The Foundations of Color Measurement and Color Perception

Brian A. Wandell
Department of Psychology
Stanford University
Stanford, California 94305

March 20, 1996

Abstract

This tutorial is an introductory review of color science. The logical and experimental basis of the CIE standard observer and standard observer tools are emphasized. In addition, some aspects of device-calibration are discussed. Finally, the shortcomings of the CIE standard observer are reviewed in order to ready the student for the more advanced tutorials on color appearance.

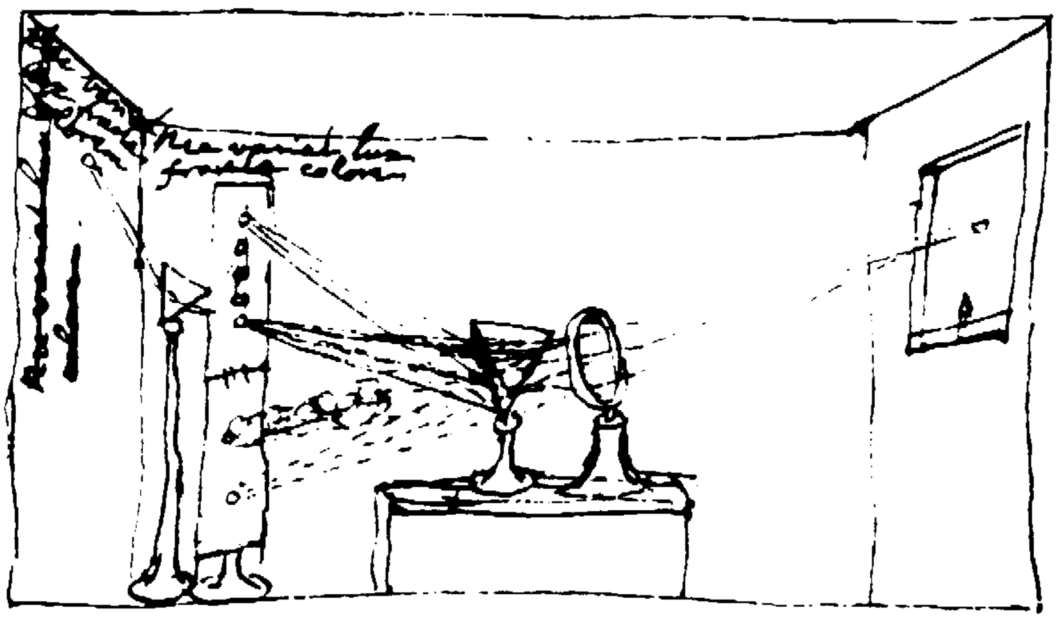

Figure 1: This figure is Isaac Newton's sketch summarizing various experiments he performed.

Introduction

Some of the most beautiful experiments in science are summarized in the drawing in Figure 1. The figure contains a sketch, by Isaac Newton, of the apparatus he used to investigate the properties of light.

Newton placed a shutter in the window in his room at Cambridge. He left only a small hole in the shutter. He then waited for a sunny day (which was the tedious part of the experiment in Cambridge England!). With the sun a very distant object, and the hole in the window shutter quite small, Newton had a point source to illuminate his apparatus. The key elements of the apparatus, featured prominently in the center of the picture, are the lens and prism.

Newton's experiments demonstrate that sunlight is a mixture of various fundamental components, that is components that cannot be further decomposed. By passing the light through the prism, he separated the light into the elements of the rainbow. The drawing illustrates two experiments that show that the rays of the rainbow were fundamental constituents of light. In the first experiment, shown at the top left, he proved that one did not obtain a rainbow simply by passing light through a prism.

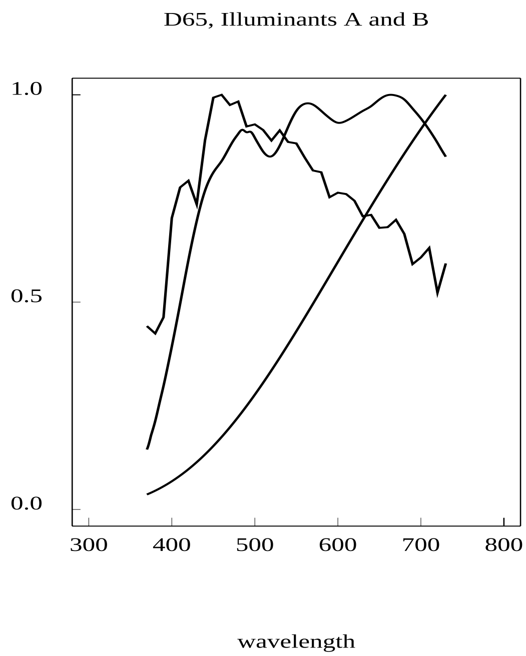

Figure 2: Relative spectral power distributions of several CIE (Committee Internationale d'Eclairage) standard illuminants.

Passing the separated rays through a second prism did not create another rainbow. In the second experiment (bottom left) he proved that the process was reversible; the basic character of the light is left intact after passing through a prism.

We now know that Newton succeeded in decomposing the sunlight into its spectral wavelength components. The prism separates the rays because the prism bends the spectral components by different amounts. Light with relatively long wavelengths, say around 700 nm, appears red when viewed against a dark background. Light with relatively short wavelengths, say around 400 nm, appears blue when viewed against a dark background. Shorter wavelengths of light are refracted more strongly than longer wavelengths.

When we measure the energy at different wavelengths in a light source, our results must be characterized by a fairly long series of numbers describing all of the sample measurements. To capture the variation of energy with wavelength, we may have to sample quite finely. If we sample from 400 nm to 700nm in 1 nm steps, sampling the visible wavelength region requires about 300 measurements. If we sample in 10 nm steps, we require about 30 measurements. How finely we sample must depend upon the goals of the measurement as well as our expectations about how rapidly the energy will be varying as a function of wavelength.

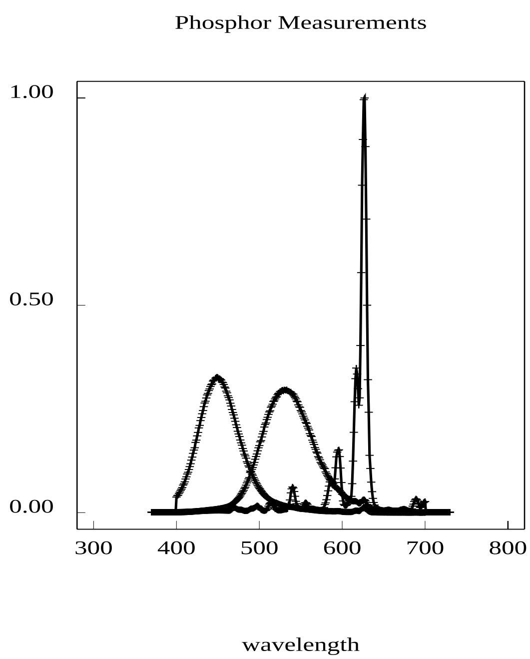

Figure 3: The relative spectral power distribution of the color phosphors of a display in my laboratory.

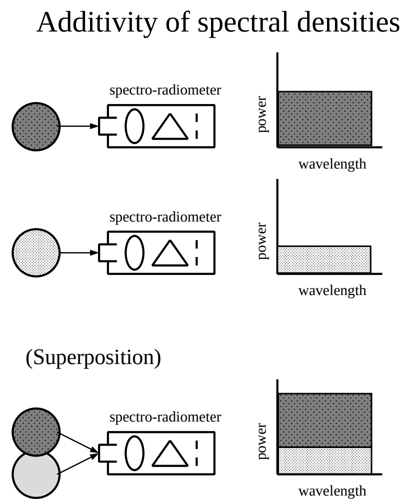

Figure 4: Measurements of the spectral power densities of illuminants satisfy the principle of superposition. If we superimpose two lights, spectral power distribution is the sum of the component measurements. The principle is particularly important for calibrating CRT displays.

Figure 2 contains measurements of the energy per unit wavelength for several standard light sources. The curves describe the relative *spectral power distributions* of these standard lights. Many naturally occurring illuminants, such as blackbody radiators, are fairly smooth functions of wavelength. [Judd et al., 1964] There are some important light sources whose spectral power distribution vary rapidly with wavelength. For example, the most commonly used red CRT phosphor (see Figure 3), contain spectral power distribution with sharp spikes.

Principle of Superposition: Lights

Perhaps the most important property of the measurement of spectral power distributions, indeed the property that makes it possible to calibrate emissive displays easily, is that the mixture of lights satisfies the principle of superposition. Recall that superposition, illustrated in Figure 4, is the key property of a linear system. To test superposition, we measure the spectral power distributions of two lights separately. We then measure the spectral power distribution of the mixture of the lights. Superposition informs us that the spectral power distribution of the mixture is the sum of the spectral power distributions of the components.

Superposition is important for monitor calibration because the light emitted from a monitor consists of the mixture of three different spectral power distributions. The three spectral power distributions are determined by the properties of the phosphors placed on the CRT face. There are two important regularities of these phosphor emissions. First, as we increase the amount of light emitted by a single phosphor, the spectral power distribution always has the same basic form; at all levels the phosphor output is a scaled copy of a single spectral power distribution. Second, when we form outputs by turning on combinations of the phosphors, the monitor output consists of weighted mixtures of basic spectral power distributions of the individual phosphors. [Cowan and Rowell, 1988] [Brainard, 1989] [Post and Calhoun, 1989] Because of these two regularities, we can characterize all the spectral power distributions emitted from the monitor by taking weighted sums of the three functions.

As a specific example, suppose we measure the spectral power distribution of each of the phosphors at their maximum level. If we measure the light using a 1 nm spacing over the interval 400nm to 700nm, we obtain 301 sample measurements. We write the measurements in column vectors, $\mathbf{r}, \mathbf{g}$, and $\mathbf{b}$ These vectors contain our sampled measurements of the red, green and blue phosphor spectral power distributions, $r(\lambda), g(\lambda)$, and $b(\lambda)$. By superposition, we know that light from the monitor screen is the weighted sum of the light emitted from the three phosphors. For example, if we set each of the phosphor intensities to a fraction of their maximal value, then the light emitted from the monitor will be

$$c(\lambda) = e_r r(\lambda) + e_g g(\lambda) + e_b b(\lambda) \tag{1}$$

where $\mathbf{e} = (e_r, e_g, e_b)^t$ is the fractional intensity of the phosphor emission. The set of possible monitor spectral power distributions is restricted to the weighted sums in equation 1 with the entries of $\mathbf{e}$ ranging between 0.0 and 1.0.

Ordinarily one additional step is required to calibrate: we must specify the relationship between the frame-buffer values in the computer and the phosphor intensities. Typically the intensity is a power function of the frame-buffer value, with an exponent near 2.2. This part of the calibration procedure is referred to as *gamma correction* of the monitor.

It is often convenient to represent the sum in equation 1 using matrix notation, as illustrated in the matrix tableau in Figure 5. We construct a matrix whose columns contain the monitor phosphor SPD's. These are the measured vectors $\mathbf{r}$, $\mathbf{g}$ and $\mathbf{b}$. The spectral density of light emitted from the monitor is always the weighted sum of these three vectors; the three weights are equal to the intensities of the phosphors. Since monitor spectral power distributions can only be weighted sums of the columns of the matrix, the set of monitor spectral power distributions is quite different compared to the spectral power distributions we observe in nature. Indeed, you can probably see that no weighted sum of the three monitor phosphors equals the smooth blackbody radiator shown in Figure 2. Yet, the colors we perceive on most television monitors can be adjusted to match the appearance of any of these illuminants. The ability to

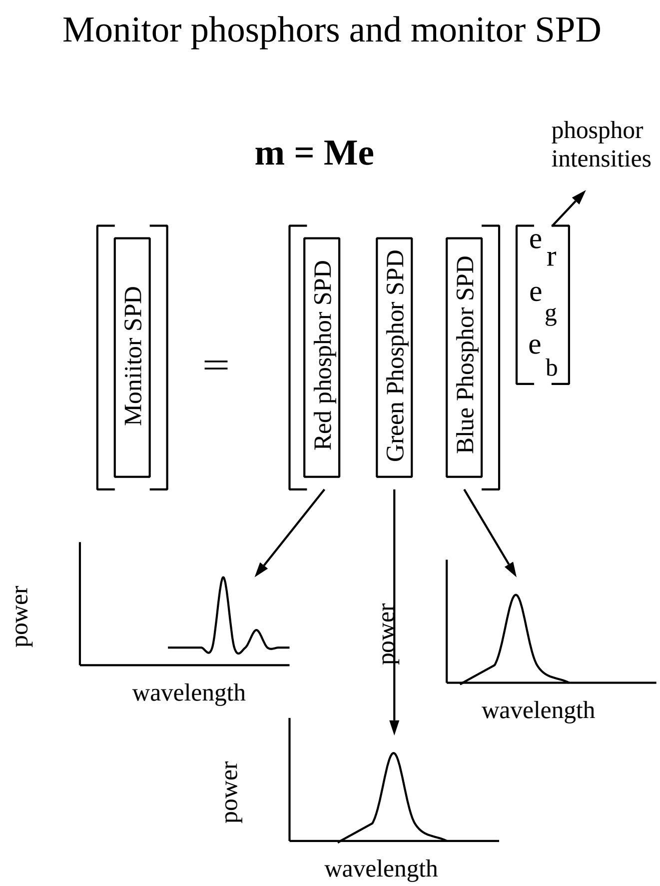

Figure 5: The spectral power distribution emitted from a small region of a video monitor is the weighted sum of the light from the three monitor phosphors. The spectral power distribution of the emitted lights, therefore, can be described using only three parameters (the fractional gun intensities) and a simple matrix product.

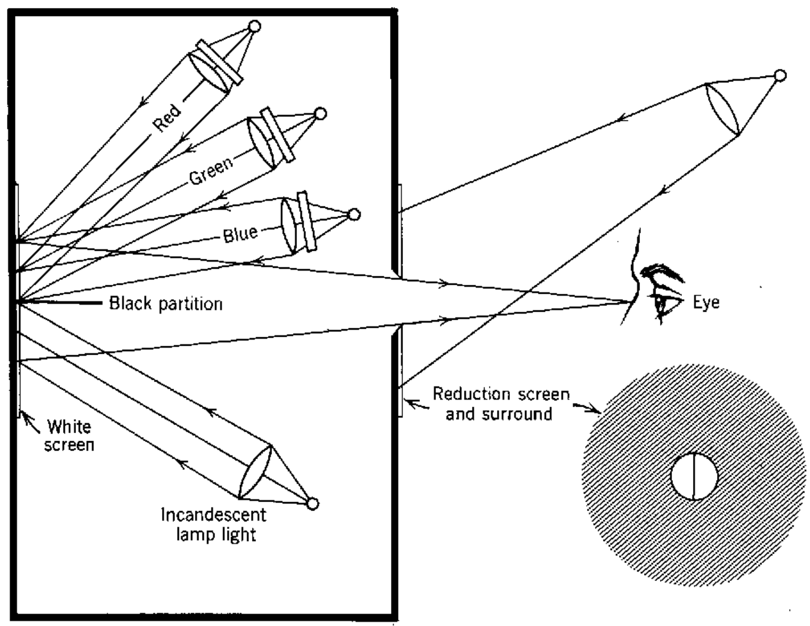

Figure 6: In the color-matching experiment the observer sees a circular field composed of two halves, shown at the bottom right of the figure. One half of the field contains an arbitrary test light. The other half of the field contains the mixture of three fixed lights called the primaries. This figure is taken from Judd & Wyszecki, figure 1.12.

arrange a perceptual match, despite our inability to arrange a physical match, is due to the wavelength encoding properties of the visual photoreceptors. These are revealed by the color-matching experiment.

The Color-Matching Experiment

The fundamental experiment of human color vision, the experiment upon which color science is founded, is the *color-matching* experiment. Among other things, this experiment provides the scientific basis for the reproducing colors on television monitors.

The spatial arrangement of the stimuli used in the color-matching experiment is illustrated in Figure 6. The observer views two adjacent visual fields, typically arranged so that from the observer's point of view these appear as the right and left halves of a circle. On one side, called the *test field*, the observer is presented with a test light. The test light is the input to the experiment. The test light may consist of any spectral power distribution, $t(\lambda)$, which we represent by the column vector $\mathbf{t}$.

The other side of the bipartite field is called the *matching field*. This matching field

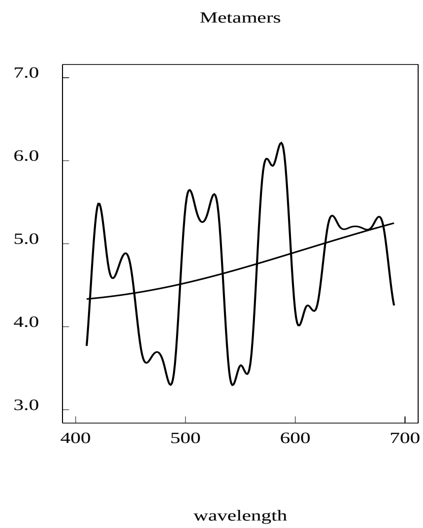

Figure 7: The two spectral power distributions are metameric, that is they appear the same to most observers. Both have the same color appearance as the disk of the sun.

consists of the mixture of a small number of *primary lights*. Just as in the design of a monitor, the relative spectral power distribution of each primary light remains constant throughout the experiment. Only the intensity of the primary light varies. We can use the matrix tableau description in Figure 5 of monitor spectral power distributions to describe the spectral power distributions possible by mixing the primary lights.

In the color-matching experiment the observer adjusts the intensity of the primary lights so that the two sides of the bipartite field appear identical. The smallest number of primary lights needed to assure that a match can be obtained depends upon the system being studied. For rod vision, a single primary light suffices. For cone vision, three primary lights are required.

In the color-matching experiment observers adjust the primary intensities so that the two sides appear identical. Although the two sides match perceptually, they can be very different physically. Since the light on the primary side of the bipartite field is the weighted sum of the primary lights, and the light on the test side of the bipartite field can be any light at all, the two sides of the field are not generally close to being physical matches.

Pairs of lights that are visual matches, but not physical matches, are called *metamers*.

Figure 7 contains a pair of spectral power distributions that match visually but differ physically, i.e. a pair of metamers. As the comparison in figure 7 illustrates, our wavelength encoding of the retinal image fails to discriminate between certain enormous differences in light spectral power distributions. Yet, we are capable of making certain very fine wavelength discriminations. Capturing which discriminations we can make, and which we cannot, is one of the important contributions of CIE standard observer theory.

Caveats

There are some additional experimental considerations that arise in the color-matching experiment. First, concerning the selection primaries, there is a matter of common sense. When we select the three primary lights, $primaryi$, we should chose lights that are *independent*: that is, no additive mixture of two of the primary lights should be a visual match to the third primary. For example, it would be unreasonable to choose the second primary light with the same spectral composition as the first, differing only in intensity. Should we do so, we could always replace the second light by simply increasing the intensity of the first primary light. This choice of a second primary would add nothing to the range of visual matches we can obtain. Similarly, a primary that can be matched by a mixture of the first two adds nothing. We must choose our primary lights so that they are independent of one another; apart from the requirement of independence, any three primary lights can be used.

Second, as we perform the experiment we will find that there are some test lights that cannot be matched by an additive mixture of the three primaries. This will be true for any set of primaries we choose. Yet, if we move one, or perhaps two, of the primary lights from the matching field to the test field, we will find that a match can be obtained.

When we move, say, the first primary light to the test field, we obtain a visual match of the form

$$\mathbf{t} + e_1\mathbf{p}_1 = e_2\mathbf{p}_2 + e_3\mathbf{p}_3 \tag{2}$$

We suspect that this match is equivalent to the match

$$\mathbf{t} = -e_1\mathbf{p}_1 + e_2\mathbf{p}_2 + e_3\mathbf{p}_3 \quad . \tag{3}$$

But we cannot instrument this match because we cannot create lights with negative intensities, $-e_1$. As a convention that is consistent with the spirit of additivity, when we are forced to mix the primary with the test we record the intensity as a negative number.

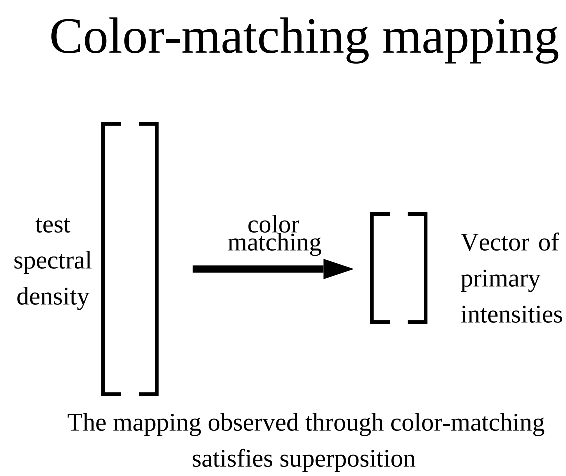

Figure 8: The color-matching experiment defines a mapping from the test light spectral power distribution to the three dimensional vector of primary intensities.

The Principle of Superposition: Color-Matching

The color-matching experiment defines a mapping illustrated in Figure 8. The input to the color-matching experiment is the spectral power distribution of the test light. The output of the experiment is the three dimensional vector of primary intensity settings.

By testing the principle of superposition, we can ask whether the experimentally defined mapping is linear. To test superposition, we follow the logic illustrated in Figure 9. We first match the test light $\mathbf{t}$ by adjusting the primary intensities to $\mathbf{e}$. We then match test signal $\mathbf{t}'$ by adjusting the three primary intensities $\mathbf{e}'$. To test superposition for the color-matching experiment we match the sum of the first two inputs, $\mathbf{t} + \mathbf{t}'$; superposition predicts that the primary intensities of the match will be $\mathbf{e} + \mathbf{e}'$.

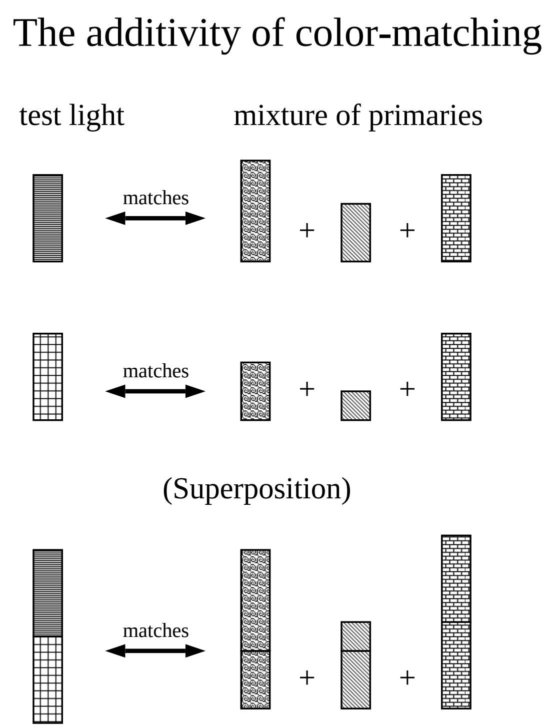

Figure 9: The color-matching experiment satisfies the principle of superposition.

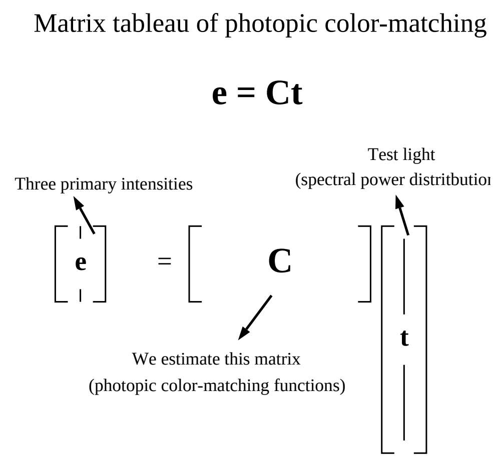

Figure 10: Because of linearity, the mapping from test signal to primary intensities can be predicted by a matrix multiplication. We call the system matrix **C**. We can estimate the matrix entries by performing a series of color matches.

The input-output relationships of photopic color-matching obey the principle of superposition. We honor the person who first understood the importance of superposition in color-matching by calling this property *Grassmann's Additivity Law*.

The color-matching experiment establishes a linear relationship between the input stimulus, **t**, and the measured primary intensities, **e**. Remarkably, then, we can characterize all of the observer's matches using a simple linear model illustrated in Figure 10. Because superposition is satisfied, we can describe the observed relationship between the input signal and the primary lights by a matrix multiplication. The system matrix, **C**, maps the input signal **t** to the output measurement, **e**. An easy way to estimate the matrix **C** is to have the observer set a series of color-matches to monochromatic lights. The match to each monochromatic test light defines a single column of the system matrix, **C**. The rows of the system matrix obtained by a series of matches to monochromatic lights are called the *color-matching functions*. The entries the system matrix define the intensities of the three primary lights needed to obtain a match at each wavelength. Each column contains the three intensities needed to match a single, monochromatic light. Each row contains the complete set of intensities of one primary required to match all the different monochromatic lights.

Each set of the color-matching functions is defined with respect to a particular set of primary lights. Figure 11 plots the color-matching functions using three

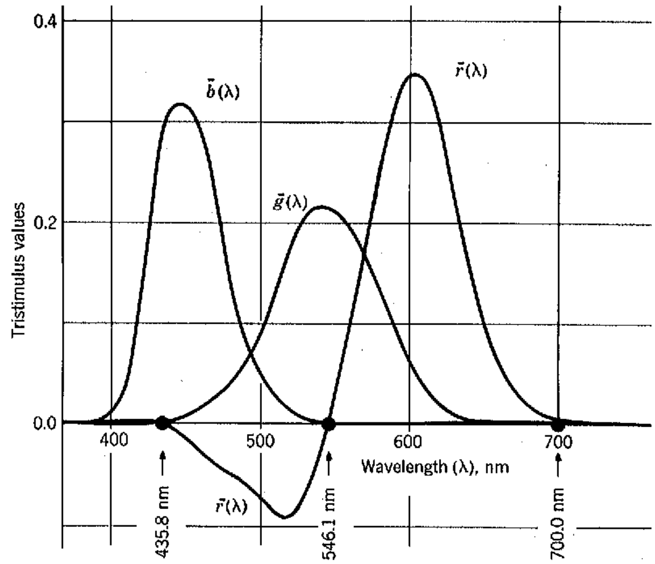

Figure 11: The horizontal axis is the wavelength of the test light. The vertical axis is the energy of a matching light. Each curve defines the energy for one of the three primary lights (wavelengths 700, 546.1, 435.8) needed to match the unit energy test light at the test wavelength. The relative power of the primaries was 72.1:1.4:1.0. This figure is taken from Wyszecki and Stiles, 2nd edition, figure 4(3.2.3) page 124.

monochromatic primary lights at 700 nm, 546.1 nm and 435.8 nm. Each curve plots the intensity of a primary as a function of the wavelength of the test light. The intensity of the red primary, at 700nm, is negative over a large region of test light wavelengths, indicating that to obtain a match over this range of test lights the 700 nm primary light was added to the test field.

Matching on Monitors

To see why the color-matching functions are extremely important in color technology, consider how we use the color-matching functions to decide how to adjust the intensities of the phosphors on a television screen to match *the appearance* of a light incident at a remote camera. Remember from the matrix tableau in Figure 5 that the set of spectral power distributions emitted from a monitor can be described by the weighted sum of the three phosphor spectral power distributions. The matrix of phosphor spectral power distributions in Figure 5 is $\mathbf{M}$. The phosphors play the role of the primary lights in the color-matching experiment, so we call them $\mathbf{e}$. The light emitted from the monitor is the weighted sum of the monitor phosphors, $\mathbf{Me}$.

We wish to match a test light, $\mathbf{t}$, by adjusting the monitor phosphor intensities. When seen in a common context, as in the color-matching experiment, light from a monitor

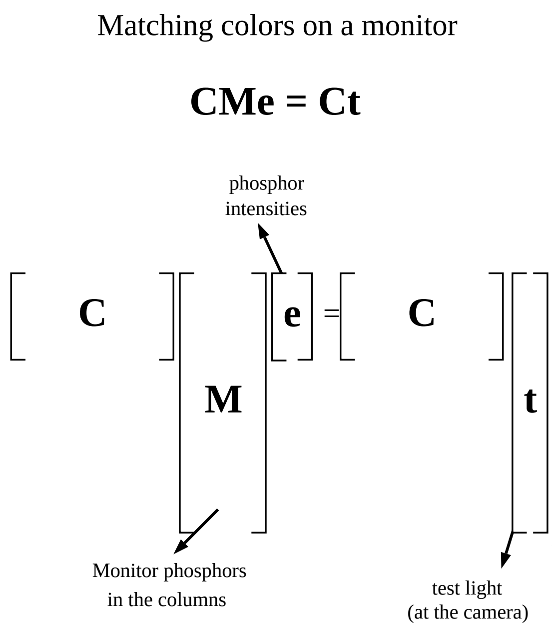

Figure 12: We can use the color-matching functions to establish color-matches between a test light and the light displayed on a monitor. We adjust the intensities of the phosphor emissions so that the color-match set to the television emission would be the same as the color-match set to the original test light.

will match the test light if both are matched by the same vector of primary intensities. The primary intensities needed to match $\mathbf{t}$ can be calculated from the color-matching functions as $\mathbf{Ct}$. The possible lights we can create on the monitor are described by $\mathbf{Me}$. We need to find the monitor intensities such that $\mathbf{C\ Me}$ equals $\mathbf{Ct}$. This relationship is illustrated in matrix tableau in Figure 12.

We can calculate the primary intensities using linear methods. The matrix $\mathbf{C\ M}$ is a three by three matrix. If we invert this matrix $[\mathbf{C\ M}]^{-1}$, we can calculate the monitor intensities needed to match any test light since $\mathbf{e} = [\mathbf{C\ M}]^{-1}\mathbf{Ct}$. Everything on the right side of the equation is known, so we can use this equation to solve for the primary intensities needed to match any test light.

Of course, we will have serious difficulties in those cases in which the solution to $\mathbf{e}$ contains negative entries or intensity values that are beyond the range of the monitor. In these cases, we say the matches are out of the *device gamut.* A good choice of monitor phosphors reduces the number of out-of-gamut colors, although no choice of phosphors can reproduce all possible colors. Various ad hoc procedures exist for handling gamut errors, though at present this is more of an art than a science.

Uniqueness

Suppose one research group measures the color-matching functions with respect to the primary lights $\mathbf{p}_i$, while a second group uses three different primary lights $\mathbf{p}'_i$. The two sets of primary lights will lead to different color-matching functions, $\mathbf{C}$ and $\mathbf{C}'$. How will the two sets of color matching functions be related?

To discover the intensities of the second group of primaries required to match the first group, we can use the same logic as in matching appearance using a television monitor. The matching field consists of the weighted sum of the three primary lights, $\mathbf{p}'_i$. We can form a matrix, analogous to the television monitor matrix, whose columns contain the spectral power distributions of these three primary lights. Call this matrix $\mathbf{P}'$. This matrix plays a role analogous to the monitor matrix.

Now, for any test light $\mathbf{t}$, the first primary intensities will be set to $\mathbf{e} = \mathbf{Ct}$. To match the appearance of the test light using the second group of primary lights, we must find a vector of three intensities $\mathbf{e}'$ such that

$$\mathbf{Ct} = \mathbf{e} = \mathbf{CP}'\mathbf{e}' \qquad (4)$$

From equation 4 we see that the primary intensities used in first primary system, $\mathbf{e}$, are related to the second primary intensities by the three by three linear transformation $\mathbf{CP}'$. The columns of this matrix contain the result of a simple experiment. Remember that each column of $\mathbf{P}'$ is the spectral power distribution of one of the primary lights

$\mathbf{p}'$. Thus, the first column of $\mathbf{CP}'$ is the vector of intensities of the first group of primaries needed to match $\mathbf{p}'_1$. Similarly the second and third columns of $\mathbf{CP}'$ contain the intensities of the first group of primaries needed to match the corresponding primaries in $\mathbf{p}'$.

Equation 4 tells us the primary intensities $\mathbf{p}$ and $\mathbf{p}'$ are related by this linear transformation. The columns of the two groups of color-matching functions, $\mathbf{C}$ and $\mathbf{C}'$, are simply the matches obtained using monochromatic lights. Therefore, the columns in these two matrices must be related by this same linear transformation.

$$(\mathbf{CP}')\mathbf{C}' = \mathbf{C} \tag{5}$$

The uniqueness result is commonly summarized by saying that the photopic color-matching functions are only unique up to a free linear transformation.

Having derived the uniqueness result from first principles, we can also see why this uniqueness result make sense. The color-matching experiment informs us that two lights $\mathbf{t}$ and $\mathbf{t}'$ match when $\mathbf{Ct} = \mathbf{Ct}'$ If we use a different set of color-matching functions, $\mathbf{QC}$ where $\mathbf{Q}$ is any invertible three by three matrix, then we will find that if $\mathbf{Ct} = \mathbf{Ct}'$, so too $\mathbf{QCt} = \mathbf{QCt}'$. The color-matching functions inform us about equality between test light representations. The additional linear transformation, $\mathbf{Q}$, does not alter any of the equalities established by the color-matching functions, and thus is permissible.

Color Representation Standards

When the members of the CIE (Committee Internationale d'Eclairage) met in 1931, they decided to adopt one particular set of color-matching functions as an international standard. The color-matching functions are denoted as $xbar(\lambda)$, $ybar(\lambda)$ and $zbar(\lambda)$. These three functions form the entries of the rows of the standardized system matrix $\mathbf{C}$. Figure 13 plots these three functions.

The three values we calculate using the CIE color-matching functions are usually referred to as X, Y and Z. Rather than specifying absolute units or precise wavelength sampling rules, the CIE advises us on how to calculate these values in various settings. Using summation notation, rather than matrices, we can express the relationship between the various quantities as follows.

$$X = k \sum_{\lambda} \bar{x}(\lambda) t(\lambda) d\lambda \tag{6}$$

$$Y = k \sum_{\lambda} \bar{y}(\lambda) t(\lambda) d\lambda \tag{7}$$

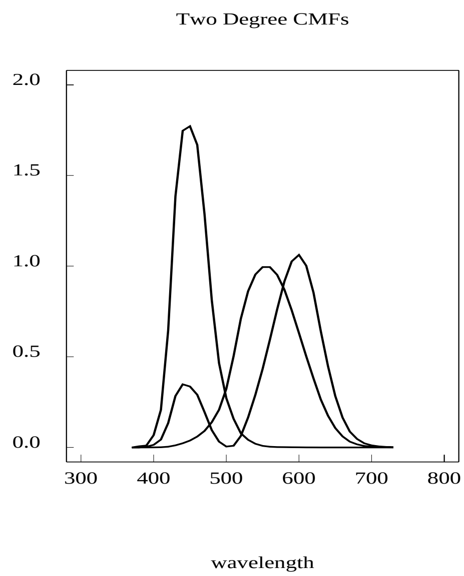

Figure 13: The CIE 1931 two degree color-matching functions.

and

$$Z = k \sum_{\lambda} \bar{z}(\lambda) t(\lambda) d\lambda \tag{8}$$

The factor $d\lambda$ refers to the wavelength sampling separation, which we will assume to be 1 and thus ignore.

The quantity k is a normalizing factor whose use depends upon whether we are measuring surfaces or lights. When we measure surfaces, k is conventionally set equal to 100 divided by the value of Y computed for a perfect white diffuser under the same viewing conditions. When measuring surfaces, the units are relative and are scaled by the intensity of the illumination source. By this definition of k, the largest value of Y for any surface will be 100.

To specify the tristimulus coordinates of an emissive source, the coordinates are normalized so that the Y coordinate can be given in terms of absolute photometric units (e.g. lumens) that are defined in other CIE standards. In the case of emissive stimuli, all the units are absolute. A discussion of these conventions may be found in the two excellent reference books by Judd and Wyszecki, and by Wyszecki and Stiles. [Judd and Wyszecki, 1975] [Wyszecki and Stiles, 1982]

In either case, the vector (X, Y, Z) is called the *tristimulus coordinates* of the test light or sample. The Y component of the vector is also called the luminance of the measured sample. This measurement is roughly correlated, with but not equal to, the perceived brightness of the sample.

The uniqueness result informs us that the XYZ representation is but one of many possible representations of the test light. These particular functions were chosen for several reasons. One reason is that the Y value is a rough approximation to the brightness of monochromatic lights of equal size and duration. A second important reason is that the curves are non-negative. Non-negativity has important consequences for the design of instruments to measure the tristimulus coordinates.

But like almost any standards decision, there are some irritating aspects of the XYZ color-matching functions as well. Empirically, it is impossible to find any three lights so that we can match all other lights using only a positive mixture of primaries. Yet, as you can see from Figure 13, the standardized curves are all positive. Thus, there is no set of physically realizable primary lights that by direct measurement will yield the XYZ color-matching functions' primary lights that yield these functions must have negative energy at some wavelengths. Such lights cannot be instrumented, so you may hear it said that the primary lights for the XYZ functions are imaginary. I prefer to say that there are no physically realizable primary lights for these color-matching functions.

Apart from the very approximate relationship between Y and brightness, there is almost nothing intuitive about the XYZ color-matching functions. While they have served us quite well as a technical standard, they have served us quite poorly in explaining the discipline to new students and colleagues or as an intuition about color appearance.

Chromaticity coordinates

Often we would like to have a way of expressing the perceptual features of a light, such as its brightness and color. How can we use the XYZ representation of the stimulus to a form that is suggestive of the color appearance of the light?

As we increase the intensity of a light, the brightness will change substantially. If one measures carefully, as we increase the intensity of a light seen on a dark background we find that its color appearance, that is its hue and saturation, change as well. But in most circumstances most observers agree that as intensity grows the color changes are not as perceptually salient as the brightness changes.

This observation suggests that we can obtain a rough characterization of the color appearance of a light, viewed against a dark background, by eliminating that part of the representation due to the intensity of the light. From the linear relationship between the test light spectral power distribution and the XYZ vector, you can verify that if we scale the intensity of a light by a factor a, so that the light $\mathbf{t}$ becomes the light $a\mathbf{t}$, the XYZ representation is scaled from (X, Y, Z) to (aX, aY, aZ). To find an

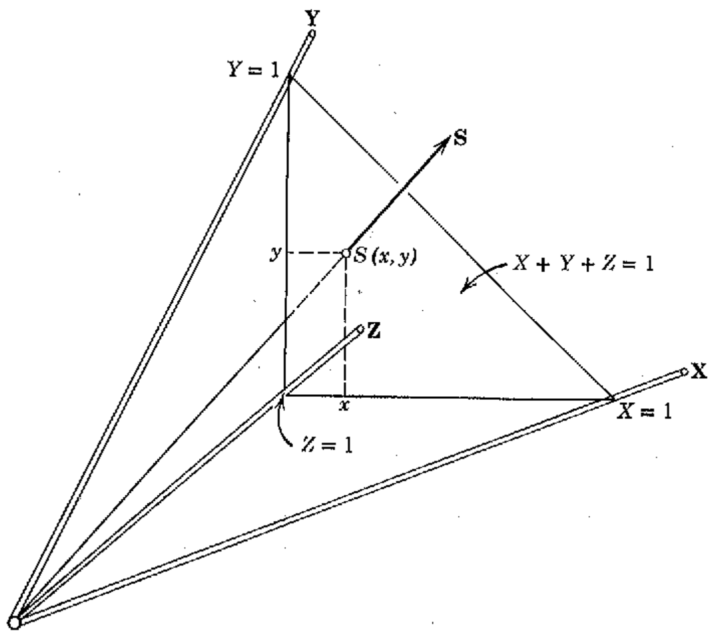

Figure 14: Lights that differ only by an intensity scale factor fall along a ray through the origin in the XYZ representation. Tristimulus points that fall along a ray through the origin are mapped to the same chromaticity coordinates. This figure is from Judd & Wyszecki, fig. 1.20.

representation of the color, we would like to arrange a transformation such that $\mathbf{t}$ and $a\mathbf{t}$ are represented by the same value. We can use a projective transformation, as in

$$x = \frac{X}{X+Y+Z} \quad , \quad y = \frac{Y}{X+Y+Z} \quad , \quad z = \frac{Z}{X+Y+Z} \quad . \tag{9}$$

This transformation defines a new set of values, (x, y, z) that we can use to represent the light. By the definition of equation 9, $x + y + z = 1.0$, implying that there are only two free values in the vector. The projective transformation permits us to specify the test light using just the pair of numbers (x, y). These values are called the *chromaticity coordinates* of the test light. The projective mapping in equation 9 assigns lights differing only in intensity to the same chromaticity coordinates.

The mapping from XYZ coordinates to chromaticity coordinates is shown graphically in Figure 14. Lights of different intensities fall along rays through the origin of the XYZ coordinate system. The projective mapping in equation 9 assigns each tristimulus coordinate to a point in the plane, defined by the equation $X + Y + Z = 1$, where the ray through the origin intersects the plane.

The chromaticity coordinate representation has some simple properties that make it useful. First, the values of the chromaticity coordinates permits us to make a rough

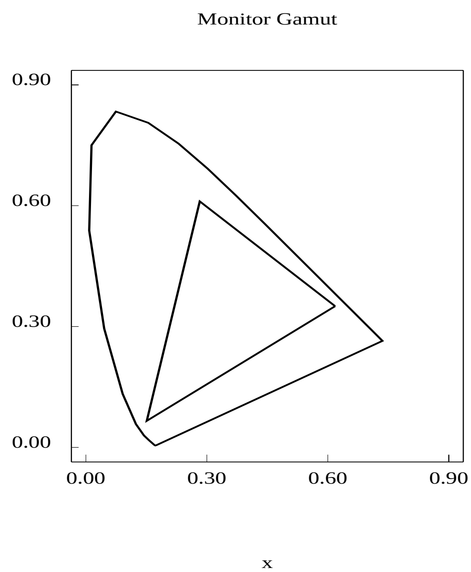

Figure 15: The chromaticity diagram represents light using two values that are more closely related to hue and saturation than brightness. Because of the linearity of the color representation, and because chromaticity is obtained by a projective mapping, the chromaticities of mixtures of lights fall on a line between the chromaticities of the individual lights. This figure shows the chromaticities of the three monitor phosphors and the range of chromaticities that can be obtained by positive mixtures of the phosphors.

guess about the hue and saturation of the test light. The coordinates do a good job in describing the color appearance when the light is a fairly large spot, of moderate intensity, viewed on a dark background. As these conditions are violated, the chromaticity coordinates become less useful as a guide to color appearance.

Second, because the chromaticity coordinates are a projective transformation they inherit some of the simple properties of the full linear representation. The most important is this. Suppose we begin with two lights with different chromaticities, say two television phosphors. The chromaticity of any light formed by mixing the two phosphor outputs will fall on a line connecting the chromaticity coordinates of the individual phosphors. Because of this property, we can graph the range of chromaticity coordinates that we can obtain by the mixture of three test lights by drawing a triangular bounding region in the chromaticity plane. This is called the *color gamut* of the television.

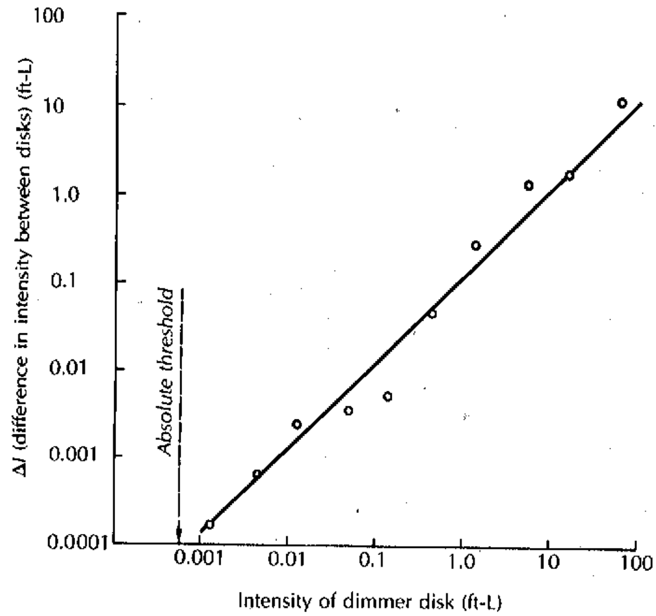

Figure 16: The increment necessary to discriminate a pair of lights depends upon the intensity levels of the two lights. The size of the increment is approximately proportional to the intensity level. This relationship is called *Weber's Law*. Under some conditions this approximation holds well, as these data from Cornsweet and Pinsker show. But the approximation fails under other viewing conditions. This figure is from Cornsweet's book, figure 4.11.

Color Difference Measurement

The XYZ representation informs us only that two lights, or two surfaces under a common illuminant, match each other. But when two lights do not match, the representation does not offer us much guidance on how large the perceptual mismatch will be. Indeed, based on the experimental methods used to obtain the XYZ representation, there is no reason to expect that a metric, such as the vector lengths between points in the representation, should be predictive of the discriminability of different lights. Nothing about the experimental procedure informs us about discriminability. To define a space with a metric that predicts discriminability requires more experiments that measure discriminability and then an elaboration of the representation.

There are a number of ways to see that equal vector differences in the XYZ representation do not represent equal perceptual differences. One way is illustrated in by the data in Figure 16. The experiment required subjects to discriminate the brightness of two lights with the same apparent color (white) but different intensities.

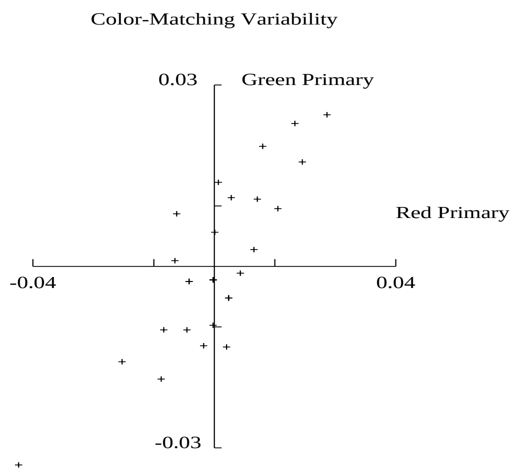

Figure 17: When we perform a color-match multiple times, the observed matches form a cloud in tristimulus space and in chromaticity coordinates.

Suppose one light has intensity I, then the experimenter determined the increment, $\Delta(I)$ necessary to add so that the subject could discriminate I from $I + \Delta(I)$. The data in the Figure show that the increment required to discriminate the two lights is not constant; rather, the increment increases linearly, $\Delta(I) = kI$. The measurements I report here were made by Cornsweet and Pinsker, [Cornsweet and Pinsker, 1965] but the basic phenomenon was discovered early in the last century and is called *Weber's Law*. The measurements show that lights separated by the same vector difference are not equally discriminable.

The data in Figure 17 illustrate a second way in which we can see that equal vector differences do not correspond to equal discriminability. This graph plots the results of a number of color-matches to a single test light. The value of the constant side of the bipartite field is plotted at the center of the graph, and the points plot the projection of the matches onto one plane in three-dimensional color space. These particular measurements are reported by Poirson and me. [Poirson and Wandell, 1990] As you can see from the repeated matches, the precision of the matches is much greater in some color directions than others.

MacAdam [MacAdam, 1942] suggested that we use the variance of the matches to assess discriminability in different vector directions. Directions where the variance of the match is large, the subject has little sensitivity. Directions where the variance is small, the subject has greater sensitivity. Macadam, Brown and their colleagues at

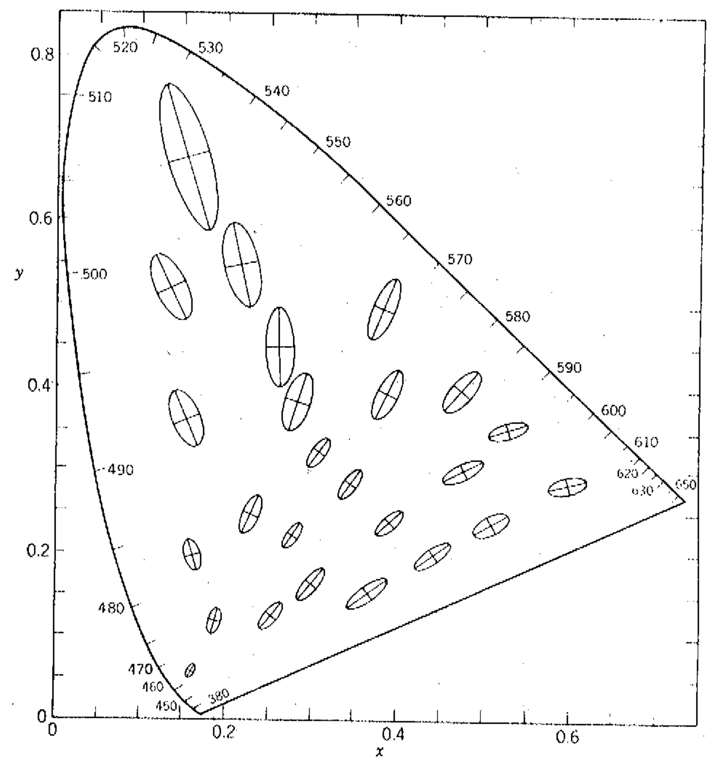

Figure 18: Equal variance contours derived from color-matches. Each ellipsoidal contour is centered around the chromaticity of the standard light. The size of the ellipsoids is ten times larger than the actual measurements. This figure is from Boynton [Boynton, 1979] Figure 8.11, page 281.

Kodak showed that we can summarize the random deviations from a perfect match fairly well using a trivariate normal distribution. Equal variance contours for these distributions form ellipsoids in XYZ space and they fall close to an ellipse in the projection onto chromaticity coordinates.

If we repeat these measurements for many different test lights, we can measure many ellipses centered at many different points in color space. The collection of ellipses summarizes color discriminability under these types of viewing conditions. One set of MacAdam's measurements are shown in Figure 18. The ellipses are plotted at ten times their measured size to make them visible. Subjects' ability to judge the difference between a pair of lights does not simply depend on the XYZ vector difference between the two lights.

CIELUV and CIELAB

Many people have contributed to the development of a new representations in which the vector difference between two lights will inform us about how discriminable the two lights will be. These representations also go somewhat beyond the fine-scale discrimination measurements. In addition to the discrimination data the spaces are also based on data that describe how similar a pair of lights appears, even though the lights are plainly discriminable.

There are two widely used representations that seek to transform the XYZ representation into a metric space, or what is often called a *uniform color space*. The transformation to a uniform color space strives to achieve two goals.

- the Euclidean distance between points in the uniform color space should be equal to the perceived color difference.
- The uniform color space ought to be calculated from XYZ data measured under any different illuminant.

Neither formula achieves either objective precisely. And neither formula incorporates any information about the spatial structure of the target. But, they are widely used in engineering practice because they improve on the metric properties of the XYZ representation and perform better across viewing conditions.

The formulae were derived by a series of comprises made in committee. The formulae were built from previously suggested formulae and an attempt to come to grips with an array of data, including the MacAdam discrimination data and perceptual scaling data (e.g. on a scale of 1 to 10, how different are these two colors).

The basic procedures for calculating CIE($L^*u^*v^*$) and CIE($L^*a^*b^*$) representations share much in common. Both are described in most standard references, so here I will only describe the steps required to calculate CIELUV. In the case of specifying the CIELUV coordinates of surfaces, we proceed in three steps.

First, calculate the XYZ representation of the light reflected from the sample of interest. You can either do this from first principles using the spectral power distribution of the light reflected from the sample and the color-matching functions, or you can buy an instrument called a *photometer* to measure the XYZ representation. Call these values (X, Y, Z).

Second, measure the XYZ representation of a white surface in the image. Call these values (X_w, Y_w, Z_w). Both the tristimulus values of the object of interest and the tristimulus values of a white point are necessary. The white point data serve to correct the formula for differences in the illumination.

Finally, to derive the three coordinates, $L^*u^*v^*$, use the sequence of formulae. Calculate an intermediate representation by

$$u' = \frac{4X}{X + 15Y + 3Z} \qquad v' = \frac{9Y}{X + 15Y + 3Z} \tag{10}$$

$$u'_w = \frac{4X_w}{X_w + 15Y_w + 3Z_w} \qquad v'_w = \frac{9Y_w}{X_w + 15Y_w + 3Z_w} \tag{11}$$

Then calculate the final values using

$$L^* = 116(Y/Y_w)^{1/3} - 16 \tag{12}$$

$$u^* = 13L^*(u' - u'_w) \qquad v^* = 13L^*(v' - v'_w) \tag{13}$$

(If $\frac{Y}{Y_w} <= 0.01$, then substitute $L^* = 903.3\frac{Y}{Y_w}$.)

The distance between two lights is simply the Euclidean distance between their $(L^*u^*v^*)$ coordinates,

$$\Delta E_{uv}{}^{*2} = \Delta L^{*2} + \Delta u^{*2} + \Delta v^{*2} \tag{14}$$

where ΔL^* refers to the difference in the L^* value of the two lights, and so forth. Lights with a $\Delta E_{uv}{}^*$ value of one are discriminable only under the best viewing conditions. Lights with a $\Delta E_{uv}{}^*$ value of less than three are very hard to discriminate.

Two features of these formulae address the shortcomings of the metric properties of the XYZ representation. First, the non-linear transformation of the XYZ coordinates stretches the distance relationships. Using the CIE 1976 $(L^*u^*v^*)$ representation, the Euclidean distance between pairs of lights is a better estimation of the perceptual differences than in the XYZ representation. Second, by including a normalization with respect to a white surface, the representation attempts to compensate for illuminant changes.

While the CIELUV and CIELAB representations are fairly widely used, there are some fundamental unanswered questions about them. First, the metrics do not take into account the spatial structure of the image. There is only limited guidance in the literature as to how we should alter the parameters of the metric as the size or spatial composition of the colors are changed.

Second, the formulae are incompletely specified for application to monitor images. The difficulty is that the white point normalization term has no obvious interpretation for many types of monitor images, such as business graphics. In these cases, practitioners commonly use the tristimulus coordinates of the monitor when the three guns are set to maximum as the white point. This procedure is easy, but not always sensible. If we display the identical XYZ representation on a pair of monitors, surely the CIELUV values should be the same. But, if the monitors happen to have different white points, common practice of white point normalization of monitor images will lead to different CIELUV values. This makes no sense. A normalization procedure that depends on the viewing field near the image needs to be worked out.

Color Appearance

The theory and data of photopic color-matching provide a remarkably complete explanation of whether two lights will match. The color-matching experiment is so important to our understanding of vision that there is a tendency to act as if they explains more than it does. But you should notice that the theory is silent about what the lights look like.

The reader who is being exposed to these ideas for the first time may be surprised that in this entire introduction I have rarely used words like brightness, saturation and hue. I have not used them because the logic of the color-matching experiment, and what the color-matching experiment tells us about human vision, does not directly inform us about color appearance. What we learn from color-matching is fundamental. But for many applications we wish to know not that two lights look alike, but rather what color name an observer will assign to a light; that is, we wish to learn what the light looks like. Conversely, there are many display applications in which we wish to know how to construct a light with a desired color appearance. Understanding color-matching is a pre-requisite for approaching the problem of color appearance, but it is not a solution to the problem.

To emphasize the difference between color-matching and color appearance, consider the following experiment. Suppose that we form a color-match between two lights that are presented as a pair of crossing lines against one background. Such a pair is illustrated on the left hand side of Figure 19. On the left, the metameric pair both appear gray. Now, move this pair of metameric lights to a new background. Since the two lights are metamers, color-matching assures us that the two lights will continue to match one another as we move them about. But we should not be assured that the appearance of the lights remains the same. For example, on the right of the figure we find that the pair of lights now have quite a different color appearance. By examining the point where the lines come together at the top of Figure 19, which was created by the artist Joseph Albers, you can see that the lines are physically identical on both sides of the image.

From this simple demonstration we find that color-matching, and the study of metamers, is different from the study of color appearance. Color-matching assures us

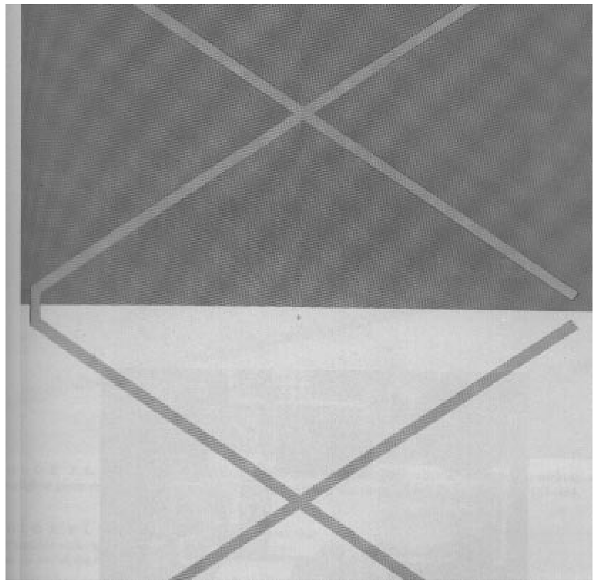

Figure 19: This figure is here to remind you of the a color illusion from the work of Joseph Albers.

that a pair of lights look the same as one another. Since the two lights stimulate the photoreceptors in the same way, it makes sense that the pair of lights will continue to look the same viewed even as they are viewed in other contexts, such as against different backgrounds in the Albers painting. But as the context changes, their color appearance may change.

From the point of view of color appearance, then, The Albers illusion points out that when two lights with identical XYZ representations are viewed in different contexts, their color appearance will not be the same. Color matches across context are called *asymmetric color-matching*. When the viewing context is changed, the XYZ representation cannot be used to predict whether or not two lights appear the same.

The spatial structure of a pattern can also influence its color appearance. Color variations seen at finer resolution appear far less saturated than the same color variations seen at lower resolution. This is illustrated in the various colored sweep frequency patterns shown in Figure 20. This perceptual principle is embedded in the NTSC encoding of television images; color components of the signal are only transmitted at low spatial resolution. Yet another objective, then, is to extend our color representations to include some of the spatial structure of the image in our calculations.

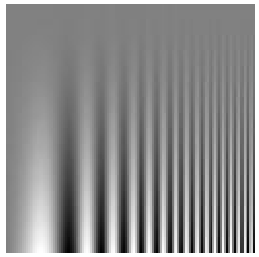

Figure 20: This figure is here to remind you of the color appearance of the sweep frequency patterns.

Summary

The XYZ representation of lights predicts color-matching when two lights or two surfaces are viewed in a common context. Since its introduction in 1931, the XYZ representation has been useful in many different industries.

Many modern applications require us to predict whether a pair of samples, which might be seen together in various viewing contexts, will match. For these applications we require a theory of color discriminability that extends across viewing contexts, including illumination changes. In 1976 the CIE introduced to additional representations, CIE 1976 ($L^*u^*v^*$) and CIE 1976 ($L^*a^*b^*$) that serve as a useful, if imperfect, guide for predicting discriminability of samples under different illumination conditions.

References

[Boynton, 1979] Boynton, R. M. (1979). *Human Color Vision*. Holt, Rinehart and Winston, New York.

[Brainard, 1989] Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *Col. Res. Appl.*, 14:23–34.

[Cornsweet, 1970] Cornsweet, T. N. (1970). *Visual Perception*. Academic Press, New York.

[Cornsweet and Pinsker, 1965] Cornsweet, T. N. and Pinsker, H. M. (1965). Luminance discrimination of brief flashes under various conditions of adaptation. *J. Physiol. (London)*, 176:294–310.

[Cowan and Rowell, 1988] Cowan, W. B. and Rowell, N. (1988). A fast inexpensive spectroradiometer. *Personal communication*.

[Hunt, 1987] Hunt, R. W. G. (1987). *The Reproduction of Colour*. Fountain Press, Tolworth, England.

[Hunter and Harold, 1987] Hunter, R. S. and Harold, R. (1987). *The Measurement of Appearance*. Wiley, New York.

[Judd and Wyszecki, 1975] Judd, D. and Wyszecki, G. (1975). *Color in Business, Science, and Industry*. Wiley, New York.

[Judd et al., 1964] Judd, D. B., MacAdam, D. L., and Wyszecki, G. W. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, 54:1031.

[MacAdam, 1942] MacAdam, D. L. (1942). Visual sensitivities to color differences in daylight. *J. Opt. Soc. Am.*, 32:247–274.

[Poirson and Wandell, 1990] Poirson, A. B. and Wandell, B. A. (1990). Task-dependent color discrimination. *J. Opt. Soc. Am. A*, 7:776–782.

[Post and Calhoun, 1989] Post, D. L. and Calhoun, C. S. (1989). An evaluation of methods for producing desired colors on crt monitors. *Color Res. and Appl.*, 14.

[Wyszecki and Stiles, 1982] Wyszecki, G. and Stiles, W. S. (1982). *Color Science*. John Wiley and Sons, New York, second edition.