Color Matching with Amplitude Not Left Out

James A. Worthey
Gaithersburg, Maryland, USA

Abstract

Amplitude for color mixing is different from other amplitudes such as loudness. Color amplitude must refer to a light's ability to look different from other lights, to express its redness or other chromatic intensity, so that its color is not lost during transduction. To reveal independent stimulus dimensions, a set of orthonormalized color matching functions is derived, similar to opponent color primaries. Following an idea of Jozef B. Cohen, it is then assumed that a light of unit power varies in wavelength through the spectrum. The track of that light in the orthonormal color space gives a curve that Cohen called "the locus of unit monochromats," after he found it by different steps. The locus defines a surface that is interesting but not complicated, which Cohen called "butterfly wings." Projecting the locus into a plane normal to the achromatic axis gives a boomerang shape with 3 well-defined local extreme points. The extrema are William A. Thornton's Prime Colors, so a few steps reveal the inner workings from which the Prime Colors arise. The results can explain color mixing to beginners, but are also quantitative and ready for practical use.

Introduction

With regard to things that stimulate our senses, we all know what amplitude means—usually. A high-amplitude noise SOUNDS LOUD. Summer sunlight looks bright. When adaptation state is controlled, sensation is often a simple—if nonlinear—function of physical amplitude. Color mixing is different, and we all know this too. In a color match, lights add linearly, and the match does not depend on adaptation conditions. The CIE teaches us to predict color matches by certain algebraic steps, but the XYZ formulas lose the idea of "color amplitude," which has practical importance and in fact lurks within a set of color matching functions, such as $\{\bar{x}, \bar{y}, \bar{z}\}$.

Consider the sensitivities of the three cone systems,[1] Fig. 1. Whatever color amplitude is, it should vanish at the ends of the spectrum and show a local minimum in the blue-green region of the spectrum. We also see why the idea of amplitude in color mixing is not as simple as it might be. The cone sensitivities overlap! This is to our benefit in catching plenty of photons and in finely discriminating hues in the red to green range, but it complicates the discussion of color mixing. Consider, for instance, the traditional instructions for transforming color mixing data to new primaries. That formulation is really about overlap, in the absence of which every experimenter would set his primaries to the eye's three well-separated sensitivity peaks and be done with it.

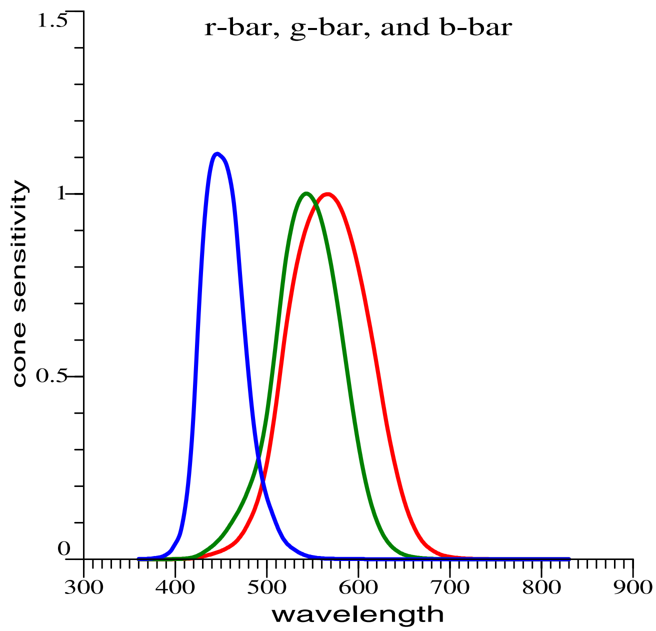

Figure 1. Spectral sensitivities of the 3 cone types.

Have these issues of amplitude and overlap been addressed before? Yes, they have, though not with these words, usually. Opponent-color models[2] emphasize the small but important difference between the overlapping red and green sensitivities (or long and middle sensitivities, if you like). For analysis of color amplitude, we can look particularly to the work of William A. Thornton[3-6] and Jozef B. Cohen.[7,8] A fresh derivation below will show what color amplitude is and how it relates to practical work. What then comes out is one of Cohen's results in a slightly different format. Out of that comes an interpretation for Thornton's Prime Colors, and confirmation of the wavelengths at which the Prime Colors occur. In short, once the calculation starts in a certain direction, it proceeds with no arbitrary steps, and everything falls into place. "Everything" includes a new appreciation for Cohen's "locus of unit monochromats," an independent confirmation of Prime Colors, and graphical presentations for further use. It will be seen that the locus of unit monochromats is the amplitude of the spectrum for color mixing. Even beginning students, without learning the full background, will find color amplitude to be a stable concept with immediate meaning for such problems as color rendering by lights and the choice of phosphors and dyes for color reproduction.

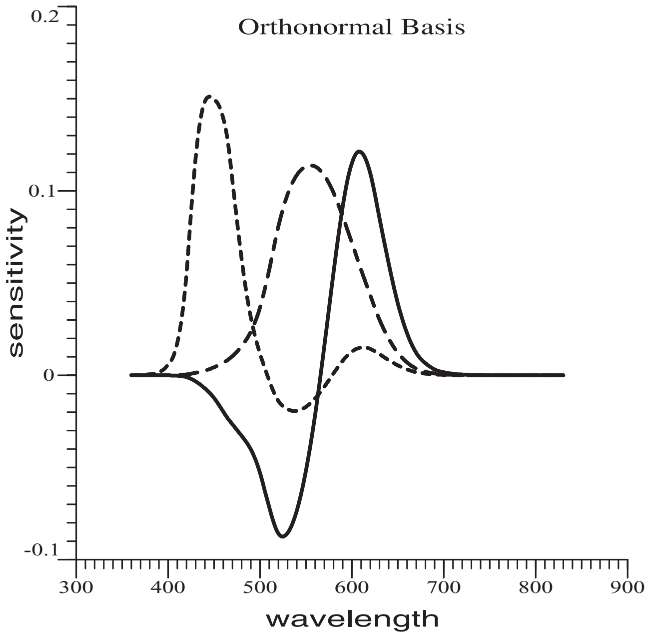

Figure 2. Orthonormal opponent color functions. Dash-dot = 1st vector = achromatic; solid = 2nd vector = red-green; short dashes = 3rd vector = blue-yellow.

Terse Approach

The formulation that we seek could be approached in many ways, but will tend to end up in the same place. Let us first derive some algebraic and graphical results. Then a further discussion can enhance understanding of color amplitude as a concept. First define a matrix $\mathbf{M}_0$ by[2]

$$\mathbf{M}_0 = \begin{bmatrix} 0 & 0.9341 & 0 \\ 0.7401 & -0.6801 & -0.1567 \\ -0.0061 & -0.0212 & 0.0314 \end{bmatrix} \qquad (1)$$

If the usual CIE color matching functions are the columns of a matrix C,

$$\mathbf{C} = [\,\bar{x}, \bar{y}, \bar{z}\,]\,, \qquad (2)$$

then a matrix **G** can be computed whose columns are opponent color functions:

$$\mathbf{G} = \mathbf{C}\mathbf{M}_0^{\mathrm{T}}\,, \qquad (3)$$

where superscript T denotes matrix transpose. The columns of **G** approximate an opponent model of Guth,[2] with columns in the sequence

$$\mathbf{G} = [\text{achromatic, red-green, blue-yellow}]. \qquad (4)$$

Now perform Gram-Schmidt orthonormalization on the columns of **G** and call the result **Ω**. The first column of **G**, the achromatic function, is the familiar $\bar{y}$ times a constant. The first column of **Ω** is $\bar{y}$ times a different constant. The term "achromatic" means that this is the sensitivity for whiteness. The columns of **Ω** retain the interpretation of Eq. (4), though they are quantitatively different. Symbol **Ω** stands for "orthonormal," meaning that its columns, three vectors $\boldsymbol{\omega}_j$, are color matching functions but with the additional property that:

$$\boldsymbol{\omega}_i^{\mathrm{T}}\,\boldsymbol{\omega}_j = \delta_{ij}\ . \qquad (5)$$

δ_{ij} is the Kronecker delta, equal to 1 if $i=j$, equal to 0 if $i \neq j$. These color matching functions can be graphed, Fig. 2. Kets $|1\rangle$, $|2\rangle$, $|3\rangle$ are synonymous with vectors $\boldsymbol{\omega}_j$. A ket $|j\rangle$ is a column vector, the bra $\langle j|$ is its transpose, a row vector. Amplitudes are fixed by the normalization requirement, $\langle j|j\rangle = 1$. To repeat, the first vector, $|1\rangle$, is a constant times the ever-familiar $\bar{y}$, $|2\rangle$ is red versus green, and $|3\rangle$ is a sort of blue versus yellow function, but with a yellow lobe that does not conform to other opponent-colors models.

Because of the way that Gram-Schmidt operates on the Guth model,[2] which used cone fundamentals derived from CIE functions,[9] both $|1\rangle$ and $|2\rangle$ depend only on the red and green cone sensitivities. Only $|3\rangle$ has input from blue cones; in fact it has input from all 3 cones. For further discussion and derivations, the orthonormal property is extremely convenient, and yet a certain linkage to physiological fundamentals is preserved. The CIE's $\bar{y}$ has the significance that it fits such data as flicker photometry, and it is considered (by Guth,[2] for example) to be a sum of red and green inputs only. Here $|1\rangle$ shares those benefits. The new vectors are color matching functions in the familiar sense that if $|N_1\rangle$ is a column vector representing a spectral radiance $N_1(\lambda)$, and $|N_2\rangle$ another such radiance, the condition for a visual match is the usual one:

$$\begin{bmatrix} \langle 1|N_1\rangle \\ \langle 2|N_1\rangle \\ \langle 3|N_1\rangle \end{bmatrix} = \begin{bmatrix} \langle 1|N_2\rangle \\ \langle 2|N_2\rangle \\ \langle 3|N_2\rangle \end{bmatrix}. \qquad (6)$$

If the source of complication was the overlap of cone sensitivities, we have met the problem head on. The direction cosines among cone sensitivities are tabulated below.

Cones	direction cosine
Red, green	0.918
green, blue	0.121
Red, blue	0.0579

The orthonormal functions, of course, have cosines of zero among themselves. Now we should notice what color mixing means, and why it is linear. Light of a certain spectral distribution falls on a patch of retina, and the three cone types "interact" in the most simplified way, not by signaling each other, but by their differing photon catches, according to the light's spectrum. The importance of the detection step derives from the information that is lost.[10] If two lights have different spectra, $|N_1\rangle \neq |N_2\rangle$ but they are metameric, meaning that they match according to Eq. (6), then the information that they differ is lost in detection. The detection step *is* the color mixing step, and color mixing amplitude must have something to do with action in color mixtures, not only with the achromatic dimension $|1\rangle$, which is essentially light-meter luminance.

Now suppose that we have a narrow-band light representing one watt of optical power.[8] Let its tristimulus vector be computed for the orthonormal basis, like either vector in Eq. (6). Then let the wavelength of the narrow band vary through the spectrum and trace the locus of the vector. The result is a curve in three dimensions. Wavelengths can be marked off along the locus. A 3-D graph has been prepared and colorized using Virtual Reality Modeling Language, VRML. The graph will be shown at the meeting and can be interactively rotated. The VRML file is also available on the author's web site, http://www.jimworthey.com . A projection of the 3-D graph is seen in Figure 3.

Figure 3 is essentially a parametric plot of the orthonormal basis functions, with the interpretation of a 1-watt light moving through the spectrum. It is exactly the curve that Cohen[8] called the "locus of unit monochromats," even though my algebra is different from Cohen's. The blue lobe of the surface is near the $|3\rangle$ axis, while the red and green directions on the $|2\rangle$ axis are easy to see—the axis itself is colorized, if you are seeing this in color. The achromatic or $|1\rangle$ axis looks more like it is coming out of the paper, and it lies close to yellow-green. Cohen didn't emphasize specific basis functions and axes, so part of my contribution is the choice of basis functions, and the axes thus created. Because amplitude is not lost, the locus falls into the origin at the short and long ends of the spectrum. Seen in three dimensions, the surface is interesting but not intricately folded. Cohen called it "butterfly wings."[8] Because the $|1\rangle$ axis measures whiteness, it is logical to think of the $|2\rangle$-$|3\rangle$ plane as the chromatic plane. Projecting a tristimulus vector into that plane loses only its whiteness component. Projecting the spectrum locus ("butterfly wings") into $|2\rangle$-$|3\rangle$ gives the boomerang shape of Fig. 4.

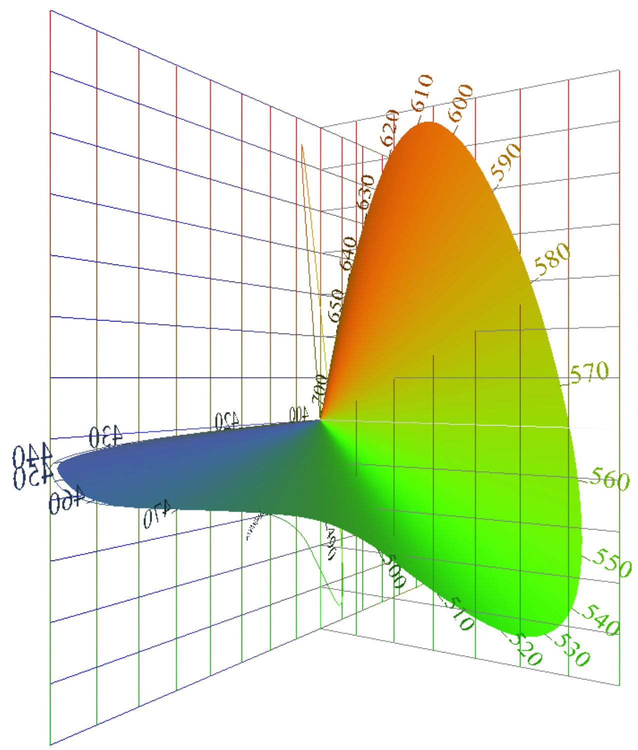

Figure 3. One view of a 3-D graph of the locus of unit monochromats. The axes are determined by the choice of orthonormal cmfs. Otherwise, the graph is the invariant one drawn by Jozef B. Cohen.

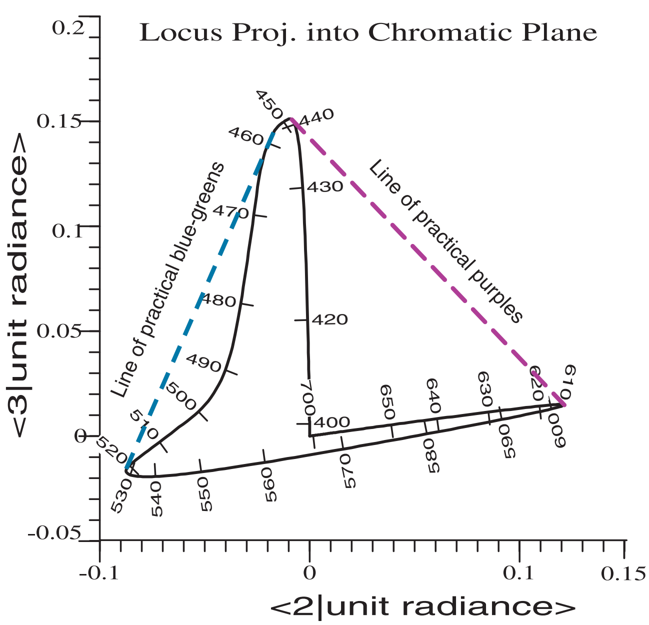

Figure 4. The locus of Figure 3, but projected into the 2-3 or chromatic plane. The loci in Figures 3 and 4 are not boundaries or gamuts in the usual sense.

Application

No Purples?

With a colleague, Michael H. Brill, I discussed this boomerang-shaped projection of the locus. It looked to be a kind of alternate chromaticity diagram, based on mathematically appealing orthogonal color matching functions. But, we asked, how can it be a chromaticity diagram without a line of purples? The answer is, it's not a chromaticity diagram, but it is what it is. It is a projection of a tristimulus diagram, showing either sensitivity, or chromatic stimulus at unit power, however you want to look at it. It shows **chromatic sensitivity with power not left out**. It shows which wavelengths will act strongly, or not so strongly in chromatic mixtures.

Prime Colors

In the $|2\rangle$-$|3\rangle$ projection, the spectrum locus has 3 extrema, 3 local maxima in its radius from the origin. Those wavelengths are Thornton's Prime Colors! The radius peaks at 445, 525, and 608 nm. Here in Scottsdale in 1998, he reported 447, 541 and 604 nm as Prime Colors for the CIE 2° observer.[3] In truth, there are two ways to extract Prime Colors from the locus of unit monochromats. Consider Fig. 5, with wavelength on the abscissa. The upper curve is radius from the origin for the locus in 3 dimensions. This is what Jozef Cohen called the square root of the diagonal of matrix **R**, and he described it as the amplitude of spectral lights. The lower curve is radius within the $|2\rangle$-$|3\rangle$ projection. The peaks for root diagonal of **R** are 445, 536, 604, while those for the projection are as just stated, 445, 525, and 608 nm.

Suppose the question is "Which lights strongly affect chromatic color?" Then the Prime Colors based on projected radius may be appropriate. If the question is, "At what wavelengths will metameric spectra cross?" then the peaks of radius in three dimensions may be better. The point is not the small numerical differences, but the idea that now we see exactly where the Prime Colors come from.

In Fig. 4, we saw immediately that the extrema are the Prime Colors, and one may quickly notice that those are for practical purposes the television phosphor colors.[11] Thornton invented his color-enhancing 3-phosphor lights based on Prime Colors, so the spectrum locus in orthonormal tristimulus space explains two common technologies. In more general terms, the locus shows the vector magnitude and direction of each narrow spectral band, describing its action in color mixing. Suppose that you bring a red apple to your teacher. The apple's school assignment is simple, to be as red as it can be. To do this, it must reflect the red Prime Color and longer wavelengths; it should absorb the green Prime Color, which would tend to cancel out its redness. It relies on the classroom lighting to supply these wavelengths, and actual school lighting may fall short. That is the classic problem of Color Rendering, not usually stated in these terms.

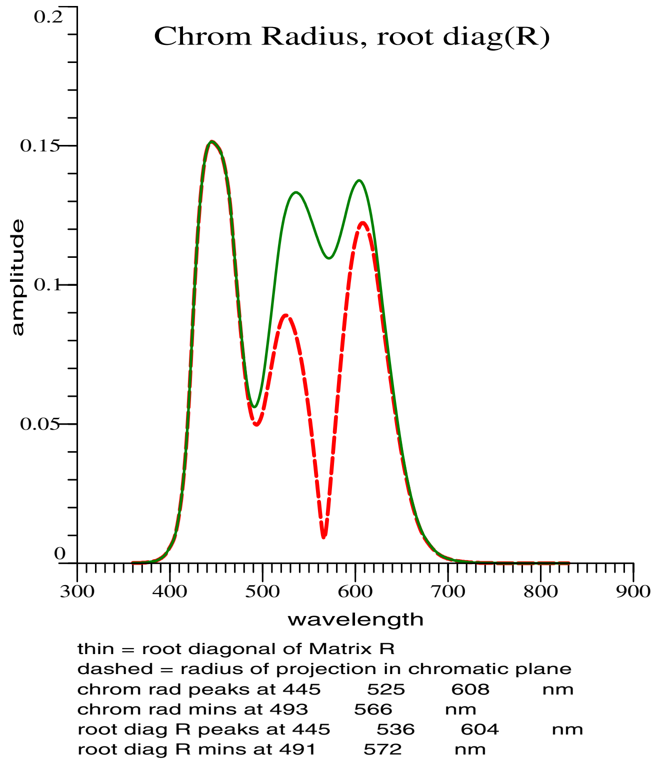

Figure 5. Radii of the loci in 3 and 2 dimensions. In Jozef Cohen's terminology, the radius in 3 dimensions is the square root of the diagonal of Matrix R. The peaks and local minima of the graph were found by the program that drew the graphs.

Power Budgets

The CIE's chromaticity diagram in (x, y) is an optics lab diagram. In an optics lab equipped with large light bulbs and the proper slits and prisms, light of 680 nm or 490 nm can be displayed and the intensity cranked up to give a bright ruby red or aquamarine. In real life, power is limited. Television and light bulb designers have power budgets. An image printed on paper has a power budget, imposed by the light under which it will be viewed. Even in daylight, the total power is ample, but the eye adapts to mean luminance. For an object to show strong chromatic color, it still must work with the daylight spectrum. A red leaf in autumn cannot reflect 300% of incident light in a narrow band at 650 nm. To be bright red, it must absorb a range of greens and reflect a range of reds, including the prime color region.

If I were to say broadly "Color images must work with the available power," the reader would say, "Well, of course!" That would not be news. The news is that the Cohen-like presentation goes to the heart of the problem, **and** it is easily calculated and invariant, **and** it reveals three spectral regions that are clearly most important, **exactly** as William A. Thornton has said. The orthonormal functions are a convenient tool that Cohen mentioned but did not emphasize. The shape of the curve is invariant to the choice of basis functions, so long as they are orthonormal.

Practical Purples!

In short, Figs. 3 and 4 speak to color mixing when power is limited, which is most practical situations. With this in mind, we see that there *is* a line of purples after all. It is the "line of practical purples" as drawn. It marks a region where approximately prime colors are efficient at making purples, more so than the extreme wavelengths of 400 and 700 nm. A line of practical blue-greens is also seen, where color vision will respond better to a mixture of blue and green than to actual blue-green light. Some detail is lost by sketching the dashed lines in two dimensions, however the full 3-dimensional graph is also available for thinking about these issues.

Why This Orthonormal Space?

One rationale for the orthonormalized color matching functions is intuitive. If detection by cones loses information, and the goal of color technology is to stimulate color vision in its 3 independent dimensions, then color stimuli should be evaluated along truly independent dimensions. Another benefit is that it gives a new specific meaning to tristimulus values. If $\mathbf{A}$ is a matrix with *any* set of color matching functions as its columns, such as perhaps $\mathbf{A} = [\,\bar{x}, \bar{y}, \bar{z}\,]$, then the projector matrix $\mathbf{R}$ can be found by[7]

$$\mathbf{R} = \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}} . \tag{7}$$

Matrix $\mathbf{R}$, discovered by Jozef Cohen, extracts from a spectral radiance $|N\rangle$ the component that lies in the vector space of the color matching functions, called $|N^*\rangle$:

$$|N^*\rangle = \mathbf{R}|N\rangle . \tag{8}$$

In words, $|N^*\rangle$ is the fundamental metamer of $|N\rangle$. $\mathbf{R}$ is invariant to the choice of color matching functions in $\mathbf{A}$. They could be cone sensitivities, or they could be $\mathbf{\Omega}$. If we let $\mathbf{A} = \mathbf{\Omega} = [|1\rangle, |2\rangle, |3\rangle]$, so the columns are the orthonormal color matching functions,

$$\mathbf{R} = \mathbf{\Omega}(\mathbf{\Omega}^{\mathrm{T}}\mathbf{\Omega})^{-1}\mathbf{\Omega}^{\mathrm{T}} . \tag{9}$$

Thanks to orthonormality, the expression $\mathbf{\Omega}^{\mathrm{T}}\mathbf{\Omega}$ is the 3×3 identity matrix, so Eq. (9) reduces to Eq. (10),

$$\mathbf{R} = \mathbf{\Omega}\mathbf{\Omega}^{\mathrm{T}} \qquad \text{<only for orthonormal cmf's!>} \ . \tag{10}$$

Keeping in mind the definition of $\mathbf{\Omega}$ and the rules of matrix multiplication, this can be written in an alternate form,

$$\mathbf{R} \ = \sum_{j=1}^{3} |\,j\rangle\langle j\,| . \tag{11}$$

The notation of Eq. (11) is sometimes used for what is called a unity operator. In any case, Eqs. (8) and (11) then lead to

$$|N^*\rangle = |1\rangle\langle 1|N\rangle + |2\rangle\langle 2|N\rangle + |3\rangle\langle 3|N\rangle . \tag{12}$$

Each complete bracket, such as $\langle 1|N\rangle$, is an inner product, and as such is a single number. These brackets are the coefficients in the approximation of $|N\rangle$ by a basis function expansion. They are also the tristimulus values of $|N\rangle$ in the orthonormal system. This fact gives a level of meaning to tristimulus values that they otherwise lack, and Eq. (11) aids in deriving formulas for such chores as converting a tristimulus vector from one basis to another[1].

Questions Answered

In the traditional teaching of color science, the early color-matching experiments are discussed. Data lead to sets of three graphs, called color matching functions, but these functions are not unique. A change of primary wavelengths will lead to different data; in fact, the change of a single primary alters all three graphs. Thus transformation of primaries is introduced, presumably just as primaries were transformed in the 1920s. The particular notation has symbols that represent lights and not algebraic variables in the usual sense. Such an approach honors the original experimenters and avoids linear algebra.

The important facts of color mixing thus present themselves in an unstable way, through sets of functions that can be transformed into alternate sets of functions. The teacher and students are unsettled by this version of reality and seek answers to practical questions such as "What color is a mixture of yellow and blue?" or "What hues can be displayed on a television?" One way to get rid of the changeable data is to choose a specific set of color matching functions, and stick with it. The CIE system is introduced and students are then speaking the world color language, and there is no turning back.

In particular, color mixing is explained through the (x, y) chromaticity diagram, in which all possible hues are displayed, but the transition to (x, y) loses stimulus amplitude. I recall a question like this coming up when I was a student: "If you have 1 watt of 580 nm, and 1 watt of 490 nm, does the mixture lie at the midpoint between the 580 and 490 points on the chromaticity diagram?" The answer is "No, not at all," and students learn to solve color mixing problems by algebraic manipulation of X, Y, and Z. No graphical scheme is offered for thinking about stimulus amplitudes, but the XYZ formulas work easily enough for predicting matches, so most work stays within the CIE framework.

Now along come David MacAdam[12] and William Thornton[4] and they ask questions such as "Which wavelengths act most strongly in mixtures?" By thought experiments, they come up with graphs like Fig. 6.

The original caption reads in part, "Alterations of the chromaticity (x = 0.333, y = 0.333) of the equal-energy SPD by the addition of the same increment of power at successive wavelengths. Scale of the inset is four times that of the surrounding ... diagram." What an interesting question to ask, "Which wavelengths act most strongly?"

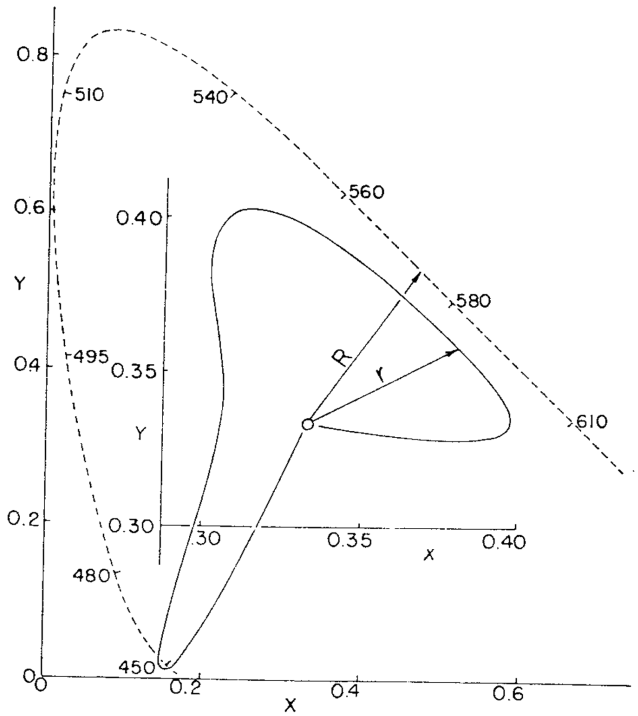

Figure 6. This graphic is taken from an article by Thornton4 concerning digital experiments in which narrow-band lights are added to Equal Energy White.

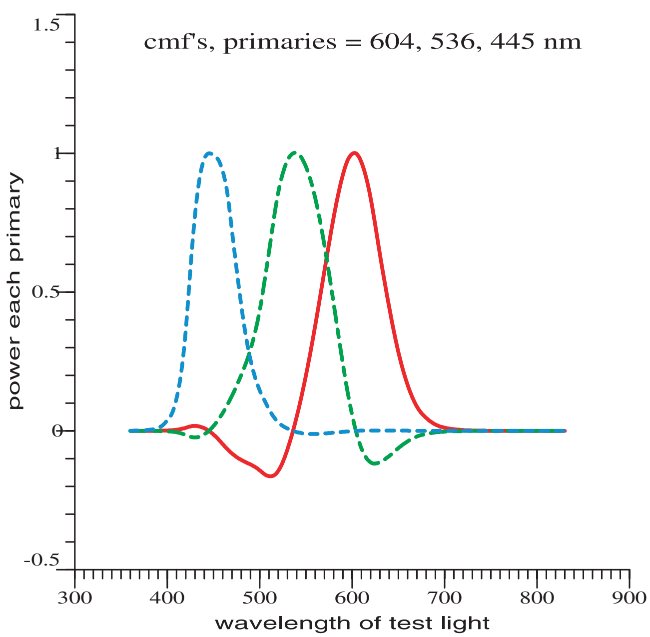

Figure 7. Color matching functions based on primaries of 604, 536, and 445 nm, and the 2-degree observer.

However, think about experimental color matching functions, Fig. 7. Suppose that you wanted to explain to a child what this set of color matching functions means. You could say that "These graphs show how strongly different wavelengths act in color mixtures. The red graph shows how strongly the lights pull in the red direction, for example." The idea of different wavelengths acting more or less strongly was right there on the paper when the color matching data were first graphed. Thornton now knows this and has looked at Prime Colors more directly in terms of color matching data.

The color matching functions in Fig. 7 are calculated based on narrow-band primaries set to Prime Color wavelengths. Theory mavens might ask if those color matching functions are orthogonal functions, if the prime colors are at right angles in color space, and other such questions. The short answer is, no, the cmf's are not orthogonal, and the Prime Color directions are not perpendicular, though they come pretty close.

Conclusion

The orthonormal basis and the locus of unit monochromats answer the question that a student might ask, whether there is a graphical presentation for color mixing with amplitude not left out. Since the orthonormal color matching functions are perpendicular in the space of spectral functions (direction cosines = 0), it is logical to use them in computing tristimulus vectors. The Prime Colors are then found to be the wavelengths for which unit-power stimuli give the longest tristimulus vectors. The 3D colorized graph of the Locus of Unit Monochromats gives a larger context to Prime Colors and other features such as the line of practical purples. For 3D graphs in virtual reality, see http://www.jimworthey.com .

References

1. James A. Worthey, "Color rendering: a calculation that estimates colorimetric shifts," Color Research and Application **29**(1):43-56, February 2004.
2. Sherman Lee Guth, Robert W. Massof, and Terry Benzschawel, "Vector model for normal and dichromatic color vision," *J. Opt Soc. Am.* **70**, 197-212 (1980).
3. Michael H. Brill, Graham D. Finlayson, Paul M. Hubel, William A. Thornton, "Prime Colors and Color Imaging," *Sixth Color Imaging Conference: Color Science, Systems and Applications*, Nov. 17-20, 1998, Scottsdale, Arizona, USA. Publ. IS&T, Springfield, Virginia.
4. William A. Thornton, "Luminosity and color-rendering capability of white light," *J. Opt. Soc. Am.* **61**(9):1155-1163, September 1971.
5. William A. Thornton, "A simple picture of matching lights," *J. Illum. Eng. Soc.* **8**(2):78-85 (1979).
6. William A. Thornton, "Three-color visual response," *J. Opt. Soc. Am.* **62**(3):457-459 (1972).

7. Jozef B. Cohen and William E. Kappauf, "Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks," *Am. J. Psych.* **95**(4):537-564 (1982).
8. Jozef B. Cohen and William E. Kappauf, "Color mixture and fundamental metamers: Theory, algebra, geometry, application," *Am. J. Psych.* **98**(2):171-259, Summer 1985.
9. V. C. Smith and Joel Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm," *Vision Res.* **15**: 161-171 (1975).
10. Cornsweet, Tom N., *Visual Perception*, Academic Press, New York, 1970.
11. James A. Worthey, "Color rendering: asking the question," *Color Res. Appl.* **28**(6):403-412, December 2003.
12. David L. MacAdam, "Photometric relationships between complementary colors," *J. Opt. Soc. Am.* **28**:103-111 (1938).

Biography

James A. Worthey has a BS in Electrical Engineering, and an MS in Physics. His PhD in Physiological Optics is from Indiana University of Bloomington, Indiana. He is particularly interested in lighting and the interaction of lights with objects and the eye. He has published work on light source size, color rendering, object color metamerism and color constancy.