

PRESIDENTIAL ADDRESS

Tell me, what did you see? The stimulus on computers

JOHN H. KRANTZ

Hanover College, Hanover, Indiana

Most psychology experiments start with a stimulus, and, for an increasing number of studies, the stimulus is presented on a computer monitor. Usually, that monitor is a CRT, although other technologies are becoming available. The monitor is a sampling device; the sampling occurs in four dimensions: spatial, temporal, luminance, and chromatic. This paper reviews some of the important issues in each of these sampling dimensions and gives some recommendations for how to use the monitor effectively to present the stimulus. In general, the position is taken that to understand what the stimulus actually is requires a clear specification of the physical properties of the stimulus, since the actual experience of the stimulus is determined both by the physical variables and by the psychophysical variables of how the stimulus is handled by our sensory systems.

Psychological research most often begins by presenting a stimulus to a subject, even though stimulus-response theories are not in vogue these days. The stimulus is that part of an experiment that the psychological researcher can most carefully and completely manipulate. Thus, the manipulation of the stimulus provides the most powerful set of independent variables that we have. However, the stimulus is not independent of the device used to present the stimulus. To take a crude example, a photograph presented on film does not appear the same when scanned and viewed on a monitor or when printed with a color printer. The impact of the display technology on the stimulus is no less true today than it was with older technologies of stimulus presentation, but it seems that we often take computer technology for granted; often, we are not aware of the impact the monitor has on the stimulus. The small variations in how the stimulus is generated, coupled with our own sensory mechanisms, can often obliterate desired differences between stimuli or magnify differences between stimuli that are not desired. This paper reviews several of the issues regarding presenting a stimulus on a computer and its attendant monitor in order to help researchers use this technology more effectively. The ease of developing experiments on computer by people not aware of the way that computers and their attendant hardware work, the growing number of computer programs for laboratory use, and the growing number of experiments

being done over the Web suggest that this topic is timely for review.

It is important to understand that all monitors sample the image. The sampling of the image by the monitor occurs in four domains: space, time, luminance, and color. The fact that monitors sample means that a simple copy of the image is not created. In other words, information from the image is lost. It should also be noted that the monitor does not perform anything approximating a probability sample. The stimulus, as presented on the screen, can differ in distinct ways from the desired nonsampled image. Some stimuli are more faithfully reproduced than others, and it is important to understand the impact of sampling in order to maximize the correspondence between the desired image and the obtained image. The idea that images can be sampled is not new. Printers also sample, but only spatially and at a much higher resolution. In addition, the eye has sometimes been considered a sampling device. In this conception, it is important to match the sampling characteristics of the stimulus generation device to the sampling characteristics of the eye (Silverstein, Krantz, Gomer, Yeh, & Monty, 1990).

Most of this review will focus on the venerable CRT, which is still the dominant display technology attached to the computer. LCDs are common on laptops, and plasma displays may appear on our desks in the future, but I doubt if I will see one soon. Where appropriate and possible, comparisons with other technologies will be made. For example, the fact that monitors sample does not depend on the technology, although the nature of the sampling may differ depending on the technology.

Each of the four sampling domains will be considered in turn. However, in the background, it is important to note that monitors are electronic devices that are not completely stable. One of the recommendations at the end of this

I would like to acknowledge Peggy Ann Johnston for introducing me to the fascinating topic of sensation and perception, Keith White for introducing me to the rigor of the methods in this field, and, finally, Lou Silverstein for showing how wonderful the applications of this field can be. Correspondence should be addressed to J. H. Krantz, P. O. Box 890, Hanover College, Hanover, IN 47243 (e-mail: krantzj@hanover.edu).

paper will be for us, as researchers, to do more about calibrating our monitors. Since they do not stay at a fixed value, these calibrations need to be done before each study.

SPATIAL SAMPLING

The Pixel

The spatial sampling element of the display is called a *pixel*, which stands for picture element. The pixel is the smallest addressable full-color element on the monitor surface. The pixel needs to be distinguished from the pattern of color dots on the surface of the monitor (Silverstein et al., 1990). These color dots represent the primaries that are used to generate the range of colors produced by the monitor (Silverstein & Merrifield, 1985). (See the Chromatic Sampling section below for further information on color reproduction on the monitor.)

However, each dot can only produce the color for that primary; therefore, the dots are not full-color elements. Thus, they cannot reproduce any part of an image unless that image is monochromatic for that primary color. To produce a sampled element of the image, at least one dot of each primary needs to be used; these primaries need to be close together so that they fall within the spatial summation of the eye to produce a full-color element of the original image. This fact is hidden on the CRT and on most LCDs because the screen matrix is not addressed directly. Thus, on your desktop or laptop computer monitor, if the resolution is 800×600 pixels, then you have 800×600 full-color pixels. In general, most monitors today meet this requirement of the visual system.

One of the basic assumptions of a good sampling technique is that each element in a sample is independent of every other element of the sample. That is largely but not completely true on the monitor. On the CRT, the vertical addressing is independent as the electron guns sweep horizontally across the screen. Thus, for 2 pixels that are positioned adjacent vertically relative to each other, what is commanded on the upper of the 2 pixels will in no way affect what the monitor displays on the lower of the 2 pixels. The same cannot be said for a pair of pixels that are adjacent horizontally. On the CRT, the ability to address a horizontal pair of pixels independently depends on the bandwidth of the drive voltage for the electron guns that activate the phosphors on the screen surface (Pelli, 1997). Translated, that means the independence of two horizontally adjacent pixels depends on how fast the voltage can swing from very high to very low voltages. Recall that the CRT is an analog system; therefore, in order to have a white pixel next to a black pixel, all the intermediate voltage values representing all of the intermediate luminance levels have to be shifted through. The greater the change in desired luminance value between the horizontally adjacent pixels, the greater the chance the luminance change between the two pixels will fall short of the desired value. As Pelli (1997) has indicated, you can

measure how severe this problem is on your monitor by putting a very fine grating on the screen. The grating should be composed of lines 1 pixel wide, with lines of the minimum luminance alternating with lines of the maximum luminance. This grating is very fine and merges into an intermediate gray from any moderate viewing distance. Make two versions of this grating, one with the lines going horizontally and one going vertically. The gratings should be the same average intensity, since they are made of 50% white and 50% black. However, for many monitors, the luminance or brightness is far greater for the horizontal lines than for the vertical lines.¹ If you do not have a photometer handy (a luminance measuring device), you can get a powerful impression of the problem by using an animated GIF that switches between the horizontal and vertical versions of the grating at the rate of about 4 Hz (Pelli, 1997). The flicker caused by the luminance change between the two gratings is very noticeable. Figure 1 is a graph of luminance ratios for the vertical grating divided by the mean luminance for the horizontal grating using the static versions of the two gratings. The percentage for the ratio should be 100 since the luminance should be the same for the two gratings. As is clear from Figure 1, the ratio can vary widely and be dramatically short of the optimal value. It is important to note that several of these monitors are exactly the same brand and model and all are recent, high-quality/high-resolution CRTs. Each display used for research needs to be measured independently. Notice, in Figure 1, that the LCD gives about the best response. Since, on the LCD and most other display technologies, pixel addressing is truly digital and, thus, the pixels are addressed far more independently, LCDs have less of a problem with pixel interactions. For displays that have problems with pixel interactions, use stimuli that do not vary greatly in luminance from the background and that are not real fine. It might be worth building a pair of gratings of the desired width and flipping between vertical and horizontal versions to see whether that width no longer causes the display to flicker.

Spatial Inhomogeneity

Spatial inhomogeneity refers to the fact that the same commanded luminance or brightness value at different screen locations could lead to different luminance output. For example, if the screen is covered with a plain white, the brightness of that white is supposed to be the same all across the screen. If your luminance is not constant, then you might have variations in contrast that could add a confound to such dependent measures as reaction time and accuracy of detection (Krantz & Silverstein, 1990; Sanders & McCormick, 1987). As Figure 2 shows, the luminance is not the same across the screen surface. The data plotted in Figure 2 are from measurements taken from one representative CRT, a standard high-resolution 17-in. monitor. The measurements are for white and the three primaries, all at full intensity in a solid figure cov-

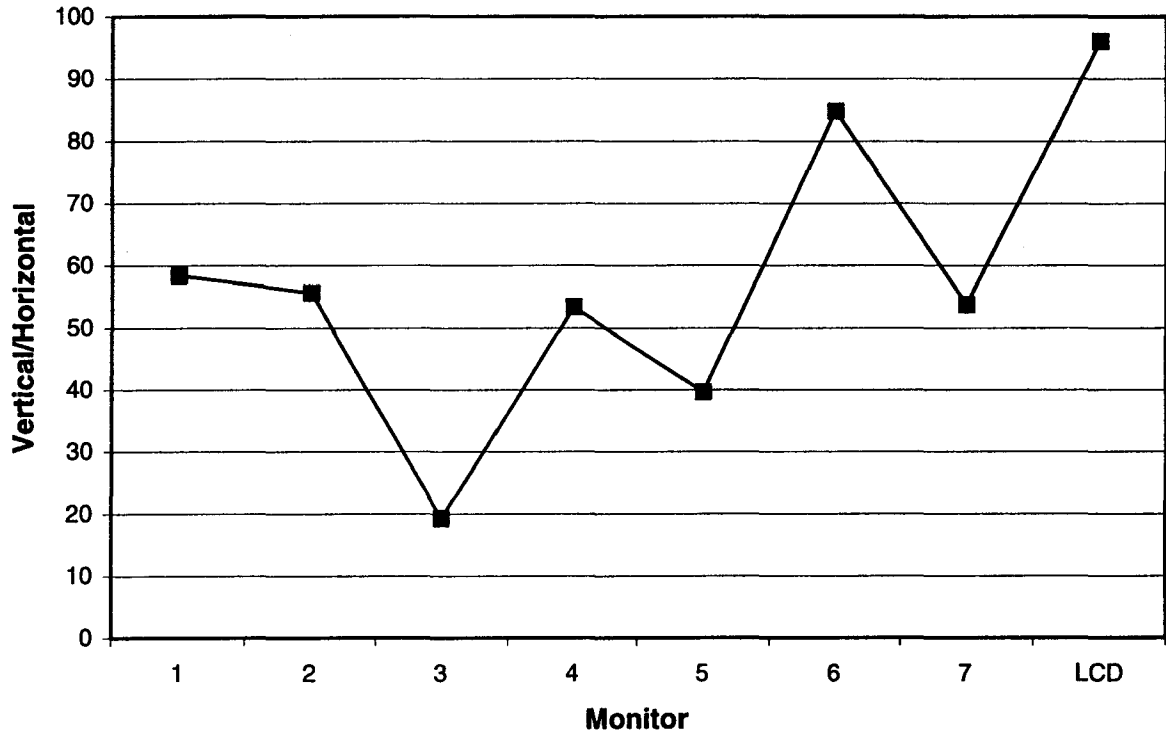


Figure 1. Ratio of luminance for a 1-pixel-wide grating when oriented vertically versus when oriented horizontally. The ratio is converted to percentages. The measurements were taken on seven CRTs (labeled 1–7) and on one LCD. The ratio should be near 100% if there are no interactions between adjacent pixels.

ering over 90% of the screen surface. These data are plotted as a percentage change in the luminance from the dimmest part of the screen. As you can see, the center of the screen is the brightest region of the screen, with the edges less bright (the blue gun is a bit more erratic). This pattern is not uncommon on the CRT. The differences on this display were up to 20% over the dimmest region of the screen.

While the LCD showed itself superior in terms of avoiding pixel interactions, this technology does not seem to have better performance in terms of spatial inhomogeneity. I took the same measurements on a laptop screen, which is a 12-in. 800×600 diagonal active matrix LCD. On this screen, the luminance varied by over 40% across the screen surface. In general, on this display, the brightest regions were near the edges and the dimmest in the center. Since LCDs are newer technology, these measures cannot be considered as standard, but certainly luminance homogeneity cannot be assumed for LCD displays.

There are two ways to minimize the problems of spatial inhomogeneity: limit your stimuli to the center of the display or use the corrections methods, such as those by Cook, Sample, and Weinreb (1993) and Hu and Klein (1994), to smooth out the spatial inhomogeneities. The spatial inhomogeneities will change with different settings of contrast and brightness on the monitor, so these settings need to be fixed during an experiment.

Aliasing

One of the most obvious differences between most computer graphics and hand-drawn lines is the presence of small shifts in position in the computer line. Since the pixels are on a square mosaic, when the line is not horizontal, vertical, or at 45° , it will have small “jaggies” where it moves its position by a 1-pixel step to make up for the mismatch between the desired line and the computers ability to draw that line. Because screen resolution has gone up as displays have gotten better, the steps in the lines have become smaller, but they have not disappeared. This imperfection in the graphic image is called *aliasing*. Aliasing arises when the desired image is beyond the sampling ability of the pixel mosaic. Technically, aliasing results from putting an image on the screen that has spatial frequencies that exceed the capability of the display. Spatial frequency is a concept that relates roughly to the size of the feature of an object or image. Features such as sharp edges or fine details are made up of high spatial frequencies. Features such as smooth gradients or large general shapes are made up of low spatial frequencies.

The aliased spatial frequencies—those in an image beyond the display’s ability to reproduce accurately—wrap around and appear in the displayed image as lower spatial frequencies, which are easily visible (Silverstein et al., 1990). Here is a nice interaction between display

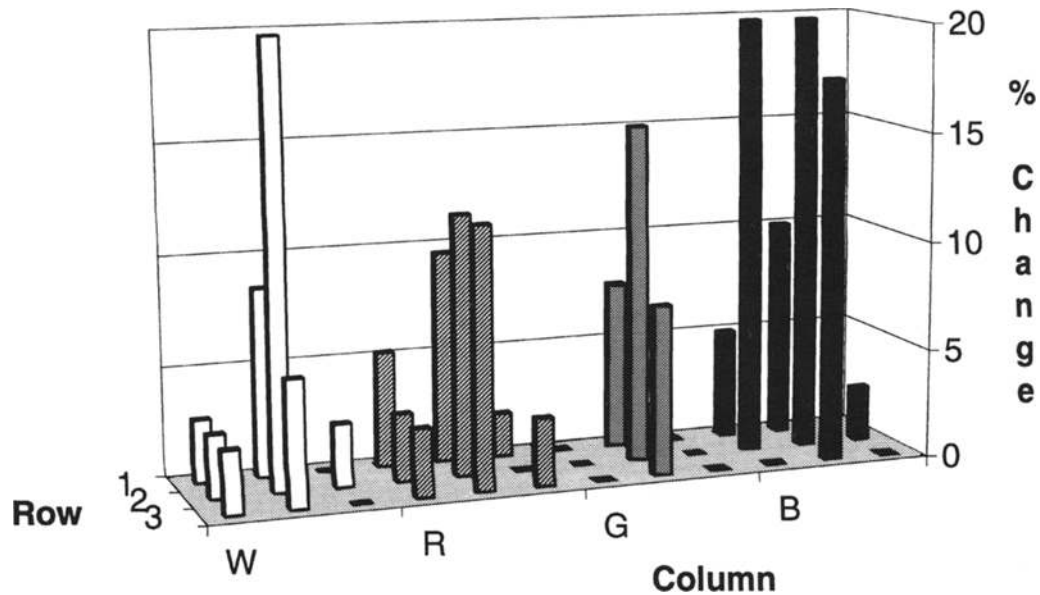


Figure 2. Relative luminance across the surface of one CRT for white (W) and the three primaries (RGB). The luminance is expressed as a percentage of change from the dimmest region of the screen.

sampling and visual system sampling that works to the detriment of the image quality of the stimulus. The jaggies make sharp edges, and sharp edges are very noticeable by our eyes. Through lateral inhibition, edges such as those in the jaggies are enhanced and stand out (Schiffman, 1996). Take the classic Mach Bands as an example (see Krantz, 1999, for several examples of this basic phenomenon). Perceptually, a great deal of aliasing appears as effects of the individual pixels. Take the jaggies of the diagonal line. The jaggies are 1 pixel in dimension, and these discrete steps make the pixel structure visible. Contrast this situation with viewing a photograph of a natural scene on a computer screen. Most natural scenes do not contain aliases, since there are usually not really high spatial frequencies in natural images to cause aliases. In this case, the pixels are not visible at all, and the edges of objects appear smooth and continuous, even when off axis (Infante, 1985).

The best way to avoid jaggies is to bandlimit or antialias the image. In practical terms, that means to blur the image a little bit. Using the technical terminology from above, the range of spatial frequencies is limited to a range that the display can produce. An antialiased line has the same pixel steps in the line, but they are rendered imperceptible because the edges of the line are dimmed and the image does not have real sharp transitions. As a result, the viewer is not made aware of the pixels when looking at the line. The technique dims the sharp edges below where they are noticeable (see the minimal contours demonstration by Krantz, 1999). Often, a gaussian filter is used to bandlimit the image. The electron beam on the CRT already does some bandlimiting of the image, and it is approximately gaussian in shape (Lyons & Farrell, 1989). Since computers and LCDs are capable of

only discrete luminance levels, the minimum number of gray levels necessary to do good bandlimiting edges of graphic images such as lines and alphanumeric characters has been determined. Depending on the resolution of the screen, a good rule of thumb is to use 3 bits or 8 levels between the line color or brightness and the background (Krantz & Silverstein, 1989, 1990; Rogowitz, 1988).

TEMPORAL SAMPLING

There are three issues related to temporal sampling of an image: rendering apparent motion, eliminating flicker, and timing of a stimulus. Rendering apparent motion is possible only because of the fact that our visual system will respond to a quick sequence of static images in the same way as for real motion (Schiffman, 1996). All that is required is that the update of the still images meets the needs of the visual system. Rendering apparent motion is not a significant issue on modern displays. Theater films update at 24 Hz and NTSC video at 30 Hz, and both give very good apparent motion. The computer monitor's frame rate of 60 Hz or better is more than enough for getting good apparent motion. So no further discussion will be given on this issue.

Flicker

The light levels of most artificial luminance sources are not continuous but instead flicker on and off or at least fluctuate between higher and lower levels of luminance. Generally, this rate of flicker exceeds the capability of our visual systems to detect it. This visual threshold is called the *critical flicker frequency* or *critical fusion frequency* (CFF; Brown, 1965). The perception of flicker

is different from the update needed for apparent motion showing that these are indeed separate perceptual functions. As noted above, update for movies is 24 Hz, but the standard rate for the CFF is usually cited to be 60 Hz (Brown, 1965). As is true of most artificial light sources, the computer monitor is not on continuously but is on briefly and off most of the time between frames. The typical duration of a phosphor in a CRT is 4 msec of the 16.7 msec between frames of a 60-Hz monitor (Bridgeman, 1998). So the monitor flicker rate should be fast enough to eliminate the perception of flicker; however, this standard threshold value does not take into account many variables that affect our perception of flicker, such as image size or image location. Generally, as the size of the flickering stimulus increases in size, so does our sensitivity to flicker increase. Thus, as computer monitors get larger, they will need to flicker faster than 60 Hz, which has been the standard for so many years. It is recommended that the flicker rate of a computer monitor be between 66 and 120 Hz (Bridgeman, 1998).

Timing a Stimulus

It is often important to be able to state precisely how long a stimulus has been displayed and when it appears and is removed. The first timing issue results from the fact cited in the last section that pixels are not on continuously during a frame. In part, this blank period between frames is necessary for the perception of clean motion so that we do not integrate or average too much of the two successive frames. If you watch motion on a slow LCD or recall the scrolling of text on the slow green monochrome monitors of not many years ago, you can appreciate how having a pixel on too long disrupts the perception of motion.

However, researchers often do not take this blank period into account when reporting stimulus timing. Stimulus duration is usually calculated as $d = n/f$ (Bridgeman, 1998), where d is the stimulus duration, n is the number of frames the stimulus is on, and f is the frame rate. However, this equation does not take into account the decay of the screen phosphor or the blanking between frames on other display technologies. This equation assumes that the stimulus is on the entire frame. A more accurate determination of stimulus duration is given by $d = [(n-1)/f] + p$ (Bridgeman, 1998), where d , n , and f are all the same as above, and p is the persistence of the phosphor (Bridgeman, 1998). Using 4 msec as the persistence and a 60-Hz monitor, a couple of examples will show the potential magnitude of the errors involved. If the stimulus lasts four frames, the traditional equation gives a result of $d = n/f = 4/.060 = 67$ msec, and the corrected equation gives a timing of $d = (n-1)/f + p = 3/.060 + 4 = 54$ msec, which is a ratio of 1.24 or a 24% error in the timing. In a tachistoscopic presentation of one frame, the proportional error is much larger. The traditional equation gives a stimulus duration of 16.7 msec, but the actual stimulus duration is 4 msec. The error is over 400%. For long stimuli, the traditional calculation gives suffi-

ciently accurate results, but many psychological experiments use a brief presentation of stimuli in which the duration of the pixel needs to be used in determining the duration of the stimulus.

Some researchers might argue that, with very short stimuli, the duration is not meaningful to measure. They would argue that, for stimuli shorter than a certain duration, only the total amount of light energy in the stimulus is relevant. They are referring to what is known as *Bloch's law* or the *Bunsen-Roscoe law* (Barlow & Mollon, 1982). This limit is about 30 msec for cones and about 100 msec for the rods, according to many sources. The interpretation of this finding is that all temporal information below this level is lost as a result of the temporal summation by an early level of the sensory system. However, Zacks (1970) shows that this strict interpretation of Bloch's law may not necessarily be the case. He was testing in the dark, when the rods are active. Under these conditions, he found that threshold stimuli of 4 and 81 msec of equal energy were discriminable. In other words, the two stimuli were of equal energy and below the limit for Bloch's law. According to the strict interpretation of this law, they should look identical, and yet they could be distinguished from each other. This finding should not be too surprising in that flicker is perceptible at a frequency below that predicted by Bloch's law, even when the cones are considered. Thus, it seems that there may be many different perceptual functions, each with its own time constants. If this is the case, providing detailed temporal information seems advisable at this point in time, as advised by Bridgeman (1998).

The other potential source of error in timing a stimulus is a result of the multitasking environment of the modern computer (Diehl, 1995). Windows and the Macintosh operating system do not give exclusive time to any job but share time and resources of the central processing unit across several tasks, many of which are background tasks. Normally, in an experiment where it is important to carefully time a stimulus or measure reaction time, the program determines on which frame the stimulus is presented and then begins the timing function with the next command. However, since computer instructions are done sequentially, the timing is not begun exactly at the same time as when the stimulus is presented; however, if the experiment is the only task being done by the computer, this error of a fraction of a millionth of a second is safely ignored. In the multitasking environment, it is possible for adjacent commands in one program to be separated by a much larger period of time, and the frame during which a stimulus is presented and the timing begun can be separated by a much larger interval of time. Figure 3 reports some selected results from Myors (1999). In this task, the computer program simply timed the counting of 35 frame refreshes on a 70-Hz monitor. In other words, this timing task should last 500 msec. Each value represents the mean of 1,000 trials on that system, and the error bars are 2 standard deviations around the mean. The same computer was used but with different operat-

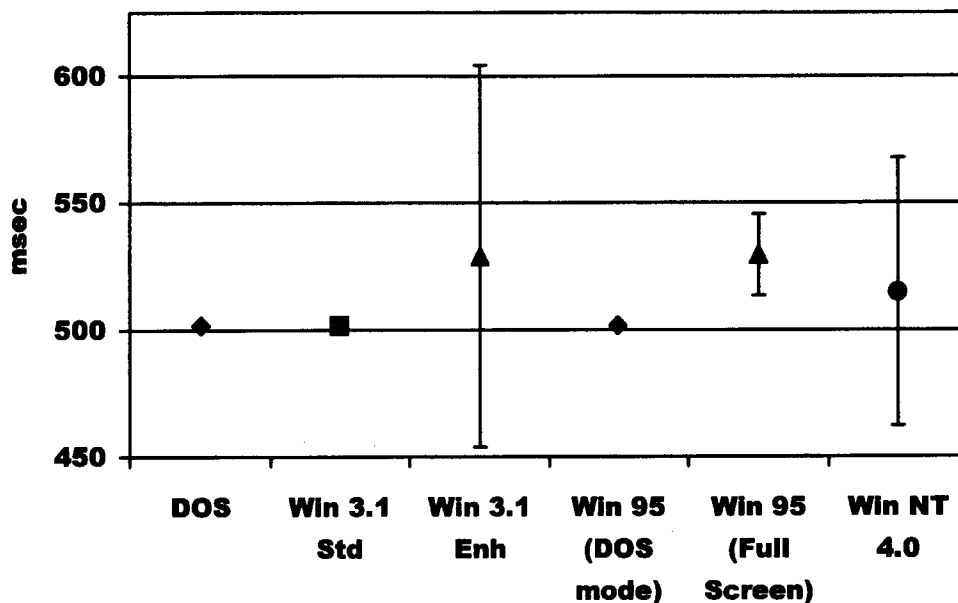


Figure 3. Timings obtained from a routine counting 35 frames on a 70-Hz monitor. The time period should be 500 msec. The means represent the means for 1,000 repetitions of the routine. For each operating system, the mean and 2 standard deviations each side of the mean are indicated. DOS is DOS version 6.2, Win 3.1 std is Windows 3.1 in standard mode, Win 3.1 enh is Windows 3.1 running in enhanced mode, Win 95 (DOS mode) is Windows 95 rebooted into DOS mode, Win 95 (Full Screen) is the routine running in a DOS window set to the full screen but booted in the Windows 95 operating system, and Win NT 4.0 is the Windows NT 4.0 operating system. In all cases, the same computer was used. The figure is developed from data in "Timing Accuracy of PC Programs Running Under DOS and Windows," by B. Myors, 1999, *Behavior Research Methods, Instruments, & Computers*, 31, pp. 326-328. Copyright 1999 by The Psychonomic Society, Inc. Adapted with permission.

ing systems installed. As can be seen, when this task is run on DOS or Windows 3.0 BASIC mode (Win 3.1 std) or Windows 95 in DOS mode [Win 95 (DOS mode)], the timing is accurate enough not to cause worry. The mean errors are about 1–2 msec, and the width of the 2 standard deviations about these means fall inside of the symbol (they actually are plotted). These operating system modes are not multitasking, so the job of the counting is the only task being performed by the computer. In these cases, the accuracy of knowing when the stimulus might be presented can be very accurate. However, the true multitasking operating systems of Windows 3.1 enhanced (Win 3.1 enh), Windows 95 [Win 95 (Full Screen)], and Windows NT 4.0 (Win NT 4.0) all show large constant and variable errors. All three operating systems mistimed the interval on average and had large standard deviations showing great trial-to-trial variability. Of the two, the wide standard deviations are most disturbing, making individual trial's measurement quite suspect. Myors's (1999) study did not manipulate the priority of the jobs, which could vary these results. However, these data show a serious stimulus timing issue. As a result, it is recommended that DOS mode be used when accurate time is necessary. In addition, those experimental packages built for Windows and Mac systems need to report

both method and results, showing their ability to time stimuli in their manuals. The same should be required of method sections of studies run on multitasking computers. It should be noted that these timing errors are only for the presentation of a stimulus. These errors will be compounded by other possible errors in measuring reaction time as well (Myors, 1999).

LUMINANCE SAMPLING

On the computer monitor, discrete values are used to determine luminance values. Values usually ranging from 0 to 255 (8 bits) are used to indicate the luminance desired on the screen. There are a separate 8 bits for each of the primaries on most high-end monitors and occasionally even more bits per primary. It would make life much easier if these 256 values were constant luminance steps. It would be easy to determine the luminance and contrast of your stimulus, and, as mentioned above, these are important values to know since they affect many behavioral responses. However, the values of 0–255 do not relate directly to luminance. The first several values above 0 usually do not generate much luminance and are essentially lost. After that point, an exponential function is often used to model the relationship between the computer val-

ues and the luminance for that color primary. The relationship between the bit value and luminance for a primary is often described by the function $L = L_{0+a}(V - V_0)^g$, where V is the computer bit value, and V_0 is the baseline computer output. V_0 is usually larger than 0 to account for the fact that the first several bit values do not contribute to the luminance on the display. The value of a is a constant used for unit conversion, and g is referred to as the gamma for that primary (g is typically near 2.3). However, it is important to note that all of these constants are affected by the settings of brightness and contrast on the display and on the video card. In addition, these values will be different for each of the primaries, so each primary needs to be calibrated independently (see Olds, Cowan, & Jolicoeur, 1999, for a procedure for determining the gamma of a system). These values are not fixed until the display has warmed up, usually about 20–30 min (Metha, Vingrys, & Badcock, 1993).

CHROMATIC SAMPLING

The production and reproduction of color on the CRT is based on the trichromatic theory of color. This model

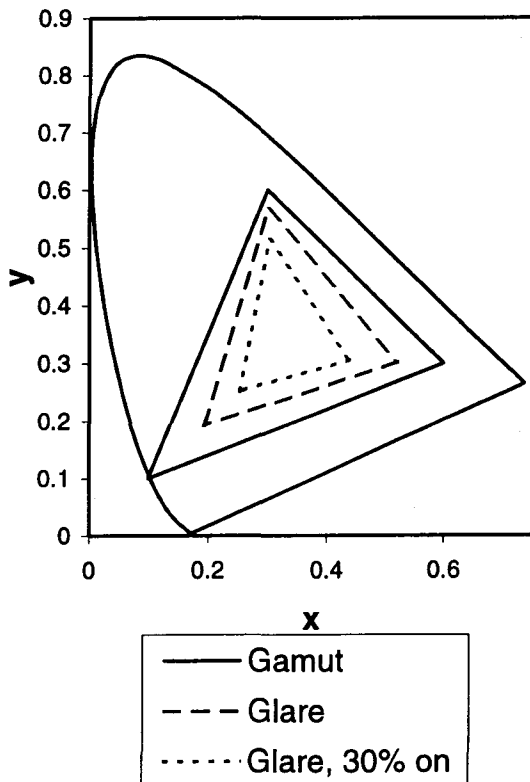


Figure 4. The color gamut for a CRT shown on the CIE 1931 system. The large triangle is the color gamut for a CRT in the dark. This represents the range of colors that the CRT can reproduce with no other light sources around. The middle triangle is the gamut in a moderately lit lab with no external windows. The small triangle is the gamut for the monitor with the same lighting but when each color gun is at 30% intensity. Notice how in a lab with lighting, the primaries are no longer constant.

works as it matches the trichromatic front end of the visual system, the three cone classes. The trichromatic theory of color is modeled in a set of equations developed by the CIE, an international commission that sets standards for lighting and illumination. The first set of equations was derived in 1931 with updates in 1960 and 1976 (Silverstein & Merrifield, 1985; Wyszecki & Stiles, 1967). These equations are the basis of most systems of color reproduction and not just CRTs.

Part of the sampling of color on the CRT is the result of the luminance sampling along the luminance range for each primary. Since the luminance of each primary is a set of discrete values, the range of colors that can be reproduced is a discrete set as well.

However, a more serious limitation of the color sampling comes from the choice of color primaries for the system. Figure 4 shows color space as represented by the 1931 CIE system. The outer curve represents all possible colors that are visible to the normal human visual system. Each color is indicated by a set of x, y coordinates. The corners of the large triangle presented a fairly standard set of primaries found on CRTs (Silverstein & Merrifield, 1985). Color matching on the CIE system is much like the old color wheel, often taught to children, in which mixes fall between the colors being mixed. Thus, the triangle represents the range of colors that this color device can reproduce. This triangle is called the *color gamut* for this device. Notice that many colors are not possible to reproduce on the color CRT represented in Figure 4. This gamut is fairly representative of most CRTs.

The problem with this rather standard presentation of the color reproduction capability of the CRT that I am making in this paper is that these measurements are taken in the dark, and we rarely view the monitor in the dark. A portion of the light that falls on the display surface is reflected back to the viewer (Krantz, Silverstein, & Yeh, 1992). This reflected light blends with the CRT's light to change the color of the intended pixel in the same way any other mixture works. Since most overhead lights are generally a type of white, this screen reflection serves to wash out the color of the desired image. The middle triangle represents what happens to the color gamut for the central regions of a monitor in a moderately well lit lab without any outside windows. The sun can be far brighter than artificial illumination, so windows are potentially a greater threat to the color image on the screen than overhead lighting can be. Even in this lab, the gamut is measurably reduced toward desaturation. Now look at the inner triangle, where the problem even gets worse. Here is the color gamut in the same lighting conditions, but the primaries are only on at 30% of their full intensity. There is the same lighting in the room, but it has now made the color gamut unstable. The primaries do not have the same coordinates for each intensity of the primary as they do in the dark. Thus, the color I get for any mixture is no longer easily predictable or stable.

If precision is needed in the color of the stimulus, these changes could significantly affect the results. Moreover, recall that, on CRTs, the center of the display is nor-

mally the brightest region of the display. The gamut will be constricted even more around the edges, and stimuli thought to be the same color will have a different color appearance at the edge of the display than at the center. If possible, it is important to conduct studies using the CRT in the dark (Neri, 1990). This conclusion is reinforced by the results Agostini and Bruno (1996), who found that the accuracy of the perceived stimulus is affected by the illuminance or the amount of light falling on the surface of the display.

CONCLUSIONS AND RECOMMENDATIONS

In conclusion, the display is a sampling device that samples in four dimensions: spatial, temporal, luminance, and color. Each of these domains can alter the image and interact with the way our own visual system samples the image. In addition, the sampling does not approximate, in analogy, the optimum of the probability sample that is desired in research for generalizing from sample to population. These deviations may become magnified perceptually in the way the visual system acts. Problems in spatial sampling result from the nature of pixel addressing, bandwidth limits, spatial inhomogeneity, and aliasing. LCDs will have fewer problems on some of these dimensions, especially since their pixels are more independent of each other. The major impact of temporal sampling is on timing of stimuli, which is due to the persistence of the pixel between frames and the errors produced by the multitasking environment of the modern computer. The major issue with luminance sampling is the need for researchers to determine the gamma for each primary. The limits of color reproduction due to the primaries used and glare are the major factors to take into account regarding the chromatic dimension of the stimulus.

Thus, the most basic recommendation is to calibrate displays that are to be used for psychological research. If the data are published, these calibration routines should be better reported in the method sections. Metha et al. (1993) and Olds et al. (1999) give some examples for display calibration techniques. Brainard (1997) has a nice bibliography on psychophysics that lists many articles related to monitor calibration. The degree of calibration needed depends, to some extent, on the stimulus requirements of the study. However, display calibration needs to be much better reported in the method sections of psychological research than is often the case. Moreover, calibrations need to be done periodically, since monitors change over time and settings such as brightness can be changed. In addition to calibration, monitors should be warmed up for 30 min before the experiment. The display is not stable until the warm-up period is complete (Metha et al., 1993). Experiments should be run in the dark because of the reflectivity of the CRT's surface. LCDs do not respond to glare as much but, as a result, can appear dim in a bright environment, and so they should also be run in dark rooms.

Where calibration is not possible, such as on the Web, use relatively high contrast images and avoid stimuli that have very high spatial frequencies. You do not want images that are real small or fine. A contrast ratio of the stimulus to the background above about 3:1 does not alter many behavioral responses such as accuracy and reaction time (Krantz et al., 1992). Thus, using very high contrast images will meet this criterion on most, if not all, monitors. It is also probably a good idea to use only simple variations of color and luminance: The simpler the images, the more faithfully a noncalibrated monitor can reproduce the images. These recommendations will be especially relevant for Web-based research where calibration is impossible (e.g., Krantz, Ballard, & Scher, 1997).

REFERENCES

- AGOSTINI, T., & BRUNO, N. (1996). Lightness contrast in CRT and paper-and-illuminant displays. *Perception & Psychophysics*, *58*, 250-258.
- BARLOW, H. B., & MOLLON, J. D. (1982). Psychophysical measurements of visual performance. In H. B. Barlow & J. D. Mollon (Eds.), *The senses* (pp. 114-132). Cambridge: Cambridge University Press.
- BRAINARD, D. (1997). Tips psychophysics bibliography [Online]. Available URL: <http://color.psych.ucsb.edu/brainard/software/bib.html>
- BRIDGEMAN, B. (1998). Durations of stimuli displayed on video display terminals: $(n - 1)/f + \text{persistence}$. *Psychological Science*, *9*, 232-233.
- BROWN, J. L. (1965). Flicker and intermittent stimulation. In C. H. Graham (Ed.), *Vision and visual perception* (pp. 252-320). New York: Wiley.
- COOK, J. N., SAMPLE, P. A., & WEINREB, R. N. (1993). Solution to spatial inhomogeneity on video monitors. *Color Research & Applications*, *18*, 334-340.
- DIEHL, S. (1995). Windows 95 graphic architecture. *Byte*, *20*, 241-242.
- HU, Q. J., & KLEIN, S. A. (1994). A two-dimensional lookup table to correct the spatial nonlinearity on CRT displays. *Society for Information Display Digest of Technical Papers*, *25*, 19-22.
- INFANTE, C. (1985). On the resolution of raster-scanned CRT displays. *Proceedings of the SID*, *26*, 23-36.
- KRANTZ, J. H. (1999). Demonstration of importance of edges [Online]. Available URL: <http://psychlab1.hanover.edu/Classes/Sensation/Edges/>
- KRANTZ, J. H., BALLARD, J., & SCHER, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, *29*, 264-269.
- KRANTZ, J. H., & SILVERSTEIN, L. D. (1989). The gray-scale asymptote for anti-aliasing graphic images on color matrix displays. *Society for Information Display Digest of Technical Papers*, *20*, 139-141.
- KRANTZ, J. H., & SILVERSTEIN, L. D. (1990). Color matrix display image quality: The effects of spatial and luminance sampling. *Society for Information Display Digest of Technical Papers*, *21*, 29-32.
- KRANTZ, J. H., SILVERSTEIN, L. D., & YEH, Y.-Y. (1992). Visibility of transmissive liquid crystal displays under dynamic lighting conditions. *Human Factors*, *34*, 615-632.
- LYONS, N. P., & FARRELL, J. E. (1989). Linear systems analysis of CRT displays. *Society for Information Display Digest of Technical Papers*, *20*, 220-223.
- METHA, A. B., VINGRYS, A. J., & BADCOCK, D. R. (1993). Calibration of a color monitor for visual psychophysics. *Behavior Research Methods, Instruments, & Computers*, *25*, 371-383.
- MYORS, B. (1999). Timing accuracy of PC programs running under DOS and Windows. *Behavior Research Methods, Instruments, & Computers*, *31*, 322-328.
- NERI, D. F. (1990). Color CRT characterization solely in terms of the CIE system. *Perceptual & Motor Skills*, *71*, 51-64.
- OLDS, E. S., COWAN, W. B., & JOLICOEUR, P. (1999). Effective color CRT calibration techniques for perception research. *Journal of the Optical Society of America A*, *16*, 1501-1505.

- PELLI, D. G. (1997). Pixel independence: Measuring spatial interactions on a CRT Display. *Spatial Vision*, **10**, 443-446.
- ROGOWITZ, B. E. (1988). The psychophysics of spatial sampling. In G. W. Hu, P. E. Mantley, & B. E. Rogowitz (Eds.), *Proceedings of SPIE*, **901**, 130-138.
- SANDERS, M. S., & MCCORMICK, E. J. (1987). *Human factors in engineering and design* (6th ed.). New York: McGraw-Hill.
- SCHIFFMAN, H. R. (1996). *Sensation and perception: An integrated approach* (4th ed.) New York: Wiley.
- SILVERSTEIN, L. M., KRANTZ, J. H., GOMER, F. E., YEH, Y.-Y., & MONTY, R. M. (1990). Effects of spatial sampling and luminance quantization on the image quality of color matrix displays. *Journal of the Optical Society of America A*, **7**, 1955-1968.
- SILVERSTEIN, L. M., & MERRIFIELD, R. M. (1985). *The development and evaluation of color systems for airborne applications* (DOT/FAA Tech. Rep. No. DOT/FAA/PM-85-19). Washington, DC: DOT/FAA.
- WYSZECKI, G., & STILES, W. S. (1967). *Color science—concepts and methods, quantitative data and formulas*. New York: Wiley.
- ZACKS, J. L. (1970). Temporal summation phenomena at threshold: Their relation to visual mechanisms. *Science*, **170**, 197-199.

NOTE

1. Examples of all the test images referenced in this paper can be found at <http://psychlab1.hanover.edu/SCiP/1999/Cal.html>

(Manuscript received November 4, 1999;
accepted for publication January 26, 2000.)