An Experimental Comparison of RGB, YIQ, LAB, HSV, and Opponent Color Models

MICHAEL W. SCHWARZ
University of Waterloo
WILLIAM B. COWAN
National Research Council of Canada
and
JOHN C. BEATTY
University of Waterloo

The increasing availability of affordable color raster graphics displays has made it important to develop a better understanding of how color can be used effectively in an interactive environment. Most contemporary graphics displays offer a choice of some 16 million colors; the user's problem is to find the *right* color.

Folklore has it that the RGB color space arising naturally from color display hardware is user-hostile and that other color models such as the HSV scheme are preferable. Until now there has been virtually no experimental evidence addressing this point.

We describe a color matching experiment in which subjects used one of two tablet-based input techniques, interfaced through one of five color models, to interactively match target colors displayed on a CRT.

The data collected show small but significant differences between models in the ability of subjects to match the five target colors used in this experiment. Subjects using the RGB color model matched quickly but inaccurately compared with those using the other models. The largest speed difference occurred during the early *convergence phase* of matching. Users of the HSV color model were the slowest in this experiment, both during the convergence phase and in total time to match, but were relatively accurate. There was less variation in performance during the second *refinement phase* of a match than during the convergence phase.

Two-dimensional use of the tablet resulted in faster but less accurate performance than did strictly one-dimensional usage.

Significant learning occurred for users of the Opponent, YIQ, LAB, and HSV color models, and not for users of the RGB color model.

Categories and Subject Descriptors: I.3.6 [**Computer Graphics**]: Methodologies and Techniques—*ergonomics*; *interaction techniques*

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Color matching, color models, HSV, LAB, Opponent colors, RGB, YIQ

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Province of Ontario, Digital Equipment Corporation, the University of Waterloo, and Schlumberger-Doll.

Authors' current addresses: M. W. Schwarz, Bell Northern Research, Box 3511, Station C, Ottawa, Ont., Canada; W. B. Cowan, Division of Physics, National Research Council of Canada, Ottawa, Ont., Canada K1A 0R6; and J. C. Beatty, Computer Graphics Laboratory, Department of Computer Science, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

1987 ACM 0730-0301/87/0400-0123 $00.00

1. INTRODUCTION

A cathode ray tube (CRT) is the most commonly used device for displaying computer-generated images. Color is created by taking advantage of the fundamental trichromacy of the human eye: three differently colored images are presented which combine additively at photoreceptors in the eye to form a single perceived image whose color is a combination of the three displayed. These three images are themselves created by means of three electron guns acting separately on three differently colored phosphors; using shadowmask technology it is possible to make the images intermingle on the screen, causing the colors to mix together because of spatial proximity. The color of the phosphors determines how complete a range of color there is in the image. Well-saturated red, green, and blue colors are typically used to give the most complete color range. Thus at any point in an image the color is determined by the intensity of light at that point in each of the red, green, and blue images. These three intensities are controlled by the voltages applied to the three electron guns, which are in turn controlled by three values supplied by the application program.

Since the hardware demands three digital values, specifying an intensity for each of red, green, and blue, it is easiest to allow a user to specify these values directly. But as Smith [15] and Joblove and Greenberg [11] have pointed out, we customarily select or specify color in other ways, using other color specification systems. It therefore seems plausible that one or another of these more traditional color models may be easier to learn and to use.

Indeed, folklore argues that this is true. Unfortunately, there is little experimental evidence to support the contention that an RGB color space is difficult to use or to suggest which of the many alternative color spaces available is preferable.

An interesting experimental comparison of the RGB color model with Smith's HLS (hue, lightness, and saturation) model and a proposed color naming system is described by Berk et al. [2, 3]. The task was to name one of a discrete set of target colors, either by specifying RGB or HSL coordinates as numbers in the range [0.0, 1.0] or by giving an English-language color name such as "medium strong yellow" according to a formal syntax encompassing 627 distinct color names. Subjects were most accurate with the proposed natural-language naming system, followed by HSL coordinates, followed by RGB coordinates. While this study is suggestive, it is oriented toward the specification of colors in a programming context that is able to handle only a small number of colors; we are interested in a different problem, namely, what color model and input technique best enables a user to explore the large and essentially continuous color space provided by contemporary color CRTs in the context of a highly interactive system with immediate visual feedback.

Our approach is based on *color matching*, an experimental technique with a long and fruitful history in studying the psychophysics of vision (though not previously using computer-generated imagery). We displayed target and controlled color patches one above the other on a color monitor. Subjects were able to alter the color coordinates of the controlled patch using a digitizing tablet and in doing so were asked to cause it to match the target color. By comparing the accuracy and speed with which subjects generated a match for different color

Table I. Chromaticity Coordinates and Gammas of Phosphors in Electrohome ECM 1301 Color Monitor Used in This Experiment

Phosphor	xx	y	z	γ
Red	0.620	0.330	0.050	2.43
Green	0.210	0.675	0.115	2.35
Blue	0.150	0.060	0.790	2.54

models, using various input techniques, we hope to discover which are easiest to learn and to use. By frequently recording the color of the controlled patch we made it possible to study the process by which a match is generated.

In succeeding sections we describe the physical environment, the color models used, the input techniques used, experimental procedures, and the results obtained. Further details may be found in Schwarz [14].

2. THE ENVIRONMENT

The experiment was performed using software implemented in C, running in a UNIX environment on a lightly loaded VAX 11/780, with an Adage/Ikonas RDS-3000 raster display system for output. The matching software was run at higher than normal priority to ensure immediate response. The raster display provides a 512 by 512 image via color lookup tables generating 10-bit output (per primary) on an Electrohome ECM 1301 color monitor. The brightness and contrast controls of the monitor were set midway between their minimum and maximum levels (the default positions). Phosphor gammas and chromaticities (see Table I) were determined using procedures described by Cowan [6] so that colors could be accurately displayed and recorded.

Subjects manipulated the controlled color by means of a Summagraphics Bit Pad. One tablet and four-button puck. The puck buttons were covered with white stickers and referred to by position (left, right, top, and bottom), so that subjects would not confuse button colors with the color properties being controlled.

Instructions were displayed on an Ann Arbor Genie alphanumeric terminal having a neutral white phosphor. The contrast control of the terminal was set to its minimum level.

The experiment was conducted in a darkened 8 by 13 ft room, the layout of which is shown in Figure 1. The setup shown is for right-handed subjects. A symmetric arrangement was used for left-handed subjects.

The color monitor was positioned at eye level, approximately 2.5 m in front of the subject. This distance was sufficient to prevent subjects from resolving individual scan lines on the display. The target and controlled colors appeared in the form of two rectangles. These were positioned one above the other in the center of the monitor, separated by a distance of 0.5 mm (two scan lines). Each rectangle was 4.0 cm high and 5.0 cm wide, subtending approximately 0.9 degree vertically and 1.2 degrees horizontally. Thus when the subject fixated on the matching stimuli, most of the area of the two rectangles was focused on the fovea, that portion of the retina in which visual perception is most acute and in which the discrimination of colors is greatest [18].

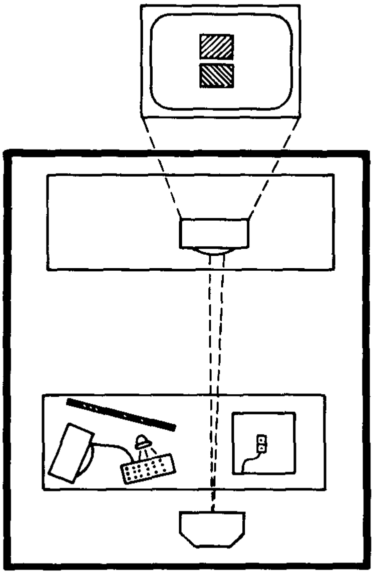

Fig. 1. Layout of the room in which the experiment was conducted.

The tablet and puck, alphanumeric terminal, keyboard, a small desk lamp, and a partition were positioned on a table in front of the subject. To minimize distractions, the terminal was placed in the periphery of the subject's field of view on the side opposite the tablet. The partition was placed to avoid reflections of the lamp and terminal in the color monitor, and for similar reasons parts of the wall, ceiling, and tables were covered with black felt.

3. COLOR MODELS

There are quite a number of color models available—more than are reasonable to include in one study—so we selected a representative few. These are discussed individually below.

For all these color models or *color spaces*, coordinates were translated into three voltage values in order to control the display. This process (explained individually for each of the color models that follow) is shown schematically in Figure 2, which summarizes the sequence of transformations used for each color model.

3.1 RGB

An RGB color space was included for several reasons. Because it corresponds directly to the hardware, it is the easiest for a programmer to implement and is in wide use. Folklore asserts that it is "user hostile," and we were curious as to whether this claim could be supported experimentally.

Gamma correction was performed so as to obtain a linear relationship between digital RGB values and the intensity of light emitted by the CRT.

3.2 YIQ

YIQ space has been in use as a television broadcast standard since its adoption by the National Television Standards Committee (NTSC) of the United States

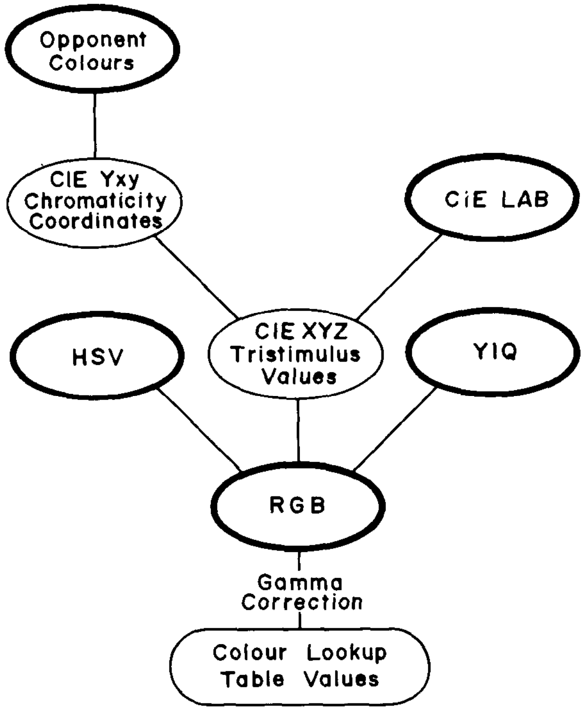

Fig. 2. The steps by which coordinates in each color model were transformed into the red, green, and blue intensities needed to generate a color on the monitor.

for broadcast television in 1953 [9]. Its primary objective was to maximize the perceptual resolution of the encoded color information using the fixed amount of bandwidth available in a broadcast signal in such a way as to be compatible with black and white transmission. One axis, denoted by Y, was arranged to be an approximation of luminance. The remaining two axes were selected so as to allot as little bandwidth as possible to chrominance information that the human eye cannot resolve. The result is an I axis encoding chrominance information along a blue-green to orange vector, and a Q axis encoding chrominance information along a yellow-green to magenta vector. Neither corresponds to a known psychophysical quantity.

The conversion from YIQ coordinates to RGB coordinates is defined by the following equation [11]:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1.00 & 0.95 & 0.62 \\ 1.00 & -0.28 & -0.64 \\ 1.00 & -1.11 & 1.73 \end{pmatrix} \cdot \begin{pmatrix} Y \\ I \\ Q \end{pmatrix}.$$

This transformation was defined by the Commission Internationale de L'Eclairage (CIE) assuming a standard set of R, G, and B phosphor chromaticities balanced so that $R = G = B$ yields gray. In principle it is necessary to insert another linear transformation to account for differences between the phosphors on our monitor and the standard NTSC phosphors. It is customary to ignore such differences, however, presumably on the grounds that the human eye adapts to the color gamuts of different phosphor sets. We conformed to this usage, transforming YIQ values directly into RGB values. Note also that the RGB space into which we transform is subsequently linearized by gamma correction.

3.3 LAB

The CIELAB color space is an international standard designed for perceptual uniformity [5, 18]. The intention is that color differences a human perceives as equal correspond to equal Euclidean distances in CIELAB space. While CIELAB space only approaches this very difficult objective, it is one of the two coordinate systems recommended by the CIE for comparing color differences, the other being CIELUV [5, 18]. We selected the CIELAB model from among the various CIE standards because of its perceptual uniformity; we chose CIELAB over CIELUV somewhat arbitrarily because of folklore suggesting that CIELAB is more appropriate over the small color differences encountered near the end of a match. Note that although the color model is called CIELAB, its coordinates are denoted by L^*, a^*, and b^*.

The CIELAB color model is defined in terms of XYZ tristimulus values derived from the 1931 CIE color matching functions. The reverse transformation, which takes L^*, a^*, and b^* to X, Y, and Z, is given by

$$\begin{aligned}
\bar{Y} &= \frac{L^* + 16}{116}, \\
X &= X_n\left(\frac{a^*}{500} + \bar{Y}\right)^3, \\
Y &= Y_n \bar{Y}^3, \\
Z &= Z_n\left(-\frac{b^*}{200} + \bar{Y}\right)^3.
\end{aligned} \tag{1}$$

X_n, Y_n, and Z_n are the tristimulus values corresponding to "the illuminant," which the eye is assumed to reference as "white." There is little detailed knowledge of how one ought to deal with the adaptive state of the visual system when it is confronted by a video monitor; we consequently took these values to be unity, which corresponds to assuming that X_n, Y_n, and Z_n specify an equal energy white.

It remains to transform the XYZ tristimulus values into voltages for the monitor. This was done following the method given by Cowan [6].

We make no attempt to recapitulate the evidence on which the CIELAB color model is taken to produce an approximately uniform color space. The interested reader is referred by Wyszecki and Stiles [18]. Of the CIELAB coordinates, L^* is a measure of lightness; a^* changes the red/green balance; and b^* changes the green/blue balance. Neither a^* nor b^* corresponds to known psychophysical properties of visual perception.

An approximate measure of the relative perceptual distance between various colors is given by CIELAB color difference formula [5], which is simply the

Euclidean distance between two colors specified in CIELAB coordinates. We refer to such distances as being measured in color distance units (cdu's). Because perceived color difference depends on viewing conditions, however, one should be cautious about inferring the absolute perceptual magnitude of color differences measured in cdu's.

3.4 HSV

The HSV color model, introduced by Smith [15], approximates the perceptual properties of "hue," "saturation," and "value." Hue and saturation are taken from common speech about color, while the term *value* was introduced by Munsell in 1939 [12], although it was defined differently. The concept of value as a perceptually uniform quantity akin to brightness was created by Munsell. Roughly speaking:

(1) *hue* associates a color with some position in the color spectrum—red, green, and yellow are hue names;
(2) *saturation* describes the "vividness" of a color, pure spectral colors being "fully saturated colors" and grays being "desaturated colors";
(3) *value* corresponds to the "lightness" of a color.

A hue-saturation slice of HSV space is derived by projecting the surface of an RGB color cube onto the $R + G + B = 1$ plane: the saturation and hue of a point on the projection are its polar cooordinates r and θ with respect to the center of the projected surface, while the value V of all the points on the projection is simply the length of the diagonal of the color cube projected. (See Smith [15] or [Foley 18] for a more detailed description.) The transformation from HSV coordinates to RGB coordinates is given by the following pseudocode:

```
F  = H * 6 - floor(H * 6);
T1 = V * (1.0 - S);
T2 = V * (1.0 - (S * F));
T3 = V * (1.0 - (s * (1.0 - F)));

case floor(H*6) mod 6 of
     0: (R, G, B) = (V, T3, T1);
     1: (R, G, B) = (T2, V, T1);
     2: (R, G, B) = (T1, V, T3);
     3: (R, G, B) = (T1, T2, V);
     4: (R, G, B) = (T3, T1, V);
     5: (R, G, B) = (V, T1, T2);
end;
```

While these computational definitions of the terms hue, saturation, and value are plausible, it is also important to realize that perceptually they are only approximations. In particular, the H and S parameters of the HSV model vary qualitatively from the psychological quantities referred to as hue and saturation. (See Figure 3, for example.) V also differs from the definition of value in the Munsell color model.

Smith [15] discusses both the HSV and HSL color spaces, advocating the former on grounds of computational efficiency. We chose an HSV model from

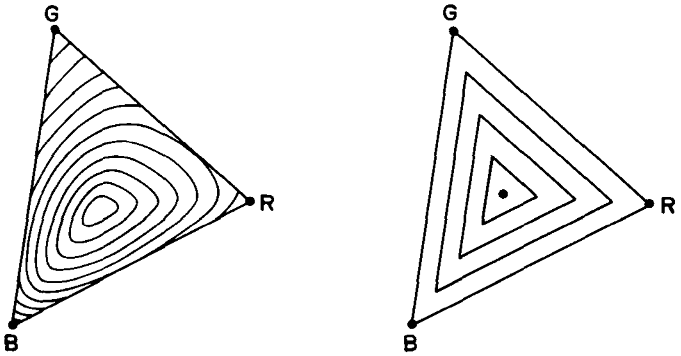

Fig. 3. The two triangular regions represent in a chromaticity diagram the gamut of colors that are obtainable using an RGB monitor. The chromaticities of the phosphors involved are the vertices of the triangles, and points within are colors that can be obtained by the combined excitation of the three phosphors. The roughly elliptical contours on the left represent psychologically determined "equal saturation" contours. The triangular contours on the right trace equal S contours in HSV space. (Adapted from [17].)

among such color spaces because of its prominence in the computer graphics literature.

3.5 Opponent Colors

According to the standard theory of color vision [4, 10], there are three cone types on the retina. The three signals which they produce in response to light stimulation are transformed as shown in Figure 4 before being passed on to the brain. In particular, the eye creates an *achromatic* channel by summing excitations from red and green cones; a *red/green* channel by differencing data from the red and green cones; and a *yellow/blue* channel by differencing data from the luminance channel (yellow = red + green) and the blue cones. This processing explains a number of visual phenomena, such as why we are able to perceive reddish-blue colors but not reddish-green colors.

A color space based on the Opponent Colors model was included in this study to see how such a physiologically based space would compare with the other models. Since the precise details of the Opponent Colors transformations are not known at present, we were not able to produce an exact Opponent Colors representation. The best we can do is to produce a representative space that has the qualitative properties of the Opponent Colors theory: three orthogonal axes for the achromatic, red/green, and yellow/blue channels. We derived such a space in the following way. Starting from the CIE tristimulus values X, Y, and Z, we transformed to chromaticity coordinates, using the standard transformation

$$x = X/(X + Y + Z)$$
$$y = Y/(X + Y + Z).$$

These two coordinates, along with the luminance Y, suffice to determine any color. It is a property of the xy plane that colors are laid out as a hue circuit,

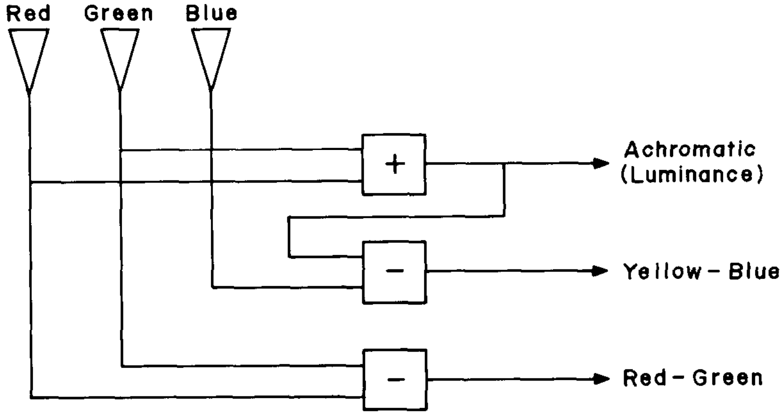

Fig. 4. A schematic representation of the Opponent Colors model of visual perception. The triangles represent data from the three cone types in the retina. The rectangles represent processing of the raw data performed by cells receiving input from many cones in a small region of the retina. "+" represents a summation and "−" a differencing of the inputs shown.

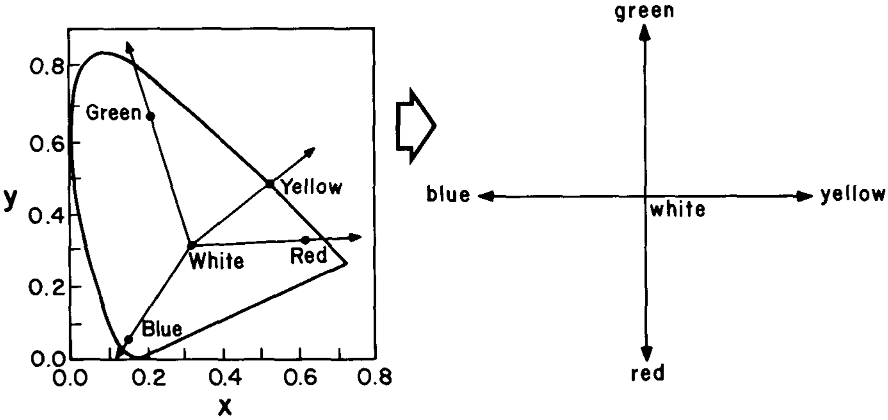

Fig. 5. The derivation of red/green and yellow/blue chromatic axes for the Opponent color model.

with white at the center and more saturated colors further away from white. Traversing a circle surrounding white yields a spectrum, with magenta (violet) falling beside red to complete the circuit.

To get an opponentlike space we used the luminance axis Y as is. Our definition of the chromatic channels is illustrated in Figure 5. First we draw vectors from the "white point" at (0.31, 0.32) to points representing the hues red, green, blue, and yellow (defined in Table II). For simplicity we chose the colors of the three phosphors in the ECM 1301 monitor as red, green, and blue. For yellow we used the most fully saturated yellow having a hue equivalent of 580 nm that we could obtain on our monitor, 580 nm being the spectral color classically identified by

Table II. Chromaticity Coordinates of Reference Points Used to Define Opponent Color Space[a]

Reference color	Chromaticity coordinates		
	x	y	z
Red	0.62	0.33	0.05
Green	0.21	0.68	0.11
Blue	0.15	0.06	0.79
Yellow	0.51	0.49	0.00
White[a]	0.31	0.32	0.37

[a] The white used is the standard CIE D65 white specified for most color TVs.

subjects as "unique yellow" [4]. A mapping is then established between these vectors and the usual right-angled coordinate system shown on the right in Figure 5. The following pseudocode implements this mapping:

```
{ Convert (LUM, YB, RG) to (X, Y, Z) tristimulus values.  }
{ (white_x, white_y) locate the white point in the        }
{   chromaticity diagram of Figure 5.                      }
{ Θ_R is the angle between the white-yellow vector and     }
{   the +horizontal axis, measured counterclockwise;       }
{   similarly for Θ_Y, Θ_G, and Θ_B.                        }
{ Θ_RY is the angle between the red and yellow reference   }
{   points; similarly for Θ_YG, Θ_GB, and Θ_BR.              }
if YB ≥ 0.0 and RG ≥ 0.0 then   { Yellow-Green Quadrant }
     Θ_OPP = arctan(RG/YB);
     Θ_Yxy = Θ_Y + Θ_OPP/(π/2) * Θ_YG;
if YB < 0.0 and RG ≥ 0.0 then   { Green-Blue Quadrant }
     Θ_OPP = arctan(-YB/RG);
     Θ_Yxy = Θ_G + Θ_OPP/(π/2) * Θ_GB;
if YB < 0.0 and RG < 0.0 then   { Blue-Red Quadrant }
     Θ_OPP = arctan(RG/YB);
     Θ_Yxy = Θ_B + Θ_OPP/(π/2) * Θ_BR;
if YB ≥ 0.0 and RG < 0.0 then   { Red-Yellow Quadrant }
     Θ_OPP = arctan(YB/-RG);
     Θ_Yxy = Θ_R + Θ_OPP/(π/2) * Θ_RY;
{  Compute chromaticity coordinates for the color specified. }
R = sqrt(RG * RG + YB * YB);
x = white_x + R * cos(Θ_Yxy);
y = white_y + R * sin(Θ_Yxy);
z = 1.0 - x - y;
{  Compute tristimulus values for the color specified.    }
X = x * LUM/y;
Y = LUM;
Z = z * LUM/y;
```

This is a crude but qualitatively correct approximation to the opponent channels of the visual system. It is not clear whether it is profitable to attempt a better approximation given what is currently known about the visual system.

4. INPUT TECHNIQUES

There are an immense number of ways in which a graphical input device could be used to provide the three scalar values needed to define a color using any model. We chose to use a tablet and puck, primarily because a tablet is one of the most flexible and common input devices [7]. In order to obtain sufficiently fine-grained control over input values, tablet data were read in "mousemode"; that is, relative movement of the puck caused the controlled coordinate or coordinates to change.

Next we had to decide how to handle scaling and clipping of the color coordinates as they were transmitted from the tablet. We decided to scale tablet input so that the maximum range that any color coordinate could take corresponded to about twice the 11-inch tablet width. Note that for color models other than RGB and HSV, the range of values that a coordinate can actually take depends on the values of the other two coordinates. For each such coordinate we determined values of the other two for which it had the largest range, and thereafter scaled the coordinate in question so that this maximal range was transformed to twice the width of the tablet.

Then, to handle the case in which tablet motions would have taken the coordinate outside the gamut of colors available on the monitor, we clipped tablet data so that whenever a tablet motion took a coordinate outside its displayable range we left all coordinates unchanged. It is clear that when inexperienced users are involved the form of the clipping algorithm will significantly affect how they are able to perform matching. Consequently, we placed the target colors well away from the boundaries of the monitor color gamut to minimize boundary encounters. The only exception to this occurred with the HSV color model at the very beginning of each trial: since the starting color (a gray) had zero saturation, a subject who began by attempting to reduce the saturation immediately experienced clipping, and altering hue had no effect until saturation had been increased.

Puck buttons were depressed to indicate which coordinate or coordinates were to be altered. Two mapping schemes were employed (see Figure 6). In the first a separate puck button was used for each axis, and only horizontal movement was significant—the 3*1d tablet technique. In the second scheme, depressing one button allowed two coordinates to be altered simultaneously by combined horizontal and vertical motion; depressing an alternative button caused changes in the third coordinate according to horizontal movement of the puck—the 1d + 2d tablet technique.

Not all mappings of two-dimensional (2d) and one-dimensional (1d) control to various combinations of the axes in the color models were evaluated. The mappings chosen are given in Table III. The axes of the color systems are given in Table IV.

For all five color models we included (1) a mapping in which each axis was controlled separately when an appropriate button was depressed by horizontal puck movement; (2) a mapping in which the luminance axis (if present) was controlled separately by one puck button and the remaining two axes were controlled simultaneously by horizontal and vertical movement, respectively, when a second button was depressed. These mappings were considered to be

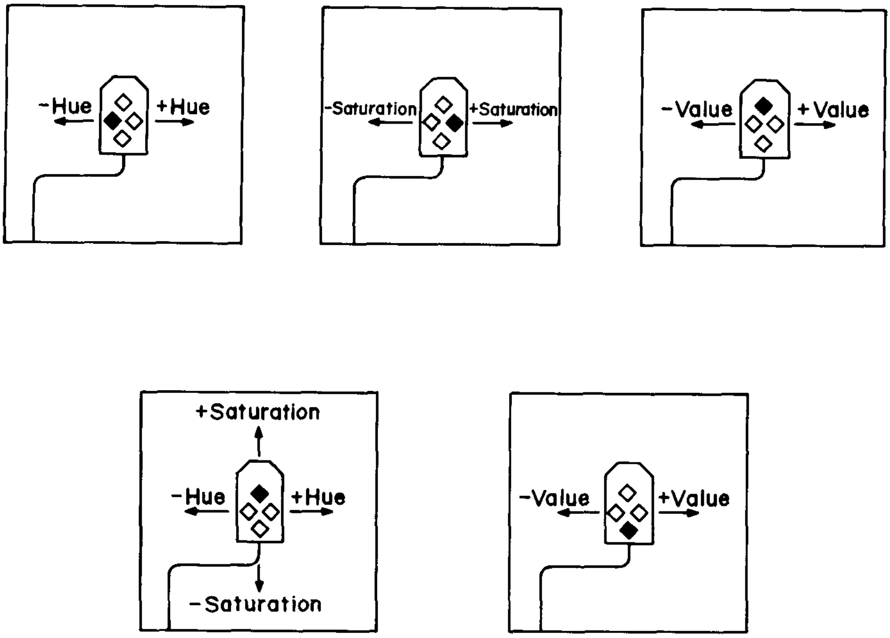

Fig. 6. The top row illustrates the scheme in which a separate puck button is used to control each axis of a color model. The bottom row illustrates the scheme in which one button is depressed so that movement of the puck controls two axes simultaneously, while a second button is depressed if horizontal movement of the puck is to control the third coordinate.

Table III. Mappings of Control Chosen

Name	Axis 1	Axis 2	Axis 3
Rgb1	L[a], H[b]	R[f], H	T, H
Yiq1	T[c], H	L, H	R, H
Hsv1	L, H	R, H	T, H
Lab1	T, H	L, H	R, H
Opp1	T, H	R, H	L, H
Rgb2	T, H	T, V	B, H
Yiq2	B[d], H	T, H	T, V
Hsv2	T, H	T, V	B, H
Lab2	B, H	T, H	T, V
Opp2	B, H	T, V	T, H
Rgb3	T, H	B, H	T, V
Opp3	T, V[e]	T, H	B, H

[a] L indicates left puck button.
[b] H indicates that horizontal movement when the corresponding button is depressed causes a change of the indicated coordinate.
[c] T indicates top puck button.
[d] B indicates bottom puck button.
[e] V indicates a vertical movement.
[f] R indicates right puck button.

Table IV. Axes of the Color Models

Name	Axis 1	Axis 2	Axis 3
RGB	R	G	B
YIQ	Y	I	Q
HSV	H	S	V
LAB	L^*	a^*	b^*
Opponent	achromatic	red/green	yellow/blue

intuitive because lightness variation is often separated from "color" variation in ordinary speech. Note that the RGB color model does not have an intuitive mapping by this criterion. However, we considered it intuitive to control red and green simultaneously because the blue phosphor of our monitor emits only short wavelength light and thus contributes a negligible amount to luminance.

In the intuitive mapping for HSV, H and S are not polar coordinates, as might be expected: our intuitive mappings of the CIELAB and opponent color spaces in fact lay out a hue circuit in the "color" plane equivalent to a polar coordinate version of HSV. That is, hues are laid out circularly on the tablet when a^* and b^* are being controlled simultaneously, as they are when the red/green and yellow/blue opponent axes are being controlled simultaneously. An explicitly polar mapping was therefore felt to be redundant.

The names introduced in Tables III and IV are used below to label the data collected. We refer to the combination of a color model and a tablet technique as a *system*. Thus Opp2 is the system in which a 1d + 2d tablet technique is mapped "in the intuitive way" to the axes of our Opponent color model.

5. EXPERIMENTAL PROCEDURES

Ninety-six "inexperienced" subjects took part in the experiment. These were students at the University of Waterloo; none had significant experience using a tablet or with color mixing techniques. Data were collected from several experienced subjects as well, although the only data of that sort reported here was generated by an expert on color technology with over 30 years of experience matching colors using the RGB color model. Unless otherwise indicated, subsequent discussion refers to the inexperienced subjects.

Each subject was required to pass a color vision test (the Standard Pseudoisochromatic Plates test, conducted under fluorescent illumination). Subjects were unaccompanied throughout the experiment, instructions being given on the terminal in the experiment room. These instructions are listed in Schwarz [14]. Points of particular interest are as follows.

(1) Subjects went through a brief terminal-driven training exercise to familiarize them with the particular color model and tablet technique to be used. The instructions were worded so as not to rely on a subject's understanding of any color terms other than "brightness" or "redness" (for RGB systems). Instead, the instructions introduced them to the use of a tablet and guided them axis by axis through the color model in question, periodically suggesting that they move the puck in particular ways and observe the effect. There were 14 instructional pages,

Table V. Initial Color of Controlled Rectangle (Gray) and Five Target Colors Used

	CIE tristimulus values			Chromaticities	
Descriptive name	X	Y	Z	x	y
Gray	0.084	0.085	0.078	0.35	0.35
Red	0.279	0.162	0.051	0.58	0.33
Green	0.126	0.306	0.116	0.23	0.57
Blue	0.091	0.055	0.335	0.19	0.11
Yellow	0.240	0.282	0.072	0.41	0.48
White	0.197	0.189	0.156	0.36	0.36

Note. The "descriptive names" red, green, blue, and yellow are used because they best describe the dominant hue of these colors, not because these are pure red, green, blue or yellow, which indeed they are not.

Slice 1	1. Red	2. Green	3. Blue	4. Yellow	5. White
Slice 2	6. Blue	7. Red	8. Green	9. Yellow	10. White
Slice 3	11. Blue	12. Green	13. Yellow	14. White	15. Red
Slice 4	16. Blue	17. White	18. Green	19. Yellow	20. Red
Slice 5	21. Green	22. White	23. Red	24. Green	25. Yellow
Slice 6	26. Red	27. Blue	28. White	29. Blue	30. Yellow

Fig. 7. The order in which the colors in Table V were matched. The term *slice*, defined here, is used in Section 6.5.

and subjects could leaf forward and backward through them during the instruction phase.

(2) With respect to the experiment itself, subjects were told to

> manipulate the color of the bottom square until it is the same as the color of the top square.
> Note that the two colors should be matched as closely as possible. This means that you should continue in your attempt to match the colors until you think they are the same or until it becomes EXTREMELY difficult to get the colors any closer to each other. Each match should take you about 1 to 2 minutes to complete. You will actually be allowed 3 minutes for each match.

Each subject matched five colors six times each for a total of 30 trials. All 30 were performed with the same system. The five colors matched are defined in Table V, and the order in which these colors were matched is given in Figure 7. Each color represents a major region of color space, lies well away from color model boundaries (to avoid clipping, as explained in Section 4), and cannot be matched trivially in any color model. (See Table VI.) A pseudorandom order was used to factor out of the data possible effects on one match of the color matched previously (see Figure 7).

Subjects depressed one terminal key to start a match and a second when they decided a match had been accomplished. Thus subjects could rest between trials.

A maximum of 3 minutes was allowed for each trial. If the limit was exceeded, the terminal beeped, an appropriate message was displayed, and the subject was asked to start the next match when ready.

Table VI. Each Row of This Table Shows the Vectors from Starting Gray to Indicated Target Color (Red, Green, Blue, Yellow, or White) in Each Color Model

		CDUs along axes		
Model	"Color"	ΔAxis1	ΔAxis2	ΔAxis3
RGB	Red	51.288	27.766	16.803
	Green	17.837	85.532	0.009
	Blue	18.681	26.915	52.617
	Yellow	35.430	59.486	17.008
	White	27.241	29.928	17.782
HSV	Red	2.128	38.275	35.046
	Green	11.490	35.901	34.629
	Blue	5.147	37.896	34.007
	Yellow	8.292	34.156	23.900
	White	1.862	5.119	18.283
YIQ	Red	12.132	101.289	52.372
	Green	26.520	55.995	255.347
	Blue	1.745	54.135	71.076
	Yellow	24.463	44.992	102.129
	White	16.014	15.551	8.051
Opponent	Red	12.249	6.350	50.328
	Green	27.201	9.647	55.812
	Blue	6.898	92.801	1.448
	Yellow	25.099	28.352	13.183
	White	15.587	1.913	4.626
LAB	Red	12.233	55.014	32.375
	Green	27.165	85.403	34.750
	Blue	6.889	35.615	65.327
	Yellow	25.066	16.295	45.467
	White	15.567	4.858	4.628

Note. The distances along each coordinate axis are measured in color difference units. This is not an entirely legitimate use of the CIELAB color difference formula, which is designed for smaller color differences, but no better means is available for indicating the amount of perceptual movement along each axis that is required to match a color using a given color model. Thus to move from the initial color to the target red, the subject must move 51.288 cdu's along the red axis, 27.766 cdu's along the green axis, and 16.803 cdu's along the blue axis.

In all cases the initial color of the controlled rectangle was the gray given in Table VI. Consequently, some target colors were farther from the initial controlled color than others. This does not affect comparisons of statistics which are computed across all target colors, nor does it affect comparisons which are confined to a single target color. It does affect comparisons between target colors, and we avoided these comparisons in our analysis.

Data were logged automatically by the matching software. During the instruction phase the total duration of the introductory tutorial, the number of times each page of instructions was read, and the time spent doing each of the five practice exercises were recorded. At the end of the matching phase the target color, final controlled color, and time taken to reach the match were recorded.

In addition the color of the controlled rectangle was sampled every second, and every button push or release was logged along with the time of its occurrence, so that the process by which a match was accomplished could be analyzed.

Information forms were completed before and after the experiment by every subject. These requested the subject's name, address, sex, age, area of study, native language, level of experience with graphics input equipment, and level of experience mixing colors. Comments and complaints about the experiment were also solicited.

6. RESULTS FOR INEXPERIENCED SUBJECTS

Most of the data are presented using bar graphs. In each graph the height of a bar represents the mean of the value being reported for all matches performed using the indicated color model and tablet technique. For many of the measures 95 percent confidence intervals are shown using a vertical line extending through the top center of each bar.

Many of the statistical measures applied in this analysis are based on the premise that the data have an approximately normal distribution. For some of the data we collected this was not true. For example, the color difference between target and controlled colors after a match was positively skewed. In such cases a logarithmic transformation was applied to the data before the statistical measures were calculated. The Kolmogorov-Smirnov goodness of fit test was used throughout to verify that the distributions were normal at the 95 percent confidence level [13].

Color differences were computed and compared using the CIELAB color difference formula [5]. Although CIELAB is only approximately uniform, we do not compare data solely in one part of color space with data solely from another, and minor deviations from uniformity are therefore thought to be irrelevant.

There were 8 subjects, and therefore 240 matches, for each of the 12 systems (color model/tablet technique pairs) evaluated. Not all of the trials were completed within the time limit, however, and when appropriate those trials are excluded from the statistical computations described below.

We are looking for trends in the data that differentiate among color models or tablet techniques. Thus we always want to test the null hypothesis that some statistic (e.g., the mean time to match) is the same for all systems. Analysis of variance (ANOVA) [16] is the appropriate statistical tool. The factors considered are the five color models (RGB, HSV, YIQ, LAB, and Opponent) and the two tablet techniques (3*1d and 1d + 2d). We usually perform a two-way ANOVA, reporting F scores for the two variables individually and also for the possible interaction between them. These are written $F[4,2572] = 4.23, p < 0.01$, indicating that the F score in question was 4.23, having 4 degrees of freedom in the numerator and 2572 degrees of freedom in the denominator, and that there is less than one chance in a hundred of the means tested being identical for all the factors involved. (See Walpole and Myers [16] for details.) With respect to p, we report the most significant of the thresholds 0.05, 0.01, 0.001, and 0.0001 that the corresponding statistic reaches.

Table VII. Number of Times Subjects Failed to Complete a Color Match within the 3-Minute Time Limit

System	Totals
Rgb1	11
Opp2	13
Opp1	14
Rgb2	14
Rgb3	15
Hsv1	19
Opp3	21
Lab1	23
Yiq1	34
Lab2	34
Hsv2	39
Yiq2	52
Total	289

6.1 Task Incompletions

The number of times subjects failed to complete a color match within the 3-minute time limit (the number of "time-outs" that occurred) for each system are given in Table VII. The RGB and Opponent systems scored well, being 6 of the best 7. (The Kruskal-Wallis statistic [16] h has value 7.41, showing that the better performance of the RGB and Opponent systems is significant at the 0.01 level.) The Yiq2 system was the worst.

6.2 Time To Match

Figure 8 shows the mean time taken to complete a match for subjects using each of the systems studied. Analysis of variance indicates that there is a significant effect of color model ($F[4,2571] = 15.76$, $p < 0.0001$) and of tablet technique ($F[1,2571] = 12.92$, $p < 0.001$), but not of interaction between the two ($F[4,2571] = 2.00$). The Opponent and RGB models were best by this metric (77.44 and 78.14 seconds, respectively), followed by LAB (83.70 seconds), YIQ (89.52 seconds), and the HSV color model (94.07 seconds), which was worst. Subjects using the 3*1d tablet technique required an average of 86.54 seconds to accomplish a match, while subjects using the 1d + 2d tablet technique required an average of 80.70 seconds.

A "•" is used in Figure 8 to mark the RGB systems, and an "O" is used to mark the Opponent systems. We use this convention subsequently wherever these systems are singled out. Also, a horizontal dotted line is used to facilitate comparison. In Figure 8 the dotted line is positioned at the top of the Rgb1 bar because this is the tallest of the RGB and Opponent systems. In general such a bar is positioned so as to help make whatever comparison is discussed in the text.

6.3 Closeness of Final Match

There is a significant effect of color model ($F[4,2571] = 24.60$, $p < 0.0001$), of tablet technique ($F[1,2571] = 13.00$, $p < 0.001$), and of interaction between the

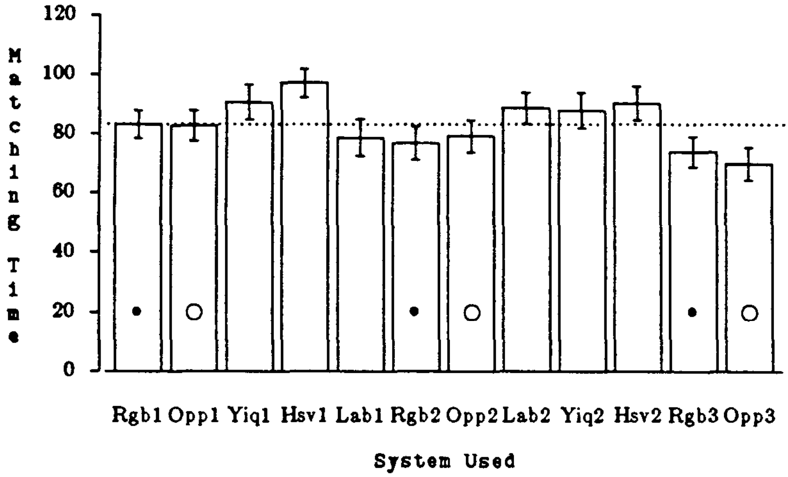

Fig. 8. The mean time in seconds taken to complete a match. Incomplete trials are omitted. Bars for the RGB and Opponent systems are marked with "•" and "O", respectively.

two ($F[4,2571] = 5.62$, $p < 0.001$) in closeness of the final match. The RGB systems were 3 of the 4 *worst* by this measure. Subjects using the 3*1d tablet technique were 12 percent closer than subjects using the 1d + 2d tablet technique: 1.95 log cdu's (7.06 cdu's) as compared with 2.09 log cdu's (8.06 cdu's). (Note that 7.06 and 8.06 are geometric, not arithmetic, means. As explained earlier we must take the logarithm of our data in order to obtain distributions sufficiently normal to allow statistical analysis.)

The CIE color difference specification [5] provides a method by which color differences can be broken down into components corresponding roughly to the subjective sensations of "lightness difference," "hue difference," and "chroma difference." (Chroma is used here, as by Munsell, to mean saturation.) We break Figure 9 into these components to see if there are significant variations between color models; it is possible that various color axes differ in their utility for minimizing distinct color attributes.

The lightness of a color (L^*, a^*, b^*) is simply L^*, which has been defined by the CIE so as to represent perceived lightness as accurately as possible. The CIE defines an appropriate measure of chroma to be

$$C^*ab = (a^{*2} + b^{*2})^{1/2}.$$

If ΔC is the difference in chroma between two colors and ΔL is the difference in their lightness, then the CIE defines their *hue difference* to be

$$\Delta H^* = (\Delta E^{*2} - \Delta L^{*2} - \Delta C^{*2})^{1/2}.$$

Figure 10 shows the color difference component corresponding to lightness. Analysis of variance shows significant effects of color model ($F[4,2571] = 24.54$, $p < 0.0001$), tablet technique ($F[1,2571] = 23.89$, $p < 0.0001$), and of interaction between the two ($F[4,2571] = 3.12$, $p < 0.05$). Note that the RGB systems are the three worst systems with respect to this component and stand out clearly.

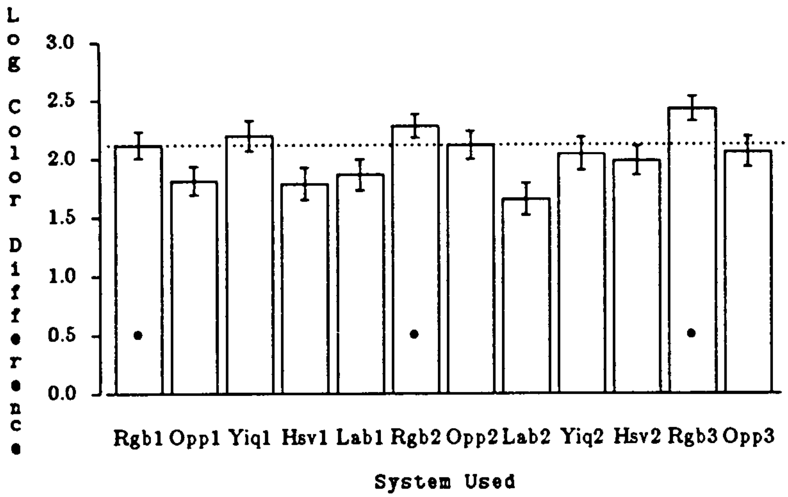

Fig. 9. The mean log color difference at the completion of a match. Incomplete trials are omitted. A log transformation (base e) was applied to obtain a normal distribution before statistical analysis was performed.

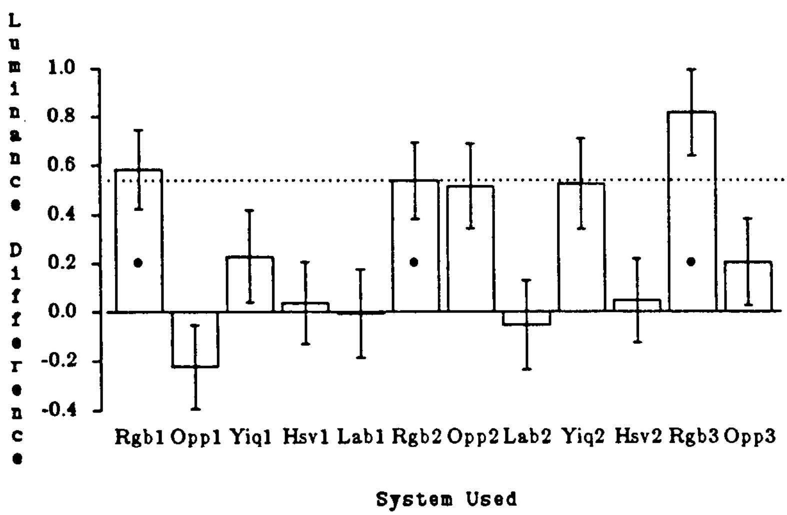

Fig. 10. The mean log lightness difference at the completion of a match. Incomplete trials are omitted.

Chroma difference, shown in Figure 11, has significant effects of both color model ($F[4,2571] = 7.95$, $p < 0.0001$) and of tablet technique ($F[1,2571] = 5.61$, $p < 0.05$), but not of interaction between the two ($F[4,2571] = 1.86$). The RGB systems are three of the worst four with respect to chroma, although only the Rgb2 and Rgb3 systems stand out noticeably from the others.

The hue difference data are shown in Figure 12. There are significant main effects of color model ($F[4,2571] = 8.80$, $p < 0.0001$) and of tablet technique

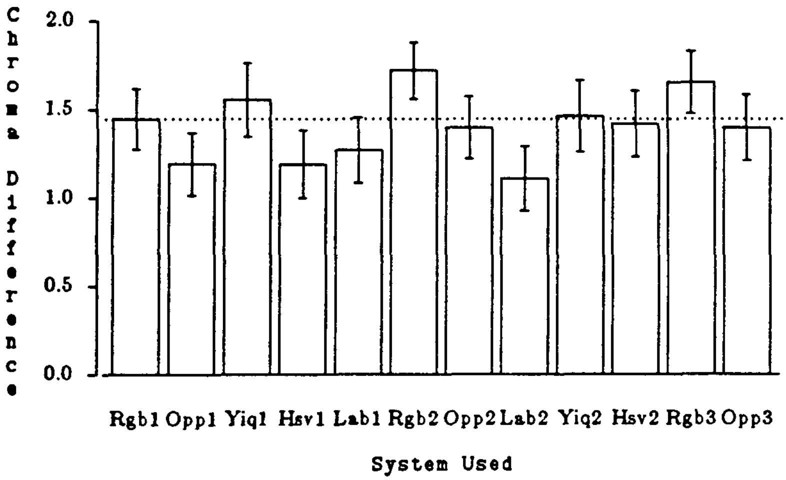

Fig. 11. The mean log chroma difference at the completion of a match. Incomplete trials are omitted.

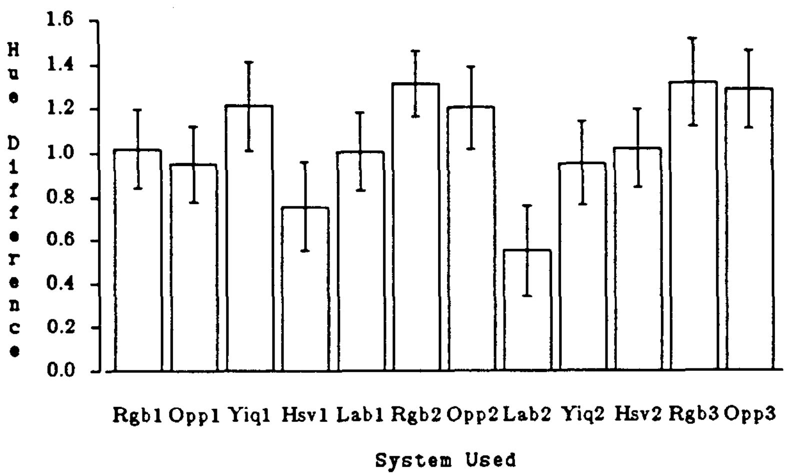

Fig. 12. The mean log hue difference at the completion of a match. Incomplete trials are omitted.

($F[1,2571] = 4.62$, $p < 0.05$), as well as significant interaction between the two ($F[4,2571] = 7.04$, $p < 0.0001$).

Lightness, chroma, hue, and overall distance order the color models in the same way (LAB, HSV, Opponent, YIQ, and RGB, from best to worst, respectively), except that the YIQ and Opponent models exchange positions for hue. If we look at the data and compare the average log accuracy of RGB systems against the average log accuracy of the other systems, all of which have a lightness-related axis, we find the RGB systems are 66 percent less accurate. However, it is difficult to see from the data whether there is a difference between using a luminance control (as in the YIQ and the Opponent models), a lightness control (as in CIELAB), or simply a lightnesslike control (as in HSV).

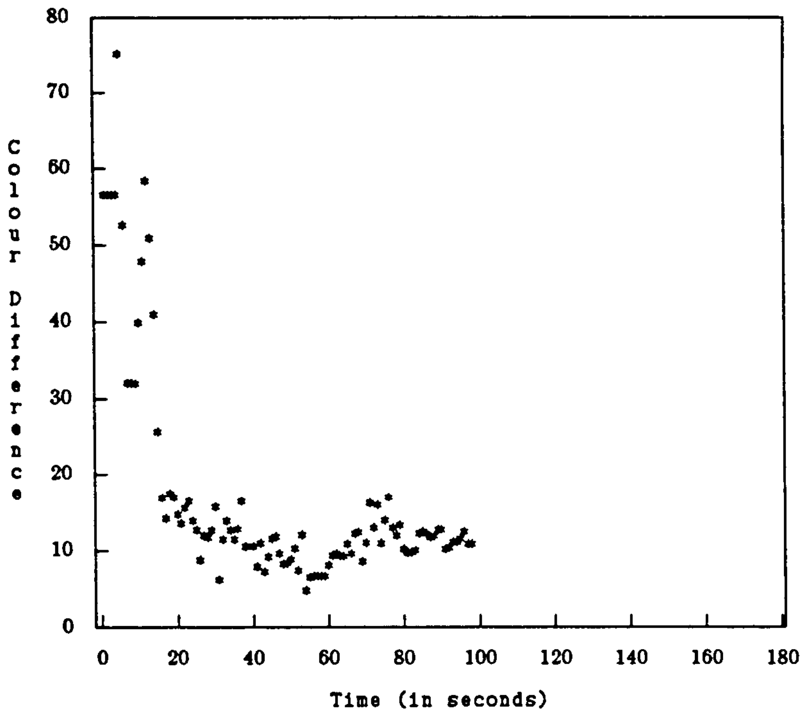

Fig. 13. Color distance of the controlled color from the target color as a function of time for a "typical" subject. Note that many different colors are the same distance from the target color and that these data are generated by sampling the subject's position in color space once per second. The subject may have approached the target more closely than shown between samples. Note also that a more or less constant color difference, as is apparent for most of the trial, does not imply a more or less constant color, since there is a whole surface of colors that is at a constant difference from the target color.

By way of comparison, note that only the HSV systems have a hue axis. However, the average log hue accuracy of the HSV systems is only 19 percent better than the average log hue accuracy of the other systems. Moreover, the LAB systems are, on average, more accurate with respect to hue than are the HSV systems.

Similarly, it is only the HSV systems that have a chroma or saturation axis, and we find that the average log chroma accuracy of these systems is only 12 percent better than the average log chroma accuracy of the other systems. Once again, the LAB systems are, on average, more accurate than the HSV systems with respect to chroma.

For all three of lightness, hue, and chroma 3*1d use of the tablet is more accurate than 1d + 2d use.

6.4 Color Difference Thresholds

By sampling each subject's position in color space it becomes possible to analyze the *process* by which a match is accomplished. In examining such traces, a typical representation of which is shown in Figure 13, we are struck by the fact that

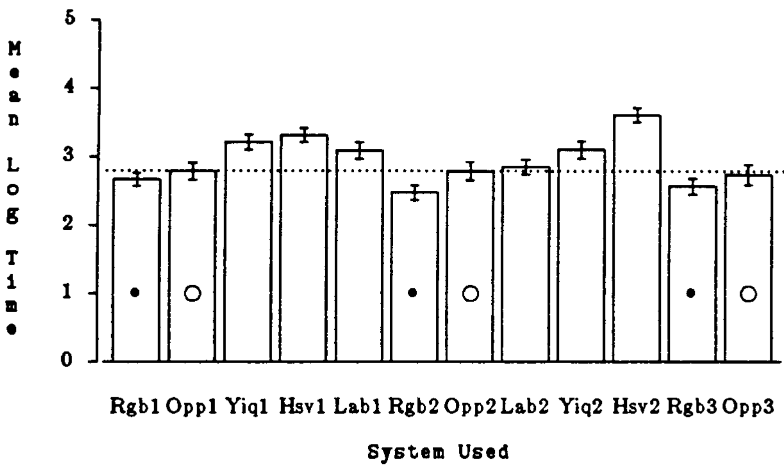

Fig. 14. Mean log time to reach the 25 cdu threshold for the first time (inexperienced subjects). Trials with white as the target color were excluded from these data. Subjects failed to reach this threshold in approximately 10 percent or fewer of the trials for all systems.

subjects appear rapidly to home in on the neighborhood of a target color, and then to spend substantial amounts of time exploring that neighborhood.

We hypothesize that there are at least two distinct phases to the matching process: a *convergence phase* characterized by a rapid decrease in the color difference, and a subsequent *refinement phase* characterized by small, almost random fluctuations in color difference in the vicinity of the target color (difference zero). The problem is to find a way of identifying the transition from one phase to the other.

Our approach is to examine the mean time taken by subjects to first reach various color thresholds (i.e., to first approach within some fixed distance of the target color, as measured in color distance units). We hypothesize that such "first-encounter" times for high thresholds reflect performance during the convergence phase, while first-encounter times for very low thresholds are dominated by performance during the refinement phase.

Fifteen thresholds ranging from 2 to 40 color difference units (cdu's) were investigated; thresholds of 25, 14, and 6 CIELAB cdu's are shown here, as they are representative of the extremes and the median. The following discussion is based on Figures 14–18, which show averages and standard deviations calculated using these thresholds. In computing the measures these figures record it was necessary to work with log data in order to obtain sufficiently normal distributions. Trials in which the match was not completed in the maximum time allowed are included, so long as the threshold itself was reached. Trials in which white was matched are not included in the 25 and 14 cdu threshold data for two reasons: (1) the white used was only 17 cdu's from the initial gray color of the controlled rectangle; (2) achromatic colors present significant potential boundary problems for users of the HSV color model, as explained earlier. We do, however, present data for the 6 cdu threshold both with and without the data for white.

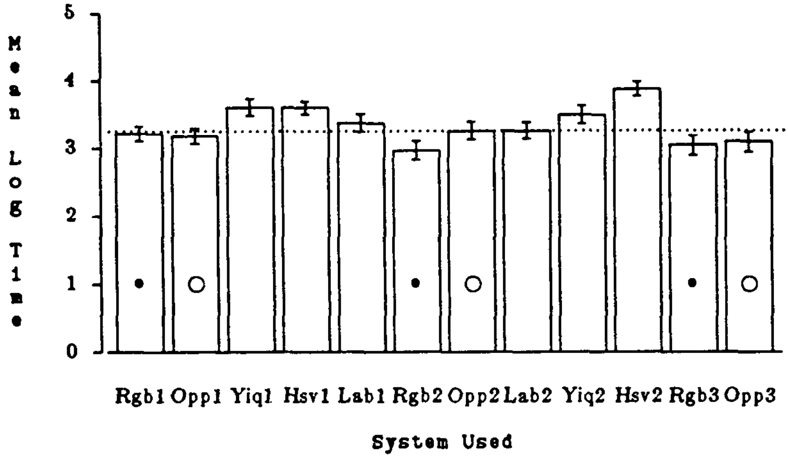

Fig. 15. Mean log time to reach the 14 cdu threshold for the first time (inexperienced subjects). Trials with white as the target color were excluded from these data. Subjects failed to reach this threshold in 25 percent or fewer of the trials for all systems.

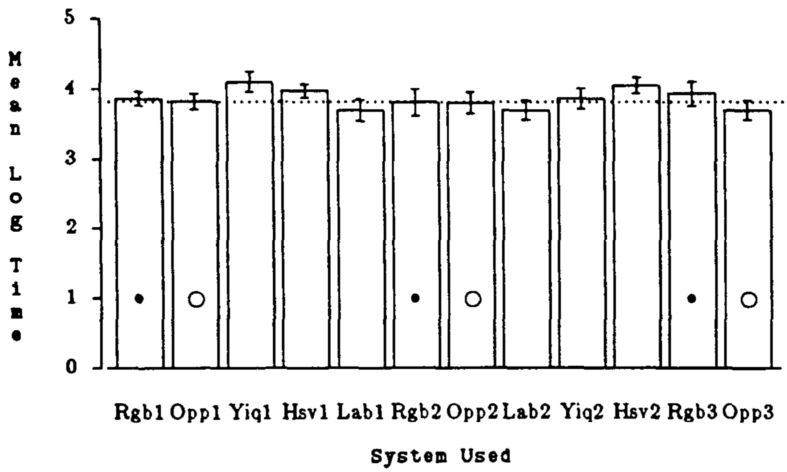

Fig. 16. Mean log time to reach the 6 cdu threshold for the first time (inexperienced subjects). Trials with white as the target color were excluded from these data.

Regarding the time required to reach the 25 cdu threshold (data for white excluded), there are significant main effects of color model ($F[4,2156] = 80.60$, $p < 0.0001$), of tablet technique ($F[1,2156] = 13.60$, $p < 0.001$), and of interaction between the two ($F[4,2156] = 2.82$, $p < 0.05$). Subjects reached this threshold fastest using RGB systems and slowest with the HSV systems (145 percent longer than RGB); the Opponent, LAB, and YIQ systems required 22 percent, 49 percent, and 80 percent more time, respectively, than the RGB

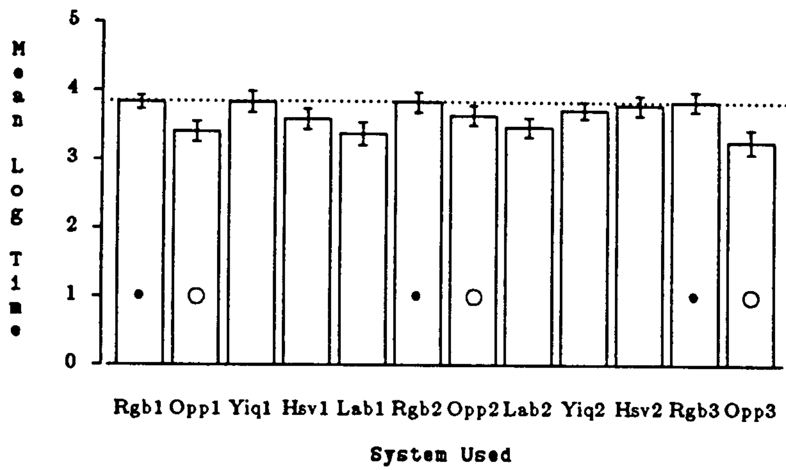

Fig. 17. Mean log time to reach the 6 cdu threshold for the first time (inexperienced subjects). Trials with white as the target color were included in these data.

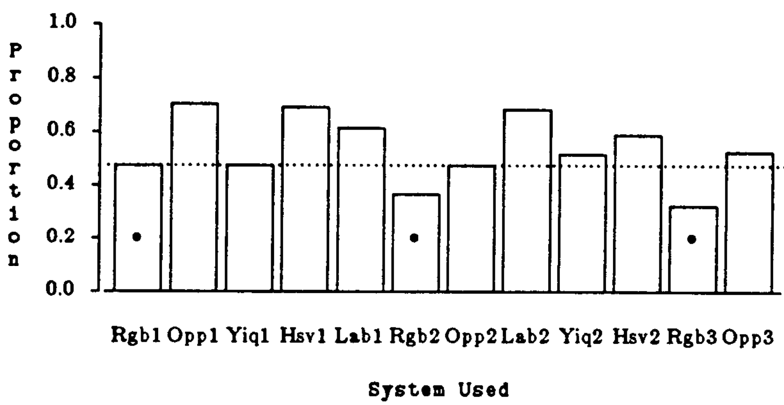

Fig. 18. The proportion of matches (including matches of white) that reached the 6 cdu threshold (inexperienced subjects).

systems. (Note, however, that users of the HSV systems may have been somewhat disadvantaged by starting from a neutral gray, as discussed in Section 4, and any such disadvantage would show up most clearly in the 25 cdu data.) Subjects required 14 percent more time to reach this threshold when using the tablet in 3*1d mode than when using the tablet in 1d + 2d mode.

At the 14 cdu threshold (data for white excluded) there are again significant main effects of color model ($F[4,1871] = 42.87$, $p < 0.0001$), of tablet technique ($F[1,1871] = 6.39$, $p < 0.05$), and of interaction between the two ($F[4,1871] = 3.46$, $p < 0.01$). The RGB systems are still the fastest and the HSV systems are still the slowest in that subjects required 93 percent longer to reach this threshold using the HSV color model than using the RGB color model. Subjects using the Opponent, LAB, and YIQ color models required 10, 26, and 60 percent longer, respectively. However, the magnitude of the difference between RGB and HSV is less for the 14 cdu threshold than for the 25 cdu threshold. Subjects required

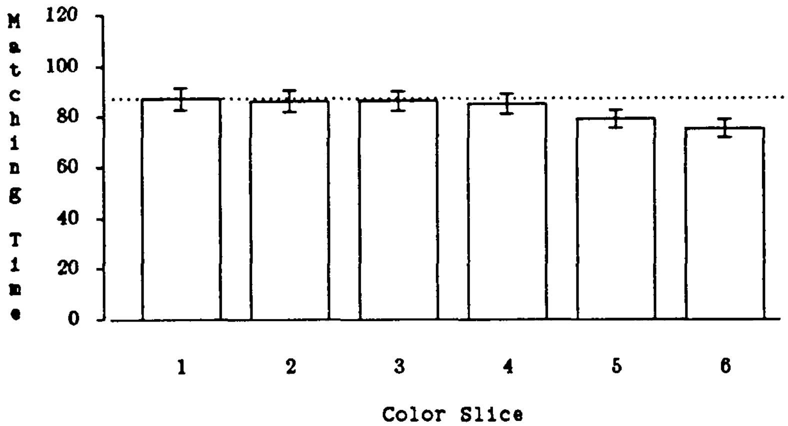

Fig. 19. The mean time taken to complete a color match by color slice (inexperienced subjects). Incomplete trials are omitted.

10 percent more time to reach this threshold when using the tablet in 3*1d mode than when using the tablet in 1d + 2d mode.

At the 6 cdu threshold (data for white excluded) there is a significant main effect of color model ($F[4,1076] = 8.98, p < 0.0001$), but not of tablet technique ($F[1,1076] = 1.64$), or of interaction between the two ($F[4,1076] = 1.40$). The variation between color models in time to reach this threshold is greatly reduced (at most a factor of 1.37, or 37 percent). No clear trends are evident, except that the HSV systems are 2 of the worst 3.

When white matches are included in the 6 cdu data (see Figure 17), there is still a significant main effect of color model ($F[4,1532] = 15.64, p < 0.0001$), and again no significant main effect of tablet technique ($F[1,1532] = 1.69$) or of interaction between the two ($F[4,1532] = 1.31$). The only noticeable difference from Figure 16 is that the RGB systems are now 3 of the worst 4, which is presumably related to the fact reported earlier that users of the RGB color model produced the least accurate matches. The percentage of trials that reached this threshold is shown in Figure 18. The RGB systems rank worst; no other trends are evident.

6.5 Learning

It is also of interest to ask whether significant learning occurred during the hour that it took most subjects to make 30 matches. When combining together data for all systems, we distinguish six successive *color slices* (see Figure 7), in each of which all five colors are matched once each. When examining color models or tablet techniques, we distinguish only the first and second *session halves* to improve the significance of the statistics, performing separate two-way ANOVA's of color model versus session half and of tablet technique versus session half. We want to know (1) whether significant learning has occurred, and (2) whether different amounts of learning occur for distinct color models or tablet techniques.

A one-way analysis of variance rejects the hypothesis that the time to match is constant across color slice ($F[5,2574] = 5.93, p < 0.0001$). In Figure 19 we see that the mean time taken to complete a match remained nearly constant across

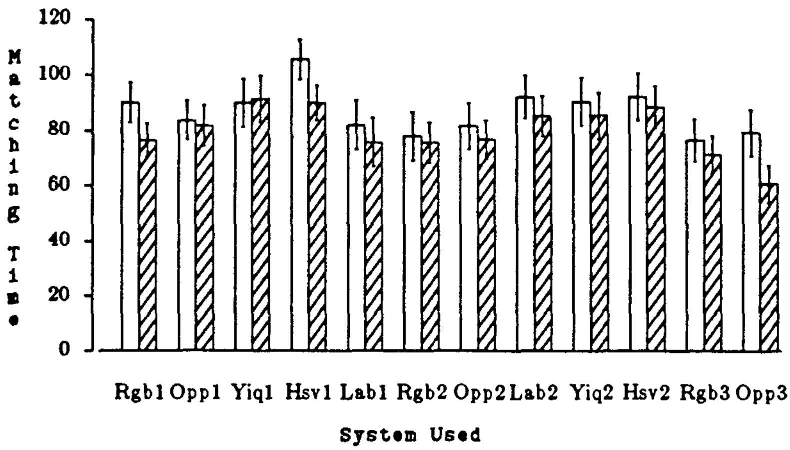

Fig. 20. The mean time taken to complete a color match by system and session half (inexperienced subjects). The unshaded bars show the mean time required for each subject to perform the first 15 color matches, while the shaded bars show the mean time required to perform the last 15 color matches. Incomplete trials are omitted.

the first four color slices, but then dropped considerably for the fifth and sixth slices.

Figure 20 shows the mean time to complete a match for individual systems, by session half. Comparing color model against session half, two-way analysis of variance shows significant main effects of color model ($F[4,2571] = 15.76$, $p < 0.0001$), and of learning by session half ($F[1,2571] = 17.12$, $p < 0.0001$). Comparing tablet technique against session half, two-way analysis of variance shows significant main effects of tablet technique ($F[1,2577] = 12.69$, $p < 0.001$), and of session half. ($F[1,2577] = 16.81$, $p < 0.0001$.) In neither case are the interactions significant ($F[4,2571] = 0.95$ and $F[1,2577] = 0.19$, respectively). Hence the average improvement in time to match of 6.65 seconds from the first to the second half is significant and represents real learning. Moreover, because the interactions are not significant, there is no evidence of differences in learning between systems or tablet techniques with respect to this measure. (Note that we already know the differences between color models and tablet techniques to be significant from our earlier analysis).

Figures 21 and 22 show the analogous data for color difference at the time subjects declared a match. In Figure 22 there are significant main effects of color model against session half ($F[4,2571] = 24.39$, $p < 0.0001$), of tablet technique against session half ($F[1,2577] = 12.47$, $p < 0.001$), and of learning by session half ($F[1,2571] = 5.36, p < 0.05$, and $F[1,2577] = 5.18, p < 0.05$). The interactions are not significant ($F[4,2571] = 1.87$ and $F[1,2577] = 0.35$). It appears from Figure 21 that most of the learning occurred at the beginning, after which there was slow improvement up until the last slice, at which point performance deteriorated, perhaps as subjects became tired or anxious to complete the experiment. The improvement between session halves was about 14 percent; the improvement from the first to the fifth slice was about 21 percent.

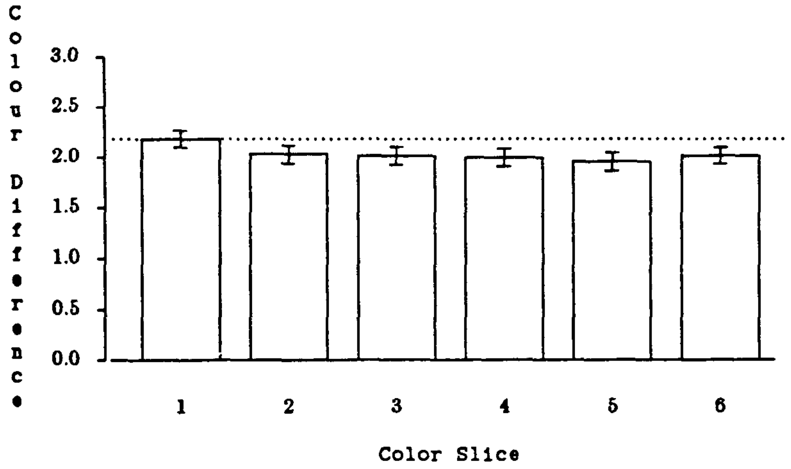

Fig. 21. The mean log color difference after a successful match by color slice and across all trials (inexperienced subjects). Incomplete trials are omitted.

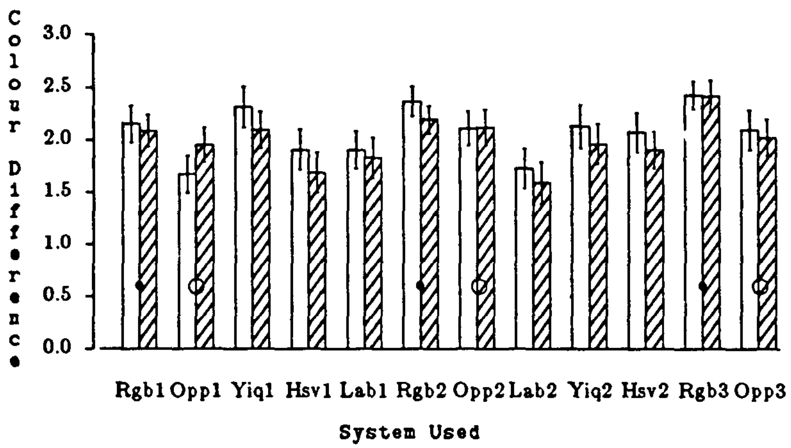

Fig. 22. The mean log color difference after a color match by system and session half. The unshaded bars represent the means of trials in the first half of each session, while the shaded bars show performance during the second half of each session (inexperienced subjects). Incomplete trials are omitted.

Note that the above data show clear improvement both of speed and of accuracy, so that genuine learning has occurred—we have not simply moved to another point on the speed-accuracy trade-off curve.

Next we turn to an examination of the time required to reach the 25, 14, and 6 cdu thresholds. In Figures 23–25 we show the time to reach these thresholds individually for each system, by session half.

For the 25 cdu threshold (Figure 23, data for white excluded) there are significant effects of color model against session half ($F[4,2156] = 82.10$, $p < 0.0001$), of tablet technique against session half ($F[1,2162] = 12.05$,

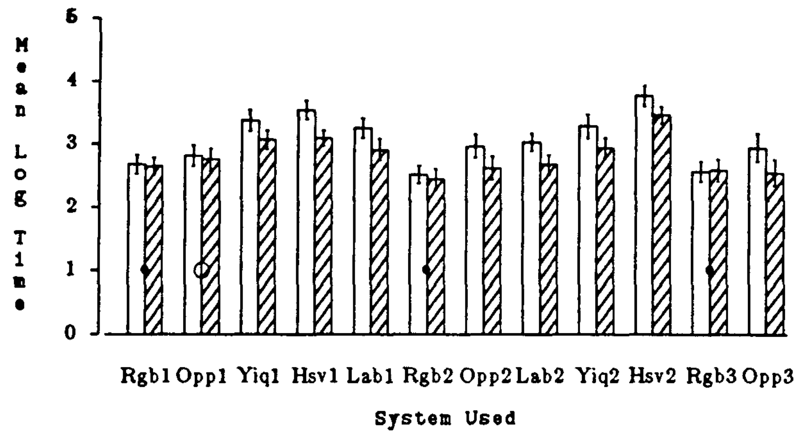

Fig. 23. The mean time required to reach the 25 cdu threshold, by system, for the first and second session halves (the unshaded and shaded bars, respectively). Trials with white as the target color were excluded from these data.

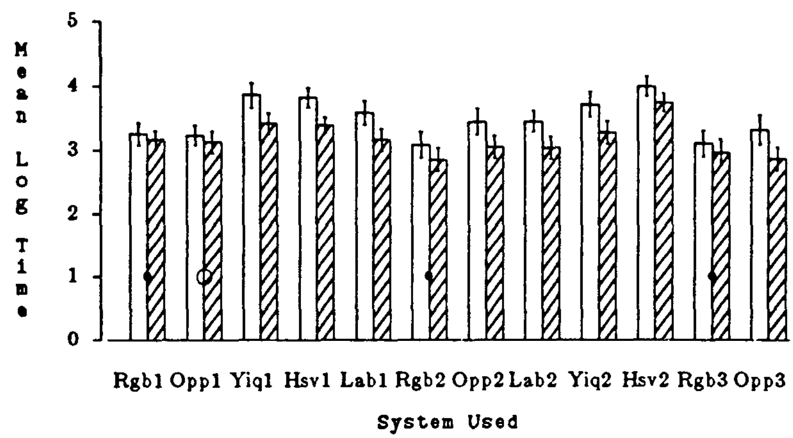

Fig. 24. The mean time required to reach the 14 cdu threshold, by system, for the first and second session halves (the unshaded and shaded bars, respectively). Trials with white as the target color were excluded from these data.

$p < 0.001$), and of learning by session half ($F[1,2156] = 49.13$, $p < 0.0001$ and $F[1,2162] = 42.71$, $p < 0.0001$). The interaction of color model with session half is significant ($F[4,2156] = 4.16$, $p < 0.01$), while the interaction of tablet technique with session half is not ($F[1,2162] = 0.15$).

The least change in time between session halves for this threshold occurs for matches involving the RGB color model, which show improvement of only 3 percent. The remaining color models show improvement of 23 percent (Opponent), 28 percent (YIQ), 30 percent (LAB), and 31 percent (HSV). Nevertheless, the RGB color model is still fastest in the second session half, and the HSV color model is still the slowest.

For the 14 cdu threshold (Figure 24, data for white excluded) there are significant main effects of color model against session half ($F[4,1871] = 44.23$,

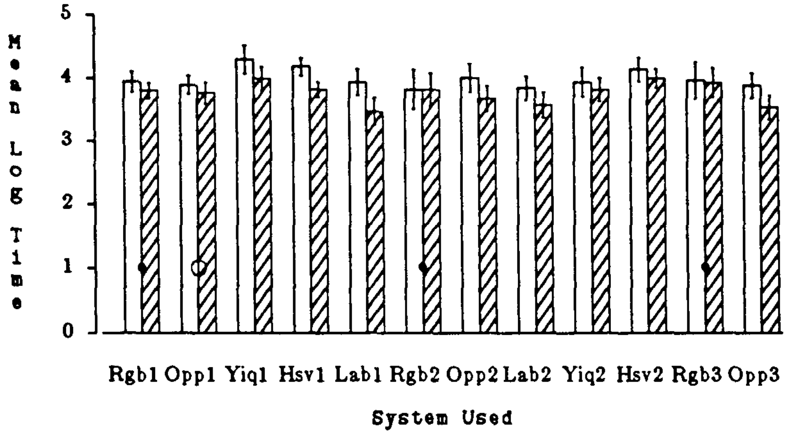

Fig. 25. The mean log time required to reach the 6 cdu threshold, by system, for the first and second session halves (the unshaded and shaded bars, respectively). Trials with white as the target color were excluded from these data.

$p < 0.0001$), of tablet technique against session half ($F[1,1877] = 6.03$, $p < 0.05$), and of learning by session half ($F[1,1871] = 70.09$, $p < 0.0001$ and $F[1,1877] = 64.16$, $p < 0.0001$). The interaction of color model with session half is significant ($F[4,1871] = 2.60$, $p < 0.05$), while the interaction of tablet technique with session half is not ($F[1,1877] = 1.24$).

Systems based on the RGB color model also show the least learning between session halves at this threshold, improving 14 percent. The other color models show improvement of 27 percent (Opponent), 29 percent (HSV), 34 percent (LAB), and 35 percent (YIQ). Systems based on the RGB color model are still fastest in the second half, though barely, and systems based on the HSV color model are still the worst.

For the 6 cdu threshold (Figure 25, data for white excluded) there is a significant main effect of color model against session half ($F[4,1068] = 9.27$, $p < 0.0001$), but not of tablet technique against session half ($F[1,1074] = 1.63$). There is a significant effect of learning by session half ($F[1,1068] = 32.81$, $p < 0.0001$ and $F[1,1074] = 31.70$, $p < 0.0001$). The interactions are not significant ($F[4,1068] = 2.18$ and $F[1,1074] = 0.41$).

At this threshold the RGB systems show 7 percent improvement from the first to the second half. The other color models show improvement of 18 percent (YIQ), 22 percent (Opponent), 23 percent (HSV), and 30 percent (LAB).

For the 25 and 14 cdu thresholds the first encounter time, averaged across all systems, decreases in small and approximately equal steps from slice to slice. At the 6 cdu threshold there is a very slight increase from the fifth to the sixth slice.

Since there is no interaction between tablet technique and session half, there is no evidence that subjects learned to use the 3*1d mapping more rapidly than the 1d + 2d mapping, or vice versa.

There is interaction between color model and session half, but only at the 25 cdu and 14 cdu thresholds. This suggests that there are learning effects that differ from color model to color model (most noticeably with respect to RGB, as indicated), but that these have to do with performance during the convergence phase, and not the refinement phase since there is no interaction between time to reach the 6 cdu threshold and session half or between total time to match and session half.

6.6 Summary for Inexperienced Subjects

For convenience we summarize here the main points for inexperienced subjects.

(1) With respect to task incompletions, users of the RGB and Opponent color models performed best.

(2) Regarding the time to final match, users of the Opponent and RGB color models were fastest, while users of the HSV color model were slowest; 1d + 2d users of the tablet were faster than 3*1d users.

(3) Use of the RGB color model produced less accurate final matches, as did 1d + 2d use of the tablet. Possession of a lightness axis aids matching lightness more than possession of a hue or chroma axis aids matching hue or chroma.

(4) Subjects reached the 25 cdu threshold fastest using the RGB color model and slowest using the HSV color model. This trend exists also at the 14 cdu threshold, but is weaker. It has largely disappeared at the 6 cdu threshold.

Subjects reach the 25 cdu threshold faster using the tablet in 1d + 2d mode. This trend exists also at the 14 cdu threshold, though again it is weaker, but not significantly at the 6 cdu threshold.

(5) Significant learning occurred with respect to matching time and closeness.

(6) Little learning occurred for users of the RGB color model during the convergence phase.

7. RESULTS FOR AN EXPERIENCED SUBJECT

Figures 26 through 30 summarize data collected over a two-day period from an expert in color science with extensive RGB color matching experience. This subject had no experience using graphics input hardware of any kind, nor had he ever performed color matching tasks with any of the other color models studied, although he had an extensive fundamental understanding of them. These data are taken under somewhat different conditions than the data from inexperienced subjects: his extensive color-matching experience led us to hypothesize that he would show little effect of learning; shortage of time prevented us from examining all the interfaces; and there are no other subjects with comparable experience to whom he may be compared. Nonetheless, the data have interesting qualitative features and are included because of their uniqueness. A one-way ANOVA shows that the variation of every mean discussed with system is significant ($p < 0.01$).

No time-outs occurred for eight of the nine systems used by the experienced subject (the Rgb3, Opp3, and Hsv2 systems were omitted); he had two time-outs while using the Yiq1 system. (Contrast this with the data in Table VII for inexperienced subjects.)

From Figure 26 it is clear that the experienced subject matched fastest using the Rgb2 and Opp2 systems. Comparing Figure 8 and Figure 26, it is clear that

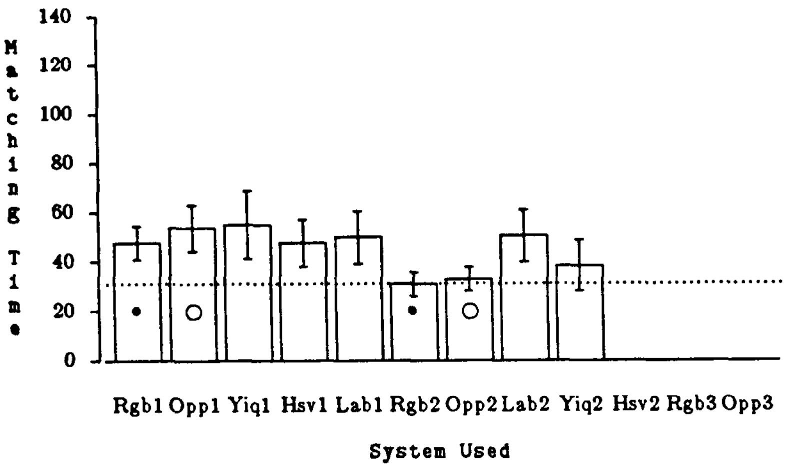

Fig. 26. The mean time taken by the experienced subject to complete a match. Incomplete trials are omitted.

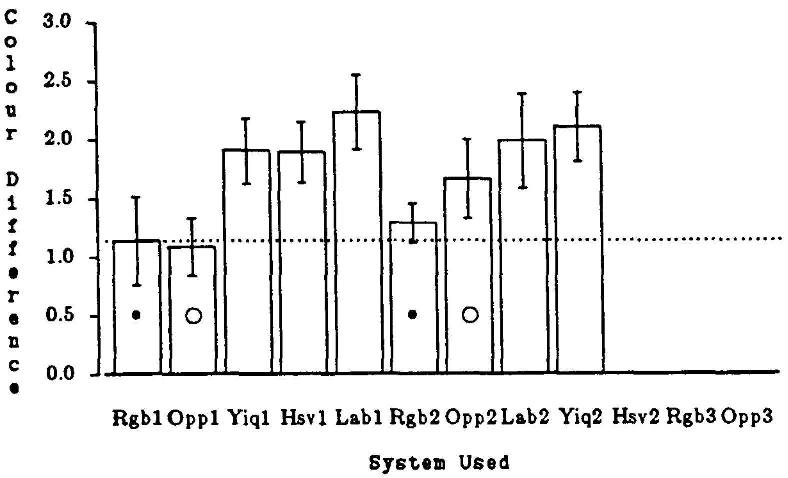

Fig. 27. The mean log color difference at the time a match was completed by the experienced subject. Incomplete trials are omitted.

he was significantly faster than the inexperienced subjects. On the other hand, he matched closest using the Rgb1 and Opp1 systems, followed by the Rgb2 and Opp2 systems. (Compare Figure 9.) It is also interesting that his matches are only noticeably closer than those of the inexperienced subjects for Rgb1, Opp1, and Rgb2, and just perhaps for Opp2.

The experienced subject's time to reach the 25 cdu threshold is best for the Rgb1 and Rgb2 systems, followed by the Opp2 and Opp1 systems (Figure 28). This is true also of the time required to reach the 14 cdu threshold, although they stand out less (Figure 29). At the 6 cdu threshold only the Rgb1 and Rgb2

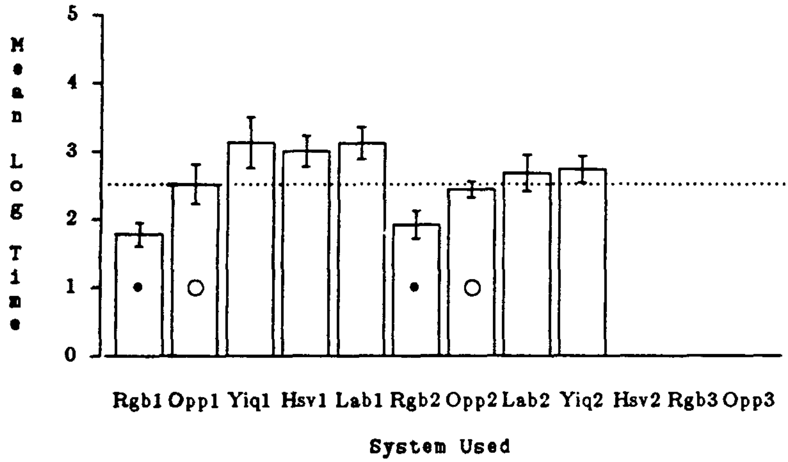

Fig. 28. The mean log time required by the experienced subject to reach the 25 cdu threshold. Trials in which white was matched are excluded.

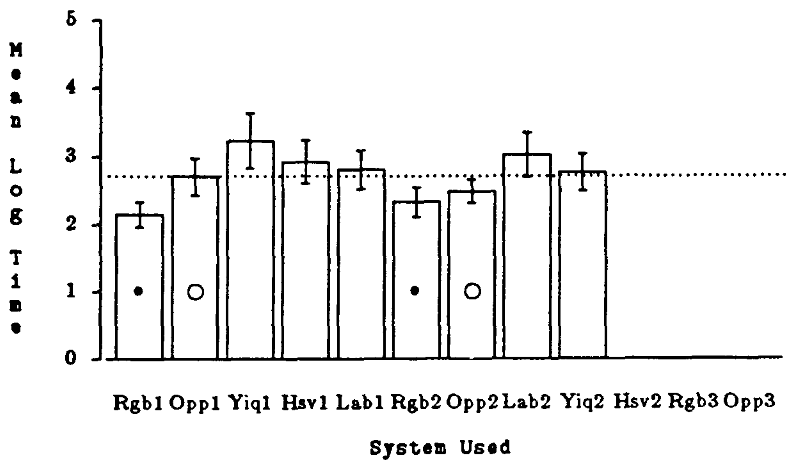

Fig. 29. The mean log time required by the experienced subject to reach the 14 cdu threshold. Trials in which white was matched are included.

systems stand out; the rest are much the same, although the 1d + 2d systems do appear to require slightly less time than the 3*1d systems.

7.1 Summary for the Experienced Subject

For convenience, we summarize here the main points for our experienced subject:

(1) He matched fastest using the RGB2 and Opp2 systems.

(2) He matched closest using the RGB and Opponent models.

(3) He was fastest reaching the 25 and 14 cdu thresholds using the RGB and Opponent models. He was fastest reaching the 6 cdu threshold using the RGB color model.

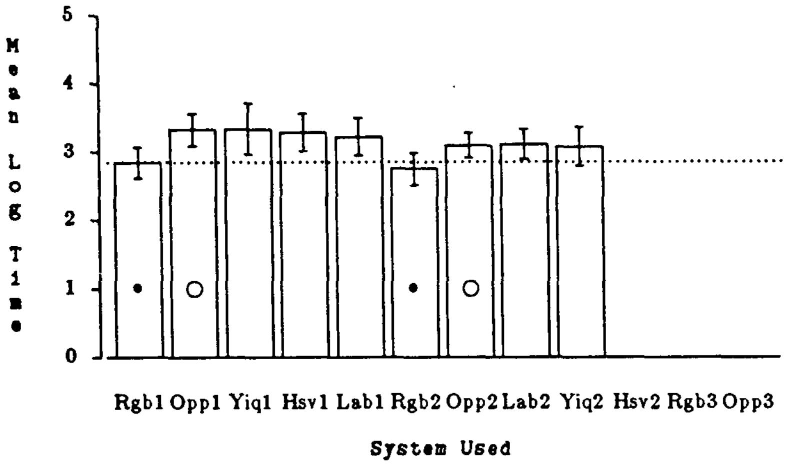

Fig. 30. The mean log time required by the experienced subject to reach the 6 cdu threshold. Trials in which white was matched are included.

8. CONCLUSIONS

The data from this experiment strongly suggest a number of interesting hypotheses, discussed below, which should be the subject of further investigation. Except where otherwise indicated, we are referring to the inexperienced subjects.

Inexperienced subjects can better match lightness (brightness) if the color model used has a luminance or lightness axis. A secondary question not answered by this study is the importance of having the chromatic axes orthogonal to the luminance axis, so as to avoid the need to constantly readjust the luminance after altering the chroma. Of the color models studied, the RGB and HSV models are probably worst in this regard.

The advantage of having an explicit hue or chroma axis is much less substantial.

The RGB and Opponent models share some desirable performance characteristics. In particular, they stand out with respect to the number of task incompletions and the overall time to match.

Subjects using the RGB color model perform best during the convergence phase relative to the other color models. Subjects using the HSV color model perform worst during the convergence phase.

Conversely, no color model is particularly helpful during the refinement phase. The most we can say about the inexperienced subjects is that those using the RGB color model appear to have been somewhat less accurate. It may be that during the refinement phase subjects are, in fact, randomly exploring a small region in color space about the target color, and that the sampled data (see Figure 13, for example) reflect their average distance from the target.

Inexperienced subjects using the RGB color model matched rapidly but inaccurately in comparison with other color models. Their speed may reflect better performance during the convergence phase.

The Opponent, LAB, YIQ, and HSV color models require learning to use effectively. Conversely, little improvement was observed for subjects using the RGB color model. The implications of this finding apply only to learning that occurs over time intervals on the order of an hour, as explored in this experiment. The combination of better performance and lack of learning for the RGB color model implies that the subject's apprehension of this model occurs quickly enough that they reach their peak performance immediately, whereas with the other color models they improved toward that level throughout the hour. Questions of learning and experience on a longer time scale are not answered in this experiment, although the data from our experienced subject (discussed below) are suggestive.

The inexperienced subjects matched faster using the 1d + 2d *tablet technique, but less accurately.* Such "speed-accuracy trade-offs" are a common phenomena in psychological experiments. It is also intriguing to note that the experienced subject was not only able to match fastest with the 1d + 2d RGB and Opponent systems, but that these were his third and fourth most accurate systems.

Learning may affect the convergence phase more than the refinement phase. The mean time taken by subjects to reach the 25, 14, and 6 cdu thresholds decreases by 31, 39, and 33 percent, respectively, between the first and fifth color slices. (We use the fifth rather than the sixth color slice here because Figure 21 suggests that fatigue or boredom may have set in during the sixth slice.) On the other hand, the mean time taken by subjects to complete a match decreases by only 18 percent between the first and fifth slices.

An immediate inference is that the proportion of time spent reaching each of these thresholds was less in the fifth color slice than in the first, so that the proportion of the total matching time needed to reach a particular threshold decreases with practice.

The practical implications of these observations, however, are not entirely clear. The significant differences between the large decreases in time spent reaching the 25 and 14 cdu thresholds and the relatively small decreases in total time spent matching suggest that most of the improvement takes place during the convergence phase. However, if this were true then one would also expect that the decrease in mean time to reach a threshold would be less at the 6 cdu threshold than at the 25 and 14 cdu thresholds. Since that is not the case it may be that there are other phenomena involved that are not evident from the current analysis. Nevertheless, the current study does indicate that learning does in fact occur during the convergence phase.

8.1 The Experienced Subject

Extensive experience with a color model does appear to help during the refinement phase. The experienced subject had participated in RGB-based matching experiments for over 30 years. Figure 27 shows that he was more accurate using the Rgb1 and Rgb2 systems, whereas Figure 9 shows that the inexperienced subjects were less accurate using RGB systems. Moreover, Figure 30 clearly indicates that he was able to reach the 6 cdu threshold faster using the Rgb1 and Rgb2 systems, whereas Figures 16–18 show that the inexperienced subjects are disadvantaged by the use of these systems. The fact that this subject's data correlate well with his known expertise increases our confidence in conclusions drawn from data generated by the other subjects.

9. FUTURE WORK

9.1 Further Data Analysis

We have collected an immense amount of data (14 megabytes worth), and further analysis seems likely to prove fruitful. In particular, we expect to

(1) analyze the path data further to see how the puck was moved by users of the 1d + 2d interaction models;
(2) look for patterns in button usage;
(3) find a better way of separating the convergence and refinement phases;
(4) look for a correlation (or a lack thereof) between time spent in the refinement phase and the closeness of the match made;
(5) test our hypothesis that the path followed during the refinement phase randomly explores a small volume of color space about the target color;
(6) compare matching data for different colors.

9.2 Additional Experiments

More data could usefully be collected to test further the hypotheses of Section 8, or to examine the questions posed in Section 9.1. More specifically:

(1) We would like to gather additional data for a few selected systems. In particular, we would like to have a few subjects perform substantially more matches in order to evaluate longer term learning.
(2) It would be interesting to gather data for other user classes (such as artists, for example) and compare their results with those reported above.
(3) Three of the colors matched in this experiment were fairly close to the RGB axes, although by design it was not possible to match any of them by movement along a single axis. It would be interesting to gather matching data for colors further from the RGB axes to see if subjects using RGB-based systems still had superior performance during the convergence phase.
(4) Use of a different initial color would make it possible to determine whether the users of HSV systems were actually disadvantaged by having to begin from gray.

We would also like to develop and test an Opponent color model that better approximates physiological reality.

ACKNOWLEDGMENTS

All the statistical measures and most of the plots were prepared using the "S" data analysis package of Becker and Chambers [1].

The authors are grateful for the support and assistance of Dr. Gunther Wyszecki, whose data are presented in Section 7, and to whose memory this work is dedicated. The suggestions of the referees, and especially of the editor, are also greatly appreciated.

REFERENCES

1. BECKER, R. A., AND CHAMBERS, J. M. *S—An Interactive Environment for Data Analysis and Graphics.* Wadsworth, 1984.
2. BERK, T., BROWNSTON, L., AND KAUFMAN, A. A human factors study of color notation systems for computer graphics. *Commun. ACM 25*, 8 (Aug. 1982), 547–550.
3. BERK, T., BROWNSTON, L., AND KAUFMAN, A. A new color-naming system for graphics languages. *IEEE Comput. Graph. Appl. 2*, 3 (May 1982), 37–44.
4. BOYNTON, R. M. *Human Color Vision.* Holt, Rinehart and Winston, New York, 1979.
5. CIE. Recommendations on Uniform Color Spaces, Color-Difference Equations, Psychrometric Color Terms. Bureau Central de la CIE (Supplement 2 of CIE Publication 15 (E-1.3.1) 1971), 1978.
6. COWAN, W. B. An inexpensive scheme for calibration of a colour monitor in terms of CIE standard coordinates. *Comput. Graph. 17*, 3 (July 1983), 315–321.
7. EVANS, K. B., TANNER, P. P., AND WEIN, M. Tablet-based valuators that provide one, two, or three degrees of freedom. *Comput. Graph. 15*, 3 (Aug. 1981), 91–97.
8. FOLEY, J. D., AND VAN DAM, A. *Fundamentals of Interactive Computer Graphics.* Addison-Wesley, Reading, Mass., 1982.
9. HUNT, R. W. G. *The Reproduction of Colour.* Fountain Press, 1975.
10. HURVICH, L. M. *Color Vision.* Sinaur, Sunderland, Mass., 1981.
11. JOBLOVE, G. H., AND GREENBERG, D. P. Color spaces for computer graphics. *Comput. Graph. 12*, 3 (Aug. 1978), 20–25.
12. MUNSELL, A. H. *A Color Notation.* Munsell Color Company, 1939.
13. OSTLE, B., AND MENSING, R. W. *Statistics in Research.* The Iowa State University Press, 1975.
14. SCHWARZ, M. W. An empirical evaluation of interactive colour selection techniques. M. Math. dissertation, Dept. of Computer Science, Univ. of Waterloo, Waterloo, Ont., Canada, 1985.
15. SMITH, A. R. Color gamut transform pairs. *Comput. Graph. 12*, 3 (Aug. 1978), 12–19.
16. WALPOLE, R. E., AND MYERS, R. H. *Probability and Statistics for Engineers and Scientists.* McMillan, New York, 1978.
17. WARE, C., AND BEATTY, J. C. Using colour as a tool in discrete data analysis. Rep. CS-85-21, Dept. of Computer Science, Univ. of Waterloo, Waterloo, Ont., Canada.
18. WYSZECKI, G., AND STILES, W. S. *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. Wiley, New York, 1982.

Received September 1985; revised March 1987; accepted June 1987