Computer parasites and hosts: Practical advice on how to be a successful parasite at your host's computer installation*

GEORGE SPERLING†
Bell Telephone Laboratories, Murray Hill, New Jersey 07974
and
New York University, New York, New York 10003

The article proposes a solution to the budgetary, computational, psychological, and sociological problems that are encountered by behavioral scientists who contemplate doing on-line computing.

Let us suppose you are a behavioral scientist who discovers that he has a problem that requires on-line computing. The ideal solution would be to have your own private computer. In this way, there are no scheduling problems, no personality conflicts, no multiple convertible interfaces, no useless heavy equipment that always needs to be moved, and no ghosts changing switches at night. However, it may happen that your problem requires a computer that is larger than you can afford for your exclusive use. For example, the computers described so far at this conference, with one exception, are all shared by several users.

If you are foolish enough to persist after you have discovered that you cannot have your own private computer, you are faced with two alternatives: First, you can seek out an existing computer installation that is adequate for your problem.[1] In this installation there will be one individual who is primarily responsible for the computer. He is the host. And you, as the scientist, by working on your own problem, will be a parasite at this installation. You must learn how to subvert the host's computer for your problem. The remainder of this article will deal with instructions for you, as parasite, on how to conduct your affair with the host and the characteristics of this form of interaction.

THE PRACTICE AND PROBLEMS OF PARASITISM

Obtaining access to the computer is utterly trivial when it is run by an authentic host. Usually, it is sufficient to say, "My, that's an impressive computer you have here. But can it so something really complex and important, such as . . . ?" All the speakers who had medium-size computers (about \$70k to \$300k) sounded like ideal hosts to me and are certainly worth investigation.

The main advantage of parasitism is that the host's computer really exists, is functioning, and conceivably could do the job. The disadvantage is that it may not be the first choice of computer. If any of you are disturbed by this, just reconsider the computational and organizational problems that you have heard described in the lectures so far. If not having the ideal computer still disturbs you, I recommend you leave the room because either you have poorly suppressed host tendencies (in which case you are certain to be offended) or you have serious personality difficulties (in which case you are at a meeting of the wrong kind of psychologists).

A second advantage is that the system programs exist and may even be considerably debugged. The disadvantage is that the system programs usually are not adequately documented and are subject to change without notice. The solution is to "freeze" the system at a particular stage. That is, to take a particular version of the system, document it, and maintain it as your own system.

A third advantage is that interfaces exist. Here again, the disadvantage is that the interfaces are almost certainly not documented and are subject to change without notice. This is serious because it means that you may come to the computer and find that it completely defeats you for some unknown reason. Yesterday it worked perfectly. In the meantime, someone has changed a single, unknown one of ten thousand wires, and the computer is completely unusuable for your problem.

No amount of interaction with the machine can solve this particular problem. As behavioral scientists, we recognize immediately that the solution depends on the selective application of reward and aversive control. Unfortunately, aversive control is the more available. It is safest to apply it directly to the guilty one of the host's slaves rather than to the host. However, undocumented changes also can inconvenience the host, and you may actually reward him (slightly) by notifying him of such a change. If he, himself, is the perpetrator, use the occasion for a delicately administered punishment; mention that his best assistant has just wasted the entire morning trying to figure out what went wrong.

A fourth advantage of parasitism is that maintenance is provided. The disadvantage is that maintenance may be unbelievably unenthusiastic when the difficulty doesn't interfere with the host's uses. For example, I found that ground noise caused slight perturbations in the visual displays that disturbed me but not the host. This kind of problem requires the parasite ultimately to supply his own maintenance. This is a far more expensive solution than it seems because the person who does the maintenance must first gain the confidence of the host in order to gain access to the computer, and confidence is acquired only by having him initially apprenticed to the host. This is expensive, and it is one of the real disadvantages of this form of interaction.

What can a parasite offer a host? Token symbiosis in several forms. The first of these is primary capital. For example, we finally bought one-third of the computer we were using. Here one can take advantage of accounting methods. For example, it often is very difficult to buy expensive equipment because it may be taxed as property by the municipality in which it is located and for other reasons. On the other hand, it often is possible to pay a considerable yearly sum in rent. When the computer has been rented for 3 or 4 years, it is vastly depreciated and can be bought for a fraction of the original cost. In fact, the rental payments should be viewed as installment payments with a final lump sum due at the end. Thus, one can buy a third or a half of a 4-year-old system for a 10th of its initial value. In practice, therefore, the primary capital that a parasite can supply is a relatively insubstantial contribution to a host, and owning even a fraction of a computer may alter a successful parasitic relationship.

A far better contribution is "secondary" capital. A true host seldom has time to use his own computer for his own research. When

---

*Originally presented at the North Atlantic Treaty Organization, Advanced Studies Institute on: On-Line Computing for Behavioral Science, Department of Psychology, Sheffield, England. July 16. 1969.

†Requests for reprints should be sent to: G. Sperling, Room 2D518, Bell Telephone Laboratories, Murray Hill, New Jersey 07974.

he needs capital for a bigger, more-fun computer, he must cite the research of his parasites. The true host is happy when his computer is overloaded because he wanted a bigger one anyway, and the overload gives him the excuse to buy the machine he really wants. The pathway to a new computer is the ultimate benefit a parasite can confer on a host.

A third advantage for the host is that the parasite may write programs that are available to the host. Actually, this is irrelevant. If he likes a parasite's program, a host will write his own equivalent version. Other people's programs are never adequate for a true host, and this dissatisfaction extends to system programs as well.

The final basic principle of parasitism is never to pain a host unnecessarily. When stealing time, you must do it painlessly. For example, at all costs try to minimize the frequency of your name on the schedule sheets; just appear when the computer is vacant. Of course, every species of parasite has developed its own devious ways of camouflaging its activity; adaptation to local conditions is the best advice.

In all fairness, I must say that at the Bell Telephone Laboratories I have been fortunate to find a most generous host, Peter Denes. He is difficult to get along with, but so am I, and the fact that we have had successful interaction for over 3 years points to the great stability of the parasite-host relationship. This stability may seem to be the greatest advantage of the parasite-host relation. The parasite thinks that when the host finally discards the computer, he, the parasite, can buy the depreciated computer at a bargain price. Of course, you immediately recognize the fallacy: when he does this, the parasite becomes the host—the ultimate catastrophe.

If ever you are presented with a computer, you must search for a host—a larval host who has not yet found his own computer. Like nature, the computer universe abounds with prospective hosts. Once you have discovered a larval host and brought him near the computer, he'll metamorphose and build his nest there. And you will continue to be able to enjoy the use of the computer for research without knowing anything about how it works inside. This is the ultimate application of behavioral science.

NOTE

1. The second alternative will be dismissed later.

The description and luminous calibration of cathode ray oscilloscope visual displays*

GEORGE SPERLING†
Bell Telephone Laboratories, Murray Hill, New Jersey 07974
and
New York University, New York, New York 10003

A description of a CRO display should include descriptions of: (1) typical display contents (e.g., a photograph), (2) CRO output parameters (e.g., refresh rate), and (3) luminous measurements. Luminous calibrations are unorthodox because CRO displays are discontinuous in space and in time, and because they are sources, not reflectors, of light. The appropriate luminous quantities are luminous *intensity* and the integral of luminous intensity *(luminous directional energy, LDE)*; the appropriate measurements are of *LDE per point* and of *LDE per unit line length.* A simple calibration procedure is described, and the formulas relating these quantities to luminances are given.

A description of a cathode ray oscilloscope (CRO) visual display should be sufficiently detailed to enable a reader to produce a visually equivalent display, either on a CRO or by other means. Such a description need not be insufferably detailed; on the contrary, by judicious selection of the relevant aspects it can be both short and useful. The description should include three kinds of information: (1) a description of the display contents; (2) a description of the CRO output parameters (refresh rate, etc.); and (3) luminous measurements.

The display contents are best described by a photograph of a typical frame with sufficient resolution to define the points, vectors, characters, or scan lines of which the frame is composed. The photograph should contain a size reference and an insert enlargement, if necessary, to show the point or line structure. The accompanying description should include a statement of the viewing distance. Black-on-white pictures reproduce well with ordinary printing methods, whereas white-on-black pictures generally require expensive photographic reproduction to produce acceptable copies. Therefore in preparing figures, it often is desirable to use a film that gives positive transparencies directly (e.g., Polaroid 146L) so that the resulting prints will be negatives of the display.

The description of the output parameters should include the refreshment rate, the rate of plotting points or characters, the total time taken to paint a frame, and the phosphor's time constant and chromaticity (e.g., color temperature, if known).

Luminous measurements should include the *luminous directional energy per point* of intensified points (explained below), the unintensified background screen *luminance*, and the approximate room *illuminance*.

It is in their treatment of luminous quantities that most CRO descriptions fail. Difficulties in luminous calibrations of CRO displays arise for two reasons: (1) whereas most natural objects are reflectors of light and best characterized by their *reflectances* or *luminances*, CRO displays are sources of light and usually are best characterized by their *intensities* or *directional energies*; and (2) whereas most naturally occurring objects are illuminated continuously in time and their luminance changes relatively continuously in space, CRO displays typically are pulsed at discrete instants in time and at discrete points in space. Calibration difficulties are overcome by measuring the appropriate luminous quantities of CRO displays, namely the *luminous directional energy* per point and the *luminous directional energy* per line length. The measurement techniques are described below.

LUMINOUS DIRECTIONAL ENERGY PER POINT

Point Source

A visual point source is defined as a luminous area whose maximum diameter is sufficiently small that the visual response does not depend on the actual distribution of light within the area, but only on the total flux. The largest allowable diameter for a point source varies with overall illumination and with the direction of view, being

---

*Originally presented at the North Atlantic Treaty Organization, Advanced Studies Institute on: On-Line Computing for Behavioral Science, Department of Psychology, Sheffield, England, July 19, 1969.

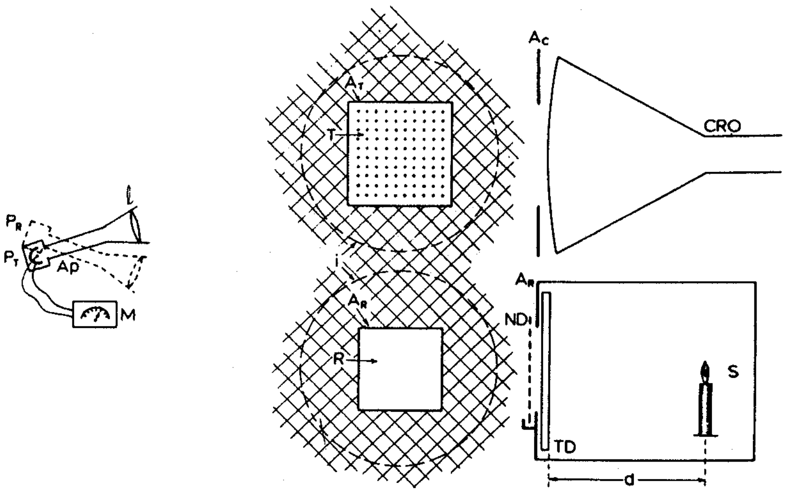


Fig. 1. Reference source (R) and CRO test pattern (T) for calibration of luminous energy per point. The reference surface of known luminance (R) is seen through the aperture A_R; it is formed by a source of known candle power (S) illuminating a plate of uniformly transmitting-diffusing glass (TD). The reference luminance may be varied by altering the distance (d) or by interposing neutral density filters (ND). The test pattern is composed of a 10 by 10 point matrix, seen through an aperture A_T that excludes all other points. The luminous intensity of T relative to R is determined by a photometer which may be in one of two positions: P_R and P_T. The photometer consists of a lens (ℓ), an aperture (Ap), a photodetector (⊂), and a meter (M). The conjugate image of the smaller of the aperture, Ap, and the photosensitive surface, ⊂, is the acceptance region (i) of the photometer and is shown as a dotted circle against A_R and A_T.

smallest in the 2 deg of central vision. As a rule of thumb, 4 min of arc (about 1 mm at 1 m viewing distance) may be taken as the limiting size of a visual point source. When a source is bigger than about 4 min, the spatial distribution of light as well as the total intensity must be given. The spot diameter of a CRO beam (plus the viewing distance) usually is a sufficient spatial characterization; this is given implicitly in a photograph of the display.

For photometric purposes, "photometric point sources" can be much bigger; even 20 deg is permissible when the entire source lies within the uniform acceptance region of the photometer. Obviously, photometric point sources need not be visual point sources.

Luminous Directional Energy per Point (LDE)

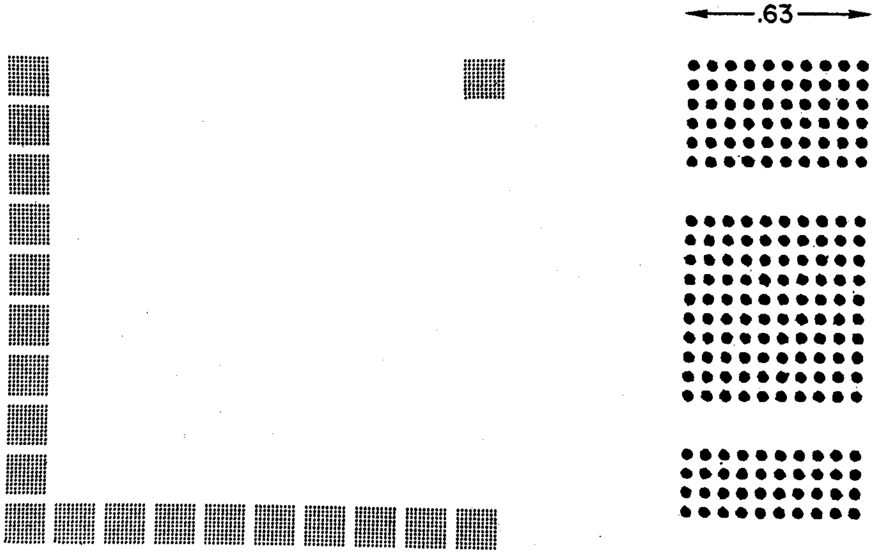
LDE is defined as the luminous intensity (in the direction of view) from a single intensification of a single point, integrated over time. For example, a point intensified for 1 microsec will be quite visible in a single flash under typical room viewing conditions when it produces a luminous output of 1 candle-microsec, that is, when the point produces the same luminous output as a standard candle would have produced if it were exposed for 1 microsec. As a rule of thumb, the temporal distribution of light from a single intensification can be ignored when either or both: (1) the luminous output is confined to 20 msec, and (2) the refresh rate exceeds about 20 Hz. In all these cases, only the integrated luminous intensity (total amount of light) resulting from a single intensification need by known.[1] This quantity is called the *luminous directional energy per point.*

Measurement of Luminous Directional Energy per Point

The strategy of this calibration is to compare a reference source of known luminance with a comparably sized dot pattern on the CRO. The basic tenet of absolute luminance calibrations is to avoid big errors, little errors being unavoidable but of little practical importance in vision. By making the reference source the same size and luminance as the CRO dot pattern, the eye can be used as a null meter. At the least, this is helpful in avoiding gross errors of calibration, and it is convenient for a quick check when a system malfunction is suspected.

The reference source is made from: (1) a surface of known luminance ("photometric brightness"), e.g., 1 fL; and (2) an aperture of known area, e.g., a square 1-in. hole.

Together, the luminous surface and the aperture form a standard source of luminous intensity (Fig. 1). Of course, if a standard source of luminous intensity already is available, it may be used instead, but unequal areas make visual comparison unreliable.



(a) (b)

Fig. 2. (a) A negative picture of a CRO test pattern. Each individual square is 5/8 in. wide and composed of a 10 by 10 point matrix. (b) Enlarged detail showing point structure; spot diameter is about 0.03 in.

Note that a perfectly reflecting-diffusing surface (approximated by white paper), placed 1 ft away from a standard candle (candela) will have a luminance of 1 fL. A surface may also be considered as a source of light. The 1-fL surface emits π^{-1} cd/ft^2 (when considered as a point source, i.e., when viewed perpendicularly from afar) so that a 1-in. square area is equivalent to

$$\frac{1\ \text{cd}}{\pi\ \text{ft}^2} \times \frac{1\ \text{ft}^2}{144\ \text{in.}^2} \times 1\ \text{in.}^2 = 2.210 \times 10^{-3}\ \text{cd} \qquad (1)$$

The CRO Calibration Display

Figure 2 shows a convenient calibration display composed of 20 "squares," each square consisting of a 10 by 10 point pattern. Because of its lateral and vertical extent, the display in Fig. 2 also is convenient for calibrating the x- and y-axis scale amplifiers. The display is refreshed n times per second. An aperture is used to exclude all except one 10 by 10 square from view.

Photometer

A null meter can be used to test the equivalence of the luminous intensity from the CRO aperture and the reference surface aperture. *The measurement must be made at the same distance from each surface!* Of course, a photometer with a continuous scale will make the comparison easier. All the light from the reference source must enter the uniform acceptance area of the photometer, and so must all the light from the test aperture (Fig. 1). The refresh rate should be fast enough to give a steady reading on the meter; usually, 20 Hz will suffice.

From the ratio (or equivalence) of the luminous intensity of m points × n exposures per second to the luminous intensity of the reference aperture, the luminous energy in one exposure of one point is readily calculated. Since luminous units can be very confusing, an example is given here. When the area of the reference aperture is a sq in., and the luminance of the reference surface is b fL, and there are m points exposed n times per second, the *luminous directional energy per point*, e, is

$$e = 2210\ (ab)/(mn)\ \text{candle-microsec.} \qquad (2)$$

Application to Vectors, Scan Lines, and Characters

The luminous calibration of vectors and of scan lines is essentially similar to that of points. The critical quantity is the LDE per unit line length. For example, the 100 points of Fig. 2b can be replaced by 10 parallel 1-in. vectors within the aperture or by 10 1-in. segments of scan lines. From the calibration, one obtains the LDE in 10 in. of line, i.e., the LDE per inch.

The LDE per character can be measured by substituting characters for points in the determination of the LDE. For many purposes it is also useful to know the ratio of LDE per character to LDE per point and to LDE per unit line length.

High Refresh Rates

When viewing of CRO displays is restricted to such high refresh rates that apparent flicker is negligible, then they may be treated as "continuous" displays. In this case, the critical luminous quantities are *luminous intensity per point* and *luminous intensity per unit line length*. These intensity quantities are defined identically to the corresponding energy quantities except in that the factor of time is implicit. The *point* units are candelas or microcandelas, the *line* units are candelas per inch or microcandelas per inch.

HOW TO USE LUMINOUS DIRECTIONAL ENERGIES PER POINT AND PER LINE

When a field is filled uniformly and densely with points or scan lines, it appears similar to a continuously illuminated surface, and its intensity is appropriately characterized by its luminance (photometric "brightness"). Luminance is defined as the luminous intensity per unit area. (Equivalently, luminance can be defined by a perfectly reflecting-diffusing surface illuminated by a known luminous flux.) Luminance measurements are particularly useful because they do not depend on the distance from which the surface is viewed, and this distance need not be known. (The intensity calibrations required careful equating of distance.)

The luminance of a CRO display depends both on the LDE per point (or LDE per line) *and* on the number of points per unit area per unit time (or on the number of inches of line per unit area per unit time). Once the LDEs per point and per line are known, it is merely a matter of multiplication to calculate luminance. For example, a uniform surface of 1 fL luminance emits π^{-1} cd/ft^2. If 1 sq ft of display surface is filled with enough points or lines to appear approximately uniform, if they are refreshed at > 20 Hz, and if the total point energy from all the intensifications is π^{-1} cd-sec, then the display surface will appear to have a luminance of 1 fL. When the refresh frequency is doubled, the same array of points will appear to have a luminance of 2 fL because it now produces twice as much light per unit area.

To convert to foot-lamberts from the total candle-microseconds produced by all point or line intensifications within an area, a, in 1 sec, one needs merely to reverse Eq. 2, i.e.,

$$b = 0.00452\ e/a \qquad (3)$$

where b is luminance in foot-lamberts, e is the LDE in candle-microseconds, and a is the area in square inches.

LINEARITY PROBLEMS

An important characteristic of phosphors to consider in luminance calibrations is that their response can be nonlinear and can depend on how frequently, how intensely, and how many neighboring points are intensified. However, a phosphor becomes more linear the less output is required within any small area in any small time, so it is possible to set up conditions under which the response is approximately linear. The purpose of the 20 squares in the calibration display of Fig. 1 is to provide ready comparison with calibration displays having only 10 or 2 squares in all, and repeated at proportionately greater rates, i.e., producing correspondingly greater output in the remaining square seen through the aperture. When the response of the phosphor is being measured in the linear range, then halving the refresh rate should exactly halve the luminous flux output through the aperture. This halving should be tested when calibrating luminous energies.

Linearity considerations are important when it is necessary to calibrate a display whose intensity is set to a level high enough to permit clear visibility of a single flash. In this case, refreshing the display as frequently as 20 times per second might damage the phosphor. This danger can be ameliorated by intensifying different points—within the same aperture—on successive exposures. The problem does not arise in calibrating continuously refreshed displays. In this case, the calibration refresh rate should be chosen the same as the viewing refresh rate.

One word of advice: Never trust calculations more than necessary. For example, though luminance is readily calculated from LDEs, it also is easily measured directly. Measure it to confirm the calculation. Finally, computers usually generate frames so rapidly that the number and timing of frames cannot be checked by eye. Before considering a display program finished, it always is worthwhile to use a photocell and an oscilloscope to monitor the display and to verify that the program is generating the intended number of frames in the intended sequence.

REFERENCES

LeGRAND, Y. *Light, colour and vision.* (2nd ed., translated from French.) London: Chapman & Hall, 1968.

WALSH, J. W. T. *Photometry.* (2nd ed.) London: Constable, 1953.

NOTE

1. There is no conventionally agreed upon photometric term for the integral of luminous intensity (LeGrand, 1968; Walsh, 1953); it is the *quantity of light* per solid angle, emitted in a particular direction. Possible units are (lumens x seconds)/steradian, talbots/steradian, and candles x seconds. Integrated luminous intensity is perhaps the most accurate name, but luminous directional energy better carries the meaning in the present context. In any case, the choice of units (candles x microseconds) makes it perfectly clear what is being measured.

Flicker in computer-generated visual displays: Selecting a CRO phosphor and other problems*

GEORGE SPERLING†
Bell Telephone Laboratories, Murray Hill, New Jersey 07974
and
New York University, New York, N.Y. 10003

The visual system's response to flicker is approximated by 6-9 RC stages in series. The CRO phosphor can be represented by one additional RC stage. Therefore, increasing the refresh rate by a factor of k can be k^5 to k^8 times more effective in reducing apparent flicker than increasing the phosphor time constant. "Slow" phosphors impair the display of rapid movement and cause undesirable persistence of old picture contents after they have been altered. Behavioral scientists usually should choose fast, efficient phosphors. Display programs should be written so that spatially adjacent points of a display are intensified as close together in time as possible.

A visual display on a cathode ray oscilloscope is composed of points, vectors, characters, or scan lines. A single picture *(frame)* can be *painted* in a relatively short time, depending, of course, on the number of elements of which it is composed, as elements are painted sequentially. In order to give the viewer the illusion that the display is present continually, the frame must be repainted *(refreshed)* about 20 to 40 times per second. The more frequently it is refreshed, the less it appears to flicker.

In contemporary CRO displays, each individual point (of a frame) typically is electronically pulsed *(intensified)* for 1 microsec or less. At a refresh rate of 40 frames/sec, the point will be reintensified 40 times per second, i.e., every 25 msec. The transduction of the electron intensification beam to light is accomplished by a *phosphor* painted on the inside of the CRO tube. Depending on the particular phosphor, the light output in response to the brief intensification pulse may either persist long after the pulse or it may follow the shape of the intensification pulse. The phosphor is a means of prolonging, or "spreading out," the light in time and thereby reducing the apparent flicker.

The light output of a phosphor in response to an exceedingly brief intensification pulse *(impulse)* typically follows an exponential curve of the form illustrated in Fig. 1. This exponential decay curve is characterized by a single parameter, its decay time constant (τ), which is defined as the time taken to decay to .368 of the initial value.[1] When the time constant is long, apparent flicker is minimized; when it is short, flicker is maximized.

THE PHOSPHOR PROBLEM: FAST OR SLOW?

In "continuous" CRO displays it is desirable to minimize apparent flicker. This could be accomplished either by (1) choosing a phosphor with a long time constant (a "slow" phosphor), or by (2) increasing the refresh rate.

One disadvantage of choosing a phosphor with a long time constant is that it makes the display system unsuitable for applications requiring brief visual presentations, because brief exposures are impossible with slow phosphors. Furthermore, when the display contents are altered, a slow phosphor causes the old contents to persist visibly under the new picture. This difficulty makes slow phosphors unsuitable for presenting rapid motion.

Alternatively, apparent flicker can be reduced by increasing the refresh rate. However, it may be impossible to increase the refresh rate when frames are composed of large numbers of points or characters. Even when it is possible, a high refresh rate places an added burden on the computer. The following sections evaluate these factors in terms of the flicker sensitivity of the human visual system. The conclusion (for behavioral scientists) will be that phosphors with short time constants should be chosen, except when display systems are dedicated to static displays.

Flicker Sensitivity of the Visual System

The response of the visual system to flickering light can be predicted from the model illustrated in Fig. 2. It is a good first-order mathematical approximation to a more complicated model (Sperling & Sondhi, 1968) in the frequency range of interest, i.e., for frequencies $>$ 10 Hz. The basic elements of this approximation are RC stages, linear elements whose response to an impulse input is an exponential output of the form already described in Fig. 1.[2]

In the dark, the response of the visual system to flicker can be represented by three RC stages in series (each having a time constant of about 25 msec) plus six more RC stages in series (each having a time constant of about 5 msec). In very bright light, the visual system's response to flicker is represented simply by the final six stages (each with $\tau = 5$ msec). That is, in very

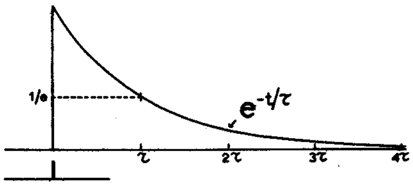


Fig. 1. Light response of a phosphor to an intensification impulse (theoretical). Ordinate: light output; abscissa: time; the intensification impulse is indicated below. The response illustrated is an exponential decay function with a time constant of τ.

---

*Presented at the North Atlantic Treaty Organization, Advanced Studies Institute on: On-Line Computing for Behavioral Science, Department of Psychology, Sheffield, England, July 19, 1969.

†Requests for reprints should be sent to: G. Sperling, Room 2D518, Bell Telephone Laboratories, Murray Hill, New Jersey 07974.

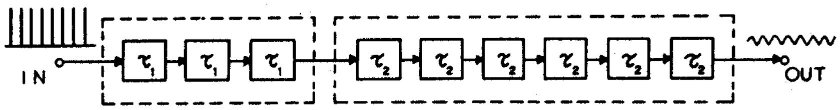


Fig. 2. First-order linear approximation to the visual system's response to flicker. There are nine RC stages in series in two sets; for all stages within a set τ(= RC) is the same. The value of τ_1 is 25 msec in the dark, 5-25 msec at room illuminations, and less than 5 msec in extremely intense light. The value of τ_2 is constant and equal to about 5 msec. When the input stimulus is a pulsed flickering light, as illustrated, the flicker in the response is markedly reduced. Here, only the first Fourier component is visible. (After Sperling & Sondhi, 1968.)

bright light, the time constant of the initial three stages becomes so small that it is negligible. At intermediate light levels (e.g., room illumination), the initial three stages can be assumed to have time constants between 5 and 25 msec, depending on the intensity of the illumination and on many other factors.

Except in very dim light, the visual system has what engineers would call excellent *gain control*. That is, its response is proportional to the percent of flicker (relative to the steady background level) rather than to the absolute amount of flicker. The *gain control* mechanism itself begins to affect the responses to flicker at frequencies below 10 Hz; therefore, the linear model is accurate only at frequencies of 10 Hz or higher.[3] Fortunately, this covers the frequency range of interest. In this range, only the fundamental frequency of a flickering stimulus will be significant for flicker perception; higher harmonics can be neglected. Thus the effective input stimulus can be regarded as the percent modulation of the background by the fundamental frequency of the stimulus. (On this scale, the sensitivity of the visual system diminishes in dim light, but in practice this change is small relative to the change in frequency response caused by the increased time constant of the first three stages.)

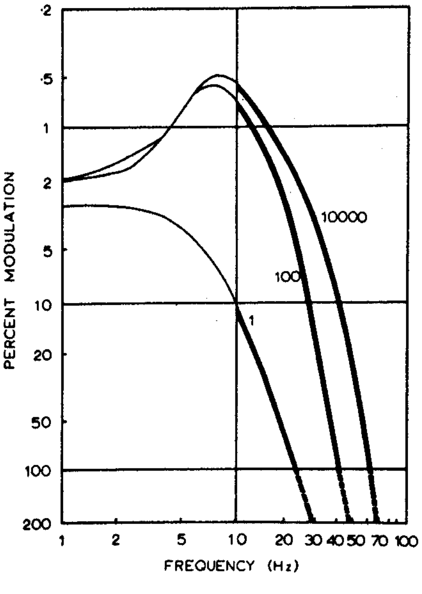


Fig. 3. Threshold sensitivity of the visual system to sine-wave flickering light. Abscissa: frequency (log scale); ordinate: threshold modulation in percent of background luminance (log scale). The indicated parameter is the mean retinal illuminance in trolands of the light being modulated: 10,000 (high intensity); 100 (moderate intensity); 1 (low intensity). (After DeLange, 1958).

If we assume impulse intensification pulses and neglect the effect of the phosphor (i.e., assume a fast phosphor), then the depth of the sinusoidal modulation produced by the refresh frequency is easily calculated: it is 200%. Impulse intensification produces effectively the same flicker as 200% modulation by a sine wave at the refresh frequency. As negative values of light are impossible, pure sinusoidal modulation is limited to 100% modulation depth, but the predictions based on the 200% fundamental component have been verified experimentally (Kelly, 1964).

The response of the visual system at typical room illuminations to flicker at frequencies greater than about 10 Hz can be summarized as follows. (1) At 10 Hz the system is extremely sensitive to flicker; a modulation of less than 1% can be detected. (2) At 20 Hz the sensitivity of the system to flicker is vastly reduced. (3) At 40 Hz the limit of detection is approached, even at 100% (or 200%) modulation. Where the limit of flicker detection lies (i.e., between 40 and 70 Hz) depends on the particular individual, the luminance level, the size and chromaticity of the flickering field, the region of the retina being used, and other factors. Some typical measurements that describe quantitatively the threshold sensitivity of the eye to sinusoidal flicker are illustrated in Fig. 3.

The Phosphor as an Added RC Stage

Mathematically, the effect of a phosphor in reducing apparent flicker is equivalent to adding a single RC stage in series to the system described above. It is immediately obvious that when the time constant of the phosphor is 5 msec or smaller, it will make no significant difference to the overall system; there are already at least six stages having a time constant of 5 msec or more. When the time constant of the phosphor is 25 msec, it can make a noticeable difference but probably not a big difference. When the time constant is 250 msec, the phosphor can reduce flicker (at refresh rates greater than about 10 Hz) by 10X relative to $\tau = 25$ msec.

The Tradeoff Between Phosphor Time Constant and Refresh Rate

The flicker reduction of an RC stage is given by its *filter* or *attenuation characteristic*. Figure 4 illustrates the attenuation characteristics (for sine waves) of a single RC stage. The "corner" or "cut-off" frequency (Fig. 4) is related to the time constant by $f_c = 1/(2\pi\tau)$. For example, when $\tau = 5$ msec, $f_c = 32$ Hz; when $\tau = 25$ msec, $f_c = 6.4$ Hz. Figure 4 shows that frequencies much larger than the cut-off frequency (high frequencies) are attenuated in direct proportion to their frequency. Equivalently, increasing τ by a factor of k ($\tau' = k\tau$) increases attenuation of high frequencies by a factor of k. It is this principle that was used to calculate the 10X reduction in flicker produced by the 10X change in time constant from 25 to 250 msec.

When there are n RC stages in series (as in the model of the visual system, the attenuation of high frequencies is proportional to k^n. Because there are effectively six to nine stages in the visual system, doubling the refresh rate can increase flicker attenuation not

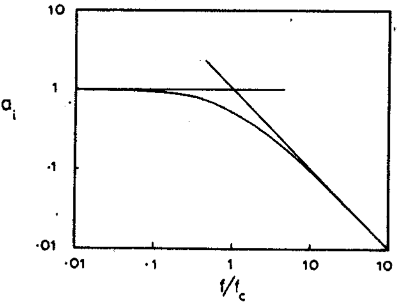


Fig. 4. Attenuation (filter) characteristic of a single RC stage. Abscissa: frequency f of an input sine wave relative to the cut-off frequency f_c (log scale); ordinate: amplitude of output a_o/amplitude of input a_i (log scale). The cut-off frequency f_c is related to RC by: $f_c = (2\pi\tau)^{-1}$ where τ = RC.

merely by a factor of 2 but by a factor of 2^6 to 2^9; i.e., it can reduce flicker by 64 to 512 times. At high frequencies, doubling the refresh rate is hundreds of times more effective than doubling the time constant of the phosphor. Unfortunately, the full benefit of these tremendous attenuations occurs only at high frequencies, i.e., frequencies greater than about 30 Hz. At refresh rates between 10 and 30 Hz, the effect of an increase in rate is smaller; at still lower frequencies, other factors come into play.

Exceedingly slow phosphors—even assuming they were available—would be impractical because the old picture would persist for an exceedingly long time under the new one. The requirement that the old picture fade within a second or two would eliminate phosphors with time constants greater than about 250 msec. This limit, in turn, means that producing flicker-free (or even "low-flicker") displays will require fairly high refresh rates, i.e., over 20 Hz. These are the very conditions in which it is far more efficient to increase the refresh rate than the time constant of the phosphor.

Behavioral scientists who use on-line visual displays frequently have only one CRO unit and many users, and some users occasionally may desire to display rapid motion or to have precise stimulus control. For these installations, the moral is: Unless the CRO display system is committed to static pictures, the reduced flicker in visual displays achievable by a slow phosphor does not justify sacrificing the stimulus control achievable with fast phosphors; i.e., by phosphors with $\tau = 5$ msec or less. A corollary is that one generally should not buy "general purpose" phosphors with time constants greater than 5 msec; these provide neither precise stimulus control nor a significant lowering of the refresh rate required for freedom from flicker.

OTHER CONSIDERATIONS

Efficiency of a Phosphor

The most important characteristics (apart from fast decay) of a phosphor for behavioral scientists are efficiency, maximum output, and durability. Efficiency refers to the efficiency of conversion of electronic energy to luminous energy. Whether a display will be bright or dim is determined by the efficiency of a phosphor, but even an inefficient phosphor can be useful if powerful electronic driving circuits (and a toleration for intense driving—"durability") enable the phosphor to reach a high maximum output. High output is important in experiments where a frame is painted only once, when the display is viewed in daylight or in a brightly lighted room, and when color or Polaroid filters are used with the display.

Durability refers to the susceptibility of a phosphor to "burning out" when it is pulsed intensely. Burn-out leaves a permanently inactive patch in the display surface. The burned-out patch usually is discolored, as well as being inactive. In currently available phosphors, unfortunately, efficiency and fast decay are correlated with a tendency to burn out.

Painting Sequence

Another problem arises when it takes a long time to paint a frame, that is, tens of milliseconds. The eye, even when it attempts to maintain steady fixation, makes small involuntary, saccadic movements several times per second. If some line segments of a frame happen to be painted before a movement and others after the movement, these segments will appear to lie in inappropriate spatial relations to each other—displaced by the amount of the eye movement (Sperling & Speelman, 1965). In viewing such a nominally unchanging display, segments of the display appear to jitter back and forth relative to each other. The effect is very convincing and, not infrequently, the cause is erroneously attributed to a hardware failure. Slow phosphors minimize the effect. The way to minimize the effect by software is to paint adjacent parts of a frame at the same, or nearly the same, time. For example, the dot above an i should be painted at the same time as the i; never wait until a whole line is painted to dot the is and cross the ts.

Dual Phosphors

One inexpensive method that has been tried to obtain the advantages of both a fast and a slow phosphor in the same CRO has been to use both phosphors simultaneously in the same CRO tube. These two phosphors usually are of different colors. By making the slow-decay phosphor more efficient, it will predominate in the normal viewing of static displays and thereby apparent flicker will be reduced. By use of selective color filters, either phosphor may be seen alone.

The problem with dual phosphors is that the separation achieved by the color filters is not perfect and the better the separation required, the denser the filters must be. The denser the filters, the more inefficient is the system as a whole. An inefficient highly chromatic display is particularly undesirable for showing unrefreshed frames (one of the main uses for which a fast phosphor is required). And, even a faint residual trace from the slow phosphor can utterly defeat a fast-phosphor experiment.

The real solution to the problem of fast or slow phosphors is to have a display device with a variable-speed phosphor. The *direct view storage tube* potentially is such a device. (A DVST displays a frame continuously—without requiring refreshing—until it is "erased" by a control signal.) The DVST is the logical display device for static displays. Its widespread use today (1969) is limited only by its high cost and low spatial resolution, but we can reasonably anticipate improvement in both factors. Ultimately, we hope for DVSTs which allow selective erasure, so that the DVST will control directly the precise time for which any portion of a frame is displayed.

REFERENCES

DeLANGE, Dzn. H. Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. Journal of the Optical Society of America, 1958, 48, 777-784.

KELLY, D. H. Sine waves and flicker fusion. In H. E. Henkes and L. H. van der Tweel (Eds.), *Flicker*. The Hague: Junk, 1964. Pp. 16-35.

SPERLING, G. Linear theory and the psychophysics of flicker. In H. E. Henkes and L. H. van der Tweel (Eds.), *Flicker*. The Hague: Junk, 1964. Pp. 3-15. (Also, Doc. Ophthalmologica, 1964, 18, 3-15.)

SPERLING, G., & SONDHI, M. M. Model for visual luminance discrimination and flicker detection. Journal of the Optical Society of America, 1968, 58, 1133-1145.

SPERLING, G., & SPEELMAN, R. G. Visual spatial localization during object motion, apparent motion, and image motion produced by eye movements. Journal of the Optical Society of America, 1965, 55, 1576-1577.

NOTES

1. Common phosphors have exponential decay only to about 10%-30% of peak emission; below this, decay is slower than exponential. As apparent flicker is determined mainly by the initial decay, the exponential is a valid approximation.

2. See Sperling (1964) for an introductory discussion of linear systems, RC stages, etc., in relation to vision.

3. The linear approximation to Sperling and Sondhi's nonlinear model requires a high-pass RC stage in series with the low-pass stages. The high-pass stage improves the accuracy of the approximation at low frequencies but is an unnecessary complication for the present purpose and therefore is omitted here.

Stereoscopic visual displays: Principles, viewing devices, alignment procedures[*]

GEORGE SPERLING[†]
Bell Telephone Laboratories, Murray Hill, New Jersey 07974
and
New York University, New York, New York 10003

The principle of binocular (stereoscopic) depth perception is that the visual system interprets the slight differences between the views seen by the two eyes as depth cues. In computer-generated displays, two slightly different images are produced on the left and right halves of the display surface and viewed by a prism, mirror, or binoculars system that delivers the appropriate image to each eye. The prism system is the simplest, the mirror system gives the best optical quality, and the binoculars system is useful for producing large apparent images from small display surfaces. All three systems can be adapted for group viewing and all require careful alignment (null adjustment of accommodative distance and vergence distance). Objective and subjective methods of alignment are described.

THE PRINCIPLE OF STEREOSCOPIC DEPTH PERCEPTION

In normal vision, the two eyes receive slightly different views of the world, and the differences between the views are interpreted by the visual system as visual depth. The name given to the depth perception that arises from differences between the images viewed by the eyes is *stereoscopic depth perception*; the two corresponding images are called a *stereo pair (of images)* or a *stereogram*, when they are produced artificially. The principle of stereoscopic depth perception is illustrated in Fig. 1.

In normal vision there are many other important cues to depth, such as those provided by the change in view produced by a head movement, the interposition of objects, perspective transformations, sizes of known objects, distance haze, vergence of the eyes, etc. Of these, only the depth cue produced by head movements is not readily simulated by a visual display system which presents an appropriate stereogram to the two eyes. In fact, most stereoscopic viewing devices require the head to remain still; even when they do not, head movements are relatively easy to measure. As this head-movement information readily can be incorporated into a stereoscopic display system, the simulation of normal binocular vision by a computer display system introduces virtually no limitations that are not already inherent in monocularly viewed displays, i.e., the limitations of image quality and of information content.[1]

VIEWING DEVICES

I shall be concerned here not with the psychological problem of how to compose a stereogram in order to provide the desired illusions of depth,[2] but rather with the simple technical problem of how to use a cathode ray oscilloscope (CRO) display system economically and efficiently to provide two different views. Basically, there are only two methods of producing the two images of the stereogram: (1) to use two different CROs, one for each image, and (2) to use two different subareas of a single CRO display surface (e.g., the left and right halves), and to view each subarea separately with each eye.

The method of two CROs is more expensive and has inherent in it the problem of adjusting the two CROs to be electronically equivalent. The two-CRO solution is useful mainly when it is necessary to separate the stereogram members (not readily possible on a single display surface because the two display areas are constrained to be adjacent to each other) and when independent hardware control of such parameters as x- and y-axis magnification is desired. For example, to control independently the stimulus area in a large annular neighborhood of each computer-generated display, it is convenient to separate the stereogram members. In principle, however, the problems of viewing two CROs are the same as those in viewing two subareas of one CRO, and the examples will refer only to the latter case.

The three main methods of viewing subareas of a display surface separately with each eye are illustrated in Fig. 2; they are (1) a prism viewing system, (2) a mirror viewing system, and (3) binoculars. All three systems are improved by providing a viewing hood to position the O's head in front of the device to reduce extraneous stimulation. The viewing hood and prisms (or mirrors) usually are attached to one end of a sturdy, lightweight frame. The other end of the frame should be quickly and conveniently attachable to the display device itself in order to provide rapid changeovers between normal and stereoscopic viewing of the visual display. The binoculars system also

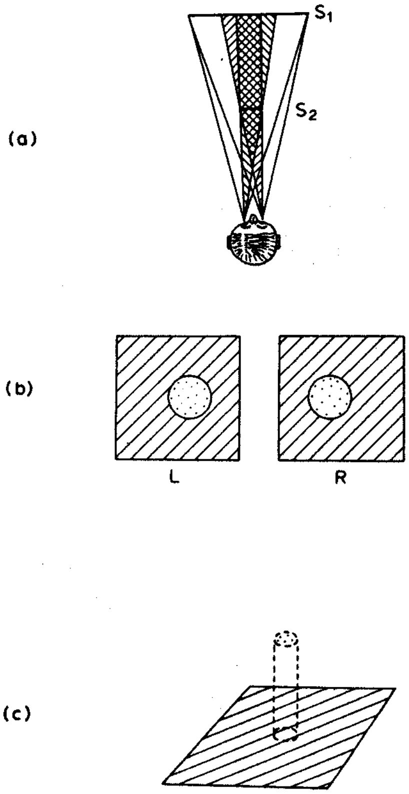


Fig. 1. The principle of stereoscopic depth perception. (a) Top view of an O viewing two surfaces (S_1, S_2). The two eyes see the nearer surface against different parts of the background, as indicated. The displacement (between the two views) of the foreground relative to the background is the stereoscopic depth cue. (b) The stereogram, i.e., the views seen by the left and right eye. The stereo pair of images may, of course, be produced either on a visual display surface or by viewing a natural object. (c) Perspective drawing of the illusion of depth that results from viewing the stereogram in (b). Note that the projection of S_2 against S_1 in (b) exaggerates its size; when S_2 is seen in stereoscopic depth (c), the apparent size of S_2 approximates its actual size (a) rather than its projected size (b).

---

[*] Originally presented at the North Atlantic Treaty Organization, Advanced Studies Institute on: On-Line Computing for Behavioral Science, Department of Psychology, Sheffield, England, July 19, 1970.

[†] Requests for reprints should be sent to: G. Sperling, Room 2D518, Bell Telephone Laboratories, Murray Hill, New Jersey 07974.

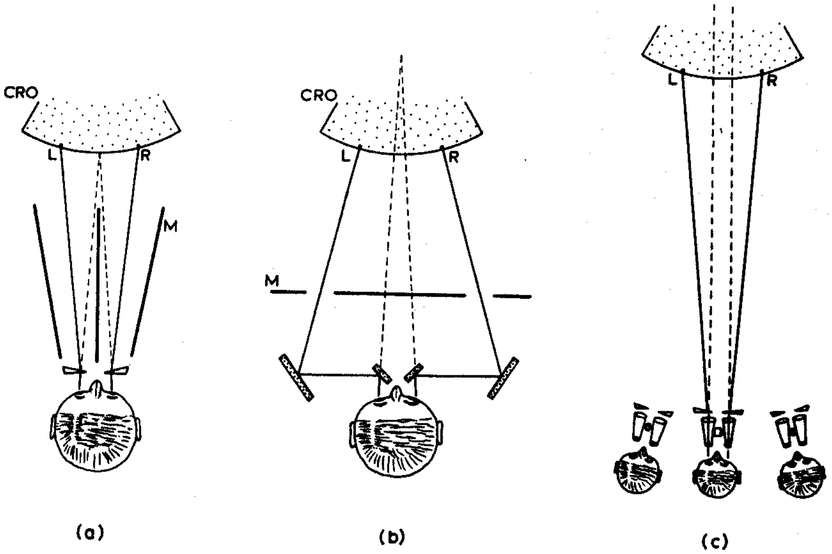


Fig. 2. Three stereoscopic viewing systems: (a) prisms, (b) mirrors, (c) binoculars. L and R indicate the nominal centers of the left and right images, respectively, of the stereogram: M = mask, CRO = cathode ray oscilloscope (the visual display surface). Dashed lines indicate lines of sight extended to their intersection. In (a) and (b) their intersection occurs at the accommodative distance of the display indicating correct alignment of the systems. In (c), the accommodative distance is assumed to be infinity, and the lines of sight are parallel, again indicating correct alignment.

can be made detachable, but as the binoculars can be located far from the display, they do not interfere with normal viewing and can be left permanently in place.

Prism System (Brewster, 1856)

The advantage of the prism viewing system is that it is the easiest to implement. A prism is needed only in one eye, and this may even be held by hand. In fact, with practice, Os can learn to deviate their eyes (either diverged or crossed) so that the prism is superfluous. However, unnatural deviation of the eyes is undesirable for reasons to be discussed later.

The disadvantages of the prism method are: the prism produces a distortion of straight lines ("prism distortion"); the power of a prism varies depending on the eye's direction of gaze and on their vergence, and this effect interacts with any lenses the O may be using (Fry, 1937; Ogle, 1951); the prism reduces the optical quality of the image being viewed (particularly by introducing chromatic aberration, which is disturbing against dark backgrounds); and the centerline-separation shield, which extends forward from the nose to the center of the display, is an awkward restriction of the image being viewed.

The optical quality of the image can be improved by using two low-power prisms, one in front of each eye, instead of one strong prism, and by using achromatic prisms. (Achromatic prisms are composed of two different kinds of glass, combined so as to minimize the chromatic aberration inherent in the use of an elementary prism.) In a useful variation of the prism method, the single prism (in either or in both eyes) is replaced by a pair of equal prisms. The prisms in such a pair can be rotated against each other to produce—in effect—a single prism of variable power. This is a convenience, but it adds to optical distortion.

The prism method for group viewing. With some slight modifications, the prism method can be adapted for group viewing. The prism (or prisms), preferably adjustable, is fitted on spectacle frames that are worn either alone or immediately in front of the viewer's usual spectacles (when these are used). Because there is now no centerline-separation shield, each eye will see both images of the stereogram, i.e., the entire display. This extraneous stimulation may create serious viewing problems. The simplest way to restrict the viewing of each eye to its intended member of the stereo pair is by means of a corresponding pair of crossed Polaroids, one pair of which is fixed to the display surface and the other to the spectacle lenses (Fig. 3); the remaining extraneous stimulation is removed by darkening the room.

The more persons who wish to view a display, the farther away they must stand. (The adjustable prism permits the O to vary his viewing distance.) A distant viewer not only sees a "smaller" stereogram than a near viewer, but he perceives differently shaped objects because perceived depth relations do not vary linearly with viewing distance. Whereas the shape of a two-dimensional object remains invariant with changes in viewing distance, the three-dimensional shape of an object defined by a stereogram changes as the viewing distance changes. In producing stereograms, the absolute retinal size at which they will be viewed must be taken into account, and the viewer must also consider this.

Mirror Method (Wheatstone, 1838)

The advantages of the mirror method are: it is virtually free from optical distortion when good quality front-surface mirrors are used; it is extremely easy to adjust by rotating the mirrors slightly; and it allows convenient placement of masks to limit the field of view, thereby giving good control of the overlap area (the shared display area between the two subareas which serves as a buffer zone for each subarea). The disadvantage of the mirror system relative to the prism system is that it is more complicated to construct. Like prisms, mirrors can be worn as spectacles to facilitate group viewing.

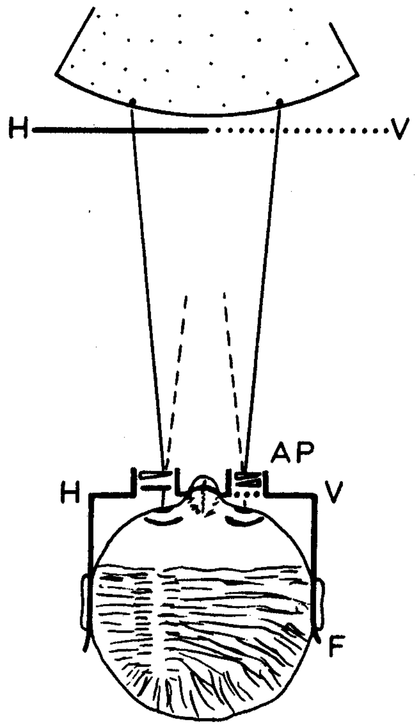


Fig. 3. Prism-plus-Polaroid system for group viewing. Spectacle frame (F) contains Polaroid filters and an adjustable prism (AP). V (vertical) and H (horizontal) indicate the direction of polarization of the Polaroid filters.

The mirror system would seem to have an advantage over the prism system in enabling each half of a curved display surface to be viewed perpendicularly, thus reducing the foreshortening that results from the nonperpendicular view in the prism system. In any viewing system, the first-order horizontal foreshortening error can be compensated by adjustment of the x-axis magnification. However, perpendicular viewing eliminates the first-order vertical foreshortening error, as well as the horizontal error. When the display surface is a true circular arc, the second-order errors (due to noncorresponding screen curvature in the two images) also are eliminated by perpendicular viewing, but these errors are unlikely, in any case, to be significant in comparison to other image defects.

Binoculars Method

The usual implementation of mirror and prism systems limits precisely controlled viewing to one O per CRO. When there are several Os, these systems require the multiplication of CROs and the associated viewing device, one system for each O. One way to overcome this multiplication is to use binoculars to view the displays. Because they magnify the images, binoculars can be placed farther from the displays than the direct viewing devices; consequently, many Os can view the same display simultaneously. If desired, they can be viewing it at different magnifications.

The disadvantages of the binoculars method are: the viewing stand with binoculars on it must be rigidly fixed relative to the display; the system is difficult to adjust (particularly centering the images) and difficult to maintain in adjustment; viewing is difficult for the O because his head must be exactly positioned vis a vis the binoculars, and the optical quality is likely to be the poorest of the methods. On the other hand, it is far cheaper to obtain a large effective display area by viewing a small CRO with binoculars than by buying a large CRO. When very small displays are required, they are conveniently produced by inverting the binoculars. The optical quality of binoculars is much less critical when they are used to minify than when they are used to magnify.

ALIGNMENT

Accommodation and Vergence

Accommodation is the technical term for the amount of focusing of the lens of the eye. Vergence refers to the angle between the two eyes that is required to place the image of a single object onto corresponding points of the two retinas.

It is not necessary here to define additive scales of accommodation and vergence (see Emsley, 1948; Sperling, 1970). It is sufficient simply to refer to the actual distance from the O to an object for which accommodation or vergence is appropriate. The stimulus dimension that is crucial to accommodation is called the *accommodative distance* of an object; the eye's response is the *accommodated distance.* Similarly, the critical stimulus dimension for vergence is the *vergence distance* of an object; the response is the *verged distance.* For example, in normal vision an object at a distance of 1 m from an O causes the lenses of his eyes to adjust accommodation until they accurately focus light rays emanating from a distance of 1 m, and it causes his eyes to *con*verge so that the lines of sight of the two eyes intersect at a distance of 1 m.

In the world of real objects, the stimulus to accommodation and the stimulus to vergence are perfectly correlated, and the visual system takes this into account. For example, when a stimulus only to accommodation and not to vergence is presented (i.e., one eye is kept covered), then the eyes nonetheless verge upon it (i.e., even the covered eye points at it). Similarly, a stimulus to vergence causes the eyes to accommodate appropriately. Fortunately, there is some flexibility in the visual system so that the eyes can be accommodated for one distance and verged upon a different distance, but only a limited amount of *accommodation-vergence disparity* can be tolerated. When this amount is exceeded, then either a failure of accommodation occurs (producing a blurred image) or a failure of vergence occurs (producing a "double" image), or both occur.

A computer can readily generate stereograms that define objects at numerous different vergence distances (and thereby give complex illusions of stereoscopic depth), but no convenient way has yet been discovered to concomitantly covary the stimuli to accommodation. For example, by varying their vergence distance, different objects defined by a stereogram may subjectively appear to be at different depths, ranging from immediately in front of the nose to infinity, even though accommodative distance remains fixed (at the distance from the O to the display). Generally, the inability to vary independently the accommodative distance of segments of a computer-generated display is not an important limitation. It is sufficient to vary the overall accommodative distance, and this is easily done by placing spectacle lenses in front of the O's eyes.

Matching Accommodative Distance and Vergence Distance (Alignment)

An important design consideration in stereoscopic displays is insuring that vergence is appropriate for the accommodative distance of the display. For example, let the accommodative distance be 1 m. Then, when the lines of sight of each eye are pointed at the nominal centers of their respective visual display, these lines of sight should intersect at 1 m. It is the task of the engineer to insure that the lateral separation of the left and right stereo images on the display surface and the optics of the viewing device are adjusted to accomplish this.

The alignment procedure may be divided into four steps which are best carried out in the following sequence: (1) determine the accommodative distance of the display; (2) alter the accommodative distance if desired; (3) measure the vergence distance of the display; (4) adjust the vergence distance to equal the accommodative distance. For many purposes, a simple subjective method that bypasses the first three steps will suffice, but it is useful also to know objective methods.

(1) Measuring accommodative distance. The easiest way to determine accommodative distance in the prism and mirror systems is to measure the optical path with a ruler. In binoculars systems, a camera with a ground-glass viewer probably is the simplest way to determine the accommodative distance of the display. The camera lens is placed at eye position and focused to produce a sharp image. The determination depends, of course, on the previous calibration of the camera lens, and it is the more accurate the greater the focal length.

(2) Altering accommodative distance. To alter accommodative distance, an ordinary spectacle lens placed before the eye is all that is required. The power D in *diopters* of the required lens is given by the formula

$$D = \frac{1}{A_2} - \frac{1}{A_1} \qquad (1)$$

where A_1 is the initial accommodative distance (in meters) and A_2 is the desired final distance. Positive values of D indicate convex (positive) lenses and negative values indicate concave (negative) lenses.

(3) Measuring vergence distance. The best method requires constructing a device for producing two narrow collimated beams of light (e.g., two penlights with small apertures at the front). The collimators should be mounted on a board with the front ends 6 cm apart and the rear ends adjustable. The main requirement is that the dual beams should produce clearly defined spots at a distance

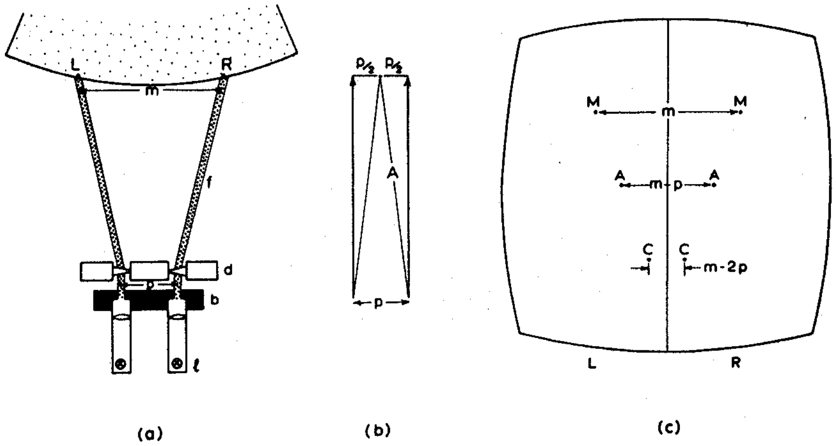


Fig. 4. Calibration of vergence distance. (a) Two parallel beams of light are produced by the collimators (ℓ). The light beams (f) pass through the stereoscopic viewing device (d) to project onto the display surface with a lateral separation of m cm. The apertures of the collimators are mounted on a board (b) and are separated by the interpupillary distance (p), nominally 6 cm. (b) Diagram to illustrate that when the lateral separation is diminished by p (twice p/2), the vergence distance changes from infinity (parallel lines) to a vergence distance of A (the accommodative distance). (c) The display surface, showing a stereogram (L, R). The Ms are assumed to be separated by a lateral separation of m, corresponding to a vergence distance of infinity—e.g., as in (a). The lateral separation between the As is m − p, corresponding to a vergence distance of A (the accommodative distance), e.g., as in (b). The Cs are laterally separated by m − 2p, corresponding to a vergence distance of A/2. In a stereoscopic view, the Cs would appear to be in front of the screen, the As in the plane of the screen, and the Ms behind it. The stereogram is correctly aligned when vergence distance equals accommodative distance; for prism viewing this occurs when the nominal centers of the L and R subfields have the same lateral separation as the As in (c).

equal to accommodative distance of the display surface.

The dual beams are placed at the same vantage point as the eyes and pointed at the display; there they produce two spots on the display surface (Fig. 4a). When the angle between the beams has previously been adjusted to the desired vergence distance (the point at which the beams intersect), the distance between the two spots *(lateral separation)* will define this vergence distance on the display surface.

As a practical matter, it is useful to set the collimated beams parallel because their projections on the display surface then define the lateral separation of objects at infinite vergence distance. This is the maximum lateral separation of "real" objects. When the lateral separation of two images on the display surface is decreased by the interpupillary distance (typically 6 cm) then the vergence distance equals the accommodative distance.

The relation between lateral separation on the display surface (S) and vergence distance (V) can be expressed in general terms.[3] Let the lateral separation on the display surface between points at infinite accommodative distance be m, the interpupillary distance be p, and the unaltered accommodative distance of the display be A. Then the vergence distance of any object on the screen is related to the lateral separation of the two images by

$$V = \frac{Ap}{m - S} \text{ and } S = m - \frac{Ap}{V} \quad (2)$$

Equation 2 relates the lateral separation of any two corresponding points in the stereogram to their vergence distance provided A, p, and m are known. Determination of the accommodative distance, A, was outlined above. The interpupillary distance, p, may be taken as .06 m. (The average population value is about 6.2 cm.) The dual parallel light beams project on the screen with the lateral separation of m (Fig. 4a).

It may be noted here that perceived stereoscopic depth depends far more on relative values of lateral separation, S, for two different objects in the stereogram than it does on the absolute value of S. Also, exact alignment (and perceived depth) will vary slightly between Os because of their different interpupillary distances.

(4) Adjusting vergence distance. In prism and mirror systems, vergence distance is easily adjusted by varying the power of the prism and the angle of rotation of the mirror. Binoculars systems also require a prism adjustment. The prism is placed on the large side of the binoculars (i.e., away from the eye); because of the larger distances, very weak prisms will suffice. Fine adjustments of vergence distance usually are made by varying the lateral separation, S, of the stereogram directly under program control. The final value of vergence distance should normally be set equal to the accommodative distance, assuming an interpupillary distance of 6 cm.

Subjective Method of Aligning Vergence Distance (Haploscopic Method)

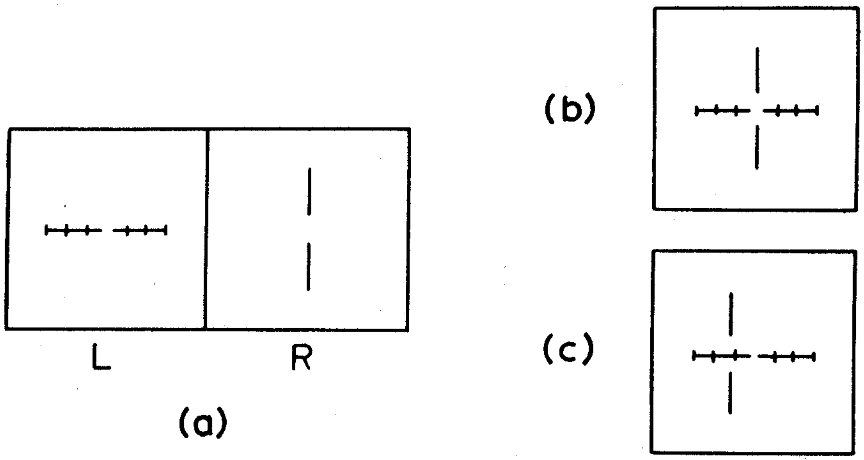


Fig. 5. Haploscopic method of adjusting vergence. (a) The test patterns. View of the display surface showing the left (L) and right (R) members of the stereogram. Each line is centered in its field. (b) Stereoscopic appearance when the viewing system is properly adjusted. (c) Stereoscopic appearance of a maladjusted system.

In this method, a stereogram such as that illustrated in Fig. 5 is presented. The engineer adjusts the prisms, the mirrors, or the stereo pair's lateral separation on the display surface until the O reports that the vertical line (which is seen only by the right eye) falls in the middle of the horizontal line (which is seen only by the left eye).

Haploscopic adjustment is virtually incumbent in the binoculars system. It is also useful in group viewing, as it allows Os to make accurate alignments, and it is particularly helpful in diagnosing failures to perceive stereoscopic depth.

Other Considerations

In all three systems, the viewer should be encouraged to wear spectacles if he normally uses them for objects at the accommodative distance of the display. Only in his normal viewing mode will his accommodation and his vergence be in their normal relation, i.e., appropriate to real objects.

When viewing continuously visible stereograms in which the two images are very similar, there is great tolerance for alignment error, and successful vergence occurs even with large accommodation-vergence disparities. Some experiments, however, require presentation of basically different stimuli to each eye. To insure that the eyes are correctly verged in such cases, the engineer should do two things: (1) He should provide a "visual frame" that is seen by both eyes, within which the diverse stimulus materials appear; and (2) he should adjust the vergence distance individually for each O (subjective method) under conditions as similar as possible to those of the experiment.

REFERENCES

BREWSTER, D. *The stereoscope, its history, theory and construction.* London: Murray, 1856.

EMSLEY, H. H. *Visual optics.* (4th ed.) London: Hatton Press, 1948.

FRY, G. A. An experimental analysis of the accommodation-convergence relation. American Journal of Optometry, 1937, 14, 402-414.

OGLE, K. N. Distortion of the image by prisms. Journal of the Optical Society of America, 1951, 41, 1023-1028.

SPERLING, G. Binocular vision: A physical and a neural theory. American Journal of Psychology, 1970, 83, 461-534.

WHEATSTONE, C. Contributions to the physiology of vision. Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision. Philosophical Transactions of the Royal Society, 1838, 8, 371-394.

NOTES

1. The divergence of light from an object—the stimulus to accommodation (focusing) of the eye—is exceedingly awkward to simulate in any kind of display; see section on Accommodation and Vergence.

2. See the section entitled, (3) Measuring vergence distance.

3. The treatment is simplified by assuming small angles, i.e., that the unaltered accommodative distance (the "viewing distance") is large relative to the interpupillary distance and relative to the picture width of each member of the stereogram. In practice, angles almost always are small.