to work with, making it easy to produce arbitrary spatial patterns. Furthermore, reloading the framestore's lookup table is an easy way to do simple real-time temporal manipulations, such as fades and counterphase flicker.

The main problem to be overcome is that an 8-bit DAC cannot adequately represent a near threshold (barely detectable) image. For example a 3 c/deg sinusoidal grating has a threshold contrast of 0.3%, where contrast is defined as $(L_{max} - L_{min})/(L_{max} + L_{min})$. The minimum and maximum luminances of the grating differ by only 0.6% of the mean luminance. But the smallest step that an 8-bit DAC can make is $1/255 = 0.4\%$ of its range. This is too large for accurate pixel-by-pixel rendering of this grating.

There are several solutions to this problem, but they are all tradeoffs. The most direct approach is to turn down the monitor's contrast knob, but in our experience the knob's range is too small to have much effect. A brute force solution is to add a large background luminance to the display, either by projecting light onto the face of the monitor, or by using a half-silvered mirror to combine the monitor's image with a

*To whom correspondence should be addressed.

†The number of bits per pixel is a digital limitation; 8 bits per pixel restricts the image to 256 different colors. This paper is concerned solely with analog limitations, and explains how to extend the palette from which those colors are chosen.

uniform field (Savoy, 1986). The added background reduces the contrast of images on the monitor. The drawback of adding light is simply the inconvenience of the bulky equipment required to produce the uniform background, and, unless the background is produced by a second monitor, the difficulty of controlling it by computer.

Halftoning is the way printers have long obtained fine gradations between the blackness of ink and the whiteness of paper. This sacrifices spatial resolution (using many pixels to represent one intensity) in order to improve luminance resolution (Ulichney, 1987; Mulligan & Stone, 1989). Halftoning requires a longer viewing distance to guarantee that the observer cannot resolve the halftone pattern, so we end up with a smaller display.

Until recently, cathode ray tubes for vision research were driven at much lower video frequencies to produce patterns varying in only one direction (Schade, 1956; Robson as described in Campbell & Green, 1965). At these lower frequencies (d.c. to 100 kHz) it was possible to use feedback techniques to linearize the luminance-to-voltage relationship of the monitor (as in the Joyce Electronics display, Cambridge, U.K.), and it was common to control contrast by a programable attenuator of the video signal. This is similar to the way that hearing studies control the amplitude of auditory signals. High frequency programable video attenuators exist, capable of handling the d.c. to 30 MHz or so bandwidth of the video signals produced by today's framestores (e.g. Kay Elemetrics, Pine Brook, NJ). Such attenuators become much more complicated if they must pass unattenuated synchronization signals (Savoy, 1986), so it is preferable to use a monitor capable of receiving a separate synchronization signal. The principal liability of the programable attenuators is the relays they use to do the switching. Relays are relatively slow (several msec to switch) and unreliable. Relay contacts can stick together or their contact resistance may rise. This unreliability is very troublesome in experiments because the failure is not obvious. You just get the wrong contrast.

The rest of this note describes an inexpensive scheme that achieves better contrast control either by sacrificing color or by using several framestores to drive one monitor. A framestore's three DACs are ordinarily used to drive a color RGB monitor, each DAC driving one gun. Since many experiments do not need color, Watson, Nielsen, Poirson, FitzHugh, Bilson, Nguyen and Ahumada (1986) suggested combining the signals from two DACs to produce a single monochrome signal. They recommended combining a coarse (unattenuated) d.c. voltage controlling the mean luminance with a fine (attenuated) a.c. voltage representing the low-contrast pattern.

The key idea is that vision experiments usually care more about contrast—luminance *differences*—than luminance *per se*. Attenuating a DAC signal and adding it to an unattenuated DAC signal has no effect on the attainable luminance accuracy, which is still determined by the unattenuated DAC. However, the *contrast* accuracy is much improved because luminance *differences* within the range of the attenuated DAC can be produced with an accuracy determined solely by the attenuated DAC.

The Methods section presents an extension of the Watson et al. circuit, now combining all three DACs by a passive resistor network. We show how to choose resistor values that will give the desired gains and input and output impedances. We explain how to calibrate the video attenuator and monitor, and we explain the algorithms that must be implemented in software to program the DAC lookup tables optimally for each contrast range. We present experimental results, precise measurements of the contrasts produced by this scheme. We also explain how to use this scheme with a color monitor, by combining the outputs of several video cards. Finally, the Discussion section presents a conceptual model of the video monitor and its limitations, to guide its use in vision experiments.*†

METHOD

Precision and accuracy of DACs

Accuracy refers to the difference between a measured quantity (voltage, luminance, or contrast) and what it ideally would be. *Precision*

*The algorithms presented in this paper, and routines to directly access video card lookup tables, have been implemented in C for use on a Macintosh II microcomputer. This *VideoToolbox* is available free of charge for use in research to anyone sending a blank Macintosh diskette.

†A limited number of video attenuators, as specified in Fig. 4, for the Apple High Resolution Monochrome monitor are available from Electronics Shop, Institute for Sensory Research, Syracuse University, Syracuse, NY 13244-5290, U.S.A.

is the size of the smallest step. We will specify both accuracy and precision relative to the full range, which may exceed the range of a single DAC when we combine several DACs.

Let us begin by illustrating the problem. Figure 1 shows the luminance profile of a 0.3% contrast sinewave grating rendered by an ideal 8-bit DAC, an ideal 9.6-bit DAC, and an ideal 12-bit DAC on an ideal monitor with linear luminance response and a luminance range of 0–1. These ideal DACs produce an output voltage proportional to the input number. We assume that the sampling along the horizontal axis is so fine as to be effectively continuous. For 8-, 9.6- and 12-bit DACs the input numbers can range from 0 to $2^8-1=255$, $2^{9.6}-1=765$, or $2^{12}-1=4095$, respectively. We assume they all have the same output voltage range. The equation for the continuous sinewave (solid line in Fig. 1) is:

$$L = 0.5(1 + 0.003 \sin 2\pi x). \tag{1}$$

The mean luminance of this sinewave is half the maximum luminance that these DACs can produce. For each DAC we plotted the available DAC level that would be nearest to the desired luminance. The 8-bit DAC renders the sinewave as a uniform luminance, producing zero contrast. The 9.6-bit DAC does a crude rendering, with only three levels, and the 12-bit DAC does a good job, with 13 levels. Remember that this sinusoidal grating has a contrast of only 0.3%, so the small errors in the 12-bit rendering will be invisible.

The errors in Fig. 1 result solely from the limited precision of the ideal DACs. Real DACs introduce additional errors due to their limited accuracy. For example, the most commonly used 8-bit video DACs are made by Brooktree (1989). They guarantee that their DACs are monotonic. For a b-bit DAC they further specify that if one measures and plots the output voltage versus input number and draws a line connecting the minimum-number (zero) point with the maximum-number (2^b-1) point, then no point will deviate from that line by

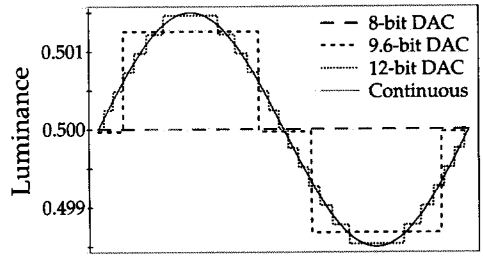

Fig. 1. Rendering of a 0.3% contrast sinusoid by ideal 8-, 9.6- and 12-bit DACs, compared with a true sinewave.

more than one part in 2^b-1 of the whole voltage range:

$$\epsilon_V = \frac{V_{2^b-1} - V_0}{2^b - 1}. \tag{2}$$

We will call this a *one-step error*.* We will consider these errors further when we estimate achievable limits on contrast errors.

We will see later that for low contrasts a typical monitor's nonlinear luminance response yields roughly twice the contrast as an ideal monitor with a linear luminance response. This doubles the contrast of a one-step DAC error, so that we end up with one less bit of accuracy than we started with. An 8-bit DAC and a typical monitor can produce images with a luminance accuracy of only about 7 bits.

Driving a monochrome monitor

Figure 2 shows the usual way to connect a color video card to a monochrome monitor. One DAC drives the monitor and the other two DACs are wasted. The current I_1 generated by the DAC is equally divided between the two 75 Ω termination resistors, so the output voltage is:

$$V = I_1 \frac{75\,\Omega}{2}. \tag{3}$$

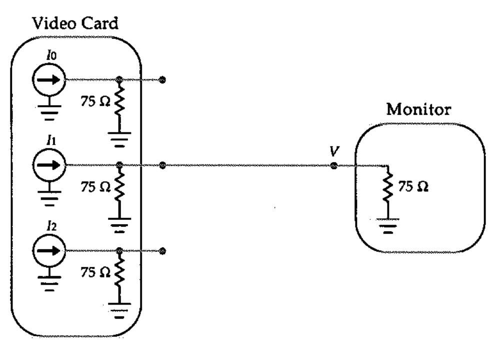

Fig. 2. The usual connection between a color video card and a monochrome monitor. Within the video card only the DACs are shown since the rest of the circuitry is digital. The three DAC outputs are typically labeled red, green, and blue. Ground and synch connections are not shown.

*In this paper we only consider DACs that meet this one-step error criterion, although there are some older video systems that use 8-bit DACs that have much larger than one-step errors.

The video card and monitor both have 75 Ω impedance, as indicated by the internal resistors, to match the impedance of the 75 Ω cable connecting them. Video work is usually done with 75 Ω cables. Since the cable may be long and the frequencies are high it is important that the cable be a transmission line terminated in its characteristic impedance at each end. If the monitor does not provide the correct input impedance then it will reflect energy back to the video card, and if the video card does not provide the correct output impedance then a part of the reflection will be reflected again and will appear on the monitor as a faint ghost. A slight mismatch of 1% or so at each end will not matter, because the reflection will be doubly attenuated, making the ghost invisible.

While proper termination is something that users ought to be able to take for granted, we have found that at least one commercial monitor (Princeton Graphic Systems Max-15) has an input impedance that is far from 75 Ω at some frequencies in the signal band (e.g. 1000 Ω at 3.1 MHz), making it unsuitable for critical work. A frequency-dependent input impedance will also make contrast frequency-dependent, so measure the input impedance of your monitor. Drive it with an oscillator through a 1 kΩ resistor and measure the voltages on either side of the resistor. (Make the connection from resistor to monitor shorter than 5 cm.) The ratio of those voltages is $Z/(Z + 1000\,\Omega)$, where Z is the monitor's input impedance. Z should be independent of the oscillator frequency, from d.c. to 20 MHz.

A video averager

Rather than wasting two of the three DACs, we can improve the precision by combining the three DAC voltages. Figure 3 shows the simplest three-input video attenuator, a video averager. It consists of four identical 37.5 Ω resistors. The resistor value was determined by the requirement that the input and output impedances of the video attenuator must equal 75 Ω. The output voltage V is the average of the three input currents:

$$V = \frac{I_0 + I_1 + I_2}{3}\frac{75\,\Omega}{2}. \qquad (4)$$

This gives us the same range as a single DAC [compare with equation (3)], but our step size is reduced to one-third its former value. It is convenient to think of a dimensionless nominal voltage, going from 0 to 1. The precision of our nominal voltage has increased from 8 bits (one part in 255) to 9.6 bits (one part in 3 × 255). The performance of an ideal 9.6-bit DAC is illustrated in Fig. 1, and provides a significant improvement over 8 bits.

While it might appear to be a trivial exercise to program this simulated 9.6-bit DAC, in practice some care is required because each DAC that is varied to create the image contributes a nominal-voltage error of up to $\pm 1/(3 \times 255)$, making it desirable to vary as few DACs as possible. Furthermore, the gains of the three DACs will be different, since the manufacturers, e.g. Brooktree (1989), only guarantee that the gains will match to within ±5%. In the next section we will explain how to measure the three gains, and take them into account in programming, so as to minimize the contrast errors. Because the averager of Fig. 3 is no easier to use than the general-purpose attenuator to be introduced in the next sub-section, we mentioned the simple averager only as an introduction to the general-purpose attenuator.

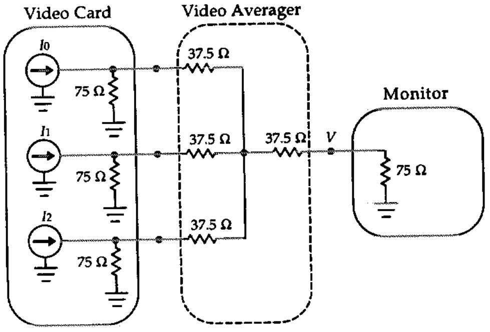

Fig. 3. The electronic schematic for a video averager.

The video attenuator

Precision can be substantially improved by applying unequal attenuations to the three DACs. Figure 4 shows the electronic schematic for a three-input attenuator. The six resistors provide enough degrees of freedom to allow for different gains on all three channels, and still satisfy the input and output impedance requirements. A formal derivation is presented in Appendix A. The output voltage V is a linear combination of the input currents, which we may write as:

$$V = (g_0 I_0 + g_1 I_1 + g_2 I_2) g \frac{75\,\Omega}{2} \qquad (5)$$

where we introduce the overall gain term g so that we may constrain the g_i to add up to 1,

$$g_0 + g_1 + g_2 = 1. \qquad (6)$$

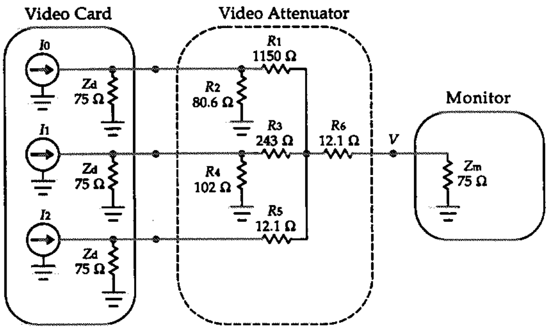

Fig. 4. The electronic schematic for the most-generally-useful three-input attenuator. Appendix A provides analytic solutions for the resistances to produce any desired combination of gains; a subset is listed in Table 1. The resistor values shown are the commercially available 1% precision resistor values nearest to the exact solution to the gain and impedance equations: $g_0 = 1/32$, $g_1 = \sqrt{1/32} - 1/32$, $g_2 = 1 - \sqrt{1/32}$, and 75 Ω input and output impedances. These commercially available values yield the specified gains and input and output impedances to better than ±1%. Low-power metal-film 1%-precision resistors are recommended; no resistor in this circuit dissipates more than 1/75 W.

The designer of a new attenuator only needs to decide how accurately low-contrast patterns should be rendered; the rest is more or less determined by that choice, the accuracy of the DACs (typically 8 bits), and the monitor's contrast gain [typically about 2, as explained below, equation (11)]. We believe that a contrast accuracy of 12 bits is a good all-around choice, resulting in the resistor values in Fig. 4, with channel gains $g_0 = 1/32$, $g_1 = \sqrt{1/32} - 1/32$, $g_2 = 1 - \sqrt{1/32}$, and overall voltage gain $g = 0.88$. The resistor values in Fig. 3 yield gains $g_0 = g_1 = g_2 = 1/3$, and $g = 1$. The overall gain g is unimportant because it merely reduces the whole voltage range by an amount easily compensated for by turning up the "contrast" knob on the monitor.

Analysis

In analyzing the performance of this network it is useful to relate everything to a new scale, to reduce the three-dimensional space of possible DAC values to one dimension, corresponding to the output voltage, and at the same time to do away with the need to measure DAC voltages. Ultimately we only want to deal with the numbers that we load into the DACs and the screen luminances that they produce and that we measure.

To this end we define a *nominal voltage* that represents, on a dimensionless range of 0 to 1, what our DACs and attenuator would produce in the absence of any accuracy errors:

$$v = g_0 \frac{n_0}{2^b - 1} + g_1 \frac{n_1}{2^b - 1} + g_2 \frac{n_2}{2^b - 1}. \quad (7)$$

The real output voltage V will be linearly related to the nominal voltage v, except for DAC inaccuracies:

$$V = (V_{2^b-1} - V_0)v + V_0 + e_V \quad (8)$$

where V_n is the output voltage produced by loading n into all three DACs and the error e_V is within the tolerance interval $\pm\epsilon_V$.

At the lowest contrasts the attenuator's contribution to contrast accuracy will be determined solely by the lowest channel gain g_0. A gain of 1/32 improves contrast accuracy by a factor of 32—5 bits—improving an 8-bit DAC to 13-bit precision and accuracy. Of course the range of this DAC will also be reduced by the same factor. The nominal voltage goes from 0 to 1. Of this range, if DACs 1 and 2 are fixed, DAC 0 can only cover a part g_0, e.g. 1/32. Images whose contrast is sufficiently low can be represented entirely by variations in DAC 0, fixing DACs 1 and 2 to appropriate values to produce the desired mean luminance.

Calibration

Having built a video attenuator, it is necessary to calibrate it, the three DACs (whose gains may differ by ±5%), and the monitor. All these calibrations can be done at the same time, measuring only the luminance of the monitor. The first thing we want to know is the nonlinear relationship between nominal voltage and luminance. We measure the luminance L of a suitable test target (e.g. a small square, large enough to fill the field of view of the photometer) as a function of the nominal voltage v resulting from loading the same number n into all three DACs. The luminance of the rest of the screen, which is not measured, should be set to mimic the conditions of the ultimate use. The results of such measurements are shown in Fig. 5, and represent the "gamma" function $L(v)$. We will need this function's inverse, $v(L)$.

The luminance measurements shown in Fig. 5 are well fit by a power law:

$$\frac{L(v)}{\text{cd/m}^2} = \begin{cases} \alpha + (\beta + \kappa v)^\gamma & \text{if } \beta + \kappa v \geq 0 \\ \alpha & \text{otherwise} \end{cases} \quad (9)$$

with a root-mean-square error less than 0.1 cd/m². It is easy to rearrange equation (9) to find the inverse function $v(L)$:

$$v(L) = \begin{cases} \dfrac{1}{\kappa}\left[\left(\dfrac{L}{\text{cd/m}^2} - \alpha\right)^{1/\gamma} - \beta\right] & \text{if } \dfrac{L}{\text{cd/m}^2} \geqslant \alpha \\ \dfrac{-\beta}{\kappa} & \text{otherwise} \end{cases} \quad (10)$$

The exponent γ is 2.284. The minimum luminance α is small (0.16 cd/m²). To at least a first approximation, the gain parameter κ in equation (9) is controlled by the monitor's "contrast" knob, and the d.c. offset parameter β is controlled by the monitor's "brightness" knob. Unfortunately the power law [equation (9)] is not a good fit for some monitors. In those cases we resort to an eighth-order polynomial. However, it is not necessary to do any curve fitting at all. The table of luminance measurements at 256 nominal voltages from 0 to 1 is easily interpolated as needed, and since it is ordered, inverse values can be found quickly by bisection and interpolation.

The straightforward way to measure the gains g_0, g_1 and g_2 of the three channels is to attach an oscilloscope probe to V (the monitor's input) and to measure the (relative) amplitude of full-scale square waves produced by each DAC. Relative measurements suffice since the three gains are normalized to add up to 1. There is no need to know the absolute voltage gain g.*

John Robson (personal communication) points out that not all photometers are suitable for measuring the luminance of a video monitor. The problem is that any point on the monitor is

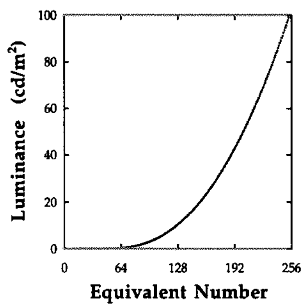

Fig. 5. A gamma function. This is the measured luminance L of an Apple High Resolution Monochrome Monitor as a function of the nominal voltage $v = n/255$ resulting from loading the same number n ($n = 0, 1, \ldots, 255$) into all three DACs in the Apple Color Video Card. (The video attenuator of Fig. 4 was in place but had little effect, other than reducing the voltage range by the factor $g = 0.85$.) The best fitting parameters of equation (9) are $\alpha = 0.16$, $\beta = -2.040$, $\kappa = 9.589$ and $\gamma = 2.2840$. *Method*: The fit was made by a multidimensional minimization of root-mean-square error using Powell's method (Press, Flannery, Teukolsky & Vetterling, 1988). The fit has a root-mean-square error of only 0.1 cd/m². This is much better than the best fitting quadratic ($\rho^2 = 0.997$), which has an unacceptably high root-mean-square error of 2 cd/m², which would result in significant errors in estimation of the slope of the luminance curve, and thus inaccurate control of contrast. The luminance measurements were made with both "contrast" and "brightness" knobs turned to maximum, for a 144 × 144 pixel test patch while the rest of the 640 × 480-pixel screen was at 48 cd/m² background. The Macintosh II computer's built-in gamma-correction table was replaced by a linear gamma table. Each of 256 luminances was measured by a UDT 61 photometer (United Detector Instruments, Atlanta, GA) digitized by a 12-bit analog-to-digital converter at 5 kHz for 120 msec (exactly 8 frames), and then averaged. (Averaging for a non-integral number of frames would introduce extra variance.) This sequence of measurements was repeated and averaged.

*A less straightforward but more convenient way to measure the channel gains is to use the measured gamma function to transform luminance measurements into nominal voltages, so as to do both the gamma and the channel gain calibrations in the same session, solely by measuring the screen luminance. For this it is desirable to interleave the measurement of the three channel gains with the luminance calibration, to make sure that exactly the same gamma function applies to both. Conceptually, the gains may be calibrated by measuring four more luminances. Let $L(n_0, n_1, n_2)$ represent the luminance measured when the three DACs are loaded with the triplet of numbers (n_0, n_1, n_2). The channel gains may be determined by the following formulae (assuming 8-bit DACs):

$$g_0 = v[L(255, 255, 255)] - v[L(0, 255, 255)]$$
$$g_1 = v[L(255, 255, 255)] - v[L(255, 0, 255)]$$
$$g_2 = v[L(255, 255, 255)] - v[L(255, 255, 0)]$$

where $v(L)$ is the inverse of the measured gamma function. The gain measurements are made relative to the maximum luminance condition (255, 255, 255) because most monitors have highest luminance gain there, so the measurement is more sensitive. A check on the accuracy of the measurements is provided by making sure that the channel gains add up to 1. If the error is acceptably small then they should all be normalized by dividing by the sum. The error should be small enough to leave several significant figures in the smallest gain, g_0, as any error in channel gain will later result in a proportional error in contrast.

In practice, gain calibration requires more than four measurements, because the DAC inaccuracies (1/256, for 8-bit DACs) limit the accuracy of calibration of the gamma function, compromising the calibration of the lowest gain (1/32, for the schematic in Fig. 4). We reduce the error (averaging many different DAC errors) by repeating the gain measurements relative to 16 triplets, not just the triplet (255, 255, 255) suggested in the equation above. Our error in gain measurement is less than ± 0.001.

very bright for a very brief time, and then quickly decays to dark until the next frame. The photometer's input amplifier must handle the peak luminance without saturating (which would clip the peak), even though this may be 100 times higher than the mean luminance. Since the lines on the monitor are swept successively one can smear out the peak in luminance by pulling back the photocell to integrate light from more of the monitor's screen. If the photometer is not clipping then the measured luminance of a uniform screen should be unaffected by changing the distance of the photocell (and thus what fraction of the screen it sees).

High contrasts

Thus far we have emphasized minimizing the error at low contrasts, assuming implicitly that we could tolerate larger errors in higher-contrast patterns. However, sometimes we want to present a low-contrast test pattern in the presence of a high-contrast mask, or discriminate between two similar high-contrast patterns. Such demands are problematic for the scheme presented here, as they do not conform to the Watson et al. (1986) premise of adding a small accurate a.c. signal to a large inaccurate d.c. signal. The mask will inevitably be a large inaccurate a.c. signal with large contrast errors. These errors could affect the visibility of a low-contrast test pattern. Since the signal and mask will be seen at once, it would not help to display them separately (e.g. by using two displays combined optically or by interleaving the two patterns on alternate frames or lines of the display) because once the two images are combined the accuracy of the luminance differences among pixels will still be dominated by the accuracy of the rendering of the higher contrast pattern. However, in particular experiments, analysis of the characteristics of the large contrast errors may show them to be negligible. If the mask is high-contrast white noise then the contrast errors will be white too and will merely increase the noise power by a negligible amount. To take another example, if the mask is periodic, with a period equal to an integral number of pixels, then the errors will be periodic, so the errors will have contrast energy only at multiples of the mask frequency. That would be fine for a high-frequency mask, but probably undesirable for a low-frequency mask. In general, large a.c. signals will be accompanied by large errors, and the experimenter must decide whether they are negligible.

Could we measure the DAC inaccuracies and take them into account?

In theory one could eliminate all inaccuracies of the DACs by accurately measuring the output voltage V at each setting, and then taking into account these deviations from linearity. However, this is usually impractical, principally because of yet another kind of DAC error, one which is not easily corrected for. When DACs switch from one voltage to another the transition may not be monotonic, and there may be a large brief voltage pulse, called a *glitch*, before the DAC settles to the new voltage. The glitch is usually largest for steps that involve changes in the more significant bits in the number loaded into the DAC. The worst case is usually the transition from $127 = 01111111_2$ to $128 = 10000000_2$, which is called a major carry, and involves a change in every bit, although the two output voltages are only one step apart. The glitches contribute to DC errors, but are difficult to correct for because they are transient and depend on both numbers in the transition. Manufacturer specifications for a typical DAC state only that the integral of the glitch voltage over time will "typically" be 50 pV s, which is about half the product of one DAC step ($0.7\,\text{V}/255 \approx 3\,\text{mV}$) and a 33 nsec pixel duration (Brooktree, 1989). We think that it is usually wiser to accept the one-step accuracy tolerance, which allows us to ignore the glitches.

However, some test patterns have very few transitions per raster line, e.g. uniform patches, or a horizontal grating. In these cases the glitches at the transitions may be negligible, and it might be reasonable to attempt precise calibration of the DACs.

The monitor's contrast gain

It is convenient to define the monitor's (low-contrast) contrast gain g_m with respect to nominal voltage as:

$$g_m = \frac{\mathrm{d}L/\mathrm{d}v}{2L}. \tag{11}$$

The numerator is the slope of the gamma function of luminance L. The contrast gain of the monitor in Fig. 5 is 1.96 at $L = \frac{1}{2}L(1) = 48\ \text{cd/m}^2$. The monitor gain g_m is affected slightly by the presence of the video attenuator because the voltage gain g reduces the parameter κ in equation (9), as though we had turned down the "contrast" knob slightly.

Programing

The video card's lookup table contains one entry for each possible pixel value (256 for 8-bit pixels). Each entry consists of three numbers that will be loaded into the three DACs whenever a pixel whose value equals the entry number is displayed. An 8-bit video card's lookup table contains only 3×256 bytes and can easily be reloaded on every frame to temporally modulate a spatial pattern. The image memory of the framestore, however, will contain on the order of 640×480 bytes, or more, which is too much for most microprocessors to reload in the 15 msec or so that each frame lasts.

In programing the lookup table there is obviously a many-to-one correspondence between the DAC triplet (n_0, n_1, n_2) and the resulting nominal voltage or luminance. In order to minimize the error in the luminance *differences*, i.e. contrasts, the best triplet to use to produce a given target luminance depends on the *range* of luminances that the stimulus will contain.

In order to calculate the best DAC triplet to produce a given luminance L within a pattern that spans a particular range, L_{min} to L_{max}, we need to take two steps. The first step is to decide how many DACs we will need to vary in order to cover the luminance range from L_{min} to L_{max}, and the second step is to decide what triplet values we will need to produce L.

The first step is begun by converting the luminance range into a nominal-voltage range:

$$\Delta v = v(L_{max}) - v(L_{min}). \qquad (12)$$

Then we must decide how to use the DACs. DACs that vary—to produce the pattern—will each contribute errors in the desired luminance differences; fixed DACs will produce no luminance difference at all. So, to minimize the error in luminance differences, we want to vary the DACs with lowest gain—the *finest* DACs—and keep the *coarsest* DACs fixed.

Define g_{vary} as the sum of the gains of the variable DACs. There are three cases. If the required range is less than or equal to g_0 then vary only DAC 0 and fix the other two DACs:

$$\Delta v \leqslant g_0 \rightarrow g_{vary} = g_0. \qquad (13a)$$

Failing that, if the required gain is less than or equal to $g_0 + g_1$ then vary DACs 0 and 1 and fix DAC 2:

$$g_0 < \Delta v \leqslant g_0 + g_1 \rightarrow g_{vary} = g_0 + g_1. \qquad (13b)$$

Failing that, vary all three DACs:

$$g_0 + g_1 < \Delta v \rightarrow g_{vary} = g_0 + g_1 + g_2 = 1. \qquad (13c)$$

Next we set the values of the fixed DACs to align the range of the variable DACs with the desired range. We do this by choosing values for the fixed DACs that yield a nominal voltage as close as possible to, but not exceeding $v(L_{min})$ when the variable DACs are all zero. This completes the first step. The work we have done so far depends only on the luminance range L_{min}, L_{max} and need not be redone until a new range is requested.

The second step is simply to choose the numbers for the variable DACs so that the three DACs together yield a nominal voltage as close as possible to $v(L)$, with a precision error of at most $\pm\frac{1}{2}g_0/(2^b - 1)$, and finally to load the triplet of numbers into the appropriate entry of the video card's lookup table.

Tolerance

A useful adjunct to these necessary computations is to compute the tolerance, i.e. the maximum error of the luminance differences. This has two parts: precision and accuracy. The more obvious source of error is that the nominal voltages are discrete and cannot produce an arbitrary requested luminance. However, this error due to limited precision, $\pm\frac{1}{2}g_0/(2^b - 1)$, will be small compared to the error due to limited accuracy unless the requested luminance is outside the range of possible luminances. The main source of error is limited accuracy: the manufacturer's tolerance on the DACs. (Errors in calibration of the monitor and channel gains can be made negligible, provided the DACs and monitor are stable.*) The specifications of the DACs guarantee at most a one-step error in voltage [equation (2)], so the error in nominal voltage will be (Brooktree, 1989):

$$\epsilon_v = \frac{1}{2^b - 1}. \qquad (14)$$

According to this specification, an arbitrary voltage *difference* might be in error by as much as two steps (since both voltages may have one-step errors), but, in our experience, DACs sold with the one-step-error specification also exhibit voltage-*difference* errors of at most one

*An anonymous reviewer noted that some cheap video cards do not tightly regulate the reference voltage of the DACs, resulting in proportional variations in the DACs' output voltage with the computer's power supply voltage. This might produce a low-spatial-frequency horizontal grating on the monitor.

step. The fixed DACs produce produce no luminance difference at all. All of the variable DACs contribute to the error, so we estimate the tolerance $\pm\epsilon_{\Delta v}$ in nominal-voltage differences to be:

$$\epsilon_{\Delta v} = \frac{g_{\text{vary}}}{2^b - 1}. \tag{15}$$

The tolerance in luminance differences over the specified range will be largest at the highest luminance, at least on monitors for which luminance is an accelerating function, as in Fig. 5. The luminance-difference tolerance $\pm\epsilon_{\Delta L}$ is:

$$\epsilon_{\Delta L} = \epsilon_{\Delta v} \left.\frac{dL}{dv}\right|_{L = L_{\max}} \tag{16}$$

$$\approx L_{\max} - L[v(L_{\max}) - \epsilon_{\Delta v}] \tag{17}$$

where $L_{\max}$ is the maximum luminance of the pattern. The contrast tolerance $\pm\epsilon_c$ works out to be the one-step error amplified by the attenuator and monitor gains:

$$\epsilon_c = \frac{\epsilon_{\Delta L}}{2L_{\max}} = \epsilon_{\Delta v} g_{\text{m}} = \frac{g_{\text{vary}} g_{\text{m}}}{2^b - 1}. \tag{18}$$

Accuracy in bits

Having computed the tolerance, it is natural to ask what we have achieved. How many bits does the accuracy correspond to? Let us first convert the attenuator's and monitor's gains to bits:

$$b_{\text{vary}} = \log_2 \frac{1}{g_{\text{vary}}} \tag{19}$$

$$b_{\text{m}} = \log_2 \frac{1}{g_{\text{m}}}. \tag{20}$$

We will attain a contrast accuracy equivalent to that of a b_c-bit DAC whose output is contrast:

$$b_c = \log_2 \frac{1}{\epsilon_c} \tag{21}$$

$$= \log_2(2^b - 1) + \log_2 \frac{1}{g_{\text{vary}}} + \log_2 \frac{1}{g_{\text{m}}}$$

$$\approx b + b_{\text{vary}} + b_{\text{m}}. \tag{22}$$

Thus, the final accuracy b_c is the sum of the contributions of the DAC's b (e.g. 8 bits), the attenuator's b_{vary} (e.g. 5 bits), and the monitor's b_{m} (e.g. -1 bit), for a total of 12 bits.

Designing a new video attenuator: choosing the gains

Some applications may justify the design of a custom attenuator. Using the schematic of Fig. 3, the designer has 2 degrees of freedom, the gains: g_0 and g_1. Appendix A will then provide the values of all the resistors. Assuming b-bit DACs, the resulting video attenuator will provide $b + \log_2(1/c_0)$ bit accuracy at contrasts up to $c_0 = g_0 g_{\text{m}}$, $b + \log_2(1/c_1)$ bit accuracy at contrasts up to $c_1 = (g_0 + g_1)g_{\text{m}}$, and $b + \log_2(1/g_{\text{m}})$ bit accuracy at contrasts up to 1.

However, Appendix B shows that for any g_0 the worst-case contrast-ratio error at contrasts above c_0 will be minimized by choosing g_1 so as to produce a uniform geometric spacing of the total gains used:

$$g_{\text{vary}} = \begin{cases} g_0 & = g_0 \\ \sqrt{g_0} & = g_0 + g_1 \\ 1 & = g_0 + g_1 + g_2. \end{cases} \tag{23}$$

Adopting this principle leaves the designer with only one degree of freedom: g_0. Table 1 lists the resistor values that result from various choices for g_0.

To choose g_0 the designer must trade off the contrast accuracy attained at contrasts below $c_0 = g_0 g_{\text{m}}$ against c_0 itself, and the maximum log-contrast error at contrasts above c_0. With b-bit DACs and a monitor contrast gain g_{m} (typically around 2), the resulting video attenuator will provide $b + \log_2(1/c_0)$ bit accuracy at contrasts up to c_0. The first column in Table 1 shows the increasing accuracy with reduced gain g_0, but this benefit is at the expense of reducing the maximum contrast $c_0 = g_0 g_{\text{m}}$ at which this accuracy will be attained. Column two shows the related cost r_8: the increasing worst-case error in contrast ratio at contrasts above c_0.

We believe 12-bit accuracy is a good all-round choice, making the contrast errors invisible. For the attenuator in Fig. 4 we assumed 8-bit DACs ($b = 8$) and a monitor gain of 2 ($b_{\text{m}} = -1$). Equation (22) told us that we needed a $12 - 8 - -1 = 5$ bit attenuator ($b_{\text{vary}} = b_0 = 5$) so g_0 is $1/32 = 0.0313$ and c_0 is $1/16 = 6\%$. Appendix B told us to set g_1 to $\sqrt{1/32} - 1/32 = 0.1455$. The end result is the attenuator illustrated in Fig. 4. With three 8-bit DACs and a monitor whose contrast gain is 1.96, this attenuator provides 12-bit accuracy at contrasts up to 6%, 9.5-bit accuracy at contrasts up to 34%, and 7-bit accuracy at contrasts up to 100%. Without the attenuator we would have 7-bit accuracy at all contrasts.

Color

The same scheme may be used to drive a color monitor by the simple expedient of using three

Table 1. Resistor values for the monochrome attenuator of Fig. 4, for various desired gains g_0. These are solutions to equations (A6). The first line corresponds to no attenuation, as in Fig. 2, and the second line corresponds to averaging, as in Fig. 3. In the rest of the table the gain g_1 is set to $\sqrt{g_0 - g_0}$, to minimize the maximum contrast ratio tolerance r_b (see Appendix B). By definition, the gain g_2 equals $1 - g_0 - g_1$. Each line of the table has two figures of merit: $b_0 = \log_2(1/g_0)$ tells us how many bits the attenuator will contribute to low-contrast accuracy, and r_8 tells us the maximum contrast-ratio tolerance (using 8-bit DACs, see Appendix B). For example, 8-bit DACs with a 5.0-bit attenuator contribution and a monitor whose gain reduces the contrast accuracy by 1 bit will result in a contrast accuracy of $8 + 5 - 1 = 12$ bits, and will have a maximum contrast-ratio tolerance r_8 of 1.022. In the first line, g_0 is zero so we substituted g_2 in computing b_0 and r_8. When building the video attenuator, it is not necessary to use the exact resistor values shown here. It is enough to choose the nearest standard values of metal-film 1% precision resistors (e.g. see Appendix D of Horowitz & Hill, 1980)

b_0	r_8	g_0	g_1	g_2	g	R_1	R_2	R_3	R_4	R_5	R_6
0.0	1.004	0	0	1	1	∞	75 Ω	∞	75 Ω	0	0
1.58	1.008	1/3	1/3	1/3	1	37.5 Ω	∞	37.5 Ω	∞	37.5 Ω	37.5 Ω
2.0	1.008	0.2500	0.2500	0.5000	0.82	95.5 Ω	169.3 Ω	95.5 Ω	169.3 Ω	31.43 Ω	31.43 Ω
2.5	1.009	0.1768	0.2437	0.5796	0.81	160.0 Ω	120.5 Ω	111.3 Ω	147.9 Ω	27.06 Ω	27.06 Ω
3.0	1.011	0.1250	0.2286	0.6464	0.82	248.6 Ω	101.5 Ω	129.5 Ω	133.3 Ω	23.07 Ω	23.07 Ω
3.5	1.013	0.0884	0.2089	0.7027	0.83	372.6 Ω	91.8 Ω	150.9 Ω	122.5 Ω	19.61 Ω	19.61 Ω
4.0	1.016	0.0625	0.1875	0.7500	0.85	547.4 Ω	86.0 Ω	176.0 Ω	114.2 Ω	16.64 Ω	16.64 Ω
4.5	1.019	0.0442	0.1660	0.7898	0.87	794.6 Ω	82.4 Ω	205.6 Ω	107.5 Ω	14.12 Ω	14.12 Ω
5.0	1.022	0.0313	0.1455	0.8232	0.88	1144.4 Ω	80.1 Ω	240.4 Ω	102.1 Ω	11.98 Ω	11.98 Ω
5.5	1.026	0.0221	0.1266	0.8513	0.89	1639.8 Ω	78.5 Ω	281.6 Ω	97.7 Ω	10.16 Ω	10.16 Ω
6.0	1.031	0.0156	0.1094	0.8750	0.91	2340.9 Ω	77.4 Ω	330.4 Ω	94.0 Ω	8.63 Ω	8.63 Ω

framestores and three video attenuators like the one in Fig. 4. Each framestore and attenuator would drive one of the three inputs of the color monitor. Other than the expense of buying three framestores (and making room for them in the microcomputer) there is the practical problem that most framestores insist on being the master for purposes of synchronization (they assume the monitor will be a slave) so it may not be possible to synchronize the three framestores.

There is a simple solution using two framestores, which reduces the expense. Even if it is not possible to accurately synchronize the framestores to make all the pixels coincide, it is usually possible to get the framestores roughly in synch by resetting both at the same time. Since the framestores are driven by accurate quartz clocks they will drift out of phase with each other very slowly, e.g. 50 pixels per second. The framestores can be reset at the beginning of each trial in the experiment, so the total drift may be less than 50 pixels. A pattern drifting at this rate would be unacceptable, but one framestore producing a pattern can synchronize the monitor, and the other framestore can be used to produce a uniform background whose drift would be invisible except at the edge of the screen, which should be obscured.

Figure 6 is a schematic for an attenuator to combine two framestores to drive a color monitor. Each monitor input is driven by the combination of the corresponding outputs (red, green or blue) of the two framestores. The design of the attenuator has only one degree of freedom, the signal gain g_1. Resistor values for various gains are listed in Table 2. If the framestores are accurately synchronized then a gain of 1/4 would be a reasonable choice, improving accuracy by a factor of 4 for low contrasts. If the framestores are only roughly synchronized then a higher gain may be needed in order to allow presentation of high contrast stimuli.

RESULTS

The measured gains using the video attenuator we built according to Fig. 4 turned out to be within ±2% of their design values. This deviation is harmless and mostly due to the DAC gains, which are only guaranteed to match to within ±5% (Brooktree, 1989).

Figure 7 assesses the performance of this video attenuator with the monitor of Fig. 5. The photometry was identical to the luminance calibration, described in the legend to Fig. 5, except that the sampling duration was increased to 3 sec (exactly 200 frames). (It is difficult to measure low contrasts accurately; averaging of the photometer readings improves accuracy.)

Figure 7A shows the contrast errors, the difference between measured and nominal contrasts. Each point represents the measured error in the luminance difference between two test patch luminances nominally symmetric about the mean luminance of 48 cd/m^2. The dashed lines show the computed contrast tolerance $\pm\epsilon_c$. All the measured errors are within the computed tolerance interval, indicating that we have achieved our design goal: 12-bit accuracy

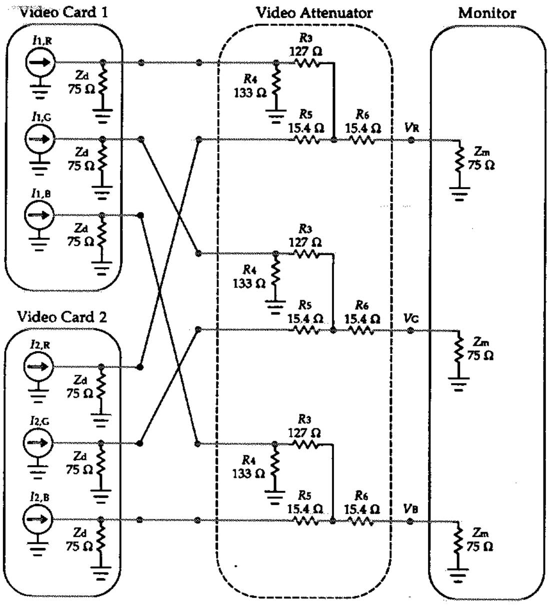

Fig. 6. The electronic schematic for a color attenuator. Ground and synch connections are not shown. The resistor values shown are the commercially available 1% precision resistor values nearest to the exact solution to the gain and impedance equations: $g_1 = 0.25$, $g_2 = 0.75$ and 75 Ω input and output impedances. See Table 2 (or Appendix A) for other gains. If the two video cards cannot be perfectly synchronized with each other, then the monitor should be synchronized to video card 1.

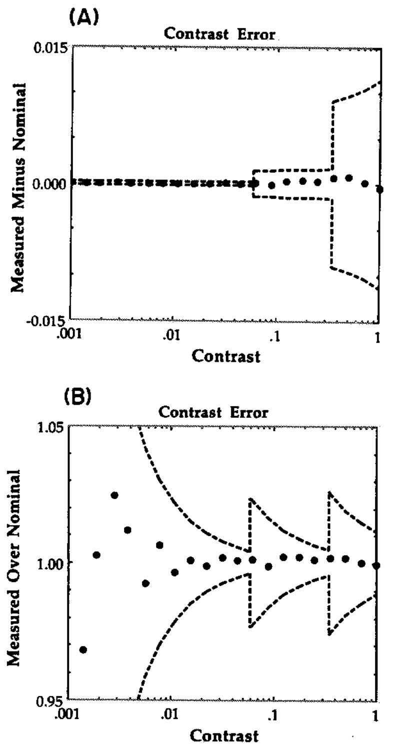

Fig. 7. (A) The contrast error (measured contrast minus nominal contrast) as a function of contrast for the video attenuator of Fig. 4. The dashed lines show the computed tolerance range $\pm\epsilon_c$ [equations (17) and (18)]. The dashed lines are sloped at high contrasts because of the monitor's nonlinear gamma function. (B) The contrast ratio (measured contrast over nominal contrast) as a function of contrast for the video attenuator of Fig. 4. The dashed lines show the computed tolerance interval $\pm\epsilon_c$. The peaks of the tolerance at contrasts of $c_0 = 0.06$ and $c_1 = 0.34$ were intended to be equal (see Appendix B), but are not quite equal because of the unequal gains of the DACs. The lowest-contrast measurements required averaging a large number of samples (at 5 kHz for 3 sec, exactly 200 frames) in order to reduce the photometer's noise and go a few bits beyond the 12-bit analog-to-digital converter's inherent precision.

at contrasts up to 6%, 9.5-bit accuracy at contrasts up to 34%, and 7-bit accuracy at contrasts up to 100%.

Figure 7B is a different view of the same results. It shows the contrast ratio, measured contrast over nominal contrast, as a function of contrast. The dashed lines represent the computed tolerance interval $(c \pm \epsilon_c)/c$, where c is nominal contrast.

The luminance and contrast calibrations (Figs 5 and 7) both measured the luminance of the same test patch in the center of the display. We wished to know whether that luminance calibration would be valid for the rest of the

Table 2. Resistor values for the color attenuator of Fig. 6, for various desired gains, g_1. These are solutions to equations (A6), with the gain g_0 set to zero. The gain g_2 equals $1 - g_1$. The attenuator applies gain g_1 to video card 1 and g_2 to video card 2. The figure of merit $b_1 = \log_2(1/g_1)$ tells us how many bits the attenuator will contribute to low-contrast accuracy

b_1	g_1	g_2	g	R_3	R_4	R_5	R_6
1.0	0.5000	0.5000	1.00	25.0 Ω	∞	25.000 Ω	25.000 Ω
1.5	0.3536	0.6464	0.88	72.4 Ω	199.6 Ω	20.503 Ω	20.503 Ω
2.0	0.2500	0.7500	0.88	125.8 Ω	133.6 Ω	15.488 Ω	15.488 Ω
2.5	0.1768	0.8232	0.89	194.6 Ω	109.6 Ω	11.488 Ω	11.488 Ω
3.0	0.1250	0.8750	0.91	287.2 Ω	97.2 Ω	8.426 Ω	8.426 Ω
3.5	0.0884	0.9116	0.93	415.0 Ω	89.8 Ω	6.128 Ω	6.128 Ω
4.0	0.0625	0.9375	0.95	593.3 Ω	85.1 Ω	4.426 Ω	4.426 Ω
4.5	0.0442	0.9558	0.96	843.8 Ω	82.0 Ω	3.180 Ω	3.180 Ω
5.0	0.0313	0.9688	0.97	1196.6 Ω	79.8 Ω	2.275 Ω	2.275 Ω
5.5	0.0221	0.9779	0.98	1694.6 Ω	78.4 Ω	1.622 Ω	1.622 Ω
6.0	0.0156	0.9844	0.99	2398.3 Ω	77.4 Ω	1.154 Ω	1.154 Ω

screen, except possibly for loss of brightness away from the center of the screen (Brainard, 1989). We divided the screen into three rows and four columns of square patches, each 160 × 160 pixels, and calibrated the contrast of each test patch. Contrast variations in the lower two-thirds of the screen were negligible ($g_m = 1.96 \pm 0.02$ = mean ± SD), but contrasts in the top third of the screen were slightly lower ($g_m = 1.86 \pm 0.02$). This error exceeds the tolerances in Fig. 7B, but in practice very few experiments would be harmed by a fixed factor of $1.86/1.96 = 0.95$ error in contrast. We also found that the top third of the screen was 12% brighter (55 ± 1.5 cd/m^2) than the bottom two-thirds (49 ± 1.5 cd/m^2), but we could not determine whether the two effects were causally related.

DISCUSSION

The digital part of image display by a video card and monitor is easily understood. This paper is concerned with the analog part: the DAC and monitor. The Methods section modeled the DAC and showed how to build a video attenuator to overcome some of its limitations. The Results section validated the model of the DAC and the attenuator design by confirming that the measured errors are within the predicted tolerance interval. Now we turn to the monitor.

So far we have restricted our attention to the control of the luminance of a moderately large patch (small relative to the screen area yet large relative to a single pixel) on the monitor, as measured by a fixed photometer. The entire output of the monitor was described by a single number, the luminance of that patch. Most experiments will need to control the contrast of more complicated images. This requires a model of the monitor. We suggest the following six assumptions.

(1) Assume independence of successive frames. (But measure the phosphor decay time, and reject any monitor, like the Princeton Graphic Systems MAX-15, whose phosphor is slow enough to violate this assumption.)

(2) Restrict the maximum change Δv between horizontally contiguous pixels to avoid slew rate limitations in the video amplifier (Lyons & Farrell, 1989; Mulligan & Stone, 1989). This limitation may be demonstrated by comparing the luminance of horizontal and vertical gratings made up of alternating one-pixel white and black lines. The luminance difference will disappear when the slew rate limit is respected.

(3) Imagine a hypothetical proto-image, made up of proto-pixels in the same locations as the actual pixels on the display (which are not uniformly spaced, due to geometric distortions usually specified as "nonlinearity"). Assume the luminance of the proto-pixels is generated by the monitor's measured gamma function, in the same way for all proto-pixels, independent of each other.

(4) Assume that the proto-pixels are blurred by convolution with a point spread function—usually Gaussian and circularly symmetric—that tends to become larger towards the edges of the screen (Infante, 1985). The point spread function is best characterized by measuring the monitor's MTF (Schade, 1958; Keller, 1986). Briefly, we recommend drifting a high-contrast sinusoidal grating past a microphotometer monitoring a very small spot. As spatial frequency is changed, use a constant temporal frequency (vary the velocity) to eliminate the effect of the temporal frequency response of the photometer. [Because of the monitor's slew rate limitations, point spread function measurements (Morgan & Watt, 1982) are an unreliable substitute for contrast measurements of extended patterns.] Make sure the photometer has enough dynamic range to handle the luminance signal without clipping.

(5) Assume that the measured luminance is the result of optical attenuation of the blurred proto-pixels by a neutral density filter that tends to grow darker towards the edges of the display (Brainard, 1989). Brainard verified this *neutral density filter* assumption for a color monitor. (He called it the "single scale factor" assumption.) Note that this will not affect local contrast, only luminance. We tested and confirmed the neutral density filter assumption for the lower two-thirds of the screen of our Apple High Resolution Monochrome monitor. The upper third violates this assumption, but the deviation will be negligible in most experiments.

(6) Assume that the luminance of a patch (or proto-pixel) is independent of the rest of the screen only if you maintain a constant mean luminance on every raster line.

Unfortunately, inexpensive monitors suffer from inadequate d.c. restoration and inadequate regulation of the high voltage power supplies. These problems both have the effect of making a small test patch brighter when the rest of the screen is made darker. The d.c. restoration error depends on the average of the input voltage since the beginning of the frame. The high voltage droop depends on the average beam current (i.e. luminance) since the beginning of the frame. On our Apple High Resolution Monochrome monitor changing the background luminance from zero to maximum reduces the luminance of a small maximum-luminance test patch from 100 cd/m^2 down to 88 cd/m^2. This is a relatively small effect (not normally noticeable) but it could be disastrous for some quantitative experiments. If possible, avoid these problems by buying a monitor which is d.c. coupled and has excellent high voltage regulation. Otherwise, design all your stimuli to have approximately the same mean luminance, and set the rest of the screen to that luminance during your luminance calibrations of a test patch. One way to maintain a fixed mean luminance on each raster line is to obscure one side of the screen, and to vary the luminance of the hidden part of the screen to cancel the changes in mean luminance of the visible part of the screen. If this proves too restrictive, then the effect of d.c. droop should be assessed to determine whether it is negligible or can be corrected for. Because of this effect, it is essential that the gamma function be measured with the same background luminance over most of the display as in the subsequent experiments.

These recommendations imply that a high spatial frequency full-screen sinusoidal grating should be presented horizontally—parallel to the raster lines—to avoid slew rate limitation, but a low spatial frequency grating should be presented vertically (or tilted), to avoid d.c. droop. Inexpensive monitors are band-pass devices, and one should put one's stimuli into the pass band.

CONCLUSIONS

A passive video attenuator consisting of six resistors allows an ordinary, off-the-shelf color video card and monochrome monitor to produce accurate visual stimuli over the entire contrast range, from threshold up to unit contrast, with highest accuracy at the lowest contrasts. The signals from several video cards can be combined to drive a color monitor, provided the video cards can be synchronized to each other. Choose a video monitor that accepts a separate synchronization signal, that has a frequency-independent input impedance, and, ideally, that has a d.c.-coupled video input and excellent high-voltage regulation. The monitor's remaining limitations can be dealt with by careful calibration and stimulus design.

Note Added in Proof—We have repeatedly calibrated a monitor, like the one in Fig. 5, that has been turned on continuously for 6 months. The screen was darkened by a "screen saver" program when the computer was not in use. All the calibrations yielded the same gamma function over all this time, except that the "brightness" parameter β has gradually dropped. This is most conveniently expressed by noting that the cut-off nominal-voltage-β/κ has gradually increased, by about 0.027 per month, from 0.181 to 0.346, presumably due to evaporation of the heater or cathode in the cathode ray tube.

Acknowledgements—Some of these results were presented at the *Computational Modeling of Visual Processing Workshop* at Cold Spring Harbor Laboratory, NY, 26–30 June, 1989. We are indebted to Stanley Klein who, as reviewer, pointed out an error in the presentation (now corrected) and supplied the analytic solution for the resistance in Appendix A, which we had only managed to solve numerically (in *Mathematica*, Wolfram, 1988). We thank David Brainard, Andrew Derrington, Bart Farell, Michael Landy, Jeff Mulligan, Avi Naiman, Evan Relkin, John Robson, Preeti Verghese and Beau Watson for helpful comments. This work was supported by National Eye Institute grant EY04432 to DGP.

REFERENCES

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

Brooktree (1989). *Product databook*, 2nd edn. Brooktree Corp., 9950 Barnes Canyon Rd, San Diego, CA 92121. (The DAC specifications we quote refer to the Bt453KP DAC used in the Apple Mac II Color Video Card.)

Campbell, F. W. & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology, London*, *181*, 576–593.

Horowitz, P. & Hill, W. (1980). *The art of electronics*. Cambridge, MA: Cambridge University Press.

Infante, C. (1985). On the resolution of raster-scanned CRT displays. *Proceedings of the SID*, *26(1)*, 23–26.

Keller, P. A. (1986). Resolution measurement techniques for data display cathode-ray tubes. *Displays—Technology and Applications*, *7(1)*, 17–29.

Lyons, N. P. & Farrell, J. E. (1989). Linear systems analysis of CRT displays. *SID Digest*, *20*, 220–223.

Morgan, M. J. & Watt, R. J. (1982) The modulation transfer function of a display oscilloscope: Measurements and comments. *Vision Research*, *22*, 1083–1085.

Mulligan, J. B. & Stone, L. S. (1989). Halftoning method for the generation of motion stimuli. *Journal of the Optical Society of America*, *A6*, 1217–1227.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1988). *Numerical recipes in C*. Cambridge, MA: Cambridge University Press.

Savoy, R. L. (1986). Making quantized images appear smooth: Tricks of the trade in vision research. *Behavior Research Methods, Instruments and Computers, 18*, 507–517.

Schade, O. H. Sr (1956). Optical and photoelectric analog of the eye. *Journal of the Optical Society of America, 46*, 721–739.

Schade, O. H. (1958). A method of measuring the optical sine-wave spatial spectrum of television image display devices. *Journal of the Society of Motion Picture and Television Engineers, 67*, 561–566.

Ulichney, R. (1987). *Digital halftoning*. Cambridge, MA: MIT Press.

Watson, A. B., Nielsen, K. R. K., Poirson, A., Fitzhugh, A., Bilson, A., Nguyen, K. & Ahumada, A. J. Jr (1986). Use of a framebuffer in vision research. *Behavior Research Method, Instrument and Computers, 18*, 587–594.

Wolfram, S. (1988). *Mathematica: A system for doing mathematics by computer*. New York: Addison-Wesley.

APPENDIX A

The Design Equations for the Video Attenuator

Figure 4 gives the standard resistor values that are nearest to the exact solution of the design constraints for gains of

$$1/32 = 0.0313,$$

$$\sqrt{1/32} - 1/32 = 0.1455,$$

and

$$1 - \sqrt{1/32} = 0.8232.$$

We believe this set of gains is the best compromise for accurate contrast control of a wide range of visual stimuli on monitors with contrast gains of around 2. A monitor with a much lower or higher contrast gain would warrant a new attenuator with appropriate gains, and a particular application might justify its own gains, optimized for a narrower range of contrasts. The following equations can be solved for the resistor values to produce any desired gains.

The input and output impedances of the video attenuator in Fig. 4 are:

$$Z_0 = R_2 \| \{R_1 + (R_6 + Z_m) \| (R_5 + Z_d) \| (R_3 + R_4 \| Z_d)\}$$

$$Z_1 = R_4 \| \{R_3 + (R_6 + Z_m) \| (R_5 + Z_d) \| (R_1 + R_2 \| Z_d)\}$$

$$Z_2 = R_5 + (R_6 + Z_m) \| (R_3 + R_4 \| Z_d) \| (R_1 + R_2 \| Z_d)$$

$$Z_{out} = R_6 + (R_5 + Z_d) \| (R_3 + R_4 \| Z_d) \| (R_1 + R_2 \| Z_d) \quad (A1)$$

where Z_d is the DAC's output impedance, Z_m is the monitor's input impedance, and $x \| y$ is read as "x in parallel with y", representing the operation:

$$x \| y \equiv \frac{1}{(1/x) + (1/y)}. \quad (A2)$$

The voltage gains, defined as:

$$G_i = \frac{\partial V}{\partial I_i} \frac{1}{Z_d \| Z_i} \quad (A3)$$

are

$$G_0 = \frac{Z_m}{R_6 + Z_m} \frac{(Z_0 \| - R_2) - R_1}{Z_0 \| - R_2}$$

$$G_1 = \frac{Z_m}{R_6 + Z_m} \frac{(Z_1 \| - R_4) - R_3}{Z_1 \| - R_4}$$

$$G_2 = \frac{Z_m}{R_6 + Z_m} \frac{Z_2 - R_5}{Z_2}. \quad (A4)$$

Matching all the impedances:

$$Z_0 = Z_1 = Z_2 = Z_{out} = Z_d = Z_m \quad (A5)$$

we can solve for the resistances (thanks to Stanley Klein, personal communication):

$$R_1 = Z_m \frac{G_2/G_0 - G_0}{1 + G_2}, \quad R_2 = Z_m \frac{G_2/G_0 - G_0}{G_2/G_0 + G_0 - G_2 - 1}$$

$$R_3 = Z_m \frac{G_2/G_1 - G_1}{1 + G_2}, \quad R_4 = Z_m \frac{G_2/G_1 - G_1}{G_2/G_1 + G_1 - G_2 - 1}$$

$$R_5 = R_6 = Z_m \frac{1 - G_2}{1 + G_2}. \quad (A6)$$

The voltage gains G_i are related to the normalized gains g_i by:

$$G_0 = gg_0, \quad G_1 = gg_1, \quad G_2 = gg_2 \quad (A7)$$

where $g_0 + g_1 + g_2 = 1$. The normalized gains g_i are supplied by the designer. However, to compute the resistances in equations (A6) we need the voltage gains G_i, and for that we need to know the overall gain g. We have obtained an analytic solution for g in terms of the normalized gains, but it is too cumbersome to be useful (and requires the cube root of a complex number). Instead we recommend using Table 1 to make an initial guess for g and iteratively adjusting it to match an impedance, e.g. Z_2, to the impedance Z_m, which is typically 75 Ω.

Numerical values of the solution for various gain combinations are given in Tables 1 and 2. The solutions in Table 2 are restricted to two channels, i.e. $g_0 = 0$.

APPENDIX B

Choosing the Gain g_1 so as to Minimize the Contrast Ratio Error

Once g_0 has been chosen, g_1 should usually be chosen so as to mimimize the maximum error in contrast ratio at moderate and high contrasts (as illustrated by the dashed line in Fig. 7B). The contrast ratio error peaks at three contrasts: zero, just over $c_0 = g_0 g_m$, and just over $c_1 = (g_0 + g_1)g_m$. Let c^+ indicate a contrast infinitesimally higher than c. The contrast ratio error at zero contrast is infinite and we cannot do anything about it. The contrast ratio errors at contrasts c_0^+ and c_1^+ will in general be different. Their maximum r_b:

$$r_b = \max\left(\frac{c_1^+ + \epsilon_{c_1^+}}{c_1^+}, \frac{c_0^+ + \epsilon_{c_0^+}}{c_0^+}\right) \quad (B1)$$

where ϵ_c is the contrast tolerance at contrast c, is minimized by making the two contrast ratio errors equal:

$$\frac{c_1^+ + \epsilon_{c_1^+}}{c_1^+} = \frac{c_0^+ + \epsilon_{c_0^+}}{c_0^+} \quad (B2)$$

where, by equations (18) and (13), we have:

$$\epsilon_{c_0^+} = \frac{(g_0 + g_1)g_m}{2^b - 1} \quad (B3)$$

$$\epsilon_{c_1^+} = \frac{g_m}{2^b - 1}. \quad (B4)$$

This simplifies to:

$$(g_0 + g_1) = \sqrt{g_0} \quad (B5)$$

so

$$g_1 = \sqrt{g_0} - g_0. \quad (B6)$$

Substituting back into equations (B2) and (B1) yields the maximum ratio error:

$$r_b = 1 + \frac{1}{(2^b - 1)\sqrt{g_0}}. \quad (B7)$$