Marcel P. Lucassen
Jan Walraven

TNO Institute for Perception
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

Evaluation of a Simple Method for Color Monitor Recalibration

An algorithm for recalibrating a color monitor's RGB input–output relations is presented that requires only a single measurement of a properly chosen reference stimulus. For the application under concern, i.e., reproduction of 35 different colored patches that were used as stimuli for psychophysical experiments on color constancy, the reference stimulus was a white (D65) presented at a luminance corresponding to the mean of the test stimuli. Three sets of data were obtained for evaluating the algorithm's error reduction power for a given stimulus configuration. These relate to different ways in which the monitor can get out of calibration. That is, slow, but cumulative changes over time, fast changes due to gun interaction (resulting from changed stimulus conditions), and error introduced by a different setting of the monitor's brightness control. Additional experiments were performed to evaluate the effect of background intensity and color. The algorithm was found to be quite effective in dealing with the instantaneous changes (gun interaction, brightness control), and also for keeping track of the slow changes that may finally necessitate a full recalibration of the monitor.

Introduction

Computer-controlled CRTs are used for a wide range of applications, from displaying text to complex animated graphics. We use our color monitor as a stimulus generator for psychophysical studies on color constancy. Typical for this purpose is the need for a well-defined input-output calibration, e.g., the relation between the CRT's digital input (digital to analog converter value, DAC value) and the screen's light output (luminance) for each of the three R,G,B guns.

When a computer-controlled color monitor has been calibrated for a certain stimulus configuration, there is no guarantee that after a period of time, or after a change of configuration, the calibration is still valid. Depending on the application, display hardware, and photometric equipment, many adjustments may be needed to reach the desired accuracy for color reproduction.

Recently, several authors reported their findings from monitor calibration efforts.[1–5] Post and Calhoun[1,2] compared seven models for generating colors with specific CIE chromaticity coordinates and luminances on CRTs. They conclude that a piecewise linear interpolation method is most accurate, and found that 16 calibration points per gun are sufficient to reconstruct the input-output relation. However, their work does not solve the common problems of gun interaction and temporal instability. Brainard[3] focussed on finding a minimal set of assumptions that limit the number of measurement points for monitor calibration, including assumptions of spatial interaction.

A full monitor calibration can be very time consuming, so it is worthwhile to find out when recalibration really becomes necessary. For most applications, a "measure and adjust" algorithm as proposed by Post and Calhoun[1,2] may be used, but again, this involves a lot of measurements.

In this communication we report on the results obtained with a recalibration algorithm that reduces measurements to a minimum. We found that, for a given stimulus condition, a single measurement, e.g., the measurement of the average stimulus chromaticity (usually white) at an intermediate luminance level, may already result in an acceptable recalibration. Recalibration here means shifting the R,G,B input–output relations along the log luminance axis. The chromaticity coordinates of the monitor's phosphors are assumed to remain constant (as was also confirmed by measurement). In the following we shall present data that show both the need for continuous calibration and the efficacy of the method proposed.

CCC 0361-2317/90/060321-06$4.00

Method

Colorimetry

In principle, that is, assuming additive color mixing to apply, one only needs the input-output relations (luminance vs. *DAC* value) and the three phosphor chromaticity coordinates to calculate the *DAC* values for the red, green, and blue gun, required for producing specified *XYZ* (CIE 1931) tristimulus values. The colorimetric equation for deriving the monitor's luminance output is given by

$$\begin{pmatrix} \mathbf{R} \\ \mathbf{G} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} x_R/y_R & x_G/y_G & x_B/y_B \\ 1 & 1 & 1 \\ z_R/y_R & z_G/y_G & z_B/y_B \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \quad (1)$$

where x, y, and z are the 1931 CIE chromaticity coordinates with subscripts R,G,B referring to the appropriate phosphor. The assumption of phosphor constancy implies that the matrix in (1) has fixed elements. Note the conversion sign on the matrix in (1). The *DAC* values for the three guns are calculated by

$$\begin{pmatrix} \mathbf{DAC}_R \\ \mathbf{DAC}_G \\ \mathbf{DAC}_B \end{pmatrix} = \text{INTERPOLATION} \begin{pmatrix} \mathbf{R} \\ \mathbf{G} \\ \mathbf{B} \end{pmatrix} \quad (2)$$

where the INTERPOLATION operation stands for interpolating the input-output curve on a logarithmic scale. A smaller interpolation error results this way, because the logarithmic input-output curves show less curvature than the linear curves. Applying (2) after (1) will be referred to as "generating" colors, whereas applying the inverse of (1) after the inverse of (2) will be referred to as "analyzing" colors. Thus, "generating" involves transforming *XYZ* to *RGB* space, whereas "analyzing" implies the opposite transformation.

Measuring the Input-Output Relation

Before a recalibration algorithm can be used, the original set of RGB input-output relations must be known. The monitor we used was a high-resolution Hitachi 19-inch color monitor (1152 × 900 pixels, 24 bit/pixel), controlled by a Sun 3/260 computer. Measurements of the CRT's light output were performed with a SpectraScan PR-702AM (Photo Research) spectroradiometer and a Spectra Pritchard (Photo Research) photometer. The photometer was used for measuring at low luminance levels.

Following the practice recommended by Cowan[5] and Brainard,[3] the pattern we used for measuring the calibration curves (spatially) resembled the pattern that was used in the psychophysical experiments. Here, the calibration pattern displayed 35 square patches (70 × 70 pixels), arranged in a 5 × 7 array, on a black background. The patches' centers were separated by a (square) grid distance of 140 pixels. The luminance of the central patch, located at the screen's center, was measured with all 35 patches displayed in the same color. The *DAC* values were chosen so as to produce roughly equal luminance intervals on a logaritmic scale. Each R,G,B curve was measured while the other two guns were disconnected, to exclude their residual contributions.[6,7]

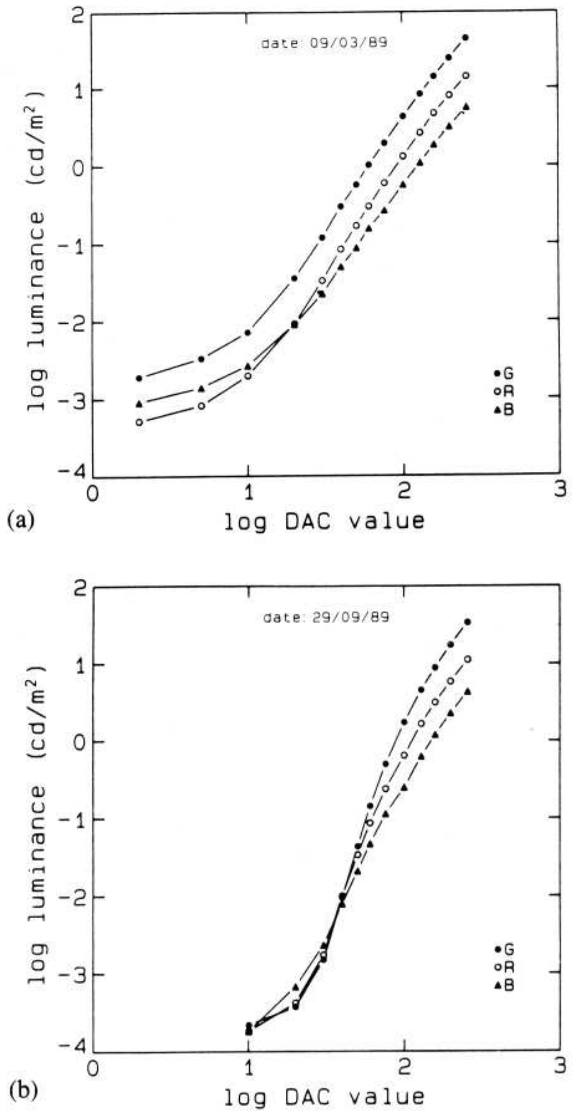

FIG. 1. (a) Luminance vs. *DAC*-value characteristics measured at installation date. (b) The same curves measured after about 6 months of display use.

Figure 1(a) shows the input-output relations, measured at the central patch, whereas Fig. 1(b) shows the same measurements six months later. Anticipating the results, to be discussed in the next section, it is clear that the monitor's calibration curves changed quite a bit over time (especially at the lower *DAC* values). This might be due to aging of the phosphors, although we found, confirming Brainard,[3] that their chromaticity coordinates had hardly changed. We initially measured, at the highest *DAC* values (255), the following set of (x,y) values for R, G, and B: (0.6312,0.3550), (0.3076,0.5957), (0.1473,0.0697), whereas 6 months later we obtained: (0.6326,0.3549), (0.3065,0.5984), (0.1459,0.0701). For sure, the two sets of curves in Fig. 1(a) and Fig. 1(b) are not related by a single scale factor and thus show the monitor's state to be complex over time.

Note that, on a log–log scale, the input-output relations show an almost linear relationship for the greater part of the *DAC* values that are used. This is the more or less expected result, considering the exponential relationship between gun voltage and beam current.

Apart from long term variations in screen luminance, also short term effects, like those following a stimulus change (gun interaction), may alter the input-output relations. These are the more day-to-day calibration problems that ask for a simple solution.

The Recalibration Algorithm

When colors are generated on a CRT screen, in a configuration that is quite different from the one used for calibrating the display, the screen voltage may not remain constant and thus affect the R,G,B beam currents. Other effects may have to be considered as well, but whatever the mechanisms involved, the net result is a change in the input-output relation. In other words, loading the *DAC* values calculated from (1) and (2) may not produce the desired luminances R, G, and B. The basic idea behind the recalibration algorithm is to compensate for such effects, in as far as they can be treated as gain changes in the *DAC*-to-luminance conversion. The adjustment consists of a vertical shift (offset) of the three input-output curves on the logarithmic scale, consistent with a scaling of the luminance (R,G,B). The adjustments are made on the basis of a single reference, i.e., an achromatic stimulus ($D65$) of medium luminance, presented in the center of the screen.

The recalibration procedure thus requires three steps:

1. Generate the white reference stimulus (x_0,y_0,Y_0) using (1) and (2), and determine the required phosphor luminances, R_0, G_0, and B_0.
2. Measure the reference stimulus (x,y,Y) which will probably deviate from its nominal values (x_0,y_0,Y_0), and calculate the required phosphor luminances, R,G, and B.
3. Calculate the correction factors C_R,C_G, and C_B, (using $C_R = R_0/R$ etc.) and correct the luminances R,G,B of the original input-output curves accordingly. That is, the original input-output relations have their outputs R,G, and B divided by the factors C_R,C_G, and C_B, respectively.

Evaluation

Constant-Configuration Case

The recalibration algorithm was evaluated in the course of psychophysical studies on color vision. Its main purpose was to correct for the gun interaction that occurred when changing from a dark background (as used for calibration) to the light backgrounds used for the stimulus pattern. In addition, the calibration provided information about the gradual change in the light output of the monitor. In order to test the precision of the recalibration, 20 colors were selected out of the 35 that build up the test grid. These 20 colors, located on two different loci of equal Munsell Chroma (see Fig. 2), were presented successively in the central patch at target luminance 10 cd/m^2 on a white (D65) background of 12 cd/m^2. From the remaining 15 colors, 10 colors were located on a third locus of equal Munsell Chroma (10 cd/m^2), whereas the other 5 were achromats in the luminance range 1–11 cd/m^2.

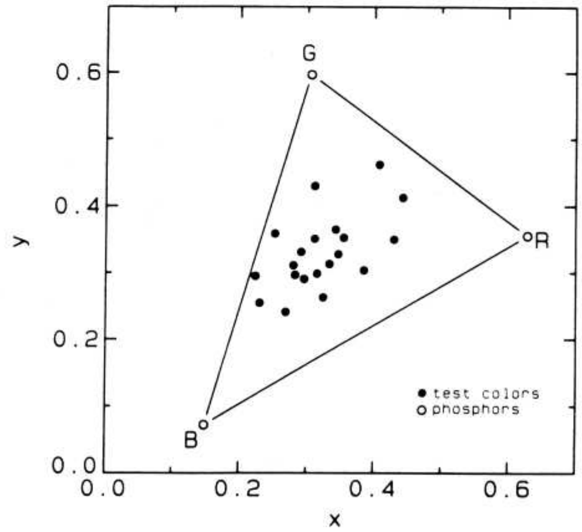

FIG. 2. Chromaticities (x,y) of the 20 test colors used for evaluating the recalibration algorithm. These colors, located on two loci of equal Munsell Chroma (value 5/), were presented at luminance 10 cd/m^2 in the central patch of the stimulus pattern.

The chromaticities (x,y) and luminances (Y) of the test colors were measured with the spectroradiometer, and then compared with their nominal values (x_0,y_0,Y_0). The chromatic error Δxy and percent luminance error $\%|\Delta Y|$ were calculated with

$$\Delta xy = [(x_0 - x)^2 + (y_0 - y)^2]^{1/2}, \quad (3)$$

$$\%|\Delta Y| = 100\,|Y_0 - Y|/Y_0. \quad (4)$$

Transformation into the uniform CIE 1976 $L^*u^*v^*$ color space enables expression of these errors in terms of a color difference ΔE^*:

$$\Delta E^* = [(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2]^{1/2}. \quad (5)$$

The errors were calculated for the set of test colors, when generated, respectively, with the original set of calibration functions or the set that resulted after applying the recalibration algorithm. The stimulus pattern for the reference measurement, as demanded by the algorithm, displayed the 35 patches at averaged chromaticity and luminance (D65, 10 cd/m²) on a white background. This background (D65, 12 cd/m²) is also the averaged background of the psychophysical experiments.

The recalibration algorithm was applied to three different sets of data. The first set (Set 1) relates to the situation where the same input-output curves are still used after a year's monitor use. It turned out, as shown already in Fig. 1, that over this period of time the gradual changes in the monitor had culminated in quite a drastic change of its input-output characteristics. The second set (Set 2) relates to the standard usage of the algorithm, that is, with up-to-date calibration curves, but not necessary applicable to the experimental condition in question (i.e., light background, rather than the dark background used during calibration). In the third set (Set 3) the data were generated in a condition where the brightness control of the monitor was deliberately changed. This is the kind of error that may be introduced when the monitor has different users.

The results obtained in the three test conditions are shown in Table I. What is shown is a comparison of the average error and standard deviation of the 20 test colors when using either the original or recalibrated (scaled) *RGB* input-output curves.

The results of Table I are plotted in Fig. 3. Note that the error reduction for data sets 2 and 3 is mainly in the luminance direction and that, exactly for that reason, the effect of recalibration is quite effective, reducing the error by 15% and 30%, respectively. The small change in chromatic error is reflected in roughly equal scale factors for *R,G,* and *B* (see Table I). The error reduction for the data of Set 1 is large in both the luminance and chromatic direction and the remaining errors cannot be neglected. Note (in Table I) that the scale factors are different now. This is the expected result in view of the change in shape of the input-output curves over a six-month period. Whether such errors are allowed depends on the application. Often, chromatic errors are compared with the size of a MacAdams ellipse, which provides an estimate of the minimum error due to the limitations of the visual system. On the basis of tabulated MacAdam ellipses,[8] we obtained a rough estimate of the average minimum error in the chromaticity space covered by the color monitor. Considering only the error in the direction of the major axis of the relevant ellipses (i.e., those located within the monitor's *RGB* space), we arrived at an average (Δxy) of 0.005. This means that, for data sets 2 and 3, the accuracy of color reproduction (obtained with interpolation of the input-output curve and the recalibration algorithm) can be in the order of a just perceptible chromaticity difference. The same conclusion can be drawn from the analysis of ΔE^* in Table I, since a just noticeable difference can be estimated to be of the order of 2 to 3 CIELUV units.[9] On the other hand, data set 1 shows that a full monitor recalibration is necessary.

Effect of Background Luminance and Color

So far we have considered only one change of stimulus configuration, that is, changing the background from dark to D65 at 12 cd/m². We also performed some additional experiments at a later time to evaluate the effects of changing the luminance or color of the background (grid). If such effects could be described by a simple relationship between scale factors and background parameters, this might possibly obviate the need for a reference stimulus for each new experimental condition. First, a full calibration, as described earlier, was performed because of our monitor's continuing decline. We then set out to measure the 20 test colors, presented as the central patch of the 35 patches, using different grid luminances. For each color, the *R,G,B* scale factors (C_R,C_G, and C_B) were calculated that would be required to exactly reproduce the nominal x_0,y_0,Y_0 values. The effect of the D65 background luminance on the obtained scale factors is shown in Fig. 4. At the time these measurements were made the monitor had still changed somewhat more, resulting in somewhat different scale factors from those shown in Table I. Figure 4 shows that the scale factors increase with increasing grid luminance, but only up to a value of about 12 cd/m². The overall pattern is regular enough to suggest a procedure for a more general recalibration algorithm, that could utilize (average) screen luminance for deriving the associated *R,G,B* scale factors.

To test whether the color of the background might also be an important variable, we compared the scale factors obtained for white light (D65 at 12 cd/m²) to those obtained for equi-luminant red ($x = 0.4150$; $y = 0.3300$) and green ($x = 0.3127$; $y = 0.4320$) backgrounds, respectively. These backgrounds introduce a change in the *R,G,B* luminance distributions in the direction of either the red or green gun, and might thus reveal a possible gun-specific effect of background on scale factor. One should observe, then, that the

TABLE I. Comparison of mean error and standard deviation of 20 test colors, either without RGB recalibration (scale factors 1) or with RGB recalibration (scale factor variable). The different data sets relate to different conditions as described in the text.

Data set	Scale factor			%|ΔY|		Δxy		ΔE*	
	R	G	B	Mean	sd	Mean	sd	Mean	sd
1	1.00	1.00	1.00	54.18	1.65	0.0265	0.0123	13.19	0.89
	1.95	2.45	2.16	31.37	5.13	0.0163	0.0094	9.18	2.90
2	1.00	1.00	1.00	15.04	0.84	0.0028	0.0019	3.09	0.29
	1.18	1.17	1.19	0.63	0.55	0.0031	0.0016	1.02	0.50
3	1.00	1.00	1.00	35.66	0.99	0.0106	0.0057	7.67	0.42
	1.53	1.54	1.52	5.15	1.21	0.0088	0.0046	3.58	1.84

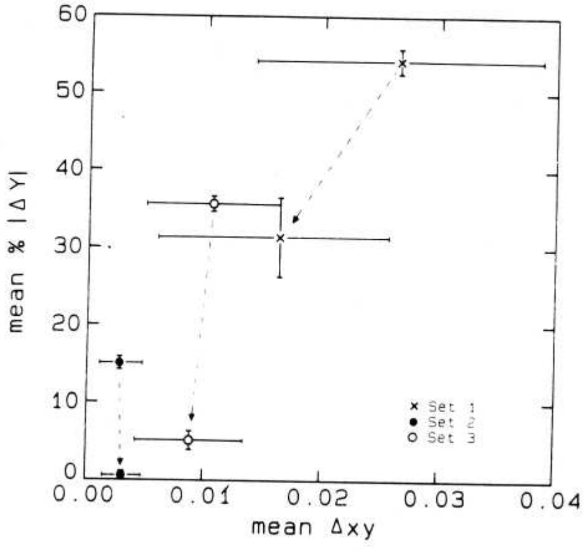

FIG. 3. Means and standard deviations for chromatic (Δxy) and luminance (ΔY) errors, measured with and without the recalibration algorithm (data from Table I). The dashed arrows indicate the error reduction due to the recalibration algorithm.

red scale factor (C_R) is more affected by the red than the green background, and vice versa.

The results we obtained with the two colored backgrounds were all in the direction of a reduction of the R,G,B scale factors relative to those obtained in the condition with a white background. That is, for the red background: 2.9% for C_R, 2.5% for C_G, and 1.8% for C_B. For the green background the reductions are: 1.9% for C_R, 3.3% for C_G, and 2.7% for C_B. These results indicate a (small) gun-specific effect (the largest reductions of C_R and C_G are found for the correspondingly colored backgrounds), but since the effect of color is small anyway (here 2–3%), a gun-specific parameterization of the CRT image does not seem very profitable. The main effect (see Fig. 4) is due to the luminance step, irrespective of the color of the background.

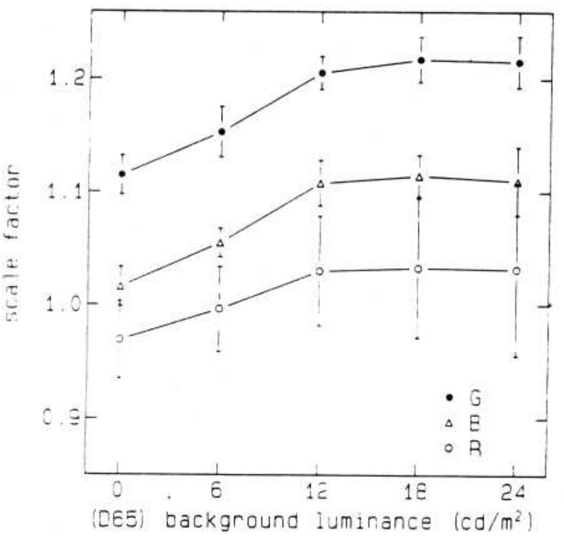

FIG. 4. Calculated R,G,B scale factors (mean values ± standard deviation) as a function of the D65 background (grid) luminance. The same procedure and the same 20 test colors (see Fig. 2) were used.

Discussion

The simple recalibration algorithm we proposed turned out to be well suited for the purpose it was developed for, that is, compensating for nonadditivity of the (separately measured) color guns. In general, this method is only suitable for correcting errors that can be described in terms of vertical translation of the log *RGB* vs. *DAC* value functions. It is of interest though, that our results show that this is the kind of error that is likely to be encountered on a CRT display.

The results from our experiments in which we varied the luminance and color of the background are too limited to allow firm conclusions. Still, they indicate that the R,G,B scale factors vary with the overall rather than the gun-specific display luminance. This might indicate an interaction between screen-voltage and beam current, that is fairly insensitive to the particular ratio of the constituent R,G,B beam currents. This should make it easier to adapt the recalibration algorithm for use in conditions with varying stimulus conditions.

If, in the course of time, the algorithm shows error reduction to be less complete, this is a warning signal. Values from 1.1 to 1.2 are normally found, but when the scale factors become too large a full monitor recalibration is needed. This is illustrated by the data of Set 1, which relate to the condition where the shape of the input-output curves had changed with time. So, regularly checking the scale factors is also effective to discover slow drifts in the monitor's output.

The fact that the scale factors are greater than 1 means that the measured output is less than would be expected from the calculations. Several factors (e.g., phosphor aging, gun interaction) may contribute to this loss in effective output, but these are nevertheless handled by the simple scaling procedure of the recalibration algorithm. This is particularly helpful when different stimulus configurations, requiring different correction factors, have to be displayed. The fact that a single measurement (of the reference white) was found to be sufficient for the recalibration procedure does not necessarily apply to all stimulus conditions. However, if it does, as can be tested in the way we have shown, much time and effort can be saved in maintaining accurate stimulus control in complex stimulus scenarios. Moreover, measuring just a single white point on the screen could be done with a simple (but reliable) chromaticity meter, which is much less expensive and cumbersome than using the spectroradiometer that would be needed for measuring colored stimuli.

1. D. L. Post and C. S. Calhoun, An evaluation of methods for producing specific colors on CRTs, *Proceedings of the Human Factors Society 31st Annual Meeting*, Human Factors Society, Santa Monica, CA, 1987, pp. 1276–1280.
2. D. L. Post and C. S. Calhoun, An evaluation of methods for producing desired colors on CRT monitors, *Color Res. Appl.*, **14,** 172–186 (1989).
3. D. H. Brainard, Calibration of a computer controlled color monitor, *Color Res. Appl.*, **14,** 23–34 (1989).
4. W. B. Cowan, An inexpensive scheme for calibration of a colour monitor in terms of CIE standard coordinates, *Computer Graphics*, **17,** 315–321 (1983).
5. W. B. Cowan, CIE calibration of video monitors (available from National Research Council of Canada, Ottawa, Ontario, K1A 0R6, Canada), 1986.
6. P. McManus and G. Hoffman, A method for matching hard copy color to display color, *SID 85 Digest*, 204 (1985).
7. J. Walraven, Colour coding for ATC displays, ECC Report No. 112, Eurocontrol, Brussels, 1988.
8. G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed., John Wiley & Sons, New York, 1982.
9. W. N. Sproson, *Colour Science in Television and Display Systems*, Hilger, Bristol, England, 1983.

Received: February 8, 1990; accepted: May 1, 1990.