6.2 HUMAN COLOR PERCEPTION

To be able to describe colors, we need to know how people respond to them. Human perception of color is a complex function of context; illumination, memory, object identity, and emotion can all play a part. The simplest question is to understand which spectral radiances produce the same response from people under simple viewing conditions (Section 6.2.1). This yields a simple, linear theory of color matching that is accurate and extremely useful for describing colors. We sketch the mechanisms underlying the transduction of color in Section 6.2.2.

6.2.1 Color Matching

The simplest case of color perception is obtained when only two colors are in view on a black background. In a typical experiment, a subject sees a colored light—the *test light*—in one half of a split field (Figure 6.5). The subject can then adjust a mixture of lights in the other half to get it to match. The adjustments involve changing the intensity of some fixed number of *primaries* in the mixture. In this form, a large number of lights may be required to obtain a match, but many different adjustments may yield a match.

Write T for the test light, an equals sign for a match, the weights—which are non-negative—as w_i, and the primaries P_i. A match can then written in an algebraic form as

$$T = w_1 P_1 + w_2 P_2 + \cdots ,$$

meaning that test light T matches the particular mixture of primaries given by $(w_1, w_2, \ldots)$. The situation is simplified if *subtractive matching* is allowed: In subtractive matching, the viewer can add some amount of some primaries to the *test light* instead of to the match. This can be written in algebraic form by allowing the weights in the expression above to be negative.

Trichromacy It is a matter of experimental fact that for most observers only three primaries are required to match a test light. There are some caveats. First, subtractive matching

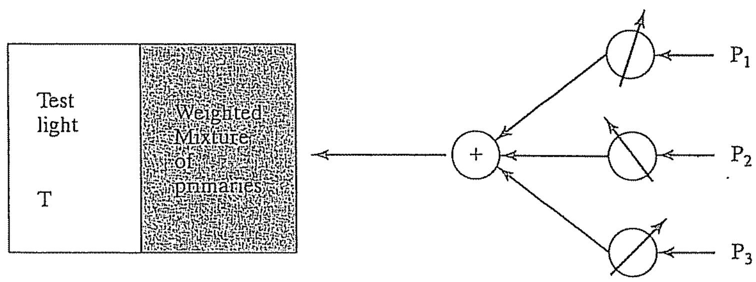

Figure 6.5 Human perception of color can be studied by asking observers to mix colored lights to match a test light shown in a split field. The drawing shows the outline of such an experiment. The observer sees a test light T and can adjust the amount of each of three primaries in a mixture displayed next to the test light. The observer is asked to adjust the amounts so that the mixture looks the same as the test light. The mixture of primaries can be written as $w_1 P_1 + w_2 P_2 + w_3 P_3$; if the mixture matches the test light, then we write $T = w_1 P_1 + w_2 P_2 + w_3 P_3$. It is a remarkable fact that for most people three primaries are sufficient to achieve a match for many colors and for all colors if we allow subtractive matching (i.e., some amount of some of the primaries is mixed with the test light to achieve a match). Some people require fewer primaries. Furthermore, most people choose the same mixture weights to match a given test light.

must be allowed; second, the primaries must be independent, meaning that no mixture of two of the primaries may match a third. This phenomenon is known as the principle of *trichromacy*. It is often explained by assuming that there are three distinct types of color transducer in the eye. Recently, evidence has emerged from genetic studies to support this view (Nathans, Piantanida, Eddy, Shows and Hogness, 1986*a*, Nathans, Thomas and Hogness, 1986*b*). Given the same primaries and test light, most observers select the *same* mixture of primaries to match that test light. This phenomenon is usually explained by assuming that the three distinct types of color transducers are common to most people. The direct evidence from genetic studies seems to support this view, too.

Grassman's Laws Under the circumstances described, matching is (to an accurate approximation) linear. This yields *Grassman's laws*.

First, if we mix two test lights, then mixing the matches will match the result—that is, if

$$T_a = w_{a1}P_1 + w_{a2}P_2 + w_{a3}P_3$$

and

$$T_b = w_{b1}P_1 + w_{b2}P_2 + w_{b3}P_3,$$

then

$$T_a + T_b = (w_{a1} + w_{b1})P_1 + (w_{a2} + w_{b2})P_2 + (w_{a3} + w_{b3})P_3.$$

Second, if two test lights can be matched with the same set of weights, then they will match each other—that is, if

$$T_a = w_1P_1 + w_2P_2 + w_3P_3$$

and

$$T_b = w_1P_1 + w_2P_2 + w_3P_3,$$

then

$$T_a = T_b.$$

Finally, matching is linear: if

$$T_a = w_1P_1 + w_2P_2 + w_3P_3,$$

then

$$kT_a = (kw_1)P_1 + (kw_2)P_2 + (kw_3)P_3$$

for non-negative k.

Exceptions Given the same test light and set of primaries, most people use the same set of weights to match the test light. This, trichromacy and Grassman's laws are about as true as any law covering biological systems can be. The exceptions include the following:

- people with aberrant color systems as a result of genetic ill fortune (who may be able to match everything with fewer primaries);

- people with aberrant color systems as a result of neural ill-fortune (who may display all sorts of effects, including a complete absence of the sensation of color);
- some elderly people (whose choice of weights differ from the norm because of the development of macular pigment in the eye);
- very bright lights (whose hue and saturation look different from less bright versions of the same light);
- and very dark conditions (where the mechanism of color transduction is somewhat different than in brighter conditions).

6.2.2 Color Receptors

Trichromacy suggests that there are profound constraints on the way color is transduced in the eye. One hypothesis that satisfactorily explains this phenomenon is to assume that there are three distinct types of receptor in the eye that mediate color perception. Each of these receptors turns incident light into neural signals. It is possible to reason about the sensitivity of these receptors from color matching experiments. If two test lights that have different spectra look the same, they must have the same effect on these receptors.

The Principle of Univariance The *principle of univariance* states that the activity of these receptors is of one kind (i.e., they respond strongly or weakly, but do not signal the wavelength of the light falling on them). Experimental evidence can be obtained by carefully dissecting light-sensitive cells and measuring their responses to light at different wavelengths or by reasoning backward from color matches. Univariance is a powerful idea because it gives us a good and simple model of human reaction to colored light: Two lights will match if they produce the same receptor responses, *whatever their spectral radiances.*

Because the system of matching is linear, the receptors must be linear. Let us write p_k for the response of the kth type of receptor, $\sigma_k(\lambda)$ for its sensitivity, $E(\lambda)$ for the light arriving at the receptor, and Λ for the range of visible wavelengths. We can obtain the overall response of a receptor by adding up the response to each separate wavelength in the incoming spectrum so that

$$p_k = \int_{\Lambda} \sigma_k(\lambda) E(\lambda) d\lambda.$$

Rods and Cones Anatomical investigation of the retina shows two types of cell that are sensitive to light differentiated by their shape. The light-sensitive region of a *cone* has a roughly conical shape, whereas that in a *rod* is roughly cylindrical. Cones largely dominate color vision and completely dominate the fovea. Cones are somewhat less sensitive to light than rods are, meaning that in low light, color vision is poor and it is impossible to read (one doesn't have sufficient spatial precision because the fovea isn't working).

Studies of the genetics of color vision support the idea that there are three types of cone differentiated by their sensitivity (in the large; there is some evidence that there are slight differences from person to person within each type). The sensitivities of the three different kinds of receptor to different wavelengths can be obtained by comparing color matching data for normal observers with color matching data for observers lacking one type of cone. Sensitivities obtained in this fashion are shown in Figure 6.6. The three types of cone are properly called *S cones*, *M cones* and *L cones* (for their peak sensitivity being to short, medium, and long wavelength light, respectively). They are occasionally called blue, green and red cones; however, this is bad practice, because the sensation of red is definitely not caused by the stimulation of red cones and so on.

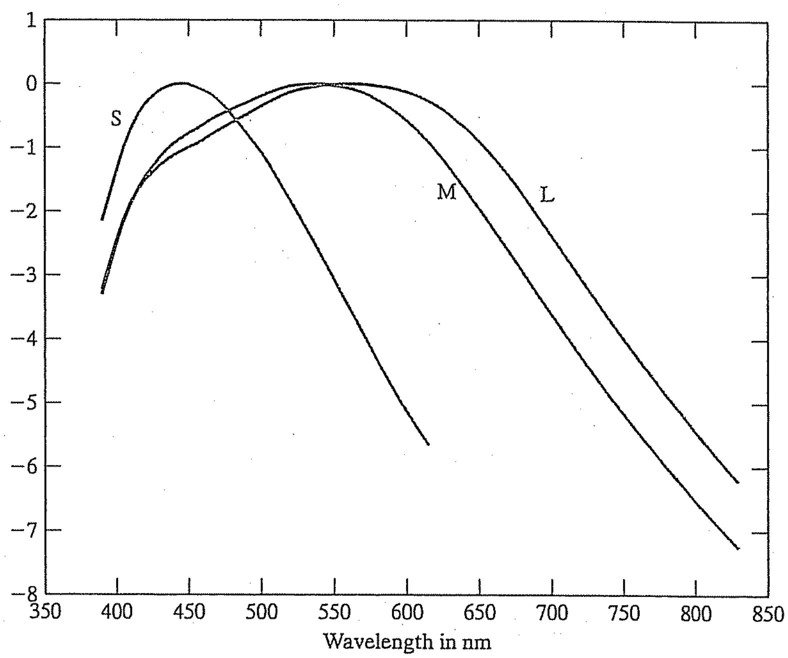

Figure 6.6 There are three types of color receptor in the human eye, usually called *cones*. These receptors respond to all photons in the same way, but in different amounts. The figure shows the log of the relative spectral sensitivities of the three kinds of color receptor in the human eye. The first two receptors—sometimes called the *red* and *green* cones respectively, but more properly named the *long*- and *medium-wavelength* receptors—have peak sensitivities at quite similar wavelengths. The third receptor (*blue* cone or, more properly, *short-wavelength* receptor) has a different peak sensitivity. The response of a receptor to incoming light can be obtained by summing the product of the sensitivity and the spectral radiance of the light over all wavelengths. *Figures plotted from data disseminated by the Color and Vision Research Laboratories database, compiled by Andrew Stockman and Lindsey Sharpe, and available at* `http://www-cvrl.ucsd.edu/index.htm`.

6.3 REPRESENTING COLOR

Describing colors accurately is a matter of great commercial importance. Many products are closely associated with specific colors—for example, the golden arches, the color of various popular computers and the color of photographic film boxes—and manufacturers are willing to go to a great deal of trouble to ensure that different batches have the same color. This requires a standard system for talking about color. Simple names are insufficient because relatively few people know many color names, and most people are willing to associate a large variety of colors with a given name.

6.3.1 Linear Color Spaces

There is a natural mechanism for representing color: Agree on a standard set of primaries and then describe any colored light by the three values of weights that people would use to match the light using those primaries. In principle, this is easy to use. To describe a color, we set up and perform the matching experiment and transmit the match weights. Of course, this approach extends to give a representation for surface colors as well if we use a standard light for illuminating the surface (and if the surfaces are equally clean, etc.).

Performing a matching experiment each time we wish to describe a color can be practical. For example, this is the technique used by paint stores; you take in a flake of paint, and they mix paint, adjusting the mixture until a color match is obtained. Paint stores do this because complicated scattering effects within paints mean that predicting the color of a mixture can be quite difficult. However, Grassman's laws mean that mixtures of colored lights—at least those seen in a simple display—mix *linearly*, which means that a much simpler procedure is available.

Color Matching Functions When colors mix linearly, we can construct a simple algorithm to determine which weights would be used to match a source of some known spectral radiance given a fixed set of primaries. The spectral radiance of the source can be thought of as a weighted sum of single wavelength sources. Because color matching is linear, the combination of primaries that matches a weighted sum of single wavelength sources is obtained by matching the primaries to each of the single wavelength sources and then adding up these match weights.

If we have a record of the weight of each primary required to match a single wavelength source—a set of *color matching functions*—we can obtain the weights used to match an arbitrary spectral radiance. The color matching functions—which we shall write as $f_1(\lambda)$, $f_2(\lambda)$, and $f_3(\lambda)$—can be obtained from a set of primaries P_1, P_2 and P_3 by experiment. Essentially, we tune the weight of each primary to match a unit radiance source at every wavelength. We then obtain a set of weights, one for each wavelength, for matching a unit radiance source $U(\lambda)$. We can write this process as

$$U(\lambda) = f_1(\lambda)P_1 + f_2(\lambda)P_2 + f_3(\lambda)P_3$$

(i.e., at each wavelength λ, $f_1(\lambda)$, $f_2(\lambda)$, and $f_3(\lambda)$ give the weights required to match a unit radiance source at that wavelength).

The source—which we write $S(\lambda)$—is a sum of a vast number of single wavelength sources, each with a different intensity. We now match the primaries to each of the single wavelength sources and then add up these match weights, obtaining

$$\begin{aligned} S(\lambda) &= w_1 P_1 + w_2 P_2 + w_3 P_3 \\ &= \left\{\int_\Lambda f_1(\lambda)S(\lambda)d\lambda\right\} P_1 + \left\{\int_\Lambda f_2(\lambda)S(\lambda)d\lambda\right\} P_2 + \left\{\int_\Lambda f_3(\lambda)S(\lambda)d\lambda\right\} P_3. \end{aligned}$$

General Issues for Linear Color Spaces Linear color naming systems can be obtained by specifying primaries—which imply color matching functions—or by specifying color matching functions—which imply primaries. It is an inconvenient fact of life that, if the primaries are real lights, at least one of the color matching functions is negative for some wavelengths. This is not a violation of natural law; it just implies that subtractive matching is required to match some lights, whatever set of primaries is used. It is a nuisance, though.

One way to avoid this problem is to specify color matching functions that are everywhere positive (which guarantees that the primaries are imaginary because for some wavelengths their spectral radiance is negative).

Although this looks like a problem—how would one create a real color with imaginary primaries?—it isn't, because color naming systems are hardly ever used that way. Usually, we would simply compare weights to tell whether colors are similar, and for that purpose it is enough to know the color matching functions. A variety of different systems have been standardised by the CIE (the *commission international d'éclairage*, which exists to make standards on such things).

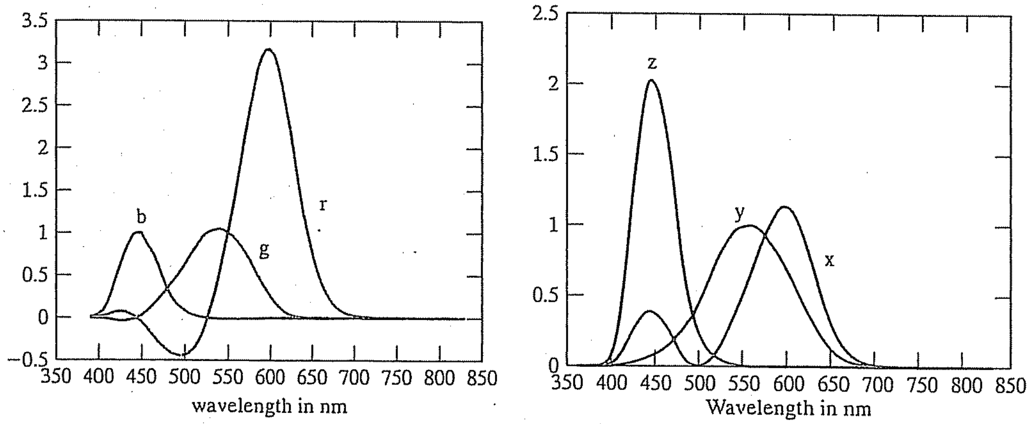

Figure 6.7 On the left, color matching functions for the primaries for the RGB system. The negative values mean that subtractive matching is required to match lights at that wavelength with the RGB primaries. On the right, color matching functions for the CIE X, Y, and Z primaries; the color matching functions are everywhere positive, but the primaries are not real. *Figures plotted from data disseminated by the Color and Vision Research Laboratories database, compiled by Andrew Stockman and Lindsey Sharpe, and available at* `http://www-cvrl.ucsd.edu/index.htm`.

The CIE XYZ Color Space The *CIE XYZ color space* is one quite popular standard. The color matching functions were chosen to be everywhere positive, so that the coordinates of any real light are always positive. It is not possible to obtain CIE X, Y, or Z primaries because for some wavelengths the value of their spectral radiance is negative. However, given color matching functions alone, one can specify the XYZ coordinates of a color and hence describe it.

Linear color spaces allow a number of useful graphical constructions that are more difficult to draw in three dimensions than in two, so it is common to intersect the XYZ space with the

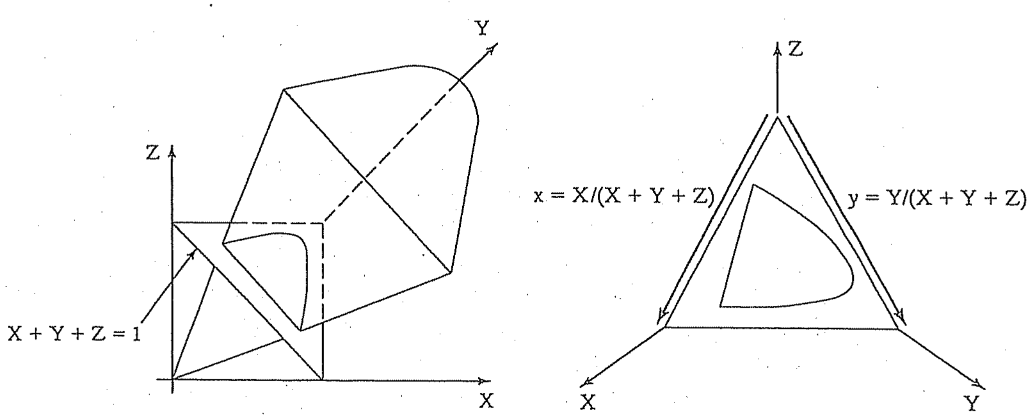

Figure 6.8 The volume of all visible colors in CIE XYZ coordinate space is a cone whose vertex is at the origin. Usually it is easier to suppress the brightness of a color, which we can do because to a good approximation perception of color is linear, and we do this by intersecting the cone with the plane $X + Y + Z = 1$ to get the CIE xy space shown in Figures 6.9 and 6.10

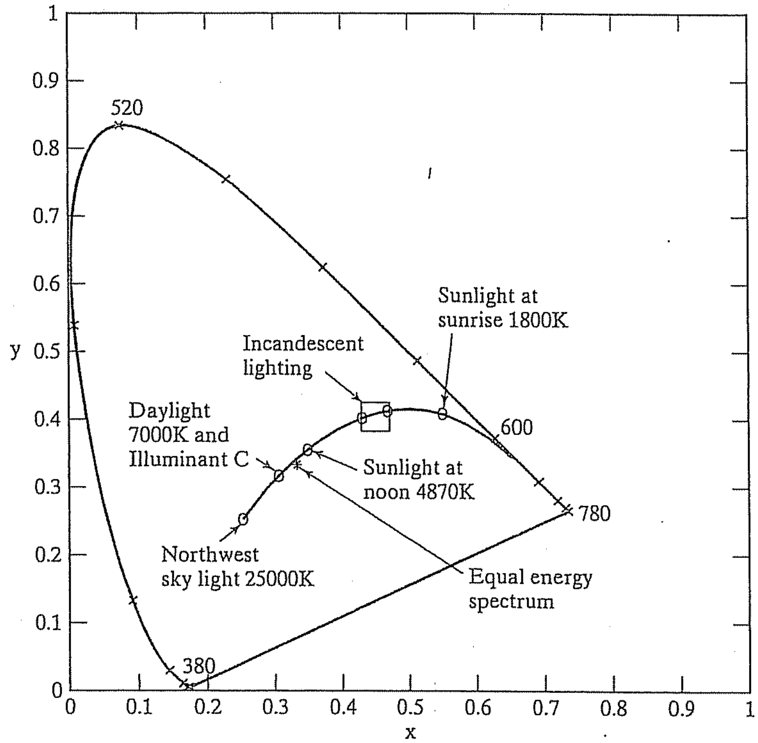

Figure 6.9 The figure shows a constant brightness section of the standard 1931 standard CIE xy color space. This space has two coordinate axes. The curved boundary of the figure is often known as the *spectral locus*; it represents the colors experienced when lights of a single wavelength are viewed. The figure shows a locus of colors due to black-body radiators at different temperatures and a locus of different sky colors. Near the center of the diagram is the neutral point, the color whose weights are equal for all three primaries. CIE selected the primaries so that this light appears achromatic. Generally, colors that lie farther away from the neutral point are more saturated—the difference between deep red and pale pink—and hue—the difference between green and red—as one moves around the neutral point.

plane $X + Y + Z = 1$ (as shown in Figure 6.8) and draw the resulting figure using coordinates

$$(x, y) = \left(\frac{X}{X + Y + Z}, \frac{Y}{X + Y + Z} \right).$$

This space is shown in Figures 6.9 and 6.10. Some more useful constructions appear in Figure 6.11. CIE xy is widely used in vision and graphics textbooks and in some applications, but is usually regarded by professional colorimetrists as out of date.

The RGB Color Spaces Color spaces are normally invented for practical reasons, and so a wide variety exist. The *RGB color space* is a linear color space that formally uses single wavelength primaries (645.16 nm for R, 526.32nm for G, and 444.44nm for B; see Figure 6.7). Informally, RGB uses whatever phosphors a monitor has as primaries. Available colors are usually represented as a unit cube—usually called the *RGB cube*—whose edges represent the R, G, and B weights. The cube is drawn in Figure 6.12.

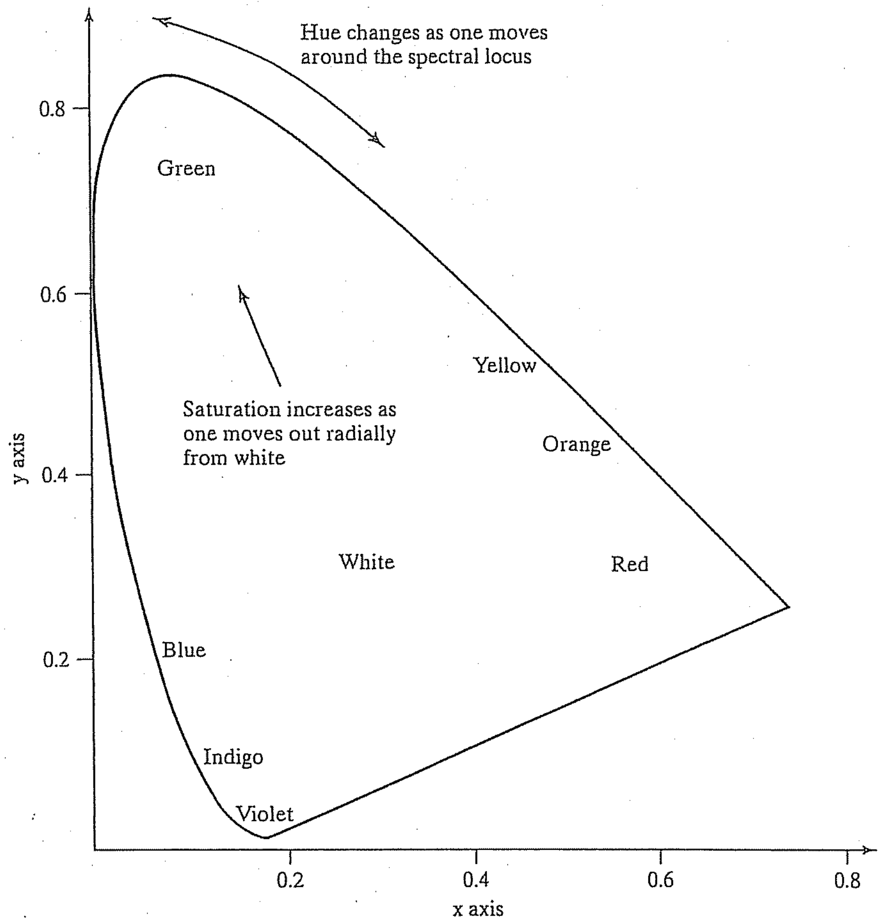

Figure 6.10 The figure shows a constant brightness section of the standard 1931 standard CIE xy color space, with color names marked on the diagram. Generally, colors that lie farther away from the neutral point are more saturated—the difference between deep red and pale pink—and hue—the difference between green and red—as one moves around the neutral point.

CMY and Black Intuition from one's finger-painting days suggests that the primary colors should be red, yellow and blue, and that red and green mix to make yellow. The reason this intuition doesn't apply to monitors is that it is about pigments—which mix subtractively—rather than about lights. Pigments remove color from incident light, which is reflected from paper. Thus, red ink is really a dye that absorbs green and blue light—incident red light passes through this dye and is reflected from the paper.

Color spaces for this kind of subtractive matching can be quite complicated. In the simplest case, mixing is linear (or reasonably close to linear) and the *CMY space* applies. In this space, there are three primaries: *cyan* (a blue-green color); *magenta* (a purplish color), and *yellow*. These primaries should be thought of as subtracting a light primary from white light; cyan is $W - R$ (white − red); magenta is $W - G$ (white − green), and yellow is $W - B$ (white − blue). Now the appearance of mixtures may be evaluated by reference to the RGB color space. For example, cyan and magenta mixed give

$$(W - R) + (W - G) = R + G + B - R - G = B,$$

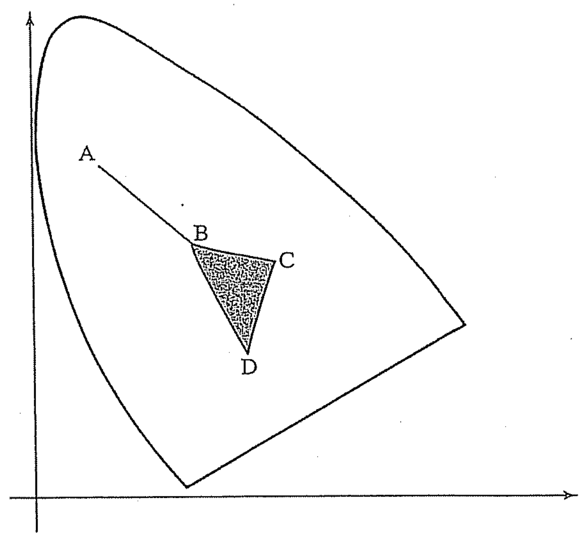

Figure 6.11 The linear model of the color system allows a variety of useful constructions. If we have two lights whose CIE coordinates are A and B, all the colors that can be obtained from non-negative mixtures of these lights are represented by the line segment joining A and B. In turn, given B, C, and D, the colors that can by obtained by mixing them lie in the triangle formed by the three points. This is important in the design of monitors—each monitor has only three phosphors, and the more saturated the color of each phosphor, the bigger the set of colors that can be displayed. This also explains why the same colors can look quite different on different monitors. The curvature of the spectral locus gives the reason that no set of three real primaries can display all colors without subtractive matching.

that is, blue. Notice that $W + W = W$ because we assume that ink cannot cause paper to reflect more light than it does when uninked. Practical printing devices use at least four inks (cyan, magenta, yellow, and black) because mixing color inks leads to a poor black, it is difficult to ensure good enough registration between the three color inks to avoid colored haloes around text, and color inks tend to be more expensive than black inks. Getting really good results from a color printing process is still difficult: Different inks have significantly different spectral properties, different papers have different spectral properties too, and inks can mix nonlinearly.

6.3.2 Non-linear Color Spaces

The coordinates of a color in a linear space may not necessarily encode properties that are common in language or are important in applications. Useful color terms include: *hue*—the property of a color that varies in passing from red to green; *saturation*—the property of a color that varies in passing from red to pink; and *brightness* (sometimes called *lightness* or *value*)—the property that varies in passing from black to white. For example, if we are interested in checking whether a color lies in a particular range of reds, we might wish to encode the hue of the color directly.

Another difficulty with linear color spaces is that the individual coordinates do not capture human intuitions about the topology of colors; it is a common intuition that hues form a circle, in the sense that hue changes from red through orange to yellow and then green and from there to cyan, blue, purple, and then red again. Another way to think of this is to think of local hue relations: Red is next to purple and orange; orange is next to red and yellow; yellow is next

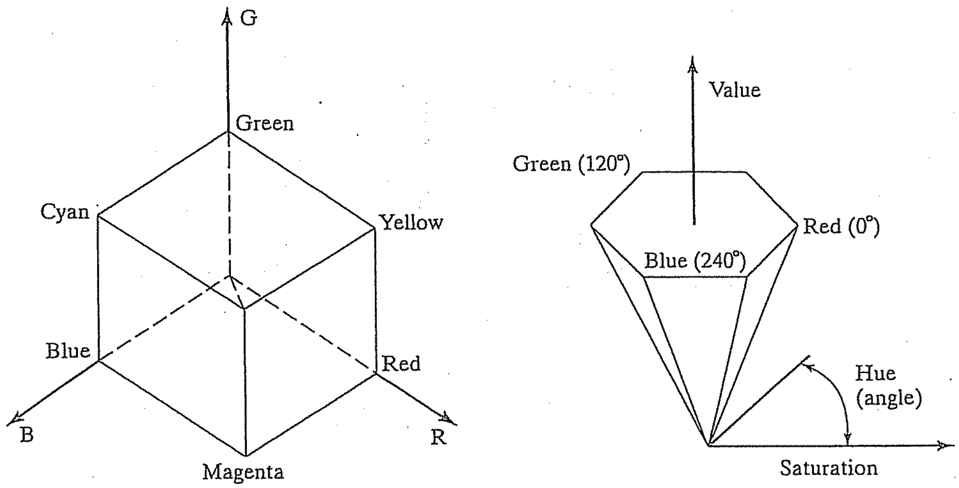

Figure 6.12 On the left, we see the RGB cube; this is the space of all colors that can be obtained by combining three primaries (R, G, and B—usually defined by the color response of a monitor) with weights between zero and one. It is common to view this cube along its neutral axis—the axis from the origin to the point (1, 1, 1)—to see a hexagon, shown in the middle. This hexagon codes hue (the property that changes as a color is changed from green to red) as an angle, which is intuitively satisfying. On the right, we see a cone obtained from this cross-section, where the distance along a generator of the cone gives the value (or brightness) of the color, angle around the cone gives the hue, and distance out gives the saturation of the color.

to orange and green; green is next to yellow and cyan; cyan is next to green and blue; blue is next to cyan and purple; and purple is next to blue and red. Each of these local relations works, and globally they can be modeled by laying hues out in a circle. This means that no individual coordinate of a linear color space can model hue because that coordinate has a maximum value that is far away from the minimum value.

Hue, Saturation, and Value A standard method for dealing with this problem is to construct a color space that reflects these relations by applying a nonlinear transformation to the RGB space. There are many such spaces. One, called *HSV space* (for hue, saturation, and value), is obtained by looking down the center axis of the RGB cube. Because RGB is a linear space, brightness—called *value* in HSV—varies with scale out from the origin. We can flatten the RGB cube to get a 2D space of constant value and for neatness deform it to be a hexagon. This gets the structure shown in Figure 6.12, where hue is given by an angle that changes as one goes round the neutral point and saturation changes as one moves away from the neutral point.

There are a variety of other possible changes of coordinate from between linear color spaces, or from linear to nonlinear color spaces (the recent book of Fairchild, 1998 is a good reference). There is no obvious advantage to using one set of coordinates over another (particularly if the difference between coordinate systems is just a one–one transformation) unless one is concerned with coding, bit rates, and the like, or with perceptual uniformity.

Uniform Color Spaces Usually one cannot reproduce colors exactly. This means it is important to know whether a color difference would be noticeable to a human viewer; it is generally useful to compare the significance of small color differences. It is usually dangerous to try and compare large color differences; consider trying to answer the question, "Is the blue patch more different from the yellow patch than the red patch is from the green patch?".

One can determine *just noticeable differences* by modifying a color shown to an observer until they can only just tell it has changed in a comparison with the original color. When these differences are plotted on a color space, they form the boundary of a region of colors that are indistinguishable from the original colors. Usually ellipses are fitted to the just noticeable differences. It turns out that in CIE xy space these ellipses depend quite strongly on where in the space the difference occurs, as the Macadam ellipses in Figure 6.13 illustrate.

This means that the size of a difference in (x, y) coordinates, given by $((\Delta x)^2+(\Delta y)^2)^{(1/2)}$, is a poor indicator of the significance of a difference in color (if it was a good indicator, the ellipses representing indistinguishable colors would be circles). A *uniform color space* is one in which the distance in coordinate space is a fair guide to the significance of the difference between two colors—in such a space, if the distance in coordinate space were below some threshold, a human observer would not be able to tell the colors apart.

A more uniform space can be obtained from CIE XYZ by using a projective transformation to skew the ellipses; this yields the *CIE u'v' space*, illustrated in Figure 6.14. The coordinates are:

$$(u', v') = \left(\frac{4X}{X + 15Y + 3Z}, \frac{9Y}{X + 15Y + 3Z}\right).$$

Generally, the distance between coordinates in u', v' space is a fair indicator of the significance of the difference between two colors. Of course, this omits differences in brightness.

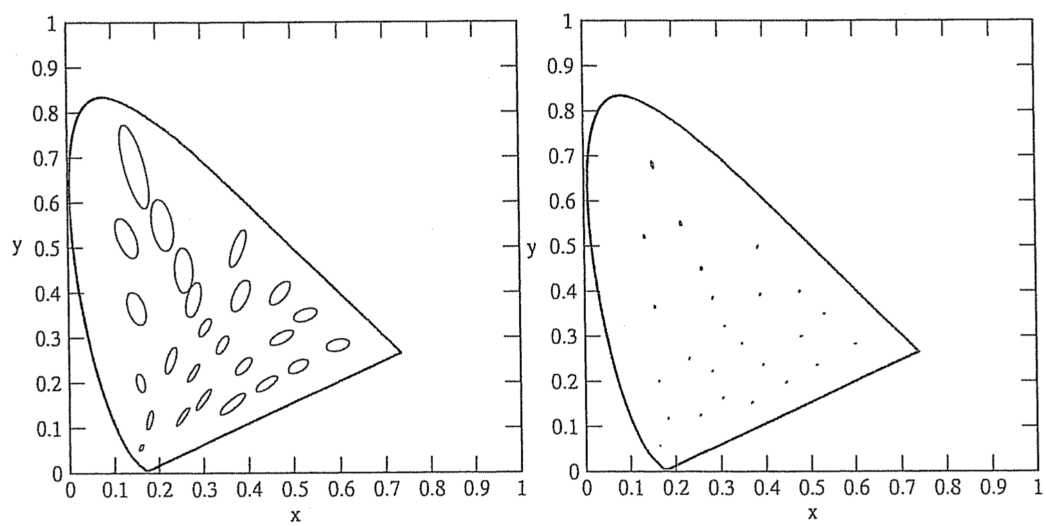

Figure 6.13 This figure shows variations in color matches on a CIE x, y space. At the center of the ellipse is the color of a test light; the size of the ellipse represents the scatter of lights that the human observers tested would match to the test color; the boundary shows where the just noticeable difference is. The ellipses in the figure on the **left** have been magnified 10x for clarity; on the **right** they are plotted to scale. The ellipses are known as MacAdam ellipses after their inventor. Notice that the ellipses at the top are larger than those at the bottom of the figure, and that they rotate as they move up. This means that the magnitude of the difference in x, y coordinates is a poor guide to the difference in color. Ellipses are plotted using data from MacAdam (1942).

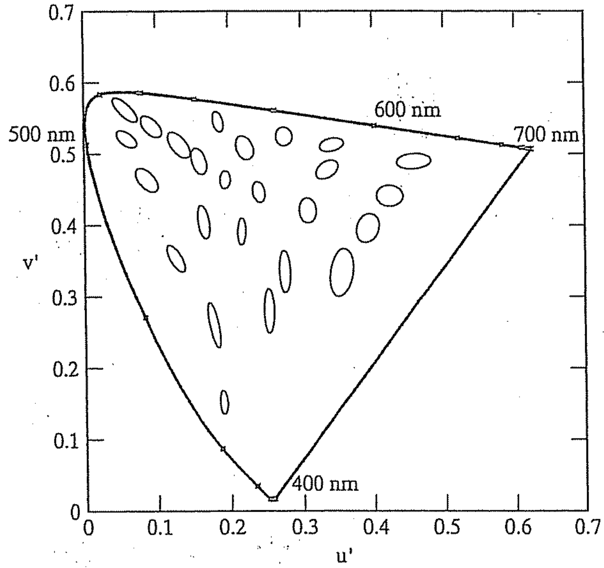

Figure 6.14 This figure shows the CIE 1976 u', v' space, which is obtained by a projective transformation of CIE x, y space. The intention is to make the MacAdam ellipses uniformly circles—this would yield a uniform color space. A variety of nonlinear transforms can be used to make the space more uniform (see Fairchild (1998) for details)

CIE LAB is now almost universally the most popular uniform color space. Coordinates of a color in LAB are obtained as a nonlinear mapping of the XYZ coordinates:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16$$

$$a^* = 500\left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right]$$

$$b^* = 200\left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right]$$

Here X_n, Y_n, and Z_n are the X, Y, and Z coordinates of a reference white patch. The reason to care about the LAB space is that it is substantially uniform. In some problems, it is important to understand how different two colors will look *to a human observer*, and differences in LAB coordinates give a good guide.

6.3.3 Spatial and Temporal Effects

Predicting the appearance of complex displays of color (i.e., a stimulus that is more interesting than a pair of lights) is difficult. If the visual system has been exposed to a particular illuminant for some time, this causes the color system to adapt—a process known as *chromatic adaptation*. Adaptation causes the color diagram to skew, in the sense that two observers, adapted to different illuminants, can report that spectral radiosities with quite different chromaticities have the same color. Adaptation can be caused by surface patches in view. Other mechanisms that are signifi-