51. The use of CRT displays in research on colour vision

J.D. MOLLON and M.R. BAKER
(Cambridge, U.K.)

Abstract

An introduction is given to the use of colour graphics monitors in visual research. The choice, setting-up, and calibration of graphics boards and displays are reviewed, and a survey is given of the non-additivities, non-uniformities and temporal instabilities that complicate the experimental use of such systems.

Introduction

A computer-controlled display offers the visual scientist a wonderful freedom to manipulate stimuli in time, space, and colour space, to present decrements as well as increments, and to hold constant the excitation of two classes of cone – even perhaps, of two putative post-receptoral channels – while modulating the third (Rodieck, 1983). It also offers a wonderful variety of snares and puzzles to entrap the innocent. The purpose of the present account is to supply some background on cathode ray tubes (CRTs) and graphics boards, to outline the setting-up and calibration of such a display system, to list the several problems that may then exercise the user, and to gather together the wisdom already available in a scattered literature. We write from the viewpoint of the consumer rather than the engineer, and we hope that our own experience may serve to guide newcomers in their choice of equipment. But ultimately, buying a graphics display is like buying a domestic appliance: one needs to own it for twelve months before one knows how to buy it.

Monitors and graphics boards

The cathode ray tube is a living fossil. It is approaching the centennial of its invention[1] and is probably in an evolutionary cul-de-sac; but the demand for

---

[1] Ferdinand Braun (1897) describes a recognizable ancestor of the modern CRT – a flask-shaped vacuum tube with cathode at one end and phosphorescent screen at the other, an anode, and electromagnetic coils for deflection in x and y dimensions.

B. Drum (ed.), Colour Vision Deficiencies XII, pp. 423–444.

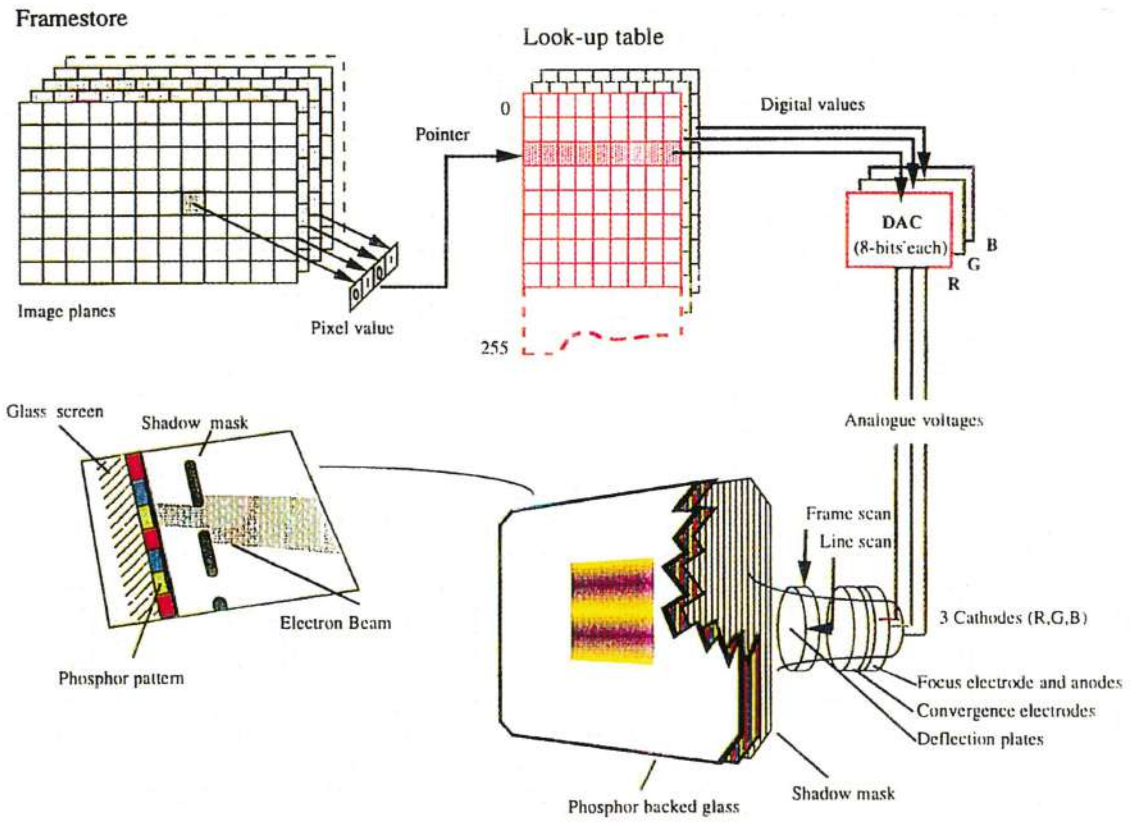

Fig. 1. Schema of a graphics card (above) and CRT monitor (below). In the framestore, the value stored for a particular pixel points to one entry in the look-up table. The look-up table depicted here has 256 entries (requiring 8 planes of framestore memory) and for each entry the gun voltage can be specified with a resolution of 8 bits. The CRT shown is based on the Sony Triniton arrangement, in that the shadow mask or 'aperture grill' consists of vertical 'wires' and the phosphor pattern consists of vertical stripes.

high-resolution computer displays has led to continued adaptations, and flat-screen displays have still not gained their long-heralded dominance. A schematic of a CRT colour raster display is shown in the lower part of Fig. 1. Within a vacuum tube, a stream of electrons is emitted from the heated cathodes of each of three guns and focussed through a shadow mask to excite phosphor dots or stripes on the faceplate of the screen. The shadow mask lies a few millimetres behind the faceplate, and the perforations in it ensure that the 'red', 'green' and 'blue' guns excite (more or less exclusively) the appropriate subset of phosphor elements. The electron beams are deflected electromagnetically from left to right, and, more slowly, from the top to the bottom of the screen, so that the modulation of the electron flux in time draws out a detailed two-dimensional image.

There exist several arrangements of guns, perforations, and phosphor elements (Wheeler and Clark, 1992). In the older 'delta gun' design, the three guns are arranged in a triangle around the neck of the vacuum tube and the perforations are circular. In the Trinitron design, introduced by Sony, the electron beams are drawn from three horizontally-aligned cathodes and the corresponding perforations are nearly continuous, vertical apertures.[2] Many other modern monitors have three horizontally-aligned guns and a shadow mask perforated with vertical slots. Clearly the purity of the colours obtained will be impaired if the shadow mask becomes misaligned relative to the triads of phosphor elements or becomes magnetized or becomes distorted through heating.

Any distortion of the temporal modulation of the electron beam – introduced by the stimulus generator, the cables, the electronics of the monitor, or nearby equipment – will necessarily be translated into a spatial distortion; and this is something that the new user should constantly keep in mind. The rate at which the entire screen is successively redrawn is the 'frame rate' and has a typical value of 70 Hz for modern display systems. An obvious point – but one sometimes forgotten when multiple targets are presented on a CRT and reaction times or eye-movement latencies are recorded – is that stimuli higher on the screen are presented several milliseconds earlier than lower ones.

Bits per pixel and bits per gun

Typically, the visual scientist drives the colour monitor either from a 'frame store' that allows picture elements (pixels) to be addressed individually, or from a specialized waveform generator that is dedicated to producing repetitive stimuli with high precision. In both cases, a stream of numerical values is transformed by three digital-to-analogue converters (DACs) into three corresponding voltage signals that specify the required outputs of the red, green and blue guns of the monitor. Digital-to-analogue converters take a finite time to settle at a new voltage level, and a longer time is required the greater the resolution with which the voltage is specified. Waveform generators may typically drive DACs with 14-bit resolution (giving 2^{14} specifiable levels), but such DACs cannot settle rapidly enough to follow pixel-by-pixel changes along a single horizontal scan line. This is why dedicated waveform generators are typically restricted to generating horizontal gratings: in the latter case the outputs of the DACs need to change only between successive horizontal lines of the raster. In contrast, the DACs at the output of a frame store must be able to follow changes from pixel to pixel across a horizontal scan, and DACs that operate at this speed typically have a resolution of 8 or 10 bits. For the user who requires better resolution at the output of a frame store, a possible manoeuvre is to add the outputs of two parallel DACs, one providing a coarsely spaced set of voltage levels and the second providing a set of smaller, finely spaced, voltages (Watson *et al.,* 1986; Pelli and Zhang, 1991).[3]

---

[2] The vertically strung wires of the Trinitron display need one or two (depending on the screen size) horizontal wires for support, and these are visible as dark lines. The experimenter will normally be able to work round them, once their presence and function are recognized.

[3] An arrangement of this kind is incorporated in the Cambridge Research Systems VSG2/2 board (Cambridge Research Systems Ltd., 80 Riverside Estate, Sir Thomas Longley Rd., Rochester, Kent, ME2 4BH U.K.)

A frame store similar to those in common use is represented in Fig. 1. The maximum number of individually addressable pixels in the image will be determined by the x,y dimensions of this store. The number of 'planes' of memory (i.e. the number of bits per pixel) will determine the number of colours we can concurrently display. If we had only one plane of frame-store memory (i.e. a single bit per pixel), we could specify only two alternative outputs at each position on the screen. To sample colour space adequately we in fact need to be able to control each of the three guns with a resolution of 8 bits (giving us 256 different output levels); and if we wish to measure colour thresholds, we need at least 10 bits per gun (Cowan, 1983a).[4] It is possible to have a frame store 24 planes deep (i.e. 24 bits per pixel) and thus directly to specify three gun values at each pixel. But until recently the computers and graphics boards available to the average visual scientist have not been fast enough to redraw the contents of such a frame store between frames.

The more usual arrangement for recent colour vision research has been that shown in Fig. 1. What is stored for each pixel is not a set of gun values but a pointer to an entry in a look-up table (Rodieck, 1983; Watson *et al.*, 1986). If there are 8 planes of frame-store memory, then there will be only 256 possible entries in the look-up table, since this is the number of addresses that can be specified by 8 bits. But each entry in the look-up table can specify with high precision the voltages to be applied to the three guns of the monitor: eight bits per gun is the very minimum that the colour scientist will wish to have available. The number of planes of frame-store memory thus determines the length of the look-up table and, in turn, the number of combinations of luminance and chromaticity that can be simultaneously present on the screen. The total number of available colours – not just those concurrently present – depends on the resolution of the three digital-to-analogue converters at the output of the graphics board. From the user's point of view, this resolution is represented by the number of 'bits per gun' with which the programmer can specify each entry in the look-up table.

Most systems allow the programmer to alter, between frames, either the contents of the frame store or the part of the frame store that is displayed or the entries in the look-up table (Watson *et al.*, 1986). Provided the graphics board has sufficient memory, rapid substitution of images can be achieved by displaying different subsections of the framestore. Alternatively, if the same entry in the look-up table is assigned to all the pixels of a complex shape or to several shapes distributed across the screen, then all these sub-areas can be simultaneously changed by changing the one entry in the look-up table. If a target area and its background field are linked to different positions in the look-

---

[4] There are, however, poor man's ways of gaining the resolution needed to measure colour thresholds with an 8-bit graphics system. One is to reduce the contrast of the display by optically superposing on it a homogeneous background field (Cole *et al.*, 1993). Another is to use 'dithering' or 'half-toning', gaining an extra bit by setting alternate raster lines to different values (Mulligan, 1986; Mulligan and Stone, 1989; Pelli and Zhang, 1991). A third solution is discussed below under Setting up and calibration.

up table, then the target can be made to appear according to whether the two entries in the look-up table contain the same set of gun values (e.g. Dixon, 1991). When a technique of this kind is used to produce apparent movement, the method is called 'look-up table animation' or 'palette animation' (e.g. Baro and Hughes, 1991).

Pixel rates and video rates

An important factor to consider in choosing a graphics board is the 'pixel rate', the frequency at which new voltages can be sent to the monitor. The highest pixel rate available on a given board is limited by the response speed of the digital-to-analogue converters, which translate the numerical values of the look-up table into voltages on the RGB lines to the monitor. The pixel rate takes a typical value in the range 10-200 MHz.

There is a systematic relationship between pixel rate and the following variables:

- The frame rate (the frequency at which the monitor is set to redraw the whole screen)
- The number of pixels per line of the raster (the limiting horizontal resolution)
- The number of horizontal lines of the raster (the limiting vertical resolution).

The pixel rate sets the limit to the product of these three variables,[5] and the user will often have to compromise between frame rate and spatial resolution. To minimize visible flicker, the visual scientist may wish to use a frame rate of 75 Hz or more,[6] but the higher the chosen rate, the more limited is likely to be the spatial resolution. Thus the speed of the graphics board, and not merely the dimensions of its frame store, often limits the total number of pixels that can be plotted within one frame.

The bandwidth of the monitor should be matched to the pixel rate of the graphics board. The 'bandwidth' or 'video rate' of the monitor is the limiting frequency at which the electron flux of the monitor's guns can be modulated. If the pixel rate is too high, in comparison with the bandwidth of the monitor, then the scanning beam of the monitor will not faithfully translate into spatial modulations the temporal modulations of the output of the graphics board.

A monitor's scan rate is the product of frame rate and number of lines, i.e. the total number of horizontal sweeps per second. For graphics monitors, typical values are in the range of 30–80 kHz, and it is scan rate that manufacturers commonly use to characterize such displays. 'Multiscan' monitors have the capacity to operate at a range of scan rates, but the range of a given monitor will still be limited, and the buyer should match this range to the properties of the

---

[5] Allowance has additionally to be made for the duration of the fly-back, the return of the beam from the end of one line to the beginning of the next. The total of the fly-back times can amount to one third of the total time per frame.

[6] A significantly higher rate is needed for electrophysiological work, since individual cells in the mammalian visual system can follow frequencies well above the human critical-flicker-fusion frequency (Rodieck, 1983).

graphics board. At a given frequency, a monitor tuned to that fixed scan rate may offer better spatial resolution than the equivalent multiscan monitor from the same manufacturer, and so may be preferred if the experimenter is sure in advance of the scan rate to be used.

The pitch of a colour monitor is the width of each triad of phosphor dots or lines. A typical value for a 19-inch Sony GDM series monitor is 0.30 mm; for Hitachi 'Accuvue' monitors, typical values are 0.28 mm for a 16-inch display and 0.31 for a 19-inch display. There is no automatic relationship between the pixels of the framestore and the phosphor triads of the monitor: the user is left to determine what this relationship should be and to ensure that no aliasing occurs between the stimulus pattern and the pattern of phosphor triads (Bauer, 1992). In most colour work it would be inappropriate to have more pixels per line than phosphor triads; for it would then be impossible to control exactly the chromaticity of a given pixel.

It is not only the pitch of the phosphor triads and the pixel rate of the graphics board that determine the spatial resolution that the user can expect from a display. The spatial profile of a single displayed pixel will depend both on the focus of the electron beams and on the extent to which the electronics faithfully reproduce the temporal waveform specified by the digitized representation in the framestore. The DACs at the outputs of the graphics board, the cables, and especially the high-voltage amplifiers of the monitor may all act as low-pass filters, and this temporal filtering will be translated into spatial filtering. Naiman and Makous (1992), measuring a monochrome monitor, found that a nominally square pixel was not radially symmetric when displayed, but rather was elongated in the direction of the horizontal scan. Cables and amplifiers may also introduce distortions other than low-pass filtering (v. infra).

Gamma functions

The most celebrated non-linearity of the CRT is the 'gamma function'. The luminous output of a given phosphor is approximately a power function of the voltage applied to the corresponding gun. The exponent typically has a value of about 2.3 and is conventionally designated γ. This non-linearity arises between the gun voltage and the electron flux, whereas the relationship between electron flux and luminous output is essentially linear – provided there is no change in the EHT (Extra High Tension) accelerating voltage present at the screen.

Setting up and calibration

Brightness and contrast

Of the many external controls on a modern monitor, two of the most important are those called 'brightness' and 'contrast', yet these are often set with little

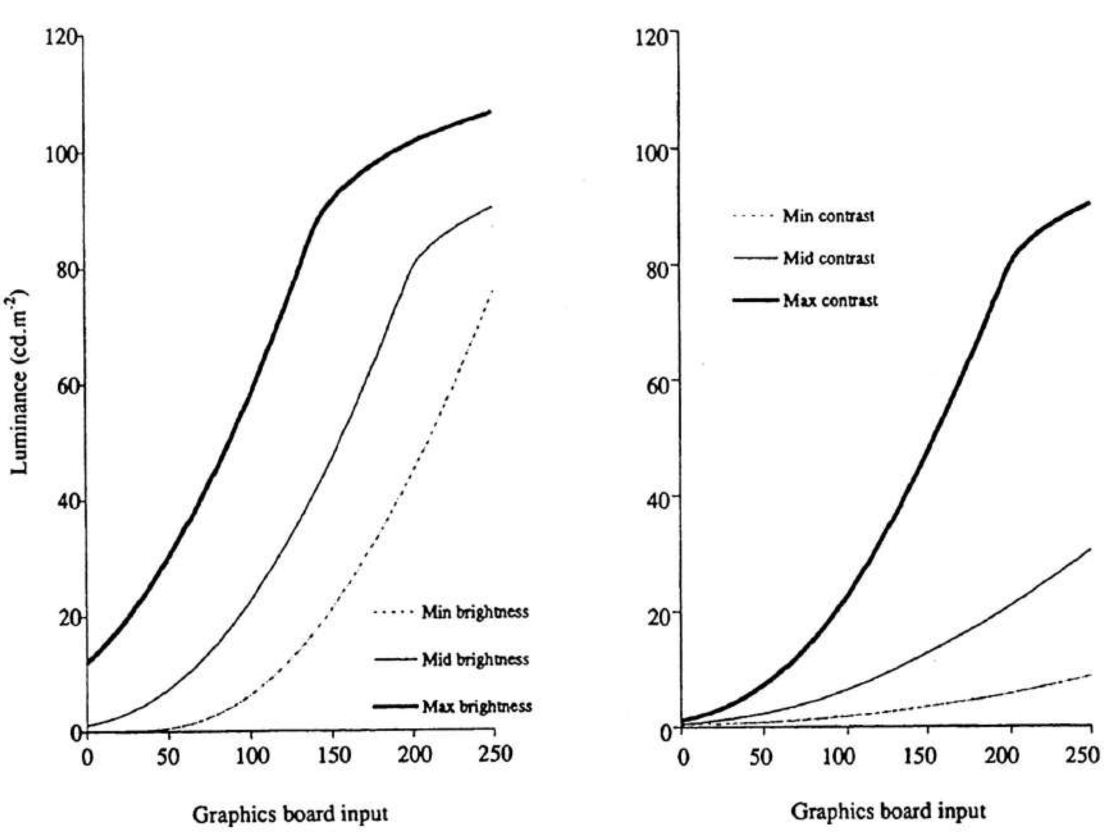

Fig. 2. Left: Gamma functions obtained from a Sony monitor with 'Contrast' set at maximum and 'Brightness' set at three different values. Right: Gamma functions obtained with 'Brightness' set at an intermediate value and 'Contrast' set at three different values.

thought by the newcomer. It is in fact crucial to set both of them carefully before investing the time and the effort demanded by a full calibration.

The very names of the brightness and contrast controls are misleading. The 'brightness' control nominally sets the offset, the outputs of the guns when the nominal input is zero, whereas the 'contrast' control sets the gain and thus the maximum luminance available. In practice, some interaction between the two controls may be observed (Fig. 2).

The visual scientist will usually wish to turn down the brightness until, with the graphics board connected, there is just no visible output when the nominal input is zero. For most purposes, the contrast setting should be the maximal setting that still gives additivity of the guns (v. infra).[7] But for some purposes, the experimenter may choose to set up the monitor differently. If thresholds are to be measured around a fixed background level, then a higher setting of the brightness and a lower setting of the contrast may offer the necessary resolution.

Once the brightness and contrast settings have been chosen, they should be sealed against interference by unthinking graduate students.

[7] An informal way to establish the appropriate level is to illuminate the full screen of the monitor and to increase the contrast setting until the illuminated area visibly expands. The expansion indicates a fall in the screen voltage, and thus in the velocity of the electron beam; the beam is then effectively less stiff and its deflection increases.

Calibration

In order to generate stimuli of known colour and luminance, the experimenter needs to know (a) the chromaticities of the three guns of the monitor and (b) the gamma function for each gun. Four main routes to calibration can be identified and are listed below from the costliest to the cheapest. The necessary apparatus varies in price over at least two orders of magnitude.

(i) Spectroradiometry. By means of a telespectroradiometer the experimenter can measure the radiant output of the monitor in a series of narrow wavelength bands across the spectrum and so calculate the XYZ tristimulus values of the CIE system. Two sources of error deserve mention. First, the light emitted from the display may be partially polarized, and a mirror or grating within the spectroradiometer may act as an analyzer. Check for this by making a second set of measurements with the spectroradiometer rotated through 90 degrees; or ensure that the light is depolarized by an integrating sphere before it enters the instrument. Second, remember that the output of a local region of the screen is a train of intense, brief pulses. The peaks of individual pulses may be 100 times greater than the mean luminance and may saturate the detector. And if these pulses are not synchronized with the sampling interval of the spectroradiometer, the total photon capture will vary from one sample to another, while the apparent chromaticity will also vary since the different phosphors typically have different time constants (v. infra). Some modern spectroradiometers can be automatically synchronized with the display being measured. Alternatively, the modulating detection system of the spectroradiometer can be by-passed and replaced by a simple integrator (Hanson and Verrill, 1991).

It is hardly worth owning a spectroradiometer if one does not also have access to a calibrated standard light source, and associated power supply, with which to calibrate the spectroradiometer itself.

(ii) Electronic colorimeters. Whereas a true spectroradiometer may cost several times as much as a complete colour graphics system, the prices of some electronic colorimeters are of the same order of magnitude as those of graphics systems. Such instruments directly estimate the CIE XYZ tristimulus levels, instead of first obtaining the spectral radiance distribution of the source. The colorimeter typically contains three detectors, which sample the light passed by three colour filters. The three detector/filter combinations have sensitivity spectra that simulate – with more or less precision – the CIE $\bar{x}$, $\bar{y}$, $\bar{z}$ colour-matching functions. Necessarily the colorimeter can be accurate only to the degree that these simulations are accurate; and different instruments vary in the extent to which they approach perfect simulations. For example, in the Minolta Chroma Meter CS100 – a relatively economical device adopted widely by visual scientists – a short-cut is taken in simulating the $\bar{x}$ function: the correspondng filter/detector combination is unimodal in its spectral sensitivity (whereas $\bar{x}$ is bimodal) and the short-wave mode of $\bar{x}$ is simulated by adding the $\bar{z}$ signal,

suitably scaled, to the long-wave lobe of $\bar{x}$. Terstiege and Gundlach (1991) have calculated – for gas-filled lamps rather than monitors – the theoretical average error of several colorimeters, calculated in terms of the CIELUV colour-difference formula. The most accurate was an LMT instrument[8] that incorporates the (expensive) device of partial filtering: small pieces of secondary filter are placed in series with the primary filters to improve the approximation to the $\bar{x}$, $\bar{y}$, $\bar{z}$ functions. In this sample of instruments, and for these sources, the Minolta was the least accurate colorimeter, while Topcon and Photo Research colorimeters were intermediate. The LMT C1200 colorimeter is also recommended in the survey of Berns *et al.,* (1993b).

Despite the inaccuracy of electronic colorimeters, they are stable instruments and they can serve well to check for drift of a research monitor in the field. The procedure used by the National Physical Laboratory (U.K.) is to calibrate a colorimeter, such as the Minolta Chroma Meter, in conjunction with a specific monitor (Hanson and Verrill, 1992). A gamut of stimuli are presented on the monitor and measurements are made concurrently with a reference telespectroradiometer and with the user's colorimeter, the same measuring geometry being adopted for the two instruments. Since each of the three phosphors of the monitor has a stable emission spectrum, the colorimeter can then be reliably used in the field to calibrate this particular monitor: the correction needed in a particular region of the chromaticity diagram will be constant. However, the corrections will not be valid for a different monitor with different emission spectra.

*(iii) Visual colorimeters.*There is much to be said for the use of older instruments, such as the Lovibond Tintometer,[9] in which a visual match is made between light that is drawn from the CRT screen and light that has been drawn from a reference source and has passed through calibrated, stable, glass colour filters (Chamberlin and Chamberlin, 1980). Although such visual measurements are relatively slow and the experimenter is unlikely to be the monozygotic twin of the CIE standard observer, it is difficult to make large errors with a visual instrument, and the calibrations are likely to be as valid as those made with a more expensive electronic instrument. One difficulty is that visual matches become very imprecise in the short-wave corner of the chromaticity diagram (Stiles, 1955), owing, probably, to the saturation of the post-receptoral channel that carries the signals of the short-wave cones (Polden and Mollon, 1980; Mollon and Estévez, 1988). So it is difficult to locate the blue phosphor directly. The way around this is to make measurements nearer the centre of the chromaticity diagram, where post-receptoral channels are more sensitive: if two different excitations of the blue phosphor are combined with two different values of the red/green ratio to give four points in the middle of the

---

[8] Lichtmesstechnik, Helmholtzstr. 9, D-10586 Berlin, Germany

[9] The Tintometer Limited, Waterloo Rd., Salisbury, SP1 2JY, U.K.

diagram, then the chromaticity of the blue phosphor can be found by triangulation.

Visual calibration of the three gamma functions would be clumsy and imprecise, and so a hybrid arrangement will often recommend itself – especially if the colour scientist needs only an approximate estimate of absolute luminance. Measure the chromaticities of the three guns with a visual colorimeter. With the same instrument measure the chromaticity of the white obtained when all guns are maximally excited. Given this information, reference to the CIE standard observer allows you to calculate the relative luminances of the three guns at maximum. The gamma function for each gun can then be measured with a relatively cheap physical detector, such as a silicon photodiode, and can be converted to relative luminance values by reference to the relative luminance at maximum. An approximate estimate of absolute luminance can be made with the visual instrument.

(iv) Use of manufacturers' values. Some of the larger manufacturers of monitors will supply chromaticities for their phosphors, and can provide 'factory-floor' values as opposed to nominal ones. The gamma functions can then be measured with an electronic photometer or a simple photodiode. If you have faith in the manufacturer's chromaticities and if your experiment needs only modest accuracy, then this approach to calibration may be an honourable one.

The calibrations should be made using the graphics board that is to be used in the experiments. Also, since the earth's magnetic field as well as those deriving from nearby apparatus may affect the electron beams, the calibration should be made with the monitor in the position and orientation in which it is to be used. Since the shadow mask and structural metalwork within the monitor may themselves have become magnetized, the monitor should be demagnetized – 'degaussed' – before any calibrations are begun. Modern monitors incorporate a degaussing circuit that operates when the monitor is turned on: an alternating current of the order of 1A, derived from the mains supply, is passed through coils mounted on the tube (Hutson *et al.*, 1990, pp. 187–189). The circuit incorporates resistors that have the property of increasing their resistance as they become hot: so the current through the degaussing coils falls to a low level within a few seconds. Some monitors additionally offer an external degaussing button. Other things being equal, the visual scientist will prefer to buy a monitor with the latter facility, since he or she may wish to leave the monitor turned on for a period to minimize warm-up effects (v. infra). It is also possible to purchase an external degaussing coil or wand, a portable electromagnet that is powered from the a.c. supply. Merely turning a warm monitor on and off may not achieve degaussing, since the 'positive temperature coefficient' resistors in the degaussing circuit must be given time to cool (Hutson *et al.*, 1990).

The 'gamma functions' of the monitor are seldom in fact found to be exactly of the form $L = V^{\gamma}$ (Cowan, 1983b), and in the older literature there is much discussion of how best to estimate the true function from a set of calibrated positions on the curve (for references, see Post (1992)). However, modern spectroradiometers, such as the Photo Research Spectrascan, and colorimeters, such as the Minolta CS100 Chroma Meter, can be interrogated via a serial interface; so it is feasible to set the host computer to measure the output of each gun at each of 256 levels. Our own recommendation is to establish a computer file containing the luminance values for each of the 256 values for each gun and to require the experimental program to seek the gun inputs closest to the values theoretically required.[10] The advantage of automated calibration is that the host computer's time is cheap and the measurements can usually be repeated or integrated until a smooth, monotonic function is obtained.

The non-additivities and non-uniformities of monitors

In principle, once one knows the chromaticities of the phosphors and the gamma function for each gun, one can calculate the gun values needed to generate any stimulus within the gamut of one's monitor (Cowan, 1983b). One's calculation will depend, however, on a number of assumptions, each of which may be mistaken (Cowan and Rowell, 1986; Brainard, 1989; Post, 1992).

(i) Phosphor constancy. This is the assumption that the luminous output generated by a given gun has the same spectral power distribution – and thus the same chromaticity – whatever the level of the signal sent to the gun. Cowan and Rowell (1986) and Brainard (1989) found that the assumption of phosphor constancy was generally valid, and that is our own experience. There are two simple ways in which an experimenter may be misled into believing that phosphor constancy does not hold. First, if the brightness knob is mis-set and a low background illumination is present, then the chromaticity of a given gun will appear to change at low gun inputs, owing to the proportionately greater contamination by the background. Second, a systematic but spurious change in chromaticity may be recorded if the Minolta Chroma Meter is used at low levels where the display is flashing and where the manufacturers explicitly do not guarantee the accuracy.

(ii) Independence of guns. This is the assumption that the luminous and chromatic output resulting from a given input to a given gun is independent of the signals sent to the other two guns. If we measure the X, Y, Z outputs when given voltages are sent to each gun separately, we can determine whether the sum of the three sets of X, Y, Z values is obtained when the specified voltages are sent to the three guns simultaneously. Cowan and Rowell (1986) found good

---

[10] This approach could be refined by not considering the guns independently but by seeking the triad of gun inputs that minimized the distance from the target combination of luminance and chromaticity, the error being computed in units appropriate to the experiment. Thus, if precision of luminance is most important, the best value for a given gun will not be independent of the values selected for the other guns.

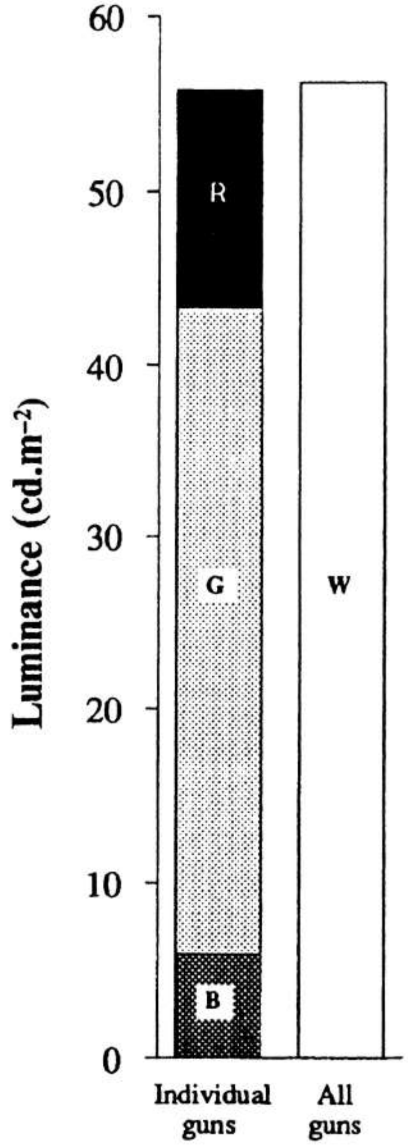

Fig. 3. Result of a luminance additivity test for a Mitsubishi HL20 monitor. The measurements were made with a Minolta Chroma Meter.

additivity for the (now obsolete) Tektronix monitor 690SR. The discrepancies were in the direction of sub-additivity and did not exceed 1%. Since the three tristimulus values were subject to similar sub-additivities, the error in chromaticity coordinates was much smaller, less than one part in 300. Brainard (1989) reported only small sub-additivities for a Barco 5351 monitor driven by a frame buffer manufactured by Number Nine Graphics Corporation: the subadditivities observed were of the order of 3%, were worst for the blue gun, and were worst at low levels. We ourselves have found good additivity for Barco CD351, Mitsubishi HL20 and Sony GVM1400 monitors driven by a Cambridge Research Systems VSG board (see Fig. 3). However, failures of additivity of guns may be observed in both cheap and expensive monitors. Berns *et al.*, (1993) found a serious failure of additivity for a Sony PVM1942Q monitor when it was driven to produce high-luminance images (186 cd.m^{-2}); at lower levels, the additivity became acceptable. In the case of cheap monitors, the non-additivity may arise from poor regulation of the accelerating voltage present at the screen (Rodieck, 1983). In the case of modern expensive monitors, the culprit may be circuits deliberately introduced to ensure compliance with regulations on X-ray emission: these circuits place limits on the total emission. In principle, non-additivities may also arise in the graphics driver board.

(iii) Spatial homogeneity. One assumption that definitely fails is that of spatial uniformity of gun outputs across the screen. Most monitors show a falling-off in the luminous output near the edges of the screen (Brainard, 1989; Post 1992; Cook *et al.*, 1993). The effect may be of the order of 20-25%, and is illustrated in Figure 4. It may be due in part to the reduced cross-section that the shadow-mask aperture presents to the electron beam. In addition to the variation in luminance level, there may also be a change in chromaticity, but any such effect is much smaller: Brainard (1989) found that most of the spatial variation could be accounted for by a single scale factor applied to all three gamma functions.

With a resolution of only 8 bits per gun and a look-up table of only 256 entries, it would be difficult to generate, by computational compensation, a truly uniform field: the quantization of the corrections would probably yield visible steps in parts of the field.

(iv) Spatial additivity. The luminous output from a given region of the screen may vary according to the illumination of other regions. Moreover, the chromaticity obtained from a given region may depend on the chromaticity of

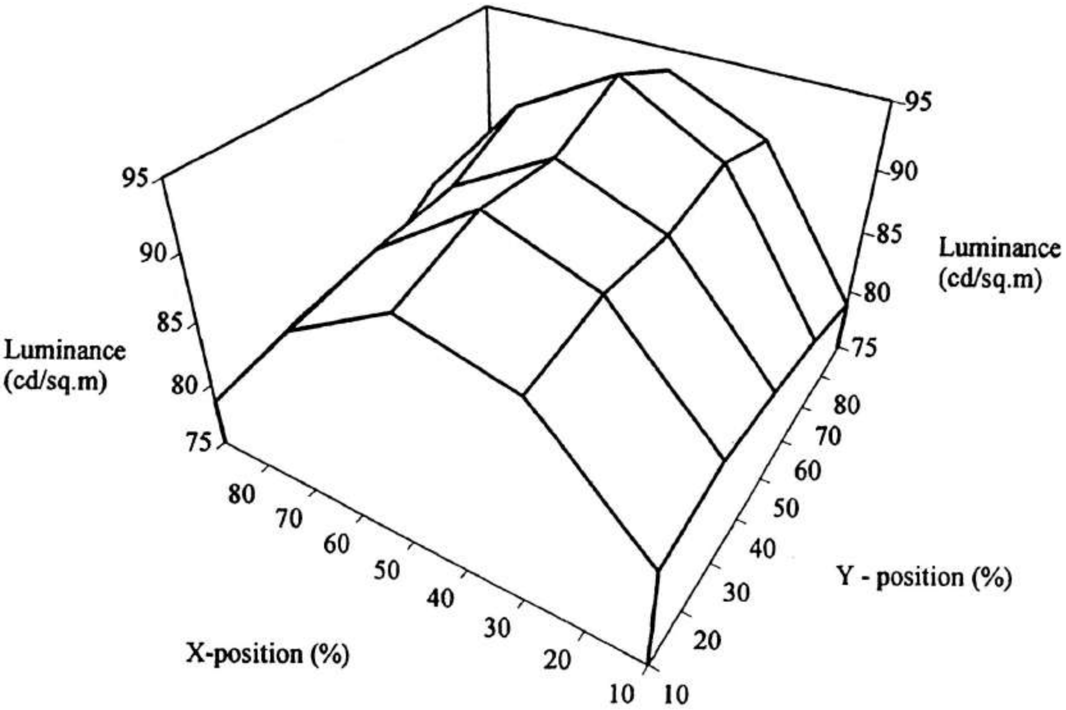

Fig. 4. Variation with screen position of the luminous output of a Sony monitor.

a surrounding region. Such effects are disastrous in experiments on colour contrast and constancy; yet they are only occasionally discussed in papers on those topics.

Brainard (1989) found only small spatial interactions in the case of a Barco 5351 monitor. He measured the output of a central 64 x 38 mm patch in the presence or absence of eight surrounding patches of the same size. The mean percentage change in fractional output was 2.5%. We have observed similarly good behaviour in the case of a Mitsubishi HL20 monitor (although the same monitor showed a spatial gradient in the absolute luminous output, increasing from the left to the right of the display). We have observed serious spatial sub-

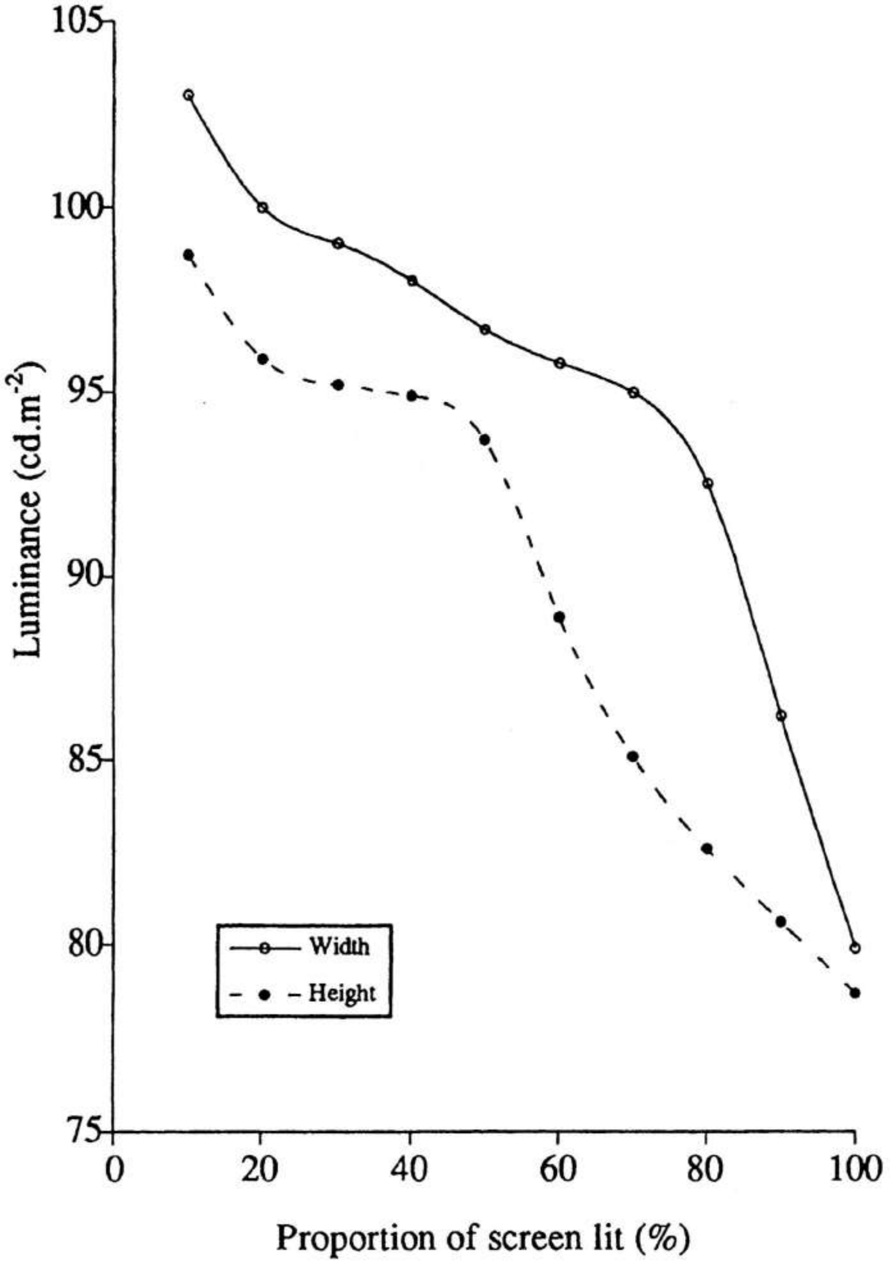

Fig. 5. Spatial sub-additivity for a Sony GVM-1400QM monitor. Luminance was measured at a fixed point in the bottom left-hand corner of the screen, and the width of an illuminated strip was increased horizontally or vertically to occupy increasing proportions of the total screen area.

additivities in the case of Sony and Eizo displays. Figure 5 illustrates the effect for a Sony GVM1400 monitor. Luminance was measured, with a Minolta Chroma Meter, at a fixed point in the bottom left corner of the display, while an increasingly large proportion of the screen was illuminated. For white light, at maximum output, there was a 20% variation in the measured luminance. Such sub-additivities may reflect limitations in the power supplies of monitors or may arise from circuits deliberately introduced to limit X-ray emission. Performing similar tests, Naiman and Makous (1992) report super-additivity in the case of a monochrome Apple display.

Spatial interactions may become more complex when the measured area and the surround area can vary independently in colour and luminance. Figure 6 shows measurements made for an Apple colour monitor by A.J. Shepherd. In this particular case, the central target was a red square and the surround was masked with matte black card. The nominal luminance and chromaticity of the target remained constant throughout the measurements. The graphs show the measured changes of the target as the surround signal is increased. Red, yellow, purple and white backgrounds (i.e. all backgrounds that require excitation of the red gun) lead to sub-additivity, whereas blue and green backgrounds give super-additivity. Analogous effects are found for other colours of the central square.

Very local interactions, those between neighbouring pixels, are likely to be super-additive in the horizontal direction. In the case of a monochrome display, Naiman (1991) and Naiman and Makous (1992) found that two vertically adjacent pixels were essentially independent, but there was a super-additive and asymmetric interaction between horizontally adjacent pixels: the luminance of a given pixel depended on that of the two preceding pixels in the raster. The super-additive component increased the total luminous flux by about 30%. Naiman and Makous suggest that the slew rate of the monitor's Z amplifier does not allow the voltage to attain full value during the time corresponding to the exposure of a single pixel in the horizontal scan. This horizontal interaction probably explains an otherwise bewildering phenomenon: if the experimenter programs two arrays of narrow bars of the same nominal luminance, one array vertical, the other horizontal (e.g. in an experiment on the McCollough effect), then the horizontal array will look much the brighter, even though the aspect ratio of the display has been adjusted to give equal numbers of pixels for a given horizontal or vertical distance. Once the artifact is recognized, it may be side-stepped in some experiments by using orthogonal arrays each at 45 degrees from the vertical.

Misconvergence and pseudo-misconvergence

Misconvergence is an error in the alignment of the landing of the 'red', 'green' and 'blue' electron beams and is apparent as coloured fringes at the edges of white regions. It is best examined by displaying a regular grid of white pixels.

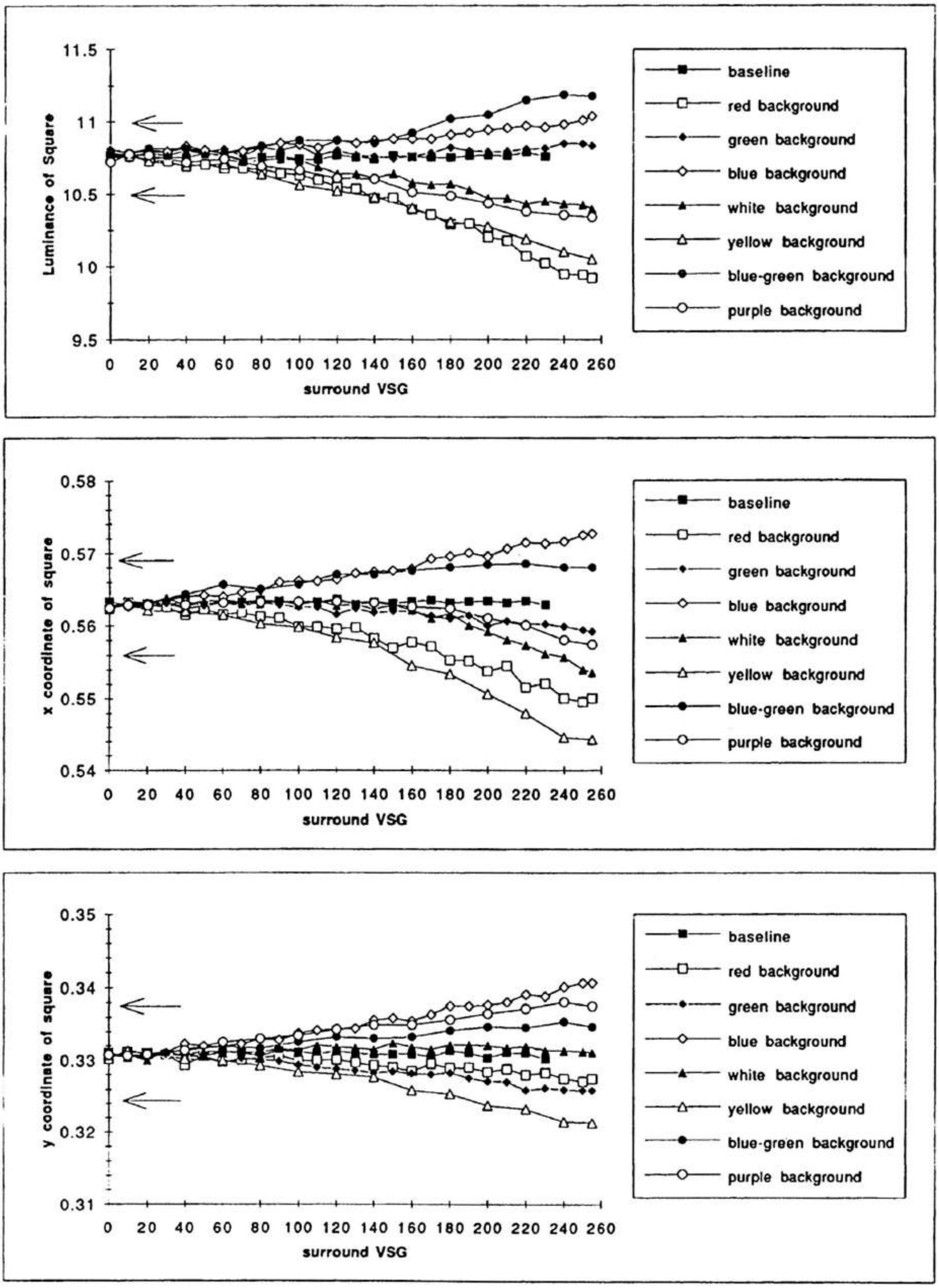

Fig. 6. Spatial interactions for an Apple colour monitor (MO401Z). A central square was illuminated (nominally) by only the red gun. The measurements show how the luminance (top panel) and chromaticity (middle and lower panels) of the central region varied as a function of the gun signals sent to a surrounding region. These measurements were made by A.J. Shepherd.

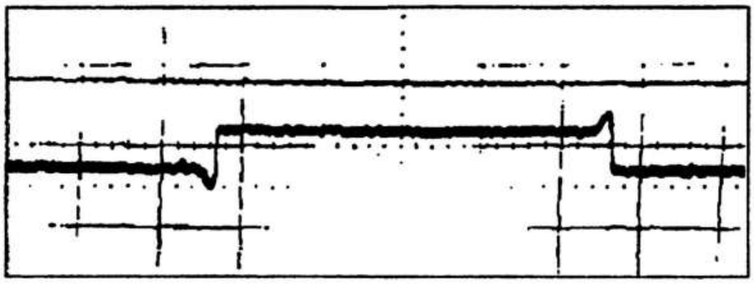

Fig. 7. Oscilloscope trace taken from the green gun input to a Sony graphics monitor from a VSG/2 graphics board under conditions of impedance mismatch. The undershoots and overshoots in the analogue signal led to edge artifacts in an 'equiluminant' random-dot stereogram.

Some monitors offer external adjustments of convergence, which adjust small direct currents that pass through an auxiliary set of electromagnets mounted within the convergence assembly. But in many cases, correction of misconvergence requires the adjustment of internal resistances or magnets. Since the adjustments notoriously interact, the user will usually wish to have such corrections carried out by the manufacturers or by an experienced technician. But is is important to distinguish true convergence errors from electronic artifacts that may simulate misconvergence. Figure 7 shows transient overshoots and undershoots recorded on the cable between a graphics board and a monitor when an impedance mismatch was present between the cable and the monitor. The image on the screen was intended to be an equiluminant red-green random-dot stereogram, and the recording was made from the cable to the green gun. The temporal transients on the line were translated into light and dark striations at the edges of individual picture elements. A pseudo-misconvergence of this kind should be suspected when (a) the primary error is in the horizontal direction and (b) it is constant across the face of the screen. Other sources of bright lines or colour shifts at borders are resonances intentionally introduced into amplifiers to sharpen their responses.

Temporal properties of monitors

In this section, working from short intervals to long, we consider several temporal aspects of the performance of monitors.

Time constants of phosphors

Long after the electron beam has hurried away along its scan, the phosphor dots or stripes continue to emit a decaying stream of photons. Some monitors have phosphors with time constants of tens or even hundreds of milliseconds: the experimenter should be alert to this problem if using simply the monitor supplied with a particular computer. When buying a specialist graphics monitor, the user will normally wish to select a screen with a short time constant, i.e. one of units of milliseconds. However, there are still problems to watch for, especially if very brief or moving coloured targets are to be used in the experiments.

First, the time constants may be markedly different for the red, green and blue phosphors. Vingrys and King-Smith (1986) have suggested how this may lead to a detectable luminance transient when a nominally equiluminant substitution is made between two stimuli.

Secondly, each of the phosphors may in fact be a mixture of different components with different time constants. The cautious experimenter may measure a trace with a photodiode and an oscilloscope and may conclude that the emission has fallen to 1% in one or two milliseconds, but he or she may miss an 'after-glow' from a second component that emits a similar total number of photons spread over hundreds of milliseconds.[11] A real example may lend substance to this warning. Jonides *et al.*, (1982) described experiments in which the subject appeared to integrate spatially, and not retinotopically, information drawn from successive fixations. They postulated a central store that maintained a single spatial representation of the visual world; but they were later obliged to admit that their results had arisen artifactually as a result of phosphor persistence (Jonides *et al.*, 1983).[12]

Beating with the host monitor

Unexplained flicker or drifting horizontal shadows on a good-quality graphics monitor may be due to the proximity of a second monitor – that of the host computer, for example. A sure way to detect such interaction is to alter the frame rate of the graphics display: the drift rate of the shadows will slow down as the frame rate of the host's monitor is approached.

Warm-up times

Monitors typically should be left on for an hour before being calibrated or used experimentally, and many experimenters leave them on continuously or arrange time switches to turn them on before the working day. Harris *et al.*, (1987) found that 12 minutes were sufficient to stabilize a Tektronix 608 monochrome monitor, but at least 3 hours were required in the case of a Digivision CD14 colour monitor. In Fig. 8 we illustrate the behaviour of a Sony monitor over time.

---

[11] In an earlier scandal, almost an entire generation of experimenters missed the long time constants of tachistoscope lamps, until it was shown (Mollon and Polden, 1978) that the fluorescent tubes in common use were coated with a fast short-wave phosphor and a much slower long-wave phosphor.

[12] A mixture of phosphor components (or alternatively a difference between manufacturers) may lie behind the very recent controversy about the persistence of the P31 phosphor. Making direct measurements with a photomultiplier, Westheimer (1993) found that the P31 phosphor decayed to 10% in 600 μsec and to 2% in 2 msec, but Groner *et al.*, (1993) were able to read the screen when a mechanical shutter was opened tens, or even hundreds, of milliseconds after the nominal display had ended. Sherr (1993, p. 91) gives 38 μsec as the time to decay to 10%.

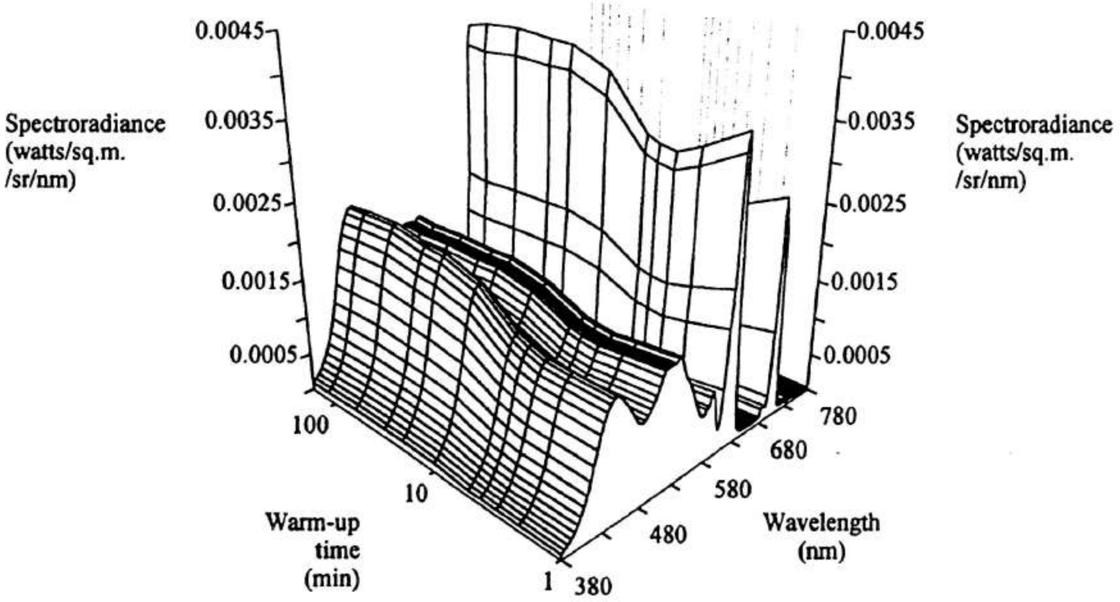

Fig. 8. Warm-up effects for a Sony GVM-1400QM monitor. The radiant output is shown as a function of time and wavelength. The measurements were made by Mr. A. Hansen, using a Photo Research Spectrascan spectroradiometer.

Long-term stability

Some colour graphics monitors show good temporal stability. Brainard (1989) made measurements on a Barco 5351 monitor at an interval of two months: there was only a small (5%) drop in luminous output and, since the three guns covaried, very little change in chromaticity. Lucassen and Walraven (1990), however, found a substantial shift in the gamma functions of a high-resolution 19-inch Hitachi monitor during the first six months of use, and the changes were not ones that could be corrected simply by translating the log gamma functions along the ordinate. They did find that the chromaticity coordinates of the three phosphors were stable.

Recommendations

Design the experiment so that the total illuminated area of the screen is the minimum necessary and remains as constant as possible. Wherever possible, confine stimuli to the central region. Use luminance and spatial noise to isolate chromatic pathways (Reffin *et al.*, 1991) rather than abutting 'equiluminant' regions. Check the initial calibrations by measuring the chromaticities of a sample of experimental colours generated from the stored gamma functions. Calibrate with the spatial array present that will be used in the experiment. Check calibrations often.

Acknowledgements

We are grateful for the advice of many colleagues, including Dr. D. Bauer, Professor C.R. Cavonius, Mr. A. Hanson, Mr. Y. Kim, Mr. B. Regan, Dr. J. Robson, Mr. T. Robson, A.J. Shepherd, Dr. S. Shevell, Dr. Y. Tadmor, Mr. J. Verrill, Dr. M. Webster, Dr. P. Whittle and Dr. S. Würger. Our examination of colour monitors has been supported in part by the SERC and in part by the National Measurement System Policy Unit, Department of Trade and Industry, U.K., through a contract from the National Physical Laboratory, U.K.

References

In addition to references cited in the text, we include here some further sources that will be of use to those who use colour monitors in visual research.

Baro, J.A. and Hughes, H.C. (1991). The display and animation of full-color images in real time on the Macintosh computer. Behav. Res. Meth. Instr. Comp. 23: 537–545.

Bauer, D. (1992). Evaluierung visueller Probleme bei Farbrasterschirmen. Wirtschaftsverlag, Bremerhaven.

Benzschawel, T. (1992). Colorimetry of self-luminous displays. In: Widdel, H. and Post, D.L. (eds.), Color in Electronic Displays: 39–92, Plenum, New York.

Berns, R.S., Motta, R.J. and Gorzynski, M.E. (1993a). CRT Colorimetry. Part I: Theory and Practice. Color Res. Appl. 18: 299–314.

Berns, R.S., Motta, R.J. and Gorzynski, M.E. (1993b). CRT Colorimetry. Part II: Metrology. Color Res. Appl. 18: 315–325.

Brainard, D.H. (1989). Calibration of a computer controlled color monitor. Color Res. Appl. 14: 23–34.

Braun, F. (1897). Ueber ein Verfahren zur Demonstration und zum Studium des zeitlichen Verlaufes variabler Ströme. Annalen der Physik und Chemie 60: 552–559.

Brettel, H. (1988). Quantitative luminance and colour representation with CRT displays. SPIE 1027 ("Image Processing II"): 107–112.

Chamberlin, G.J. and Chamberlin, D.G. (1980). Colour: Its Measurement, Computation and Application. Heyden, London.

Cole, G.R. and Hine, T. (1992). Computation of cone contrasts for color vision research. Behav. Res. Meth. Instr. Comp. 24: 22–27.

Cole, G.R., Hine, T. and McIlhagga, W. (1993). Detection mechanisms in L-, M-, and S-cone contrast space. J. Opt. Soc. Am. A 10: 38–51.

Cook, J.N., Sample, P.A. and Weinreb, R.N. (1993). Solution to spatial inhomogeneity on video monitors. Color Res. Appl. 18: 334–340.

Cowan, W.B. (1983a). Discreteness Artifacts in Raster Display Systems. In: Mollon, J.D. and Sharpe, L.T. (eds.), Colour Vision: Physiology and Psychophysics: 145–154, Academic Press, London.

Cowan, W.B. (1983b). An inexpensive scheme for calibration of a colour monitor in terms of CIE standard coordinates. Comp. Graph. 17: 315–321.

Cowan, W.B. and Rowell, N. (1986). On the gun independence and phosphor constancy of color video monitors. Color Res. Appl. 11 (Suppl.): 34–38.

Dixon, P. (1991). Tachistoscopic display using color-look-up-table display devices. Behav. Res. Meth. Instr. Comp. 23: 380–386.

Groner, R., Groner, M.T., Müller, P., Bischof, W.F. and Di Lollo, V. (1993). On the confounding effects of phosphor persistence in oscilloscopic displays. Vision Res. 33: 913–917.

Hanson, A.R. and Verrill, J.F. (1992). Standardisation in the measurement of the colour of self luminous displays. In: Proceedings of the 22nd Session of the CIE 1991, Melbourne, 87–88, CIE Central Bureau, Vienna.

Harris, J.P., Makepeace, A.P.W. and Troscianko, T.S. (1987). Cathode ray tube displays in psychophysiological research. J. Psychophysiol. 1: 413–429.

Hutson, G., Shepherd, P. and Brice, J. (1990). Colour Television. McGraw-Hill, Maidenhead.

Jackson, R.N., MacDonald, L.W. and Freeman, K.W. (1993). Computer Generated Colour – A Practical Guide to Presentation and Display. John Wiley, Chichester.

Jonides, J., Irwin, D.E. and Yantis, S. (1982). Integrating visual information from successive fixations. Science 215: 192–194.

Jonides, J., Irwin, D.E. and Yantis, S. (1983). Failure to integrate information from successive fixations. Science 222: 188.

Keller, P.A. (1991). The Cathode-Ray Tube: History and Applications. Palisades Press, New York.

Lucassen, M.P. and Walraven, J. (1990). Evaluation of a simple method for color monitor recalibration. Color Research and Application 15: 321–326. (reprinted in Lucassen, M.P. (1993). Quantitative Studies of Color Constancy, Ph.D. Thesis, Utrecht).

Mollon, J.D. and Estévez, O. (1988). Tyndall's paradox of hue discrimination. J. Opt. Soc. Am. A 5: 151–159.

Mollon, J.D. and Polden, P.G. (1978). On the time constants of tachistoscopes. Quar. J. Exp. Psychol. 30: 555–568.

Mulligan, J.B. (1986). Minimizing quantization errors in digitally-controlled CRT displays. Color Res. Appl. 11: S47–S51.

Mulligan, J.B. and Stone, L.S. (1989). Halftoning method for generation of motion stimuli. J. Opt. Soc. Am. A 6: 1217–1227.

Naiman, A.C. (1991). CRT spatial non-linearities and luminance linearization. In: Morris, R. (ed.), Raster Imaging and Digital Typography. Cambridge University Press, Cambridge.

Naiman, A.C. and Makous, W. (1992). Spatial non-linearities of grayscale CRT pixels. SPIE 1666 (Human Vision, Visual Processing, and Digital Display III): 41–56.

Pelli, D.G. and Zhang, L. (1991). Accurate control of contrast in microcomputer displays. Vision Res. 31: 1337–1350.

Polden, P.G. and Mollon, J.D. (1980). Reversed effect of adapting stimuli on visual sensitivity. Proc. Roy. Soc. Lond. B 210: 235–272.

Post, D.L. (1992). Colorimetric measurement, calibration, and characterization of self-luminous displays. In: Widdel, H. and Post, D.L. (eds.), Color in Electronic Displays: 299–312, Plenum, New York.

Reffin, J.P., Astell, S. and Mollon, J.D. (1991). Trials of a computer-controlled colour vision test that preserves the advantages of pseudoisochromatic plates. In: Drum, B., Moreland, J.D. and Serra, A. (eds.), Colour Vision Deficiencies X, Doc. Ophthalmol. Proc. Ser. 54: 69–76, Kluwer Acad. Publ., Dordrecht.

Rodieck, R.W. (1983). Raster-based colour stimulators. In: Mollon, J.D. and Sharpe, L.T. (eds.), Colour Vision: Physiology and Psychophysics: 131–144, Academic Press, London.

Sherr, S. (1993). Electronic Displays, 2nd ed. John Wiley, New York.

Sproson, W.N. (1983). Colour Science in Television and Display Systems. Hilger, Bristol.

Stiles, W.S. (1955). Interim Report to the Commission Internationale de l'Eclairage, Zurich, 1955, on the National Physical Laboratory's investigation of Colour-Matching. Optica Acta 2: 168–181.

Terstiege, H. and Gundlach, D. (1991). Characterizing the quality of colorimeters. SID 91 Digest: 641–644.

Travis, D. (1991). Effective Color Displays. Academic Press, London.

Vingrys, A.J. and King-Smith, E. (1986). Factors in using color video monitors for assessment of visual thresholds. Color Res. Appl. 11: S57–S62.

Walraven, J. (1992). Color basics for the display designer. In: Widdel, H. and Post, D.L. (eds.), Color in Electronic Displays: 3–38, Plenum, New York.

Watson, A.B., Nielsen, K.R.K., Poirson, A., Fitzhugh, A., Bilson, A., Nguyen, K. and Ahumada, A.J. (1986). Use of a raster framebuffer in vision research. Behav. Res. Meth. Instr. Comp. 18: 587–594.

Westheimer, G. (1993). Phosphor persistence in oscilloscopic displays. Vision Res. 33: 2337–2338.
Wheeler, T.R.H. and Clark, M.G. (1992). CRT Technology. In: Widdel, H. and Post, D.L. (eds.), Color in Electronic Displays: 221–251, Plenum, New York.

Department of Experimental Psychology
University of Cambridge
Downing St.
Cambridge CB2 3EB
U.K.