David H. Brainard

Psychology Department
Stanford University
Stanford, California 94305

Calibration of a Computer Controlled Color Monitor

Because of the large number of stimulus configurations that may be displayed on a computer controlled color monitor, direct measurement of the relation between the digital input values to the frame buffer and the output of the monitor is not possible. To calibrate a monitor, it is necessary to make assumptions about the monitor's performance. These assumptions allow the monitor's output for any set of input values to be predicted from the input values and the results of a feasible number of calibration measurements. This article describes a set of assumptions that are sufficient to render the problem of monitor calibration tractable. These are the assumptions of phosphor constancy, phosphor independence, spatial independence, and the single scale factor assumption. If these assumptions hold, a monitor may be calibrated by measurements of the phosphor spectral power distributions, measurements of the phosphor input-output relations, and measurements of the spatial scale factor at a number of locations. Violations of these assumptions limit the accuracy of calibration. The results of measurements that test the validity of these assumptions for a Barco 5351 monitor are presented.

Introduction

A psychophysical laboratory based upon a computer controlled color monitor is attractive because it allows software specification of the spatial, temporal, and chromatic properties of the stimulus presented to subjects. Accurate control of the stimulus is possible only if the relation between the digital input values to the frame buffer that controls the monitor and the output of the monitor is known. Because of the large number of stimulus configurations that may be displayed, direct measurement of this relation is not possible. Consider the case where the light output from the monitor is to be described by its spectral power distribution. A typical frame buffer might require the specification of the intensities for three phosphors at $512 \times 512 = 262,144$ locations on the monitor screen. If the frame buffer provides 8-bit resolution for the specification of the intensity for each phosphor, then to measure directly the relation between frame buffer input values and monitor output requires the measurement of approximately $10^{1,893,917}$ spectral power distributions.

To reduce the number of measurements needed to determine the relation between frame buffer input and monitor output, it is necessary to make assumptions about the monitor's performance. These assumptions are chosen to allow the monitor's output for any set of input values to be predicted from the input values and the results of a feasible number of calibration measurements. One common assumption is the assumption of spatial independence. This is the assumption that the monitor's output at a location depends only on the input values for that location. For the above example, the assumption of spatial independence reduces the number of spectral power distribution measurements required to calibrate the monitor from approximately $10^{1,893,917}$ to approximately 10^7. Because 10^7 is still a prohibitive number of measurements, other assumptions must be made. These typically include the assumptions of phosphor constancy and phosphor independence.[1] The assumption of phosphor constancy is that the relative spectral power distribution of the light emitted by a monitor phosphor does not vary with the intensity of stimulation of the phosphor. The assumption of phosphor independence is that the intensity of stimulation of a phosphor is determined by the input value for that phosphor and is independent of the input values for the other two. The assumptions of phosphor constancy, phosphor independence, and spatial independence allow the spectral power distribution of the monitor's output at a single location to be predicted from the input values, measurements of the three phosphor spectral power distributions, and measurements of the three phosphor input-output relations. A phosphor spectral power distribution describes the amount of light emitted by a phosphor at each wavelength for its maximum input value. A phosphor input-output relation describes how the overall intensity of the light emitted by a phosphor varies with the input value. For the above example, measurements of the three phosphor

© 1989 by John Wiley & Sons, Inc.

spectral power distributions and $2^8 \times 3 = 768$ overall intensities would be needed to calibrate the monitor at a location. Thus these assumptions render the calibration of the monitor at a single location tractable.

The exact number of measurements necessary to predict a monitor's output for any combination of input values depends on the number of assumptions made. As more assumptions are made, the number of measurements necessary is generally reduced. The accuracy of the predictions of a monitor's output depends on how closely the monitor's actual performance matches the assumptions. As more assumptions are made, the monitor's performance is generally less likely to match the assumptions. The central problem of monitor calibration is to choose a set of assumptions that allow the calibration to be performed with a practical number of measurements and that accurately model the monitor's actual performance.

The process of monitor calibration can be divided into two steps. The first is to model the monitor's performance through a set of assumptions about the relation between input values and monitor output. The assumptions determine the measurements necessary to calibrate the monitor. The second step is to assess the validity of the assumptions for the monitor of interest. If the monitor's performance violates the assumptions to an extent that reduces the accuracy of calibration to an unacceptable level, then a more accurate set of assumptions must be specified. In any case, the measurements of how well the assumptions characterize actual performance yield an estimate of the accuracy of the calibration.

The accuracy of calibration necessary for various aspects of the display depends on the application. The rest of this article describes in detail calibration assumptions and measurements for a Barco 5351 monitor used to display stimuli for psychophysical experiments.[2,3] The stimuli in these experiments consist of one or more color patches displayed on the monitor's screen. Each patch is large relative to the pixel size of the monitor, and the stimuli are displayed for several seconds. The goal of the calibration is to specify the spectral power distribution of the light coming from each patch on the screen during a steady display. Exact specification of the spatial or temporal properties of the stimulus is not required for this application. For other applications, accurate specification of different aspects of the monitor's output may be needed.

The rest of this article is organized as follows: The next section describes the assumptions made about monitor performance and shows how these assumptions render monitor calibration tractable. Then, the section after that describes the display hardware and light measurement instruments used for the calibration measurements. The last section details the measurements made to test the calibration assumptions and reports the results of these measurements.

Calibration Assumptions

Phosphor constancy. The light emitted by the monitor at a location is characterized by its spectral power distribution. This light is called the color signal. The assumption of phosphor constancy is that the color signal emitted by the monitor is always a linear combination of three primary lights with fixed spectral power distributions. The primary lights are defined by the three phosphor spectral power distributions. These functions specify the power of the light emitted by a phosphor at each wavelength when it is displayed at the center of the screen for its maximum input value. The assumption of phosphor constancy contains two components. The first is that the color signal emitted by the monitor is formed only by the mixture of the light emitted by the three phosphors. The second is that the relative spectral power distribution of the light emitted by the phosphors does not vary. This means that only the overall intensity of the light emitted by a phosphor varies with input values or location.[1]

Under the assumption of phosphor constancy, the color signal $C(x,y,\lambda)$ emitted at any location on the screen (x,y) for any set of input values has a simple form. (Note that (x,y) refers to spatial location, not to CIE chromaticity coordinates $(x,y)_{CIE}$.) Let $C_r(\lambda)$ be the spectral power distribution of the light emitted by the red phosphor for maximum input setting at the center of the screen, $C_g(\lambda)$ be the spectral power distribution of the light emitted by the green phosphor, and $C_b(\lambda)$ be the spectral power distribution of the light emitted by the blue phosphor. Then

$$C(x,y,\lambda) = r(x,y)C_r(\lambda) + g(x,y)C_g(\lambda) + b(x,y)C_b(\lambda), \quad (1)$$

where the numbers $r(x,y)$, $g(x,y)$, and $b(x,y)$ give the fraction of maximum output for each phosphor present in the emitted light. Once the spectral power distributions $C_r(\lambda)$, $C_g(\lambda)$, and $C_b(\lambda)$ are known, only the three numbers $r(x,y)$, $g(x,y)$, and $b(x,y)$ are required to determine $C(x,y,\lambda)$. For any stimulus configuration, the numbers $r(x,y)$, $g(x,y)$, and $b(x,y)$ depend on the digital input values to the frame buffer. In principle, $r(x,y)$, $g(x,y)$, and $b(x,y)$ each may depend on the input values for all three phosphors at all locations on the screen.

Phosphor independence and spatial independence. The assumption of phosphor independence is that the amount of light emitted by each phosphor depends only on the input values for that phosphor. The assumption of spatial independence is that the amount of light emitted at a location depends only on the input values at that location. Denote the digital input values for the three phosphors at a location (x,y) by $I_r(x,y)$, $I_g(x,y)$, and $I_b(x,y)$. The assumptions of phosphor independence and spatial independence restrict the form of the dependence of $r(x,y)$, $g(x,y)$, and $b(x,y)$ on the input values. They mean that when the dependence of these fractions on input values is made explicit,

$$r(x,y) = f_r^{(x,y)}(I_r(x,y)), \quad (2)$$

$$g(x,y) = f_g^{(x,y)}(I_g(x,y)),$$

and

$$b(x,y) = f_b^{(x,y)}(I_b(x,y)).$$

The functions $f_r^{(x,y)}$, $f_g^{(x,y)}$, and $f_b^{(x,y)}$ specify the input-output relations for the three phosphors at location (x,y). Dependencies of $r(x,y)$ on $I_g(x,y)$ and $I_b(x,y)$, $g(x,y)$ on $I_r(x,y)$ and $I_b(x,y)$, and $b(x,y)$ on $I_r(x,y)$ and $I_g(x,y)$ are called phosphor interactions. Dependencies of $r(x,y)$, $g(x,y)$, $b(x,y)$ on the input settings for other locations are called spatial interactions. Note that the input-output relations may depend both on phosphor type and spatial location.

Spatial inhomogeneity. Color monitors are not spatially uniform, so that the same input values does not produce the same output intensity at different locations on the screen. This is the dependency of $f_r^{(x,y)}$, $f_g^{(x,y)}$, and $f_b^{(x,y)}$ on spatial location. Let $f_r^{(0,0)}$, $f_g^{(0,0)}$, and $f_b^{(0,0)}$ be the functions specifying the input-output relation for the three phosphors at the center of the screen. (Note that (0,0) refers to the center of the screen, not to the upper left.) The single scale factor assumption is that for any other location (x,y),

$$r(x,y) = \kappa(x,y) f_r^{(0,0)}(I_r(x,y)), \quad (3)$$

$$g(x,y) = \kappa(x,y) f_g^{(0,0)}(I_g(x,y)),$$

and

$$b(x,y) = \kappa(x,y) f_b^{(0,0)}(I_b(x,y)).$$

The single scale factor assumption means that the only variation of the input-output relations with spatial location is a scaling of the function and that for any location the spatial scale factor $\kappa(x,y)$ is the same for all three phosphors.

Output specification. Under the assumptions of phosphor constancy, phosphor independence, spatial independence, and the single scale factor assumption, the output of the monitor at location (x,y) depends only on the input values $I_r(x,y)$, $I_g(x,y)$, and $I_b(x,y)$. The relation between these values and the output is

$$\begin{aligned} C(x,y,\lambda) = {} & \kappa(x,y) f_r^{(0,0)}(I_r(x,y)) C_r(\lambda) \\ & + \kappa(x,y) f_g^{(0,0)}(I_g(x,y)) C_g(\lambda) \\ & + \kappa(x,y) f_b^{(0,0)}(I_b(x,y)) C_b(\lambda). \end{aligned} \quad (4)$$

To specify the output $C(x,y,\lambda)$ for any set of input values, it is necessary to measure the phosphor spectral power distributions ($C_r(\lambda)$, $C_g(\lambda)$, and $C_b(\lambda)$, the input-output relations $f_r^{(x,y)}$, $f_g^{(x,y)}$, and $f_b^{(x,y)}$, and the spatial scale factor $\kappa(x,y)$ for all locations of interest.

The accuracy of the specification of $C(x,y,\lambda)$ depends on the validity of the assumptions described above. The next section describes the display hardware and light measurement instruments used to make measurements to test these assumptions. The following section describes the measurement procedures and results. These measurements also provide the data necessary to compute $C(x,y,\lambda)$ as in eq. 4.

Video Hardware and Measurement Instrumentation

Display Hardware

The display hardware consists of a color monitor controlled by a digital frame buffer. The monitor is a Barco 5351 RGB color monitor. The monitor input is three RS-170 video signals, one for each of the three monitor phosphors. The monitor receives its vertical and horizontal synchronization information from the green phosphor input signal. The video input to the monitor is generated by a frame buffer. The frame buffer is a model number 232808 manufactured by Number Nine Graphics Corp. The board is configured to generate video output for a rectangular image 640 pixels wide by 350 pixels high at a refresh rate of 87 frames per second. (We chose the nonstandard refresh rate of 87 frames per second to eliminate the perception of flicker during the presentation of psychophysical stimuli.) The frame buffer is controlled by an IBM personal computer XT modified to use an 8086 microprocessor. the computer interface allows the input values for the red, green, and blue phosphors at each location to be set independently at any of 256 levels. These levels are specified by an 8-bit word associated with the location. This is used as an index into a table containing three 8-bit words, one to specify the intensity of stimulation of each of the phosphors. Thus the hardware is capable of generating $(256)^3 = 16,777,216$ combinations of red, green, and blue phosphor excitation at any location, although only 256 of these combinations may used simultaneously.

Light Measurement Instrumentation

The light measurement instruments are a spectroradiometer and a photometer. The spectroradiometer is manufactured by EG&G Gamma Scientific. This instrument permits measurement of the light output power in narrow-band spectral regions. Although a spectroradiometer is necessary to measure the complete color signal emitted by a monitor, for some applications it is possible to perform satisfactory calibration without one. Cowan[4] describes techniques that may be used for monitor calibration when a spectroradiometer is not available. The second instrument is a photometer also manufactured by EG&G Gamma Scientific. This instrument permits measurement of the light output power through a broadband spectral filter. Since a spectroradiometer is available, it is not necessary to use the photometer. Because the photometer uses a broadband filter, however, accurate measurements of relative intensities of lights with known relative spectral power distributions may be obtained much more rapidly with the photometer than with the spectroradiometer.

The spectroradiometer. The spectroradiometer is based on a holographic grating monochromator (model NM-3H). The spectroradiometer is specified to have a wavelength accuracy of ± 2 nm, a wavelength repeatability of ± 0.5 nm, and a half power bandwidth of 5.8 nm. The light enters the spectroradiometer through a telescope (model 2020-31), which allows the measurement of light light from a small area of the monitor screen. All measurements were made with the telescope aperture set to measure 20′ of visual arc of the monitor screen from a distance of 180 cm. This corresponds to a 1.05-cm diameter patch of the monitor screen. The light exiting from the

telescope passes through the monochromator and is detected by a photomultiplier (model D-46AQ). The photomultiplier is controlled by an interface device (model DR-2). The interface device provides a digital output which is read using a TTL interface in the same IBM personal computer that controls the frame buffer. In addition, the computer TTL interface is used to set the monochromator wavelength through a scanning control unit (model SC-1). The software that controls the radiometer is written in the C programming language using the data structures and library support routines provided by the Color Analysis Package.[5] The spectroradiometer is calibrated to a standard light source provided by EG&G Gamma Scientific (model RS-10A irradiance head fitted with model RS-70-A diffuser and controlled by model RS-3 control unit). The source is specified to have a correlated color temperature of 2600°K and to have an output uncertainty of ±2.5% with respect to National Bureau of Standards standards of irradiance. The wavelength calibration of the monochromator is checked against the spectral line at 405 nm in an ordinary fluorescent lamp and the spectral line of a helium-neon laser at 633 nm.*

The photometer. The photometer is an EG&G Gamma Scientific model 2009JK telephotometer. The light enters the photometer through a telescope that allows measurement of light from a small area of the monitor screen. All measurements were made with the telescope aperture set to measure 20′ of visual arc of the monitor screen from a distance of 180 cm. This corresponds to a 1.05-cm diameter patch of the monitor screen. The photometer is also controlled using a model DR-2 interface device. The photometer is calibrated to an internal light source which has a correlated color temperature of 2856°K. Although the photometer has filters designed to match the CIE XYZ color matching functions, measurements were only made using its neutral density filter number 1.‡

Measurement Procedures and Results

This section describes the measurement procedures in detail and reports the measurement results. These measurements test the validity of the assumptions described earlier and provide the data necessary to predict the monitor's output for any set of input values.

Phosphor Spectral Power Distribution

The spectral power distributions of the phosphors were measured with the spectroradiometer. Each phosphor was displayed alone at its maximum power across a patch at the center of the screen. (The lookup table value for the phosphor being measured was 255.) This patch was 6.4 cm high by 3.8 cm wide, or 2.0 by 1.2 degrees of visual arc from the 180 cm distance at which the radiometer was placed. The radiometer nominally measures a 20′ of visual arc region at the center of this patch. The patch was large enough so that light from surrounding portions of the screen was not captured by the radiometer. The radiant power emitted from the patch was measured between 370 and 730 nm at 1-nm steps. Figure 1 shows the three measured phosphor spectral power distributions. These measurements provide the functions $C_r(\lambda)$, $C_g(\lambda)$, and $C_b(\lambda)$ of eq. 4.

To test the assumption of phosphor constancy, the phosphor spectral power distributions were remeasured at a lower power. (The lookup table value was 128 rather than 255.) This drove the phosphors at approximately one-fifth of their maximum power. A regression of the low power measurements for each phosphor on the corresponding full power measurements revealed a root mean squared error in fit of less than 0.5% of the peak energy value for each of the three phosphors. Figure 2 shows the low power measurements plotted against the full power measurements for each phosphor. Each point in the plots represents the two measurements at a single wavelength. If the phosphors were completely constant, data for each phosphor would fall along a line. The only significant deviations from a line are for for the red phosphor. These deviations are probably due to measurement error. Because the red phosphor has a sharply peaked spectral power distribution small variations in the monochromator wavelength settings result in larger variations in the measured energy values than for the green or blue phosphors. This point is discussed in detail below.

As a second test of the assumption of phosphor constancy, a measurement was made of the spectral power

*In general the EG&G spectroradiometer has performed satisfactorily. It has some shortcomings, however, that should be noted. First, model SC-1 scanning control unit that controls the monochromator wavelength setting has a gain error. When the input wavelength to the scanning unit is 405 nm, the wavelength setting of the monochromator is correct. When the input wavelength to the scanning unit is 633 nm, the wavelength setting of the monochromator is off by 3 nm. For this reason, the it was necessary to incorporate a correction for this problem into the software that controls the scanning unit. This correction adjusts the input wavelength passed to the scanning unit to compensate for the gain error. With the correction, measurements of interference filters with peaks between 405 and 633 nm indicated that the wavelength accuracy of the system was within the ±2 nm specification of the instrument. Second, there are a number of difficulties with the digital interface of the model DR-2 interface device that samples and holds the output from the photomultiplier. For example, the data ready line of the output interface will indicate that data is valid before all of the output data lines have settled. The interface software handles this condition as well as a number of other problems. A more detailed description of these problems and their software solutions is available from the author. Third, it is important to take great care in grounding the signal and power lines of a spectroradiometer. It is necessary to float the EG&G instruments with respect to the power lines and connect only a single signal ground wire from the IBM personal computer to the EG&G instruments. If this is not done, the readings from the instrument can be incorrect.

‡The specified accuracy of the match between the photometer's filters and the CIE XYZ color matching functions is not sufficient to provide accurate tristimulus values for lights with sharply peaked spectral power distributions. In addition, the long-term temporal stability of the filters's spectral sensitivities is suspect. For these reasons, the photometer was only used to make measurements of the relative intensity of lights with constant relative spectral power distributions.

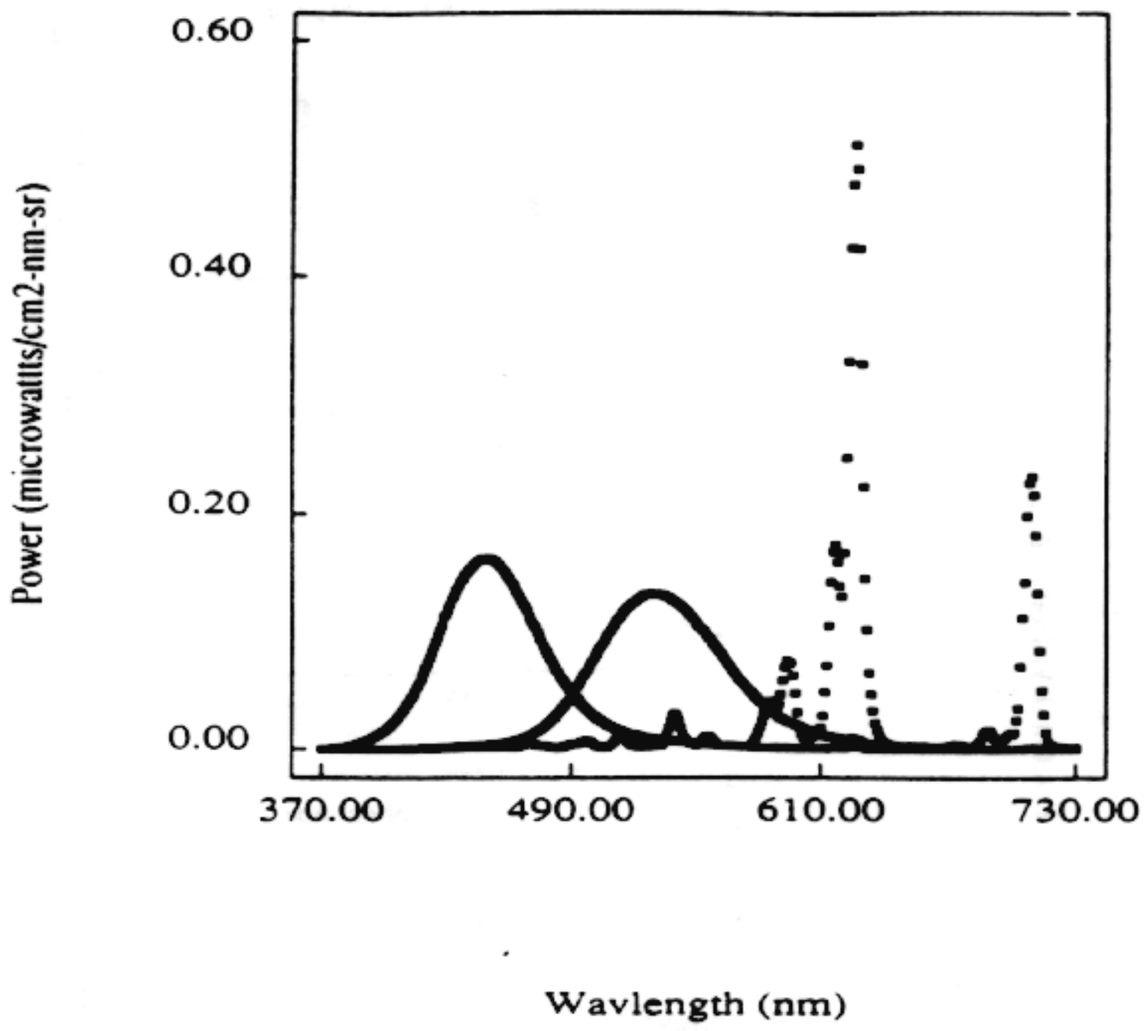

FIG. 1. Barco 5351 phosphor spectral power distributions. Measurements were made at the center of the screen.

distribution of a patch where all three phosphors were displayed simultaneously at their maximum power. (The lookup table values were 255). This measured spectral power distribution was fit with a linear combination of the

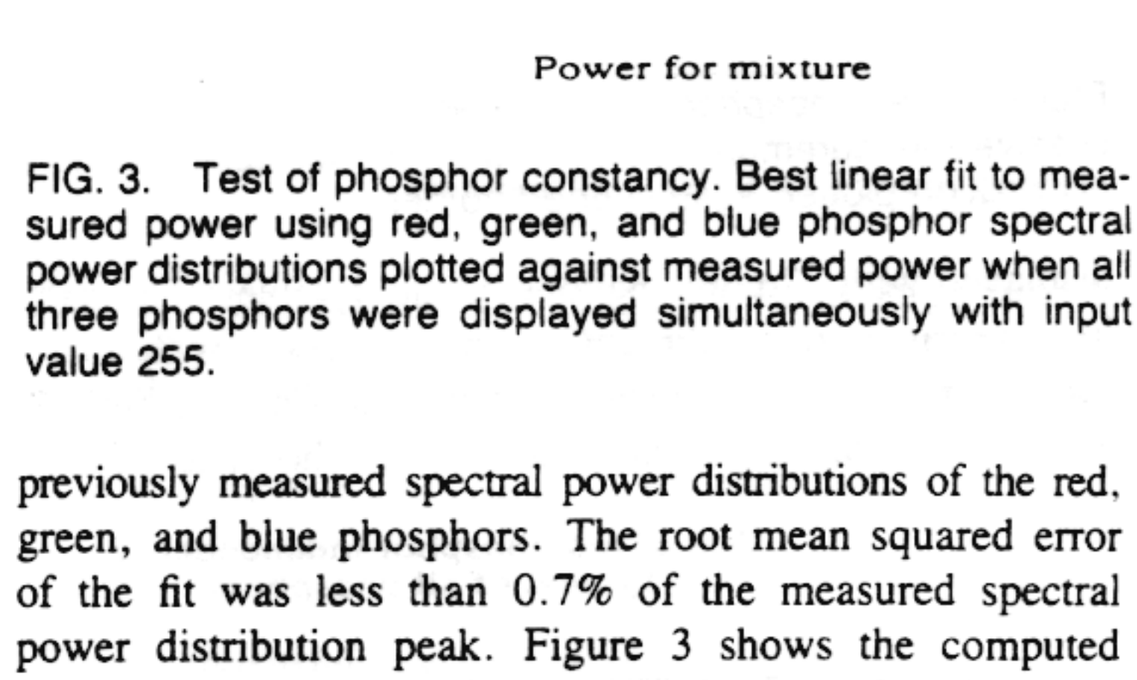

FIG. 3. Test of phosphor constancy. Best linear fit to measured power using red, green, and blue phosphor spectral power distributions plotted against measured power when all three phosphors were displayed simultaneously with input value 255.

previously measured spectral power distributions of the red, green, and blue phosphors. The root mean squared error of the fit was less than 0.7% of the measured spectral power distribution peak. Figure 3 shows the computed power plotted against the measured power for all wavelengths. If the phosphors were completely constant, these points would fall along a line. The main deviations come from the measurements at wavelengths near a peak of the red phosphor spectral power distribution. The power at these wavelengths is difficult to measure accurately because the spectroradiometer is specified to have a wavelength repeatability of only ±0.5 nm. To verify that the deviations from a line shown in Fig. 3 can be accounted for by measurement error for the red phosphor, two successive measurements of the red phosphor spectral power distribution were made. Figure 4 shows the results of these two measurements plotted against one another. Under the assumption that the monitor's output is constant over the period of time required to repeat the measurement (about one-half hour), the deviations from a line in Fig. 4 can be attributed entirely to measurement error.* These deviations from a line are similar to those of Fig. 3.

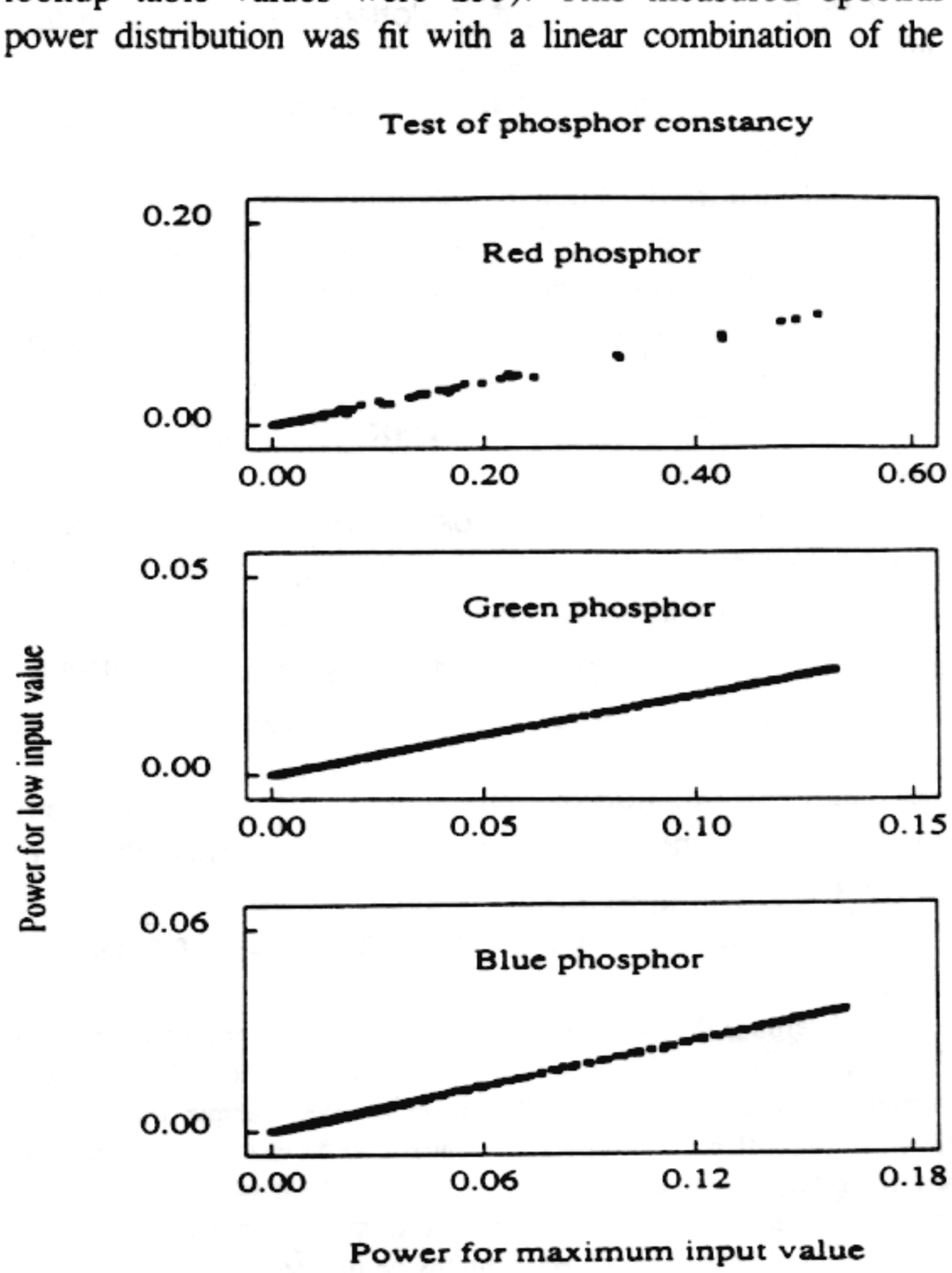

FIG. 2. Test of phosphor constancy. Phosphor power for low input value plotted against phosphor power at corresponding wavelengths for maximum input value.

Phosphor Independence

To test the assumption of phosphor independence, measurements were made of the output from a rectangular patch

*This error also limits the accuracy of the measurement of the peaks of the red phosphor spectral power distribution shown in Fig. 1. A spectroradiometer with finer wavelength sampling is desirable for the measurement of sharply peaked spectral power distributions.

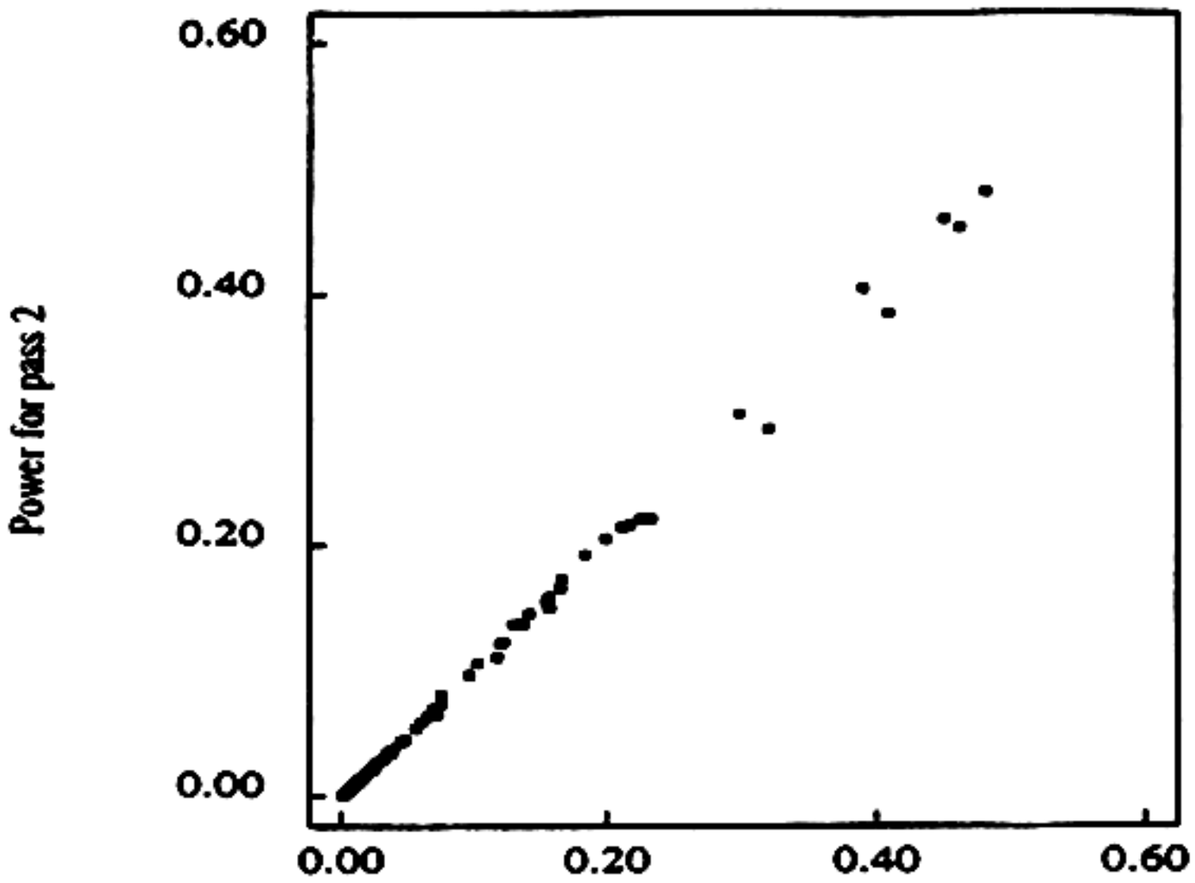

FIG. 4. Red phosphor measurement consistency. Two successive measurements of the red phosphor spectral power distribution plotted against one another.

displayed with 27 combinations of input values. The patch had the same spatial dimensions and location as the patch used to measure the phosphor spectral power distributions. The 27 combinations of input values were chosen as follows.

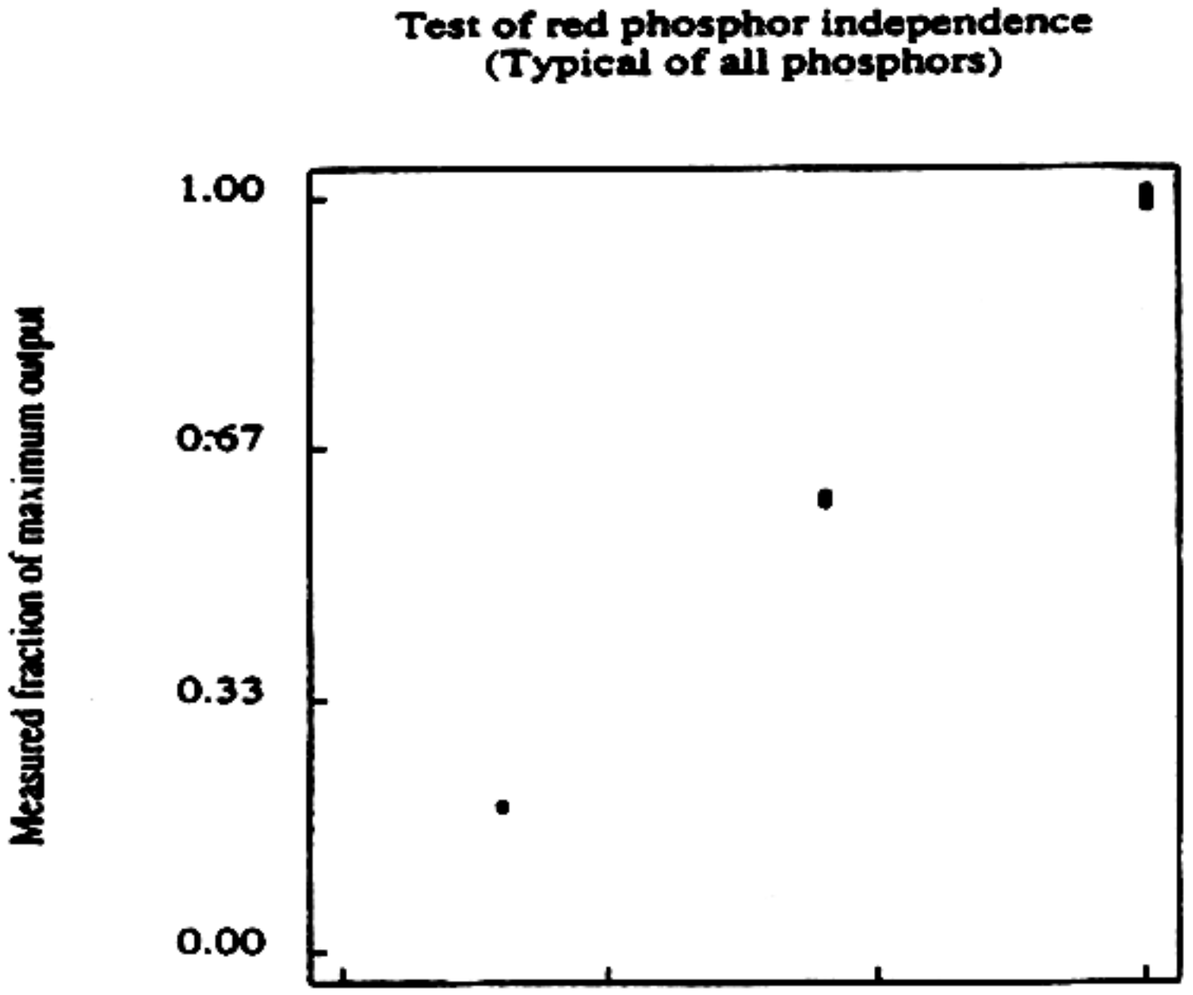

FIG. 5. Test of phosphor independence. Measured fraction of red phosphor maximum output present in output mixture plotted against desired fraction that would be present under the assumption of phosphor independence. This plot is typical of the results for all three phosphors.

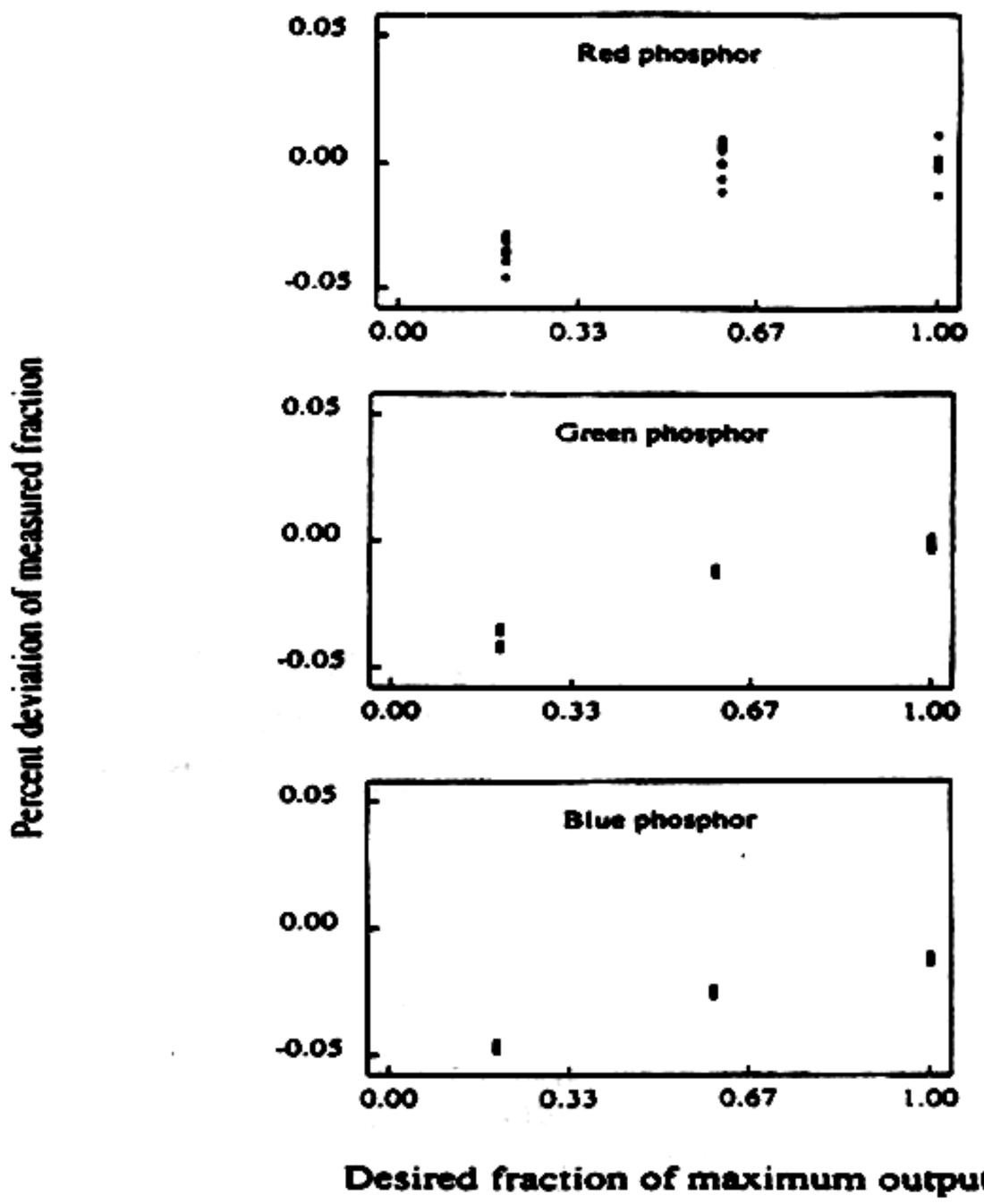

FIG. 6. Test of phosphor independence. Percentage error between measured and desired fraction of maximum phosphor output for each phosphor plotted against desired fraction.

For each phosphor, determine the three input values that would drive the phosphor at fractions 0.2, 0.6, and 1.0 of its maximum power when the phosphor was displayed alone. (Measurements of the phosphor input-output relation, described below, were used to determine these input values.) Then choose all $3^3 = 27$ possible permutations of the three input values for the three phosphors. Thus for each of its three input settings, each phosphor was displayed in nine different combinations.

Under the assumption of phosphor constancy, the spectral power distribution $C(\lambda)$ of the light output from the patch for each combination can be expressed as a linear combination of the three phosphor spectral power distributions:

$$C(\lambda) = rC_r(\lambda) + gC_g(\lambda) + bC_b(\lambda). \quad (5)$$

For each combination of input values, measurements were made to determine the fractions r, g, and b of each phosphor present in the output.** Under the assumption of phosphor independence, the fraction r of the red phosphor present in

**To determine the values of r, g, and b, it was not necessary to measure the entire output spectral power distribution $C(\lambda)$. The following measurement procedure was used: The power at seven sample wavelengths near the peak of each phosphor was measured for a total of 21 sample wavelengths. The values of $C_r(\lambda)$, $C_g(\lambda)$, and $C_b(\lambda)$ at each of these wavelengths was known from the measurements of the phosphor spectral power distributions. Let the sample wavelengths be denoted by λ_s. Then $C(\lambda_s) = rC_r(\lambda_s) + gC_g(\lambda_s) + bC_b(\lambda_s)$. A linear regression of $C(\lambda_s)$ on $C_r(\lambda_s)$, $C_g(\lambda_s)$, and $C_b(\lambda_s)$ provides the values of r, g, and b.)

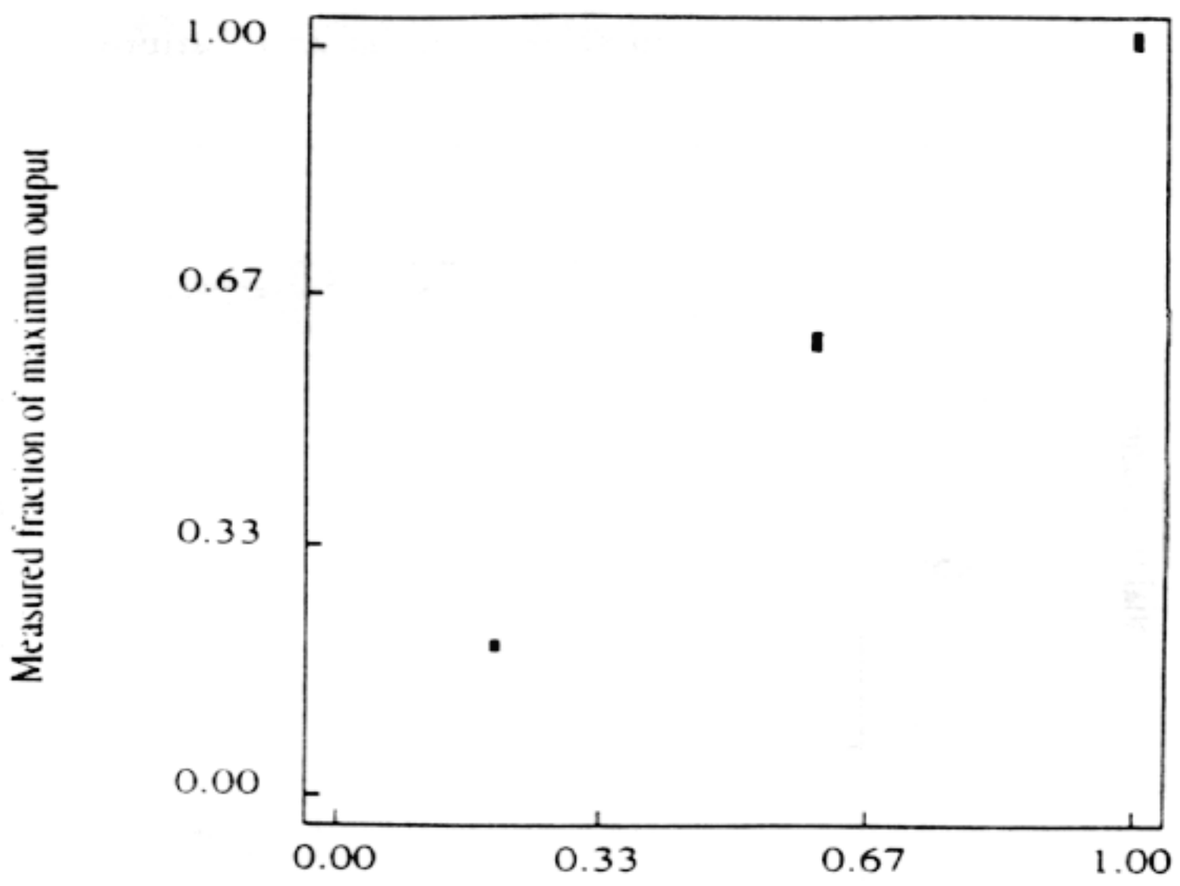

FIG. 7. Test of spatial independence. Measured fraction of maximum red phosphor output present in output mixture plotted against desired fraction that would be present under the assumptions of phosphor independence and spatial independence. The measured patch was surrounded by eight other patches. This plot is typical of the results for all three phosphors.

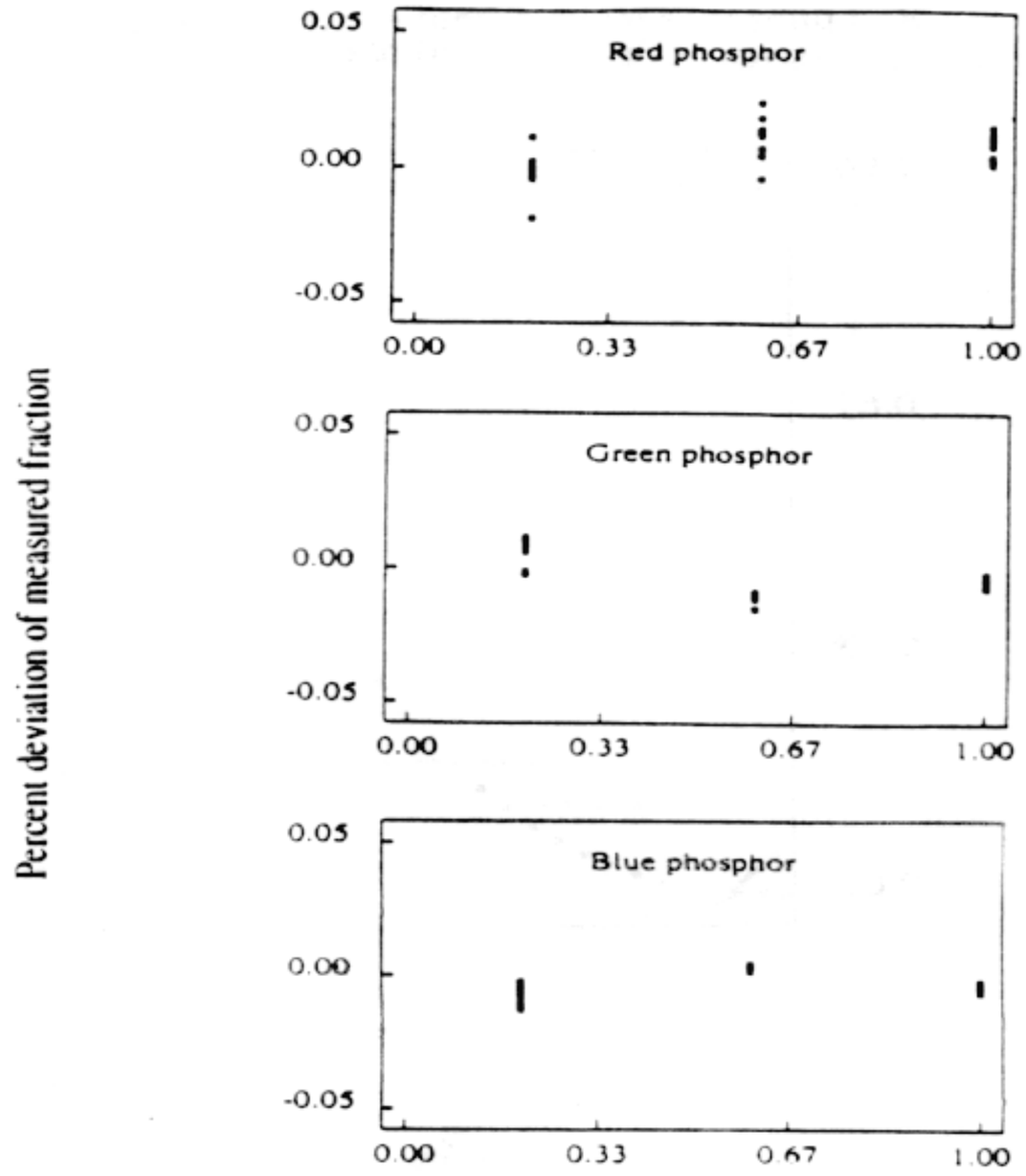

FIG. 8. Test of spatial independence. Percentage error between measured and desired fraction when measured patch is displayed with a surround of eight other patches.

the light output should be independent of the fractions g and b, and similarly for the green and blue phosphors. Figure 5 shows for the red phosphor the measured fraction of maximum phosphor output present in the output mixture plotted against the desired fraction that would be present under the assumption of phosphor independence. This plot is typical of the results for all three phosphors. Each cluster of points in the plots represents measurements for nine input combinations. The data for different combinations are sufficiently close together that it is impossible to resolve individual points. If there were no phosphor interactions, these data would all fall along the line with unit slope. The deviations from this line are small. To show the deviations, it is necessary to change the scale of the plot. Figure 6 plots the percentage error in fractional output as a function of the desired output for each of the phosphors. Denote the desired output fraction by f and the measured output fraction by m. Then the percentage error in fractional output is $(m - f)/f$. Each cluster of points in the plots represents measurements for nine input combinations. All three phosphors show departures from the desired output for low input settings of about 3%. At higher input settings, the red and green phosphor show minimal phosphor interactions, whereas the blue phosphors shows interactions at all levels. These phosphor interactions may arise from interactions in the power supplies of the monitor electron guns. The greater spread in the measurements for the red phosphor is probably due to greater measurement error for the red phosphor.

Spatial Independence

To test the assumption of spatial independence, measurements were again made of the output from a rectangular patch displayed with 27 combinations of input values. The 27 combinations of input values were the same as were used to test the assumption of phosphor independence. In this case, however, the measured patch was surrounded with eight other patches of the same size. I chose this spatial configuration because it is close to the configuration used in the psychophysical experiments for which the calibration was intended. The input settings for the surrounding patches were chosen from the 27 input combinations used for the central patch. Figure 7 shows for the red phosphor the measured fraction of maximum phosphor output present in the output mixture plotted against the desired fraction that would be present under the assumptions of phosphor independence and spatial independence. This plot is typical of the results for all three phosphors. Each cluster of points in the plot represents measurements for nine input combinations. The data for different combinations are sufficiently close together that it is not possible to resolve individual points. If there were no spatial interactions, these points would lie along the same line as the points in Fig. 5. The changes from Figs. 5 to 7 are slight. The mean percentage change

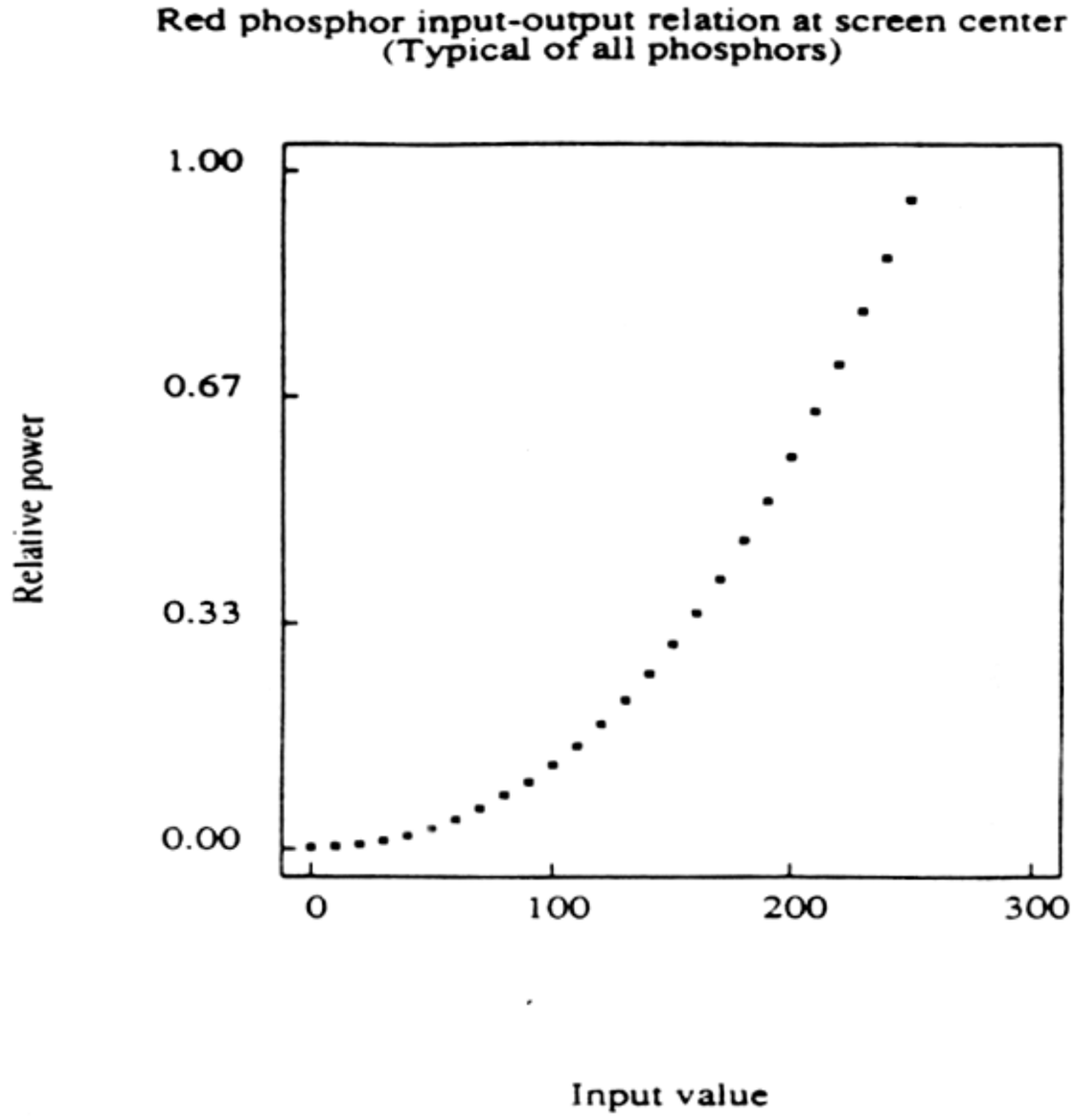

FIG. 9. Red phosphor input-output relation for center of screen. The plotted input-output relations for the green and blue phosphors are visually indistinguishable from the plot for the red phosphor.

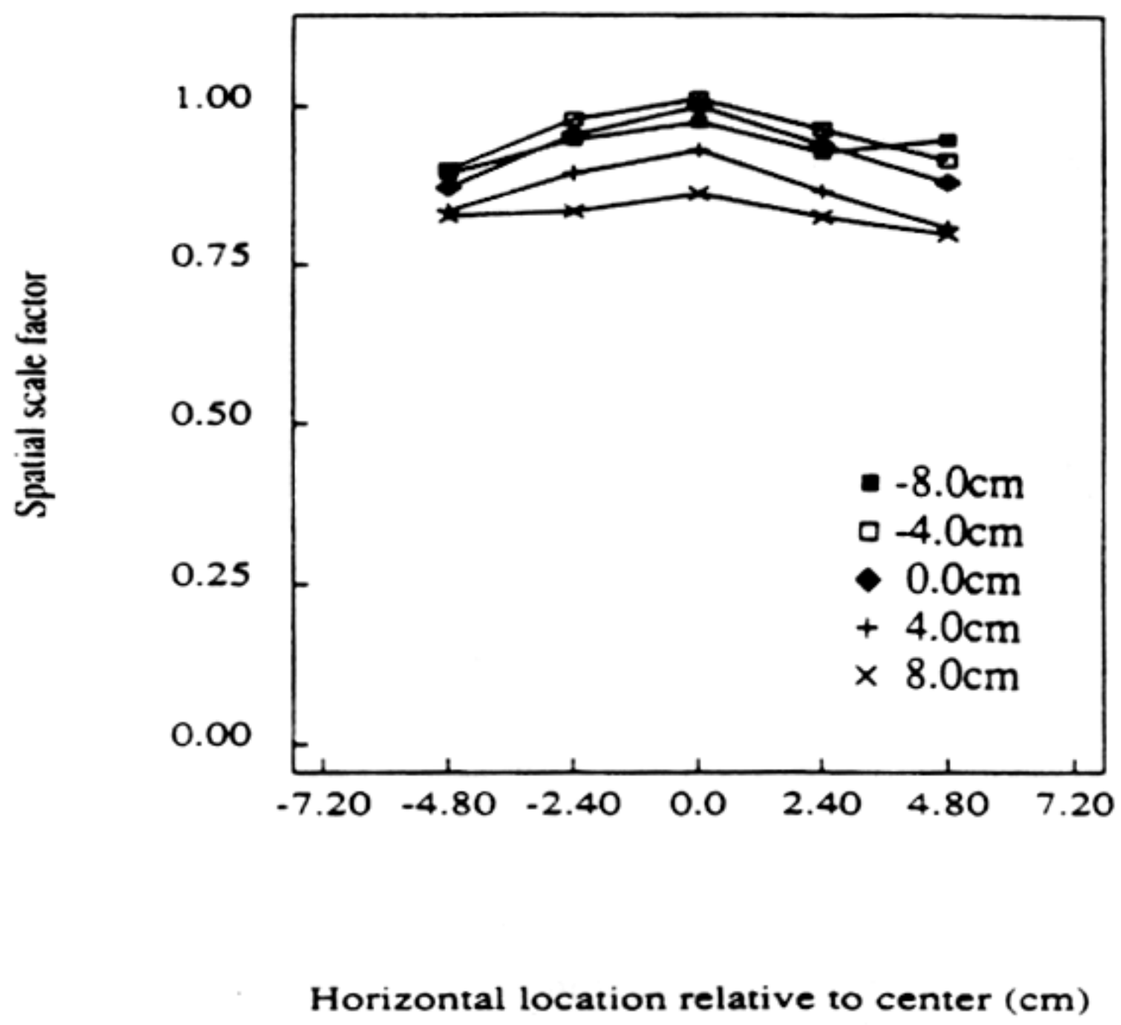

FIG. 11. Spatial inhomogeneity of Barco 5351 monitor. Each line connects data measured at a common vertical location. The *x*-axis is the horizontal location. The *y*-axis is the measured scale factor.

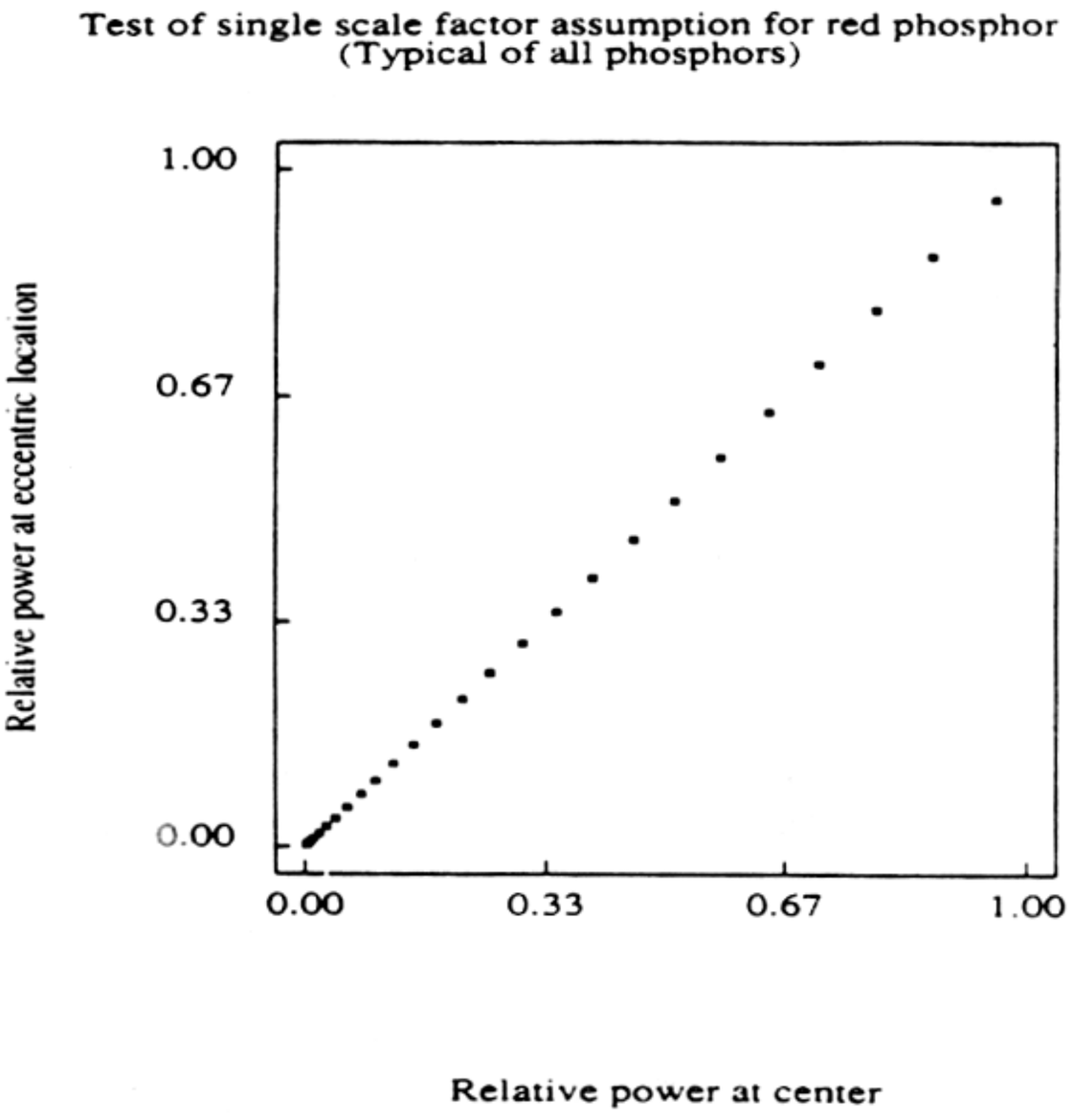

FIG. 10. Test of single scale factor assumption. Red phosphor input-output relation measured at an eccentric location plotted against the red phosphor input-output relation measured at the center of the screen. The corresponding plots for the green and blue phosphors are visually indistinguishable from this plot.

in fractional output is less than 2.5% for all three phosphors with the addition of the surrounding patches. Figure 8 plots the percentage error in fractional output as a function of the desired fractional output for each of the phosphors when the measured patch was displayed with surround. The errors are smaller than the ones shown in Fig. 6. The magnitude of the spatial interactions of the monitor has approximately the same magnitude as the phosphor interactions, but has the opposite sign. For the spatial configuration measured, the spatial interactions of the monitor have the fortuitous consequence of compensating for the general trend of the phosphor interactions.

Spatial Inhomogeneities

Two sets of measurements were made to examine the spatial inhomogeneities of the monitor. The first investigated the dependence of $f_r^{(x,y)}$, $f_g^{(x,y)}$, and $f_b^{(x,y)}$ on spatial location. These functions were measured both at the center of the screen and at eccentric location. If the spatial inhomogeneities of the monitor are accurately described by a single scale factor, then the functions $f_r^{(x,y)}$, $f_g^{(x,y)}$, and $f_b^{(x,y)}$ should only differ by a scale factor between the two locations. The second set of measurements investigated the value of the scale factor $\kappa(x,y)$ at 25 locations. This factor was measured for four different combinations of input values for the measured patch. If the single scale factor assumption holds, the measurements of the scale factor as a function of location should not depend on the combination of input values that was used when the measurements were made.

Measurement of monitor input-output relation. The input-

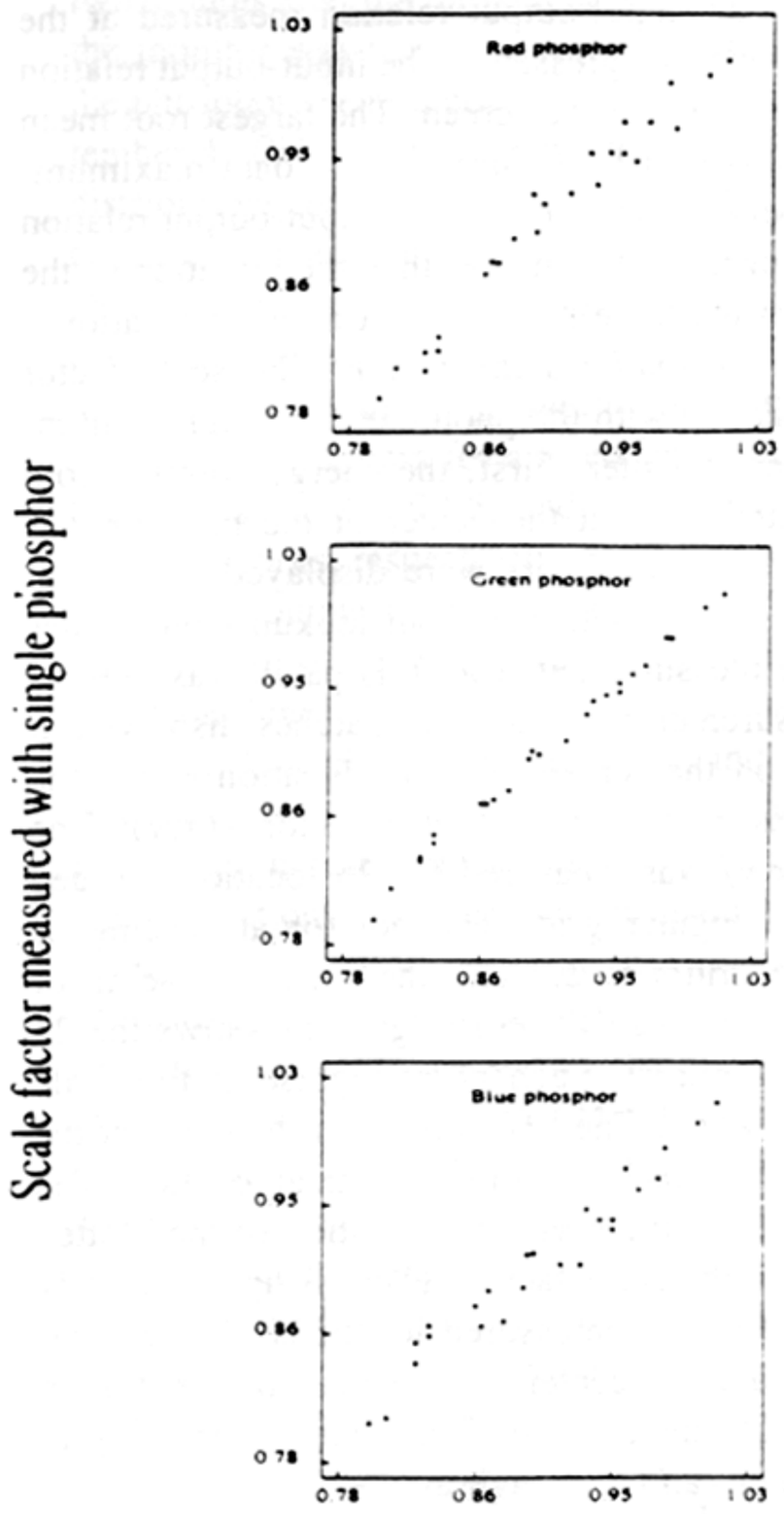

FIG. 12. Test of single scale factor assumption. Values of $\kappa(x,y)$ measured with each phosphor alone plotted against the values of $\kappa(x,y)$ measured with all three phosphors.

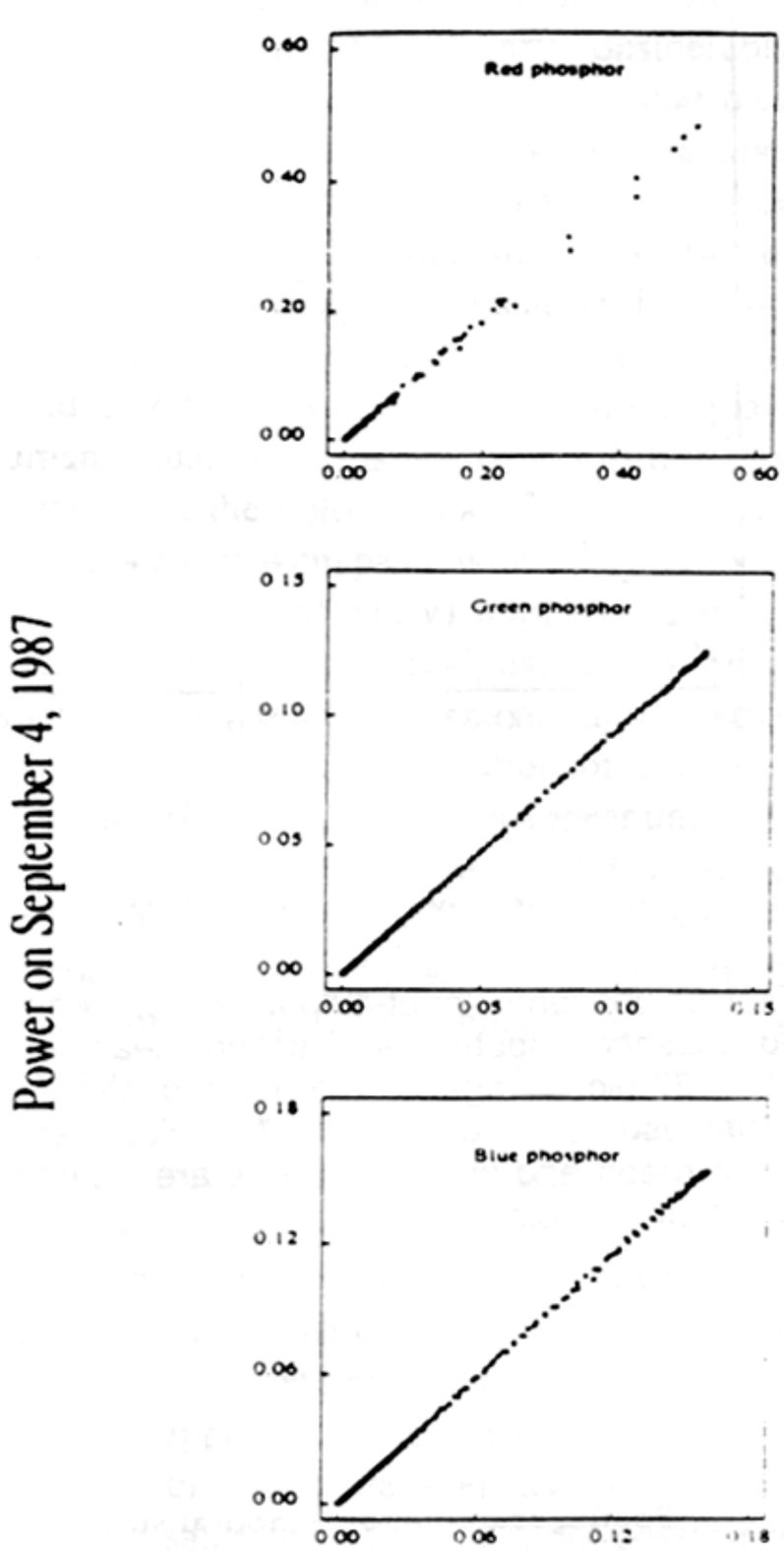

FIG. 13. Test of phosphor spectral power distribution temporal stability. Phosphor spectral power distributions measured on September 4, 1987 plotted against the same measurements on June 29, 1987.

output relations $f_r^{(x,y)}$, $f_b^{(x,y)}$, and $f_b^{(x,y)}$ were measured with the photometer through its number 1 neutral density filter. Each phosphor in turn was displayed across a rectangular patch. This patch had the same spatial dimensions as the patch used to measure the phosphor spectral power distributions. For each phosphor, a measurement was taken through the photometer with the phosphor set at its maximum input value of 255. This measurement was used to normalize the rest of the measurements. Thus the input-output functions are expressed as fractions of the maximum phosphor output at the measurement location. Measurements were made for 26 evenly spaced lookup table input values between 0 and 250. Because the functions are smoothly varying (see Fig. 9), the input-output relation was not measured at all 256 possible input values. Figure 9 shows the 26 measured values for the red phosphor at the center of the screen. The plotted input-output relations for the green and blue phosphors are visually indistinguishable from the plot for the red phosphor. A pairwise regression of the three functions on each other revealed a maximum root mean squared error of 0.012. Since the functions are expressed as fractions of maximum output, this is an error of 1.2% of the function maximum. For this monitor, it would be reasonable to characterize the input-output relation for all three phosphors with a single function.

To test the single scale factor assumption, the input-output relations were also measured at a location 8 cm above and 4.8 cm to the left of the center of the screen. Figure 10 shows the red phosphor input-output relation measured at the eccentric location plotted against the red phosphor input-output relation measured at the center of the screen. The corresponding plots for the green and blue phosphors are visually indistinguishable from this plot. The data in the plot lie along a line, indicating that the input-output relation differs only by a scale factor between the two locations.

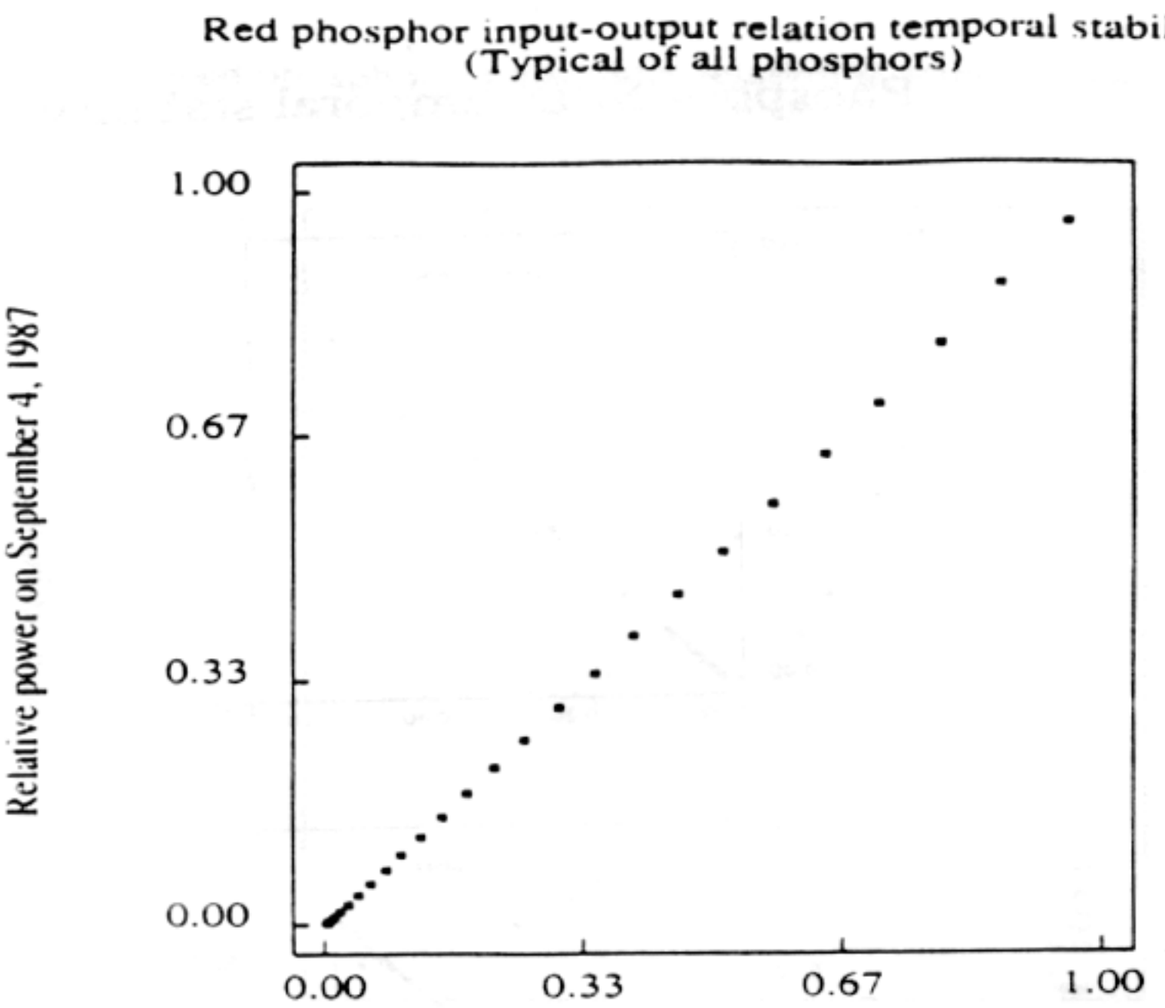

FIG. 14. Test of phosphor input-output relation temporal stability. Red phosphor input-output relation measured on September 4, 1987 plotted against the red phosphor input-output relation measured on June 29, 1987. The corresponding plots for the green and blue phosphors are visually indistinguishable from this plot.

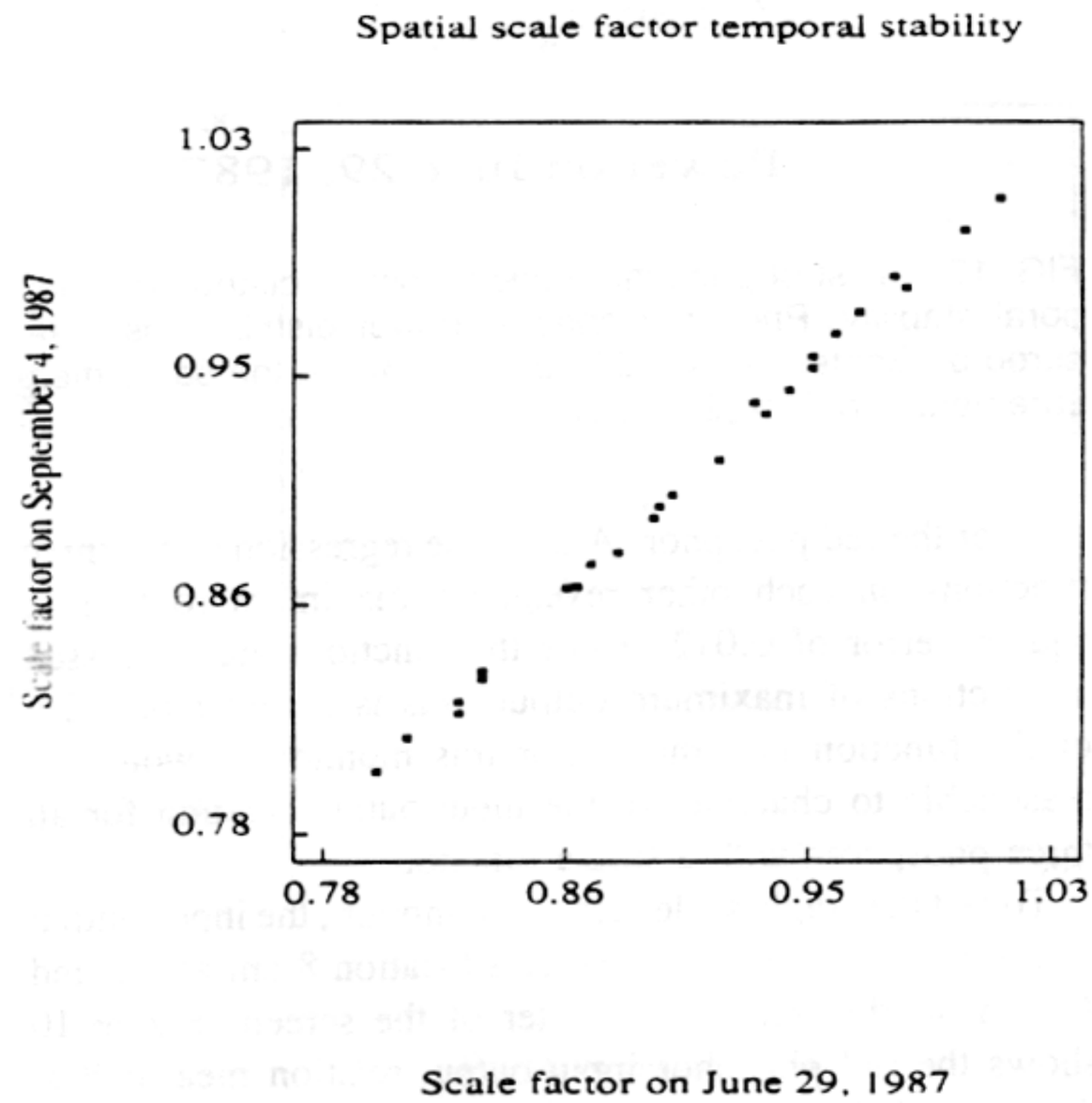

FIG. 15. Test of spatial scale factor temporal stability. Spatial scale factor for 25 locations measured on September 4, 1987 plotted against the spatial scale factor measured on June 29, 1987.

(Because the measurement procedure normalizes the output values to the maximum output at the measurement location, the slope of this line is constrained to be near unity.) For each phosphor, the input-output relation measured at the eccentric location was regressed on the input-output relation measured at the center of the screen. The largest root mean squared error was 0.001, or 0.1% of the data maximum. The variation between locations in the input-output relation is very small: considerably smaller than the variation of the relation between the three phosphors at a single location.

Measurement of spatial scale factor. The scale factor $\kappa(x,y)$ was measured with the photometer through its number 1 neutral density filter. First, the energy coming from a 50 pixel square patch at the center of the monitor was measured. All three phosphors were displayed simultaneously across this patch with the input lookup table values set to 140. The measurement from this patch was used to normalize measurements of identical patches displayed at other locations on the screen. For any location (x,y), the normalized measurement is the scale factor $\kappa(x,y)$. The scale factor $\kappa(x,y)$ was measured for 25 locations spaced on a 5-by-5 rectangular grid. The horizontal spacing of locations on the grid was 2.4 cm; the vertical spacing of locations on the grid was 4.0 cm. Figure 11 shows the 25 measurements of $\kappa(x,y)$. Each point represents the scale factor for one location. The lines on the graph connect the data for points which have a common vertical location. The x-axis of the graph is the horizontal location of the plotted data, the y-axis is the scale factor. There is approximately 20% variation across the measured locations. These locations span an area at the center of the screen that is 9.6 cm horizontally and 16 cm vertically. The change from one grid location to another can be as high as 8%.

As a second test of the single scale factor assumption, the measurements of $\kappa(x,y)$ for the grid locations were repeated using different input value combinations for the displayed patches. Three other input value combinations were used, one with the red phosphor displayed alone with its lookup table input value set to 140, one with the green phosphor displayed alone with its lookup table input value set to 140, and one with the blue phosphor displayed alone with its lookup table input value set to 140. Figure 12 plots the values of $\kappa(x,y)$ measured with the red, green, and blue phosphors alone against the values of $\kappa(x,y)$ measured with all three phosphors turned on. If a single scale factor characterized all of the monitor's spatial inhomogeneities, the data in these plots would fall along the line with unit slope. Although there is significant variation between the four sets of measurements, the variation is small compared to the overall spatial inhomogeneity of the monitor. A possible relaxation of the single scale factor assumption would be to assume that the spatial inhomogeneities of the monitor are characterized by three scale factors, one for each phosphor.

Temporal Stability

A final assumption that is implicit in the calibration of a monitor is that the calibration measurements remain valid

over a period of time. To test this assumption, the phosphor spectral power distributions, phosphor input-output relations, and the monitor spatial scale factor were remeasured two months after the initial measurements. During this time, the monitor was turned on continuously. Figure 13 shows the phosphor spectral power distributions measured on September 4, 1987 plotted against the phosphor spectral power distributions measured on June 29, 1987. The data lie close to a line for each phosphor, indicating that there is little change in the relative phosphor spectral power distributions over time. However, the slope of the lines is not unity, indicating a change in the overall intensity of the light emitted by the phosphors for a fixed input value. The slopes of the lines are 0.94, 0.95, and 0.96 for the red, green, and blue phosphors, respectively. Figure 14 shows the red phosphor input-output relation measured on September 4, 1987 plotted against the red phosphor input-output relation measured on June 29, 1987. The corresponding plots for the green and blue phosphors are visually indistinguishable from this plot. The measurements lie close to a line for each phosphor, indicating that there is little change in the input-output relations. (Because the measurement procedure normalizes the output values tothe maximum output at the time of measurement, the slope of this line is constrained to lie near unity.) Figure 15 shows the spatial scale factors on September 4, 1987 plotted against the spatial scale factors on June 29, 1987. Although these values do not lie along a line, the deviations are smaller than the deviations shown when the spatial scale factor is measured with different phosphors (see Fig. 12). The main variation in this monitor's behavior over time is a drift in the intensity of the output from each phosphor for a given input value. In addition, the drift of the three phosphors has approximately the same magnitude. This means that the CIE chromaticity coordinates $(x,y)_{CIE}$ of the monitor's output for a given set of input values do not change significantly. For example, the CIE chromaticity coordinates of the monitor white point (input values for all three phosphors 255) changed from $(x,y)_{CIE} = (0.302,0.315)$ to $(x,y)_{CIE} = (0.301,0.314)$. This change is very small: less than one just-noticeable difference.[6,7]

Discussion

This article has described the calibration of a computer controlled color monitor. The calibration procedure is based on several assumptions about the monitor's behavior. The measurements to test the assumption of phosphor constancy indicate that the relative spectral power distribution of the light emitted by each of the phosphors does not vary with input values or spatial location. The measurements to test the assumptions of phosphor independence and spatial independence indicate that under these assumptions it is possible to specify the fractions r, g, and b at a single location to an accuracy of approximately 3%. This is the magnitude of the largest measured deviation from independence. Improved accuracy of calibration might be achieved by formulating a more sophisticated model of the phosphor and spatial interactions. The spatial interactions were investigated only for one spatial configuration. It is possible that larger spatial interactions would occur for other spatial configurations.

For a fixed combination of input values, the fractions $r(x,y)$, $g(x,y)$, and $b(x,y)$ vary considerably between locations on the monitor. For a patch that measures 2.4 cm horizontally by 4 cm vertically, the fractions can vary as much as 8% across the patch. This variation can in principle be eliminated by using sophisticated software to modify the input values as a function of location. However, the software to implement this correction must be quite sophisticated to avoid introducing visible edges in the displayed patch. The human visual system is not very sensitive to smooth spatial variation in the color signal. Informal observation shows that a 2.4 cm by 4 cm patch with 8% variation in the fractions $r(x,y)$, $g(x,y)$, and $b(x,y)$ appears uniform. Suppose the patch is divided into 16 regions, each of which is corrected for the spatial variation of the monitor. The corrected patch will have a maximum variation of 2% between any two locations. However, it will not appear uniform. The changes in intensity at the boundaries of the 16 regions caused by the spatial correction will form visible edges, which are not generally desirable. In addition, for a frame buffer that allows only 256 different input combinations to be displayed simultaneously, it is not possible to choose a different combination of input values for every pixel. On the other hand, since a single scale factor provides a reasonable description of the spatial inhomogeneities of the monitor, the percentage change in color signal across a displayed edge will not vary much with the location of the edge. The percentage change in color signal determines the contrast of a stimulus, and for some applications contrast specification may be more important than absolute intensity specification.

For the same reason that it is not possible to measure the output of the monitor for all possible combinations of input values, it is not possible to test exhaustively a set of assumptions about monitor behavior: too many measurements are required. For example, the measurements described in this paper do not test the assumption of phosphor constancy with respect to variation in spatial location, nor do they measure the phosphor and spatial interactions at more than one location.

The assumptions about monitor performance and measurement procedures described in this paper are designed to calibrate the monitor for a particular application. Other applications may require different calibration procedures. For example, Lyons and Farrell[8] focus their calibration effort on specifying the fine spatial structure of displayed stimuli. The measurement results described in this article are for a single monitor. The performance of other monitors may not satisfy the same set of assumptions. Cowan and Rowell[1] have tested the assumptions of phosphor constancy and phosphor independence for a Tektronix 690SR monitor. They conclude that these assumptions are reasonably satisfied for that monitor. Cowan's[4,9] more general experience is that most of the assumptions described in this paper are violated for some monitors. As computer controlled color monitors are used increasingly in psychophysical labora-

tories, a consensus may develop about what sets of assumptions are appropriate for various applications and about what measurements are necessary to test these assumptions. To develop such a consensus, it will be necessary for different laboratories to report the results of their monitor calibrations.

Acknowledgment

This work was supported by grant 2 RO1 EY03164 from the National Eye Institute, and grant NCC 2-307 from the NASA-Ames Research Center. Some of the material in this article is based upon work supported under a National Science Foundation Graduate Fellowship. The opinions, findings, and conclusions of this paper do not necessarily reflect the views of the National Science Foundation.

I thank M. Perkins, A. Poirson, J. Ruby, B. Wandell, and two anonymous reviewers.

1. W. B. Cowan and N. Rowell, On the Gun Independence and Phosphor Constancy of Colour Video Monitors, *Color Res. Appl. Supplement*, **11,** S33–S38 (1986).
2. D. H. Brainard and B. A. Wandell, Spatial integration of chromatic information, *Investig. Ophthalm. Vis. Sci. Supplement*, **28,** 92 (1987).
3. D. H. Brainard and B. A. Wandell, Classification measurement of color appearance, *Investig. Ophthalm. Vis. Sci. Supplement* **29,** 162 (1988).
4. W. B. Cowan, "CIE Calibration of Video Monitors," *National Research Council of Canada Report*, Ottawa, 1985.
5. D. H. Brainard and B. A. Wandell, The Color Analysis Package, manuscript in preparation.
6. D. L. MacAdam, Visual sensitivities to color differences in daylight, *J. Opt. Soc. Am.*, **32,** 247–274 (1942).
7. W. R. J. Brown, The influence of luminance level on visual sensitivity to color differences, *J. Opt. Soc. Am.*, **41,** 684–688 (1951).
8. N. F. Lyons and J. E. Farrell, Linear systems analysis of CRT displays, Manuscript in preparation., 1988.
9. W. B. Cowan, Colorimetric properties of video monitors, *Notes for OSA Short Course*, 1987.

Received October 15, 1987; accepted April 30, 1988.