

Colour

The physical description of colour

Colour vision is a very complicated biological and psychological phenomenon. It can be described in many different ways, including by physics, by subjective observation, or by the tri-stimulus representation commonly used in computers. As we noted previously, light has a *wavelength* (λ), which can be measured accurately, and an *energy* or intensity.

Lasers produce a ‘pure’ light of one particular wavelength, but generally visible light is a mixture of many different wavelengths. This means that visible light can be represented as a distribution of energy into a band of wavelengths. For example, Figure 1(a) shows a distribution of light energy which we would perceive as red in colour. Figure 1(b) shows the the distribution of energy found in sunlight, which we perceive as light yellow. Light in the visible spectrum has a range of wavelengths from approximately 400 to 700 nano-meters.

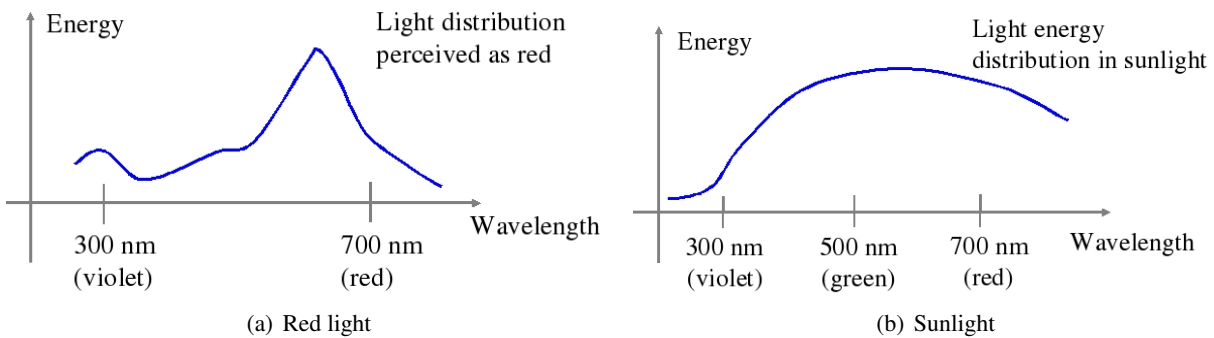


Figure 1: Energy distributions for (a) red light and (b) sunlight.

In the eye we have three distinct types of ‘cone’ cells for detecting light energy. Each type responds to one of three bands of wavelength centred around red (600 nm), green (560 nm) and blue (440 nm). The three bands overlap, so, for example green light excites all three types of cone. Typical responses of the cone cells are shown in Figure 2. The consequence of this overlap among the bands is that each type of cell may be excited similarly from very different energy distributions. A particular distribution of wavelengths will be perceived by us as a single colour, but two entirely different distributions of intensities could be perceived as the same colour.

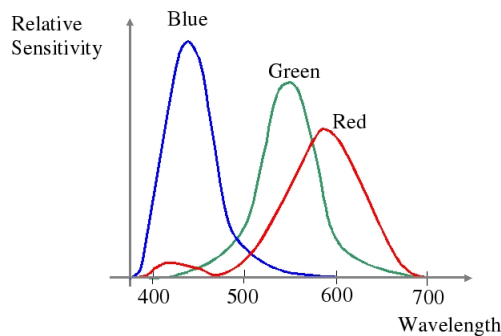


Figure 2: Typical responses of individual cone cells

Subjective observation of colour

In a subjective setting, we can say that two colours are equivalent when an observer *says* that they are equivalent. It is possible to create a large number of colours by selecting three pure light sources (R, G, B) and mixing them together, varying their respective intensities. If we represent the intensities of the red, green and blue light sources in the mix by r , g and b , and the colour they produce when mixed with these intensities as X , then we can say that we associate write this in a shorter way as ‘ $X = r + g + b$ ’. Bear in mind that this does not mean

that we are adding together three scalar values, the addition just describes how we are mixing together the three light sources with intensity contributions of r , g and b .

One way to approach subjective colour definition is through a matching experiment. For an unknown colour, a subject can compare it to a mixture X , adjusting r , g and b values until the mixture matches the unknown colour. Once they are happy that the colour is matched, the r , g and b values can be said to define the colour.

This technique is limited by the actual colours of the light sources that are mixed, and not all colours can be matched in this way. However, if an unknown colour cannot be matched, then it is always possible add one of the r, g, b light sources to it, and form a match with the other two light sources. For example, we can add a certain intensity r of the red source to the unknown colour and seek to match the result to a combination of g and b . This is, in effect, subtracting a colour from the mix. We have a choice of three possible light sources to add to the unknown colour, so matching will be possible using one of the three subtractive matching schemes:

$$\begin{aligned} X + r &= g + b \\ X + g &= r + b \\ X + b &= r + g \end{aligned}$$

The representation of colours as a mixture of three components is called the tri-stimulus representation, and is very commonly used in monitors and other active colour devices. The pure colours used are red green and blue, and are referred to as the additive primary colours.

The Standard Additive Linear System (CIE)

The matching process described above allows us to determine an rgb representation of any colour, even though the scales include negative numbers. The scales used depend on the wavelengths of the sources for matching and are not standard. In order to create a manageable standard system, a committee of scientists devised the CIE Chromaticity Diagram, in which colours are specified by a combination of three light sources normalised to the positive range $[0, 1]$ to avoid any negative values. As each colour is represented by three principal colour components, we would normally require a three dimensional diagram to represent them. This can be avoided, however, by using further normalisation and a two dimensional diagram to represent the colours can be used.

Suppose a colour is defined, as above, by mixing intensity components of r , g and b . If we divide each component by the sum of the intensities, $r + g + b$, (genuine addition this time) we get a normalised set of coordinates:

$$\begin{aligned} x &= r / (r + g + b) && \text{the normalised red component} \\ y &= g / (r + g + b) && \text{the normalised green component} \\ z &= b / (r + g + b) && \text{the normalised blue component} \end{aligned}$$

Therefore $x + y + z = 1$, so if we know two of the normalised colour components (x, y) , we can calculate the third: $z = 1 - x - y$. This means that there are two degrees of freedom and we can represent each colour in a two-dimensional xy plot.

The process by which the CIE diagram was defined is illustrated in in Figure 3. First a standard observer response was defined which is shown in Figure 3(a). This figure contains the same response curves for individual cone cells shown in Figure 3(a), but each one has been scaled according to the differing densities of each type of cell in the foveal region of the eye. The standard observer response curves in Figure 3(a) represent the typical overall cone cell responses to the light arriving in a small solid angle at the eye.

Using the curves in Figure 3(a), we can convert any wavelength in the range to a set of tri-stimulus (r, g, b) values. This is defined by the relative magnitudes of the responses of the three types of cell. For example at a wavelength of $580nm$, we can see in the graph that blue is zero and red and green are approximately equal at 0.8. If we divide these (r, g, b) values of $(0.8, 0.8, 0)$ by their sum, we get normalised (x, y, z) coordinates of $(0.5, 0.5, 0)$ with a total of one.

We can plot the normalised x and y coordinates using a 2-D cartesian graph. This normalised cartesian space is illustrated by Figure 3(b). We only need to plot x and y because we can calculate z from them. The constraint that x , y and z are all non-negative and sum to one means that all points plotted will be contained within a triangle defined by $(0, 0)$, $(1, 0)$ and $(0, 1)$. The previous example would be plotted at $(0.5, 0.5)$

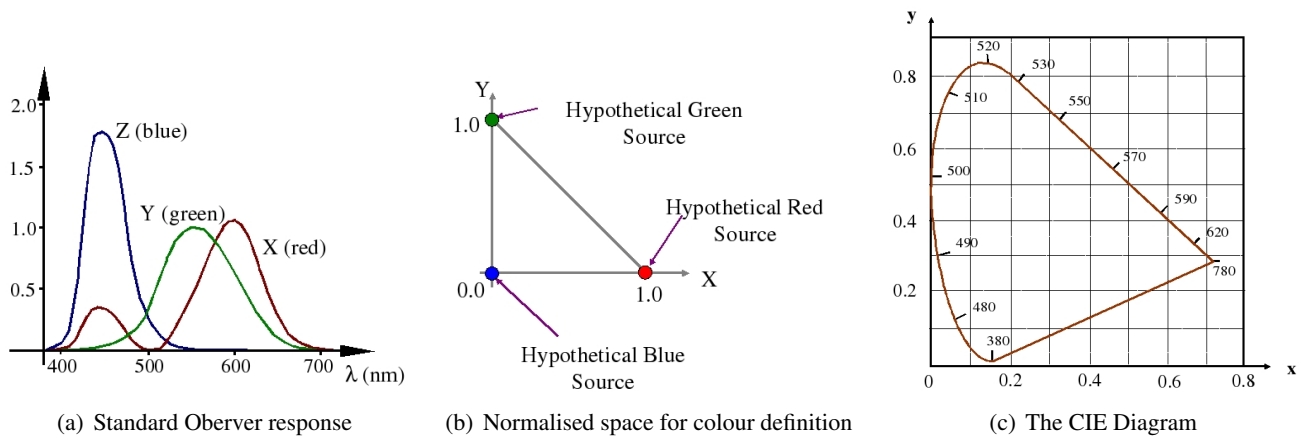


Figure 3: Defining the CIE Diagram

We can repeat this process for all the wavelengths in the range, calculating the normalised x and y coordinates and plotting them in this space. If we do this, we obtain what is known as the CIE diagram of Figure 3(c). All visible colours can be defined in this way, though only their normalised values are shown on colour prints of the diagram. As the wavelengths increase from around $400nm$ to $800nm$, the plotted points form a curve starting near $(0.2, 0)$ before moving up to the top of the space, curving around and descending towards the lower right.

The pure visible colours are distributed on the edge of this horseshoe shaped curve with their physical wavelength running from $380nm$ to $780nm$. The cell responses tail off at top and bottom ends of the visible spectrum (high and low wavelengths) and this makes the wavelengths converge very fast towards a vanishing point. For example the point for wavelength $700nm$ is virtually coincident with the $780nm$ point.

We can also see that just below $500nm$, the red component is zero and the blue and green are approximately equal which would correspond to the colour cyan. The straight line from $780nm$ to $540nm$ on the right hand side represents the condition that $x + y = 1$ (i.e. that $z = 0$) which means that the blue component is zero.

The straight line at the bottom, joining 380 to 780 is a combination of red and blue, which has no equivalent single wavelength colour. It is called the magenta line. All visible colours are contained within the horseshoe, which must be convex, since all colours are a mixture of pure wavelengths. The usefulness of this diagram is that it expresses a linear additive system and the addition of colour components can be done graphically.

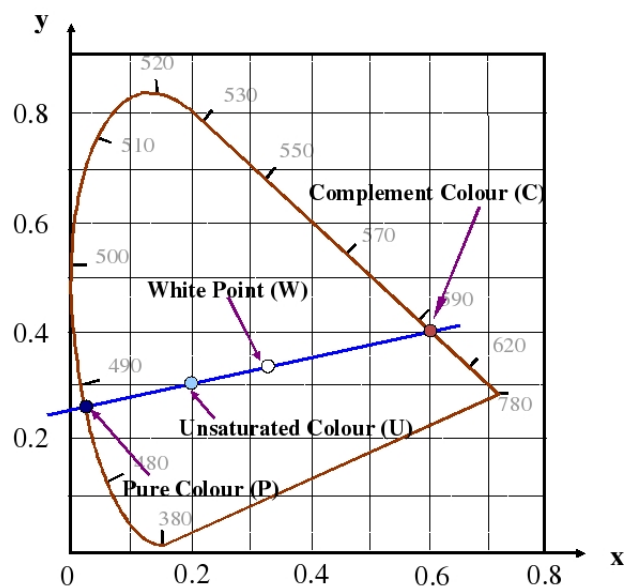


Figure 4: Calculations on the CIE Diagram

Graphical Manipulation of Colour Parameters

Some definitions are shown on the CIE diagram in Figure 4:

- The *white point*, with $x = y = z = 1/3$, has equal strength of all three basic components.
- The colours around the edge of the horseshoe, are called *pure* or *fully saturated* colours.
- If we choose a pure colour, say $488nm$, the line connecting this point and the white point contains all the possible colours which can be achieved by adding white to the chosen pure colour.
- If we extend the line, beyond the white point, it will hit the curve at another pure colour point which is called the *complement* of the original colour. We can mix a colour and its complement to make white. We can see that the complement of $488nm$ is approximately $598nm$.

‘Saturation’, or the degree of purity, of a given colour point can also be determined graphically. We say that the white point is completely unsaturated (has saturation zero) as it has no colour. The ratio of distance between the colour point U and the white point W to the distance between the pure colour P and the white point gives us saturation of U :

$$\text{saturation}_U = \frac{|W - U|}{|W - P|}$$

The saturation is always taken as 1 on the horseshoe (fully saturated) and zero at the white point. If we take any colour on the diagram, for example $U = (0.2, 0.3, 0.5)$ in Figure 4, we can measure this ratio, which in this case is approximately equal to 0.4 and this tells us the white content of the point. We can also determine the pure colour that can be used to generate U by looking at the line through the point U and the white point. From the slope of the line we can calculate that it cuts the y axis at $[0, 0.25, 0.75]$. This point is not in the visible region of the CIE diagram. The nearest pure colour will be seen to be around $[0.04, .26, .736]$.

In another calculation, we can find the complement of the colour at P . This is done by determining the colour which, when added to P , produces white. From the diagram, we can see that the complement of P will be on the $z = 0$ (or $x + y = 1$) line, i.e. the blue component is zero. So we have that

$$(0.04, .26, 0.736) + (r, g, 0) = (k, k, k)$$

from the blue component we can find that $k = 0.736$, so

$$0.04 + r = 0.736 \quad \& \quad 0.26 + g = 0.736 \quad \Rightarrow \quad r = 0.696 \quad \& \quad g = 0.476.$$

If we normalise these coordinates, we get $(0.6, 0.4, 0.0)$ as shown on the diagram.

Note that although the line joining a pure colour and its complement in the CIE diagram always goes through the white point, this does not mean that adding the co-ordinates and normalising produces white. It is, however, always possible to find a linear combination of the two colours that produces white.

Subtractive Colour representation

Paints and printing ink *absorb* colour reflecting only part of the incident light. For this reason the principal colours in paints are called *subtractive* primaries. The three principal colours of paints are magenta (purple), cyan (pale blue) and yellow. These three can be created by combining the additive colour primaries:

<i>Additive primaries</i>		<i>Subtractive primary</i>
red & blue	→	magenta
blue & green	→	cyan
red & green	→	yellow

Black is created by a blend of the three subtractive primaries, but is usually also used separately in colour printing for convenience. It reflects none of the incident light. White paint (or paper) reflects all the incident light and so does not change its colour. Black and white paints can be blended to produce different shades of grey. With grey and red paint you may make brown, and a multitude of other colours which have no

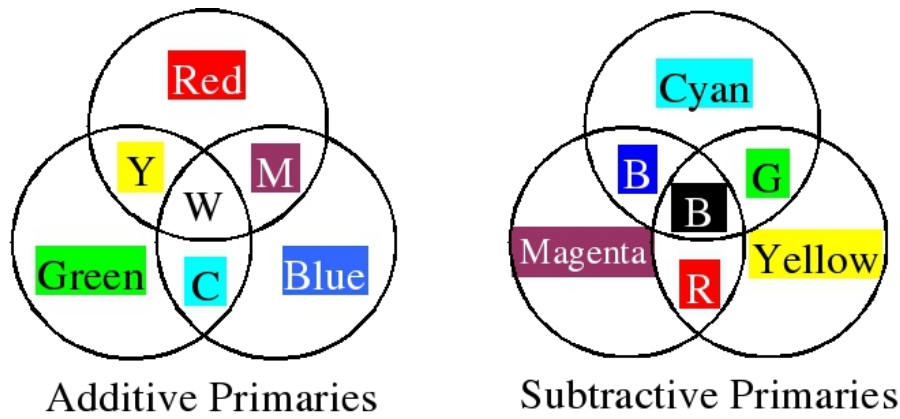


Figure 5: The Primary Colours

equivalents in the systems described in the previous sections. Paints are much richer than colours produced by red/green/blue lights.

Figure 5 indicates how the different, pure basic colours are generated by the additive (monitor) and subtractive (printer) colour systems. Each circle represents one basic additive (R,G,B), or subtractive (M,C,Y) colour. The other regions indicate pure colours generated by combining two basic colours, respectively with addition or subtraction. The centre colour (three additions or subtractions) is white in the additive system (everything reflected) and it is black in the subtractive system (everything absorbed). The two systems show additive or subtractive complementary colours in the same region: e.g. combining red and cyan light makes white while combining red and cyan paint makes black.

Practical Colours

Approximately 128 different hues may be distinguished by the human eye. For each hue around 20 to 30 different saturation values may be seen as different colours. The eye is capable of distinguishing between 60 and 100 different brightness levels. If we multiply these numbers, we get approximately 350,000 different colours. 24 bits provides us over 1.6 million colour shades, however, the eye is very sensitive to differences and under some circumstances small differences can be detected by it.

When a transformation is necessary from the CIE standard colour chart to the colour produced on a specific display device, we must express the primary colours of the display device by, for example CIE x and y values. These numbers give the colours of the individual pixels or LEDs on the monitor. Good quality monitors will be calibrated for the CIE colour chart. For example, a colour monitor may have the following primary colour sources:

	Red pixel	Green pixel	Blue pixel
x	0.628	0.268	0.150
y	0.346	0.588	0.070
z	0.026	0.144	0.780

These correspond to three points which define a triangle on the CIE chromaticity diagram, as shown in Figure 6. The display device can only reproduce colours which are points inside this triangle.

Colour Spaces and Transformation of Colours

There are many different ways of representing colour, and often they can be expressed as linear combinations of each other. For example, in the display monitor example above, the transformation from computer colour intensities R,G,B (expressed in the range from 0.0 to 1.0) may be done with matrix multiplication:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0.628 & 0.268 & 0.150 \\ 0.346 & 0.588 & 0.070 \\ 0.026 & 0.144 & 0.780 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

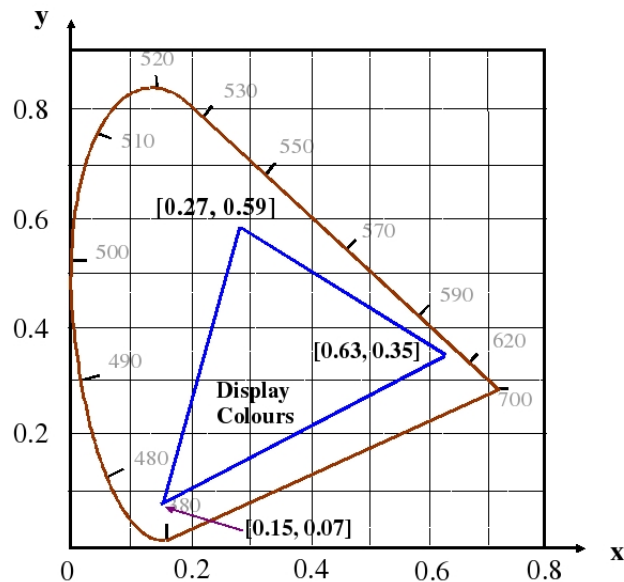


Figure 6: Display Colours

The *RGB* representation of pure red in the computer is $(1.0, 0, 0)^T$ which then produces the correct CIE values of $(0.628, 0.346, 0.026)^T$ for the monitor. In order to find out how to represent a known CIE colour on the display, we have to invert the matrix and then right multiply it with the vector $(x, y, z)^T$.

The HSV colour space

One useful colour space is called the HSV (Hue, Saturation and Value) space. It is of interest since it is thought to correspond more closely to our perception of colour than the *RGB* space we have been discussing. Like the *RGB* system, a colour is represented by three values, *H*, *S* and *V* which can be calculated from the (r, g, b) values. *V* and *S* are calculated simply as follows

$$V = \max(r, g, b)$$

$$S = (\max(r, g, b) - \min(r, g, b)) / \max(r, g, b)$$

The following procedure describes the method for calculating the hue, *H*, which is represented by an angle between 0 and 360°. Different formulas are applied depending on the values of the input *RGB* coordinates:

$$\begin{array}{ll} (r = g = b) & H \text{ is undefined} \\ \min(r, g, b) = b & H = \frac{120(g - r)}{(r - b) + (g - b)} \\ \min(r, g, b) = r & H = 120 + \frac{120(b - r)}{(g - r) + (b - r)} \\ \min(r, g, b) = g & H = 240 + \frac{120(r - g)}{(r - g) + (b - g)} \end{array}$$

In the first case, when *r*, *g* and *b* are all equal and *H* is undefined, the colour is black, white or grey.

There are other, very similar, perceptual colour spaces called HSL (lightness) and HSI (intensity) which have slightly different definitions for saturation and intensity. As in the *RGB* system colours may be adjusted by changing any of the three values. However, with systems such as HSV and HSI, the three variables each

correspond to properties of colour we can easily understand, and thus are often easier to use when adjusting colour palettes and balancing the colour in images.

White/Pure components

Note that, every colour in the RGB system can be viewed as made up of a white component whose magnitude is $\min(r, g, b)$ and a pure colour defined by the ratio of the two non-zero colours after the white component has been subtracted from them (See Figure 7). In this case, the pure colour is not a coherent wavelength as in the CIE diagram. The saturation is the proportion of pure colour.

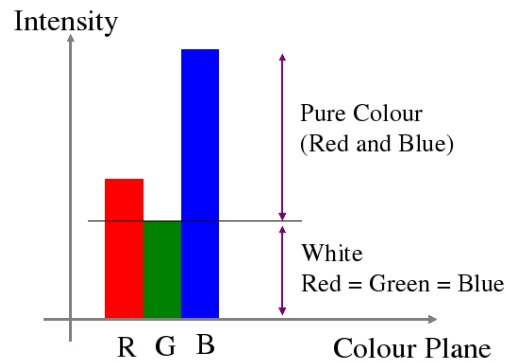


Figure 7: A tri-stimulus colour showing how the white part and pure colour pure may be separated. In this case, the green value is the smallest and defines the white component. The pure colour component is defined by the ratio of what remains of the red and blue values after green is subtracted.

Alpha Channels

Colour representations in computer systems sometimes use four components to represent each colour. The first three are R G and B and the fourth or alpha value is simply an attenuation of the intensity. This does not add anything further to the representation, but allows greater flexibility in representing colours and can avoid truncation errors at low intensities. It can also be used for providing special features such as masking certain parts of an image.