A Computer Program for Visualising the Relation between Colour and Spectral Energy Distribution

Pehr Sällström
Department of Physics, Stockholm University, Sweden
USIP Report 99-05

Abstract

A computer program has been developed for educational purposes, by help of which the relation between the spectral energy distribution of light and the corresponding perceived colour can be demonstrated and explored interactively. Thereby, among other things the phenomenon of metamerism is clarified.
Complementary colours, the colour circle, and subtractive colour mixing are treated from a physical point of view. The principles as well as the limitations of colour reproduction on CRT-displays is discussed. The program is available as an exe-file.

Introduction

On a fine summer day you may have the opportunity to enjoy and contemplate the beauty of the variegated flowers on a meadow. As you look closely at one of the flowers the question suggests itself: Why is this buttercup yellow? This seems to be a perfectly reasonable question to ask concerning the nature of colour. It can be answered in many different ways, however, according to what you are after - i.e. what kind of answer you are prepared to accept as a satisfactory explanation. Whether it is a physical, a chemical, a biological, a psychological or a philosophical elucidation you would like to get.

From the point of view of traditional physics, a common answer would run about as follows: *"Because its petals reflect those wavelengths which belong to the yellow part of the spectrum of sunlight."*

This explanation is in line with the colour theory which Isaak Newton presented in his famous Optics (1704) where he proposed that a coloured object gets its colour because it has "a Disposition to reflect this or that sort of Rays more copiously than the rest". To be sure, the rays are not coloured, strictly speaking, but each sort of ray has a certain power to stir up a sensation of this or that colour in the observers sensorium, Newton said. Thus a yellow surface looks yellow because it reflects the "yellow-making" sort of light more copiously than the rest.

On closer scrutiny this answer to our question is found untenable and even misleading in certain respects. It is true that a yellow hue is seen in a narrow region of a spectrum, at a position corresponding to wavelengths around 580 nm. And it is true that a light flux strongly dominated by photons of this wavelength looks yellow to a dark-adapted eye. But exactly the same yellow colour may be seen when the eye is stimulated with light of such a composition that there are *no* photons of wavelength at and around 580 nm. The point is that the sensation of yellow is not dependent on the presence of some kind of "yellow-making" photons, but is equally dependent on the simultaneous *absence* of certain other sorts of photons. It is first and foremost the fact that photons of short wavelengths are *missing* in the light, impinging upon the eye, that signals "yellow". A better answer to our initial question would thus be: *"The buttercup looks yellow in ordinary daylight because it is a property of its petals to absorb photons of wavelengths below 500 nm with much greater probability than above."*

In general: Single photons of any particular frequency or wavelength do not have a capacity *per se* to cause any specific colour sensation.[1] It is the distribution of photons of various wavelengths within the visible range of wavelengths that counts. Moreover, since it is found that strongly different distributions may evoke identically the same colour sensation it is not the distribution as such, but an *aspect* of it that should be correlated to colour. Various spectral distributions are equivalent with respect to this "chromatic aspect" of light. *Thus, the adequate definition of the physical correlate to perceived colour is as a set of equivalent classes of spectral distributions.*

This statement is the starting point for the discussion in this paper. There are a number of conclusions of practical importance that follow from it. Among other things it is the basis for colorimetry, as well as for existing techniques of colour reproduction.

In the following, I will go a bit further into the theoretical background of colorimetry and the definition of (physical) colour, given above. I will then describe a computer program which can be used for making these concepts and their consequences more concrete to students.

The visual situation

If you want to arrive at a thorough understanding of the phenomena of colour it is not enough to study the mechanisms of colour vision, i.e. the processes whereby radiation is detected by the photoreceptors in the retina and transformed into neuronal signals and states in the brain. You also have to consider the physical aspect of colour vision, or the stage where colours appear: the spatial configuration of material objects which the radiation encounters on its way from the light source to the observers eye.[2] From this point of view, the radiation field acts as a medium transmitting information to the observer about the material objects in the environment.

So let us consider, as point of departure for our theoretical modelling of colour vision, a situation with various objects, illuminated by ordinary daylight, and viewed by an ordinary observer. I will call this model "the visual situation" (figure 1).

The model comprises material objects at different places in a space, which might be wholly or partially closed. The objects, and the walls, are illuminated. The objects are usually opaque, but there may also be transparent objects - coloured glass plates etc. - even turbid media that the light has to pass on its way through space. The illumination is diffuse or may have a certain more or less prominent direction and in some cases the light source is counted among the objects of the situation. The situation also comprises an observer. The entrance pupil of the observer's eye selects a light flux which is a part of the total, highly complex field of radiation that builds up by reflections at the object surfaces and the walls.

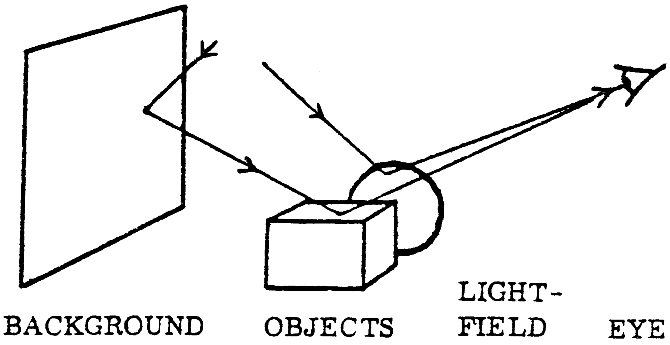


FIGURE 1 - The visual situation

The spectral distribution of the illumination, in the idealised visual situation, is essentially constant - i.e. the flux of light from the source contains about the same density of photons, independently of the photon energy, within the range of the sensitivity of the visual receptors. This is approximately true for thermal radiation at 6000 - 7000 K, as well as for sunlight or daylight. As this illumination is reflected (or re-emitted) from a surface, it is modified by selective absorption of photons, with the result that the spectral distribution of the reflected light will be different from that of the incident light. In consequence of this, the surface may look coloured to an observer. (The perceived colour quality is not uniquely determined, however, since it depends on the context. It is the layout of surfaces of different reflectances within the visual field as a whole that determines the perceived colour of the individual surfaces.)

Any visual situation, with its particular layout of object surfaces, has the capacity to generate a manifold of spectral distributions, giving rise - in its turn - to a gamut of colours, characteristic of that situation.

First order modifications of the spectral distribution

The way the spectral energy distribution of the illumination is modified is determined by the selective absorption of photons, which in turn is determined by the physical properties of the atoms, molecules and molecular aggregates that make up the material substance of the illuminated surface. Can something general be said about the possible (or in practice actually occurring) modifications?

Well, for a number of reasons the absorption resonances, representing the probability of a photon being absorbed, are quite broad compared with the width of the visual "window" 400-700 nm, with the consequence that the reflectance (and transmittance of a translucent material) can be represented by relatively smooth curves with not too drastic changes. (A study of the properties of the reflectance curves of naturally occurring materials is reported in Sällström 1977.)

Let us take a look at a typical example, the reflectances of four blue wall paints. (Fig. 2)

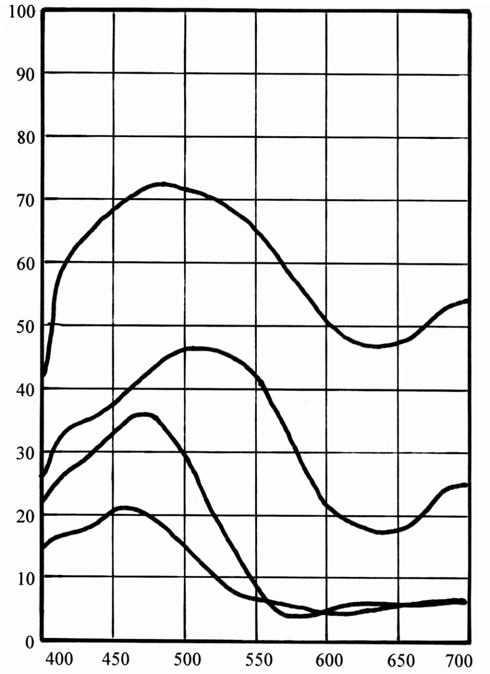


FIGURE 2 - Spectral reflectance curves for 4 blue paints.

The following general traits of these spectral curves can be observed:

1. Photons of any energy (wavelength, frequency) are absorbed. This means a general attenuation of the illumination, as it shines back from the surfaces it meets. The surfaces are more or less dark, depending on the degree of this non-selective absorption. As you see from the graphs, even the brightest of the paints does nowhere exceed 75% reflectance and the dark one just reaches 20%. This means a "black-component" in the colour of the surface.

2. Photons of any energy is re-emitted. The brightest paint, in figure 2, independently of wavelength reflects as much as 45% of the incident light, the darkest one about 5%. This non-selective reflection means a "white-component" in the colour of the surface.

3. On the whole the distribution of photons in the reflected light is accentuated towards the short wavelength side of the "window", which means a modification of the illumination that usually results in the perception of a blue colour. As the accentuation shifts towards the centre of the "visual window" the colour takes on a green hue. (The two dark blue paints have their maximum reflectance at about 460 nm. The paint that gives a blue-green colour has its maximum reflectance around 500 nm.)

Imagine a diagram, like that in figure 2, but with the curves accentuated towards the long-wave side instead. That kind of spectral distributions would represent various orange and brown colours. Interestingly enough a colour cannot be characterised as simultaneously blue *and* yellow - there are no "yellow-blue" hues - but there are yellow-red and yellow-green, as well as blue-red and blue-green hues. The qualities yellow and blue are opponent, as are red and green. When the distribution, as a general over-all tendency, is concave downwards it represents a green paint, if it is concave upwards, it represents a red or purple paint.

Consider the way a turbid medium successively modifies the spectral distribution of light passing through it. A typical example is the colour of the setting sun, which deepens to red via a sequence of yellow-red hues. The corresponding spectral distribution looks roughly as follows (figure 3):

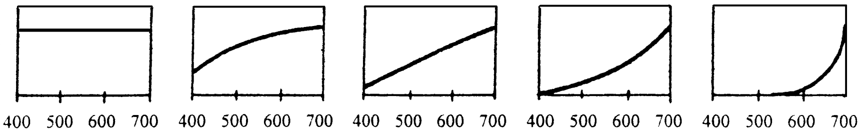
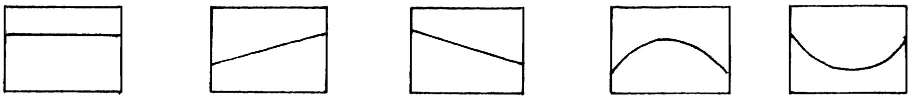


FIGURE 3 - spectral distribution of light passing through
a layer of a turbid medium of increasing thickness.

As a result of the analysis of these and similar examples of reflectance and transmittance properties of material objects it has become evident that qualitative tendencies such as lightness/darkness (or white and black components), yellowish/bluish and reddish/greenish of a colour sample can be related to general, first-order modifications of the spectral power distribution of the illumination, in which the samples are viewed, at least as long as this is perceived as colourless.



neutral yellowish bluish greenish reddish

FIGURE 4 - First order modifications of an initially constant distribution

Complementarity and circularity

A general feature of the colour world, evident from these general physical considerations, is *complementarity*. Given any colour filter with spectral transmittance $r(\lambda)$ $\{0 \leq r(\lambda) \leq 1\}$, in principle there exists a filter with transmittance $r'(\lambda) = 1 - r(\lambda)$. The two filters, placed in front of two identical spotlights, give rise to light fluxes which superimposed give colourless light, identical to the illumination from a single spotlight without filter. The two complementary light fluxes usually correspond to different colours, which could appropriately be called complementary colours. If one is blue, the other one is orange-yellow; if one is green, the other one is magenta etc. I said "usually", because the two lights must not necessarily have different hues. They may also be two shades of the colour associated with the initial light.[3] A beautiful example of (additive) complementary colours is the colours seen with double-refracting media placed between polaroids. When you rotate one of the polaroids 90 degrees each coloured area switches to its complementary colour.[4]

It follows from the theoretical discussion above that the complementary to yellow light should be blue, and that of a purple-red light green, since these hue tendencies have been related to inverse, and thus incompatible, properties of the spectral distribution. Schematically the previously mentioned first order shapes of distributions can be arranged into a circle, with complementary distributions placed diametrically opposite to each other (figure 5).

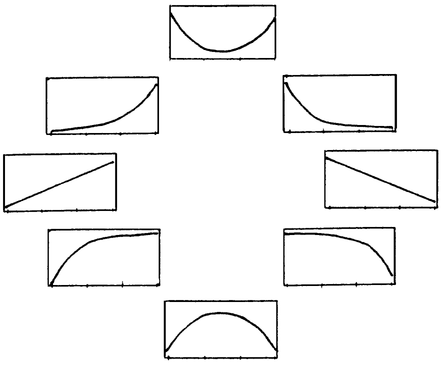


FIGURE 5 - The colour wheel, a symmetry property of the first order spectral distributions

Strong chromatic colours

The perceived colours of first order modifications of white light are quite weak and they have an appreciable amount of black and white in them. What about stronger, more saturated colours? For instance the colours of flower petals or that of mono-energetic radiation, such as laser light? I maintain that even in these cases it is essentially a question about the general shape of the distribution - i.e. the absence of photons of certain wavelengths is as important as the presence of photons of certain other wavelengths. Let us take a look at a colour wheel composed of colour samples with really strong colours, figure 6. As you see, the white and black components in the distributions are relatively small and the distributions show rapid changes between wavelength regions of strong absorption and regions of very little absorption.[5]

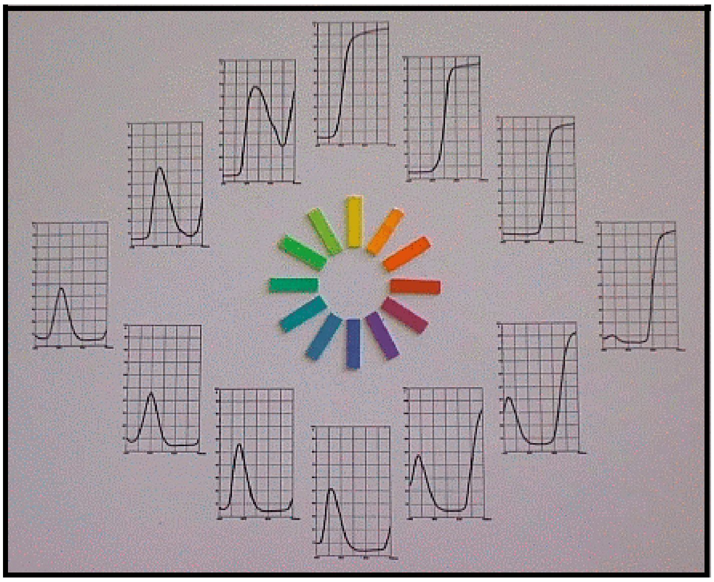


FIGURE 6 - Reflectance curves for a series of colour samples
They are arranged into a kind of hue circle, but unlike that in figure 5 this wheel has yellow at the top and blue at the bottom. Both alternatives are quite common arrangements.

One can imagine this colour wheel as being generated by the following procedure. An absorption band enters the visible window from the left, giving rise to a yellow colour, proceeds further on to the right, with orange and finally dark red as a result. As it passes further on, from this point, you get purple and violet. It is followed by a second absorption band, with the consequence that the blue peak is obtained, that further on shifts into a green peak. Finally, by successive diminution of the first absorption band, resulting in various yellow-green nuances, the series of distributions returns to yellow.[6]

What I want to make clear with this way of imagining how a series of colours around the colour wheel are generated, is that some kind of cyclic arrangement is the kind of structure one has reason to expect, given the general physical conditions of the visual situation.

Ideal colours

Contemplating this colour wheel another kind of possible idealisation, than the one in figures 4 and 5, comes into ones mind. Namely the so called "ideal colours" - or, rather, ideal reflectances, that have just two values, one or zero, and switches between these at one or two wave-lengths within the visual range. For instance, an ideal yellow filter would be one that totally absorbs light up to, say, 490 nm, and from there on is completely transparent. Or a perfect red filter absorbs everything up to, say, 610 nm, from there on being totally transparent. So, for instance a "sun-set series" of yellow-red colours, like the one shown in figure 3, can be generated by the following series of spectral distributions as well:

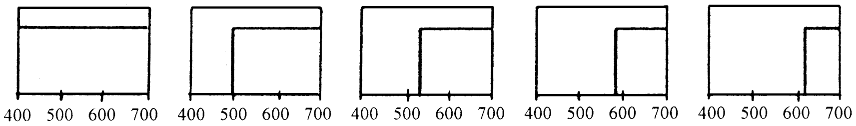


FIGURE 7 - Ideal reflectances giving a series of colours from yellow to red

White and Black

An essential feature of the visual situation model is the existence of an upper bound for the intensity of radiation. An ideal white surface, which re-emits the incident radiation to 100%, placed where the illumination is strongest, is the most luminous surface to be seen in a given situation and represents the "white-point". Likewise, the "black-point" is represented by an ideal black surface, that totally absorbs all incident radiation. In reality there are no such surfaces; a surface non-selectively reflecting 90% may very well count as white and many a black surface may actual reflect 5% of incident light. As a matter of fact, an essential feature of any model of the visual situation is that absorption is a statistical matter. No surface or filter absorbs light by subtracting a fixed amount of energy or a fixed number of photons; it is always a question of a certain proportion of the incident light. That is to say, the stronger the illumination, the more of it is absorbed by a given filter. It is essentially a question of the probability of a photon being absorbed. And this probability is in general dependent upon the characteristic frequency (or wavelength) of the photon. That is the reason why the reflectance of a surface, or the transmittance of a filter, is conveniently expressed as a percentage. It is independent of the units in which the intensity of the radiation is measured - whether you measure it in terms of traditional photometric units or in energy units or prefer to define it as density of photons.[7]

Note that according to this way of looking upon the issue, black and white are counted as colours on equal footing with the chromatic colours. To be able to distinguish a really black surface from any surface regarded in complete darkness, the visual system must however get information about the prevailing illumination of the scene. Thus a pronounced black is seen if you place a strongly absorbing surface side by side with a strongly reflecting area. The same dependence on surrounding areas is true for white. The area within the visual field that has the highest luminance is likely to take on the role of "white", even if its reflectance is far below 100% .

Serial and parallel composition of light fluxes

Let us return to the visual situation for a while. On its way from the light source to the eye the radiation passes through various media and is reflected from various surfaces. Its spectral distribution is successively modified. Likewise, any surface is illuminated from different directions by radiation that has been variously modified and in the radiation emitted from the surface these various components are superimposed. The radiation field that builds up in the room, and at each point constitutes a potential stimulus to a human eye, can be analysed and described in terms of successive modification as well as superposition of fluxes of light originating in one and the same light source. The various material media and objects work in series as well as in parallel on the link between the light source and the receptor in the observers eye. (See figure 8!)

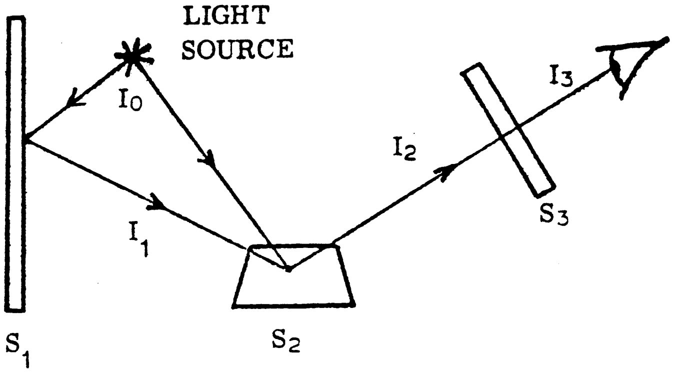


FIGURE 8 - The visual situation, once more

The classical *superposition principle* declares that superimposing light fluxes is summative, i.e. the resultant spectral power distribution is the sum of the distributions of the component fluxes, measured separately.[8]

In accordance with the superposition principle the totality of spectral distributions can be represented mathematically as vectors of an infinite dimensional linear vector space. The modifying action of material media and objects can be represented by linear operators in this vector space. Successive modification is formalised as products of operators and parallel action of media as addition of operators. Let us denote the spectral power distribution of the radiation emitted from the light source I_0, then the distribution of the radiation reaching the eye, from a certain direction, is given (in case of the example i figure 8) by an expression of the type

$$I_3 = S_3(S_2 + S_2 S_1) I_0$$

The two ways in which light is modified by material structures are used in two well known techniques: the so-called additive and subtractive colour mixing. For instance, if two colour filters are placed together they modify light successively, as when a yellow and a blue filter together modifies light the same way as a green filter would do. This is the so-called subtractive mixing. On the other hand, if two filters are placed, each in front of its own spotlight and the two light beams are directed onto the same spot, you get an additive mixture of the two coloured beams there. Typically, in this case, an orange-red and a green filter will give yellow illumination.

Equivalent visual situations

As long as ordinary colour materials (modifying light by selective absorption) are involved it is a good approximation to assume that the operators, representing translucent media and coloured surfaces, do commute; i.e. the order in which they successively modify the light flux is of no importance. (If you make a green filter by putting a blue and a yellow one together, it does not matter which filter is next to the light source). If fluorescent substances are involved, commutativity cannot be assumed to hold.

A consequence of the assumption of commutativity is the existence of *equivalent visual situations*. For instance you may put the colour filter S_3 around the light source, instead of in front of the eyes. The spectral distribution of the light I_3 reaching the eye will be exactly the same. That is, looking through a coloured glass is equivalent to having a coloured illumination of the scene.

Generality of the model

A closed visual situation, containing a single light source and a limited number of differently coloured object surfaces generates its own particular "world of colours". As soon as the spectral power distribution of the source is changed this world changes. The changes may be non-trivial: areas that had the same colour may appear with different colours and conversely, differently coloured surfaces may look alike. However, within limits, if the change only concerns the luminous flux from the source, not its relative spectral distribution, all colours would become brighter but the essential properties, hue, chromaticness and lightness, would remain unchanged. (For instance if the source is placed somewhat closer to or farther away from a number of colour samples, these will look essentially unaltered due to the visual systems capacity of adaptation. In general, a change in illuminance is related to the geometry of the situation, rather than to material properties of the objects.)

The absolute level of luminance influences colour discrimination, contrast sensitivity and the like, and to some degree even the hue aspect, but as a first approximation we may assume the hue aspect to be uninfluenced. A consequence of this "relativity principle" is that a black area, in one visual situation, may very well have the same luminance as a white surface, in another situation.

In a generalised model, we study the manifold of spectral distributions (functions of the wavelength λ) that are in principle possible to generate by selective absorption, under the condition that $0 \leq f(\lambda) \leq f_{max}$ for all values of λ within the interval, relevant to colour vision, that is 380 - 780 nm (or, in practice 400 - 700 nm). The distribution $f(\lambda) \equiv f_{max}$ represents the white-point and $f(\lambda) \equiv 0$ represents the black-point of the corresponding "colour world". Theoretically, there is no given upper limit to the intensity of thermal radiation. But in any real visual situation there is such an upper boundary, let be that it can change and that different situations may have different upper boundaries. The point is that introducing an upper limit creates a symmetry that in a way "explains" analogous symmetry features in the world of colour. As long as strict proportionality holds (and from the point of view of physics it does quite far) changing f_{max} is just a matter of scaling that does not change the structure of the system of f(λ):s generated by material substances in visual situations.

Trichromaticity and the theory of colorimetry

Let us now turn our attention to the observer and his eye. Physicists make use of instruments that measure the spectral energy distribution of a flux of light. But the photo-sensitive system of the eye does not measure spectral energy distributions. It detects certain *aspects* of the spectral energy distribution of the incident light flux. Aspects, which - as was hinted upon above - has to do with the general shape of the distribution over the visible interval 400-700 nm. Distributions which in this particular sense "have the same shape" are not distinguished by the receptors in the eye. They are *chromatically equivalent*.

To investigate this it is not enough to study one spectral distribution at a time, you have to be able to compare two lights, having different spectral distributions, side by side. Only if you realise these experimental conditions you can penetrate deeper into the connection between spectral distribution and perceived colour.

By systematically matching lights of different spectral compositions one can find out which distributions are chromatically equivalent. James Clerk Maxwell did pioneer work on this, establishing the fact of *trichromaticity* of (normal) human colour vision.
Trichromaticity means: given any set of four light fluxes, that look different in colour, you can always by superimposing in adequate proportion either three of them get a flux that matches the fourth one, or two of them get a flux that matches a superposition of the two other ones. (Theory and formalisation, see my previous publication: USIP Report 98-09.)

As a consequence the infinite dimensional linear manifold of spectral energy distributions is grouped into equivalence classes with respect to the matching criteria. Each such equivalence class is called a "chromatic valency". In view of trichromaticity, to specify the chromatic valency it is enough to give three real numbers. In the colorimetric system, standardised by CIE (Committee internationale d'eclairage) in 1931, these are the so-called tristimulus values, denoted X,Y and Z.

Two spectral distributions $f(\lambda)$ and $g(\lambda)$ have the same chromatic valency, i.e. belong to the same equivalent class, if and only if $X_f = X_g$ & $Y_f = Y_g$ & $Z_f = Z_g$.

The tristimulus values can be calculated by weighted integration, using the three matching functions $x(\lambda)$, $y(\lambda)$ and $z(\lambda)$:

$$X_f = \int f(\lambda)x(\lambda)d\lambda \qquad Y_f = \int f(\lambda)y(\lambda)d\lambda \qquad Z_f = \int f(\lambda)z(\lambda)d\lambda$$

Intuitively, this is a mathematical way of expressing the idea that the chromatic valency of a spectral distribution has to do with the shape of the distribution, more precisely three aspects of the shape. The matching functions x,y,z represent these aspects. X is a measure of the degree of conformity between f and x; Y measures its conformity with y, Z its conformity with z.

Metamerism

Two light fluxes with different spectral distributions are said to be metameric if they belong to the same chromatic equivalent class. We can denote this L_1 »« L_2 . Two different light sources emitting light fluxes that can be made metameric, by just changing their relative intensity, are said to be metameric. They will look exactly alike and the illumination from each of them onto a white paper will look identically the same (or can be made identical by just changing the distance between source and illuminated surface).

If the light fluxes reflected from two different *colour samples*, under a certain illumination, are metameric, then the two samples are said to be metameric (under this illumination). If the same two samples are illuminated by another kind of source they need not be metameric any longer. For example, even if they look exactly the same in daylight, they may look different in incandescent lamp or fluorescent tube lighting.
Denote the reflectances of the two given samples R_1 and R_2 and the two light sources I_1 and I_2 then R_1I_1 »« R_2I_1 does *not* imply that R_1I_2 »« R_2I_2.

Observe that this is true even if the two sources are metameric (i.e. I_1 »« I_2). Under such circumstances a white surface looks the same, but various chromatic samples will look different under the two kinds of illumination. This is the problem of *colour rendering* of artificial light sources. For instance, if a fluorescent tube is used to replace daylight, the colours of objects should ideally look the same. In practice this is the case only within certain tolerances. You can make a colour rendering test on artificial light sources by selecting a number of colour samples, representing typically occurring colours, and compare their colours, as seen in daylight, with their colours as seen in the artificial light. If some samples change their colour appreciably, the source is said to have bad colour rendering.[9]

Finally, in practice we often speak of metamerism in a weak sense. Two samples are said to be relatively metameric, if a change in the illuminance (the strength of illumination) on one of the samples suffices to make them match.

Colour reproduction

Various techniques for colour reproduction are based on this general knowledge of human colour vision. This concerns colour in print, as well as colour photography and colour television and colour in various video systems, analogue as well as digital. The basic principle is always the same and relies on trichromaticity and metamerism, as formalised and standardised in colorimetry.

Hence, colour reproduction need not mean true reproduction of spectral distributions. It is enough that the distributions used in the reproduction are chromatically equivalent with the ones of the original. That is, the (X,Y,Z)-values should be the same for the original and the reproduction. In the case of colour prints and photographs this can only be the case under some specified illumination. So if the reproduction and the original look quite similar in daylight, they need not do so in fluorescent tube light or incandescent light. In practice, colour reproduction is always a question of compromise.

The basic idea is, however, that a given light with spectral energy distribution $f(\lambda)$ can be represented by anyone out of a number of spectral distributions, belonging to the same equivalence class as $f(\lambda)$. In particular, it can be represented by a linear combination of the emission spectra of the three phosphors of an electronic screen: R, G, B. This is the reason why we can favourably use a computer monitor to illustrate the basic facts of colorimetry. I will now turn over to a description of how this is accomplished.

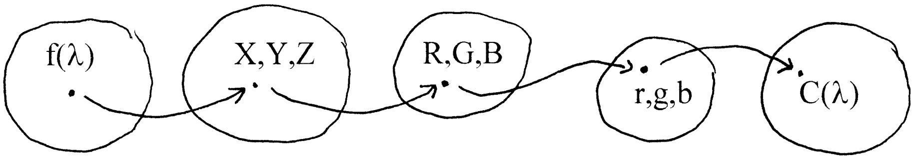
Visualising metamerism on a computer screen

If we want to visualise metamerism on a computer monitor, we need a program with the following proprerties:

- two spectral distributions are displayed as graphs
- two adjacent areas show the colours correspondig to these distributions
- each of the distributions can be independently adjusted until a match between them is found, based on visual judgement of the colours shown
- in addition, current X,Y,Z as well as r,g,b values are displayed and shown to agree when the match has been found.

The central issue is to establish as far as possible an accurate relationship between the spectral distribution and the colour currently displayed. This necessarily involves knowledge about the colour reproduction properties of the monitor.

In order to represent light of some spectral distribution f(λ) by help of the R,G,B-system of the colour monitor you have to consider a projection in four steps:



1. From spectral power distribution f(λ) to the X,Y,Z co-ordinates of the CIE colorimetric standard system.

2. From X,Y,Z to the co-ordinates R,G,B in the particular RGB-space spanned by the emission functions of the three phosphors of the monitor.

3. From R,G,B co-ordinates to the integers r,g,b of the digital code, used in the computer program that governs the monitor by setting the video signal.

4. From r,g,b code to excitation levels of the three phosphors, resulting in the emission of light with a spectral power distribution C(λ).

If this whole chain of transformations is successful, C(λ) should turn out to be chromatically equivalent with f(λ). Let us now look a bit closer on each link in the chain.

STEP 1: The first step is standard colorimetric calculation. The function f(λ) is multiplied with each one of the three given tristimulus functions x(λ), y(λ), z(λ) and each product is integrated over the visible range of λ. The result is the tristimulus coordinates X, Y and Z .

Now, in the previously described model of the visual situation, where all spectral distributions are regarded as shaped by selective absorption out of an initially given non-specific distribution, the function $f(\lambda)$ is everywhere positive and limited:

$$0 \leq f(\lambda) \leq f_{max} \qquad \text{for} \quad 400 \leq \lambda \leq 700 \text{ nm}$$

The tristimulus functions are constructed so as to be everywhere positive, hence the values X, Y and Z are positive and reach their maximal values for the equal energy function $f(\lambda) \equiv f_{max}$. Moreover, these values will be the same for X, Y and Z (this is an inherent property of the CIE XYZ colorimetric system) and we can as well put it = 100, since we are not interested in absolute photometric measures of the luminances involved. That is to say:

$$X = k \int f(\lambda)x(\lambda)d\lambda \quad \text{with} \quad k = 100/(f_{max} \int x(\lambda)d\lambda)$$

and likewise for Y and Z.

In our simplified, digital model the function $f(\lambda)$ is given as 31 numbers and the tristimulus functions are likewise given as a list of values at 400, 410 ... 700 nm. (The data file is called "cie.dat") The integration is performed as a sum of 31 terms. Moreover, $f(\lambda)$ may take on integer values between zero and 30. In the plots, the height of the bins, representing $f(\lambda)$, can be varied in 30 steps. Each time a function value is changed one step, at some wavelength, λ_ι, the three integrals are changed by adding (or subtracting, as the case may be) one elementary unit, given by $k \cdot x(\lambda_\iota)$, $k \cdot y(\lambda_\iota)$ and $k \cdot z(\lambda_\iota)$ respectively.

STEP 2: According to the fundamental laws of colorimetry any three linearly independent lights can be used to define the basis of a co-ordinate system for the manifold of chromatic equivalent classes of lights. You can easily shift between one basis and another one by help of a linear transformation. (And any such transformation implicitly defines a basis.)

In order to be able to set up the transformation matrix from (X,Y,Z) to (R,G,B) we have to know the X,Y,Z coordinates for the light emitted by the three phosphors. We make the somewhat idealised assumption that the relative spectral distribution of the emitted light is independent of the excitation level (within the range normally used on the screen). Then what we need to know is X, Y, Z at maximum emission for the three phosphors, the orange-red one, the yellow-green one and the blue one. This can be measured with a spectro-photometer, giving the emission functions $r_f(\lambda)$, $g_f(\lambda)$ and $b_f(\lambda)$ an example of which is shown in figure 9. From these the tristimulus co-ordinates for the three phosphors can be calculated as $X_R = \int r_f(\lambda)x(\lambda)d\lambda$ etc.

It should be observed that what is here meant by "maximum" emission, is the emission for an appropriate adjustment of the brightness and contrast controls on the monitor.

When all three phosphors are emitting at full strength, the screen looks white and the co-ordinates for the "white-point" of the screen are given by:

$$\begin{aligned} X_W &= X_R + X_G + X_B \\ Y_W &= Y_R + Y_G + Y_B \\ Z_W &= Z_R + Z_G + Z_B \end{aligned}$$

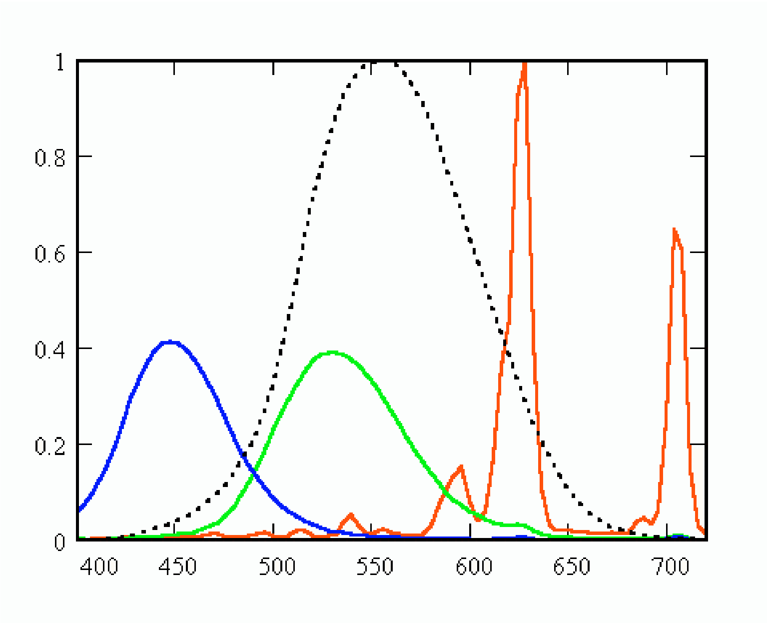


FIGURE 9 Emission functions for the phosphors of a typical display
Also shown is the brightness sensitivity curve

When the three phosphors are emitting light in a certain proportion, given by the three coefficients R,G,B, the emitted light has the spectral distribution:

$$C(\lambda) = R \cdot r_f(\lambda) + G \cdot g_f(\lambda) + B \cdot b_f(\lambda)$$

the X,Y,Z for the light of this particular composition is calculated as:

$$X = RX_R + GX_G + BX_B$$
$$Y = RY_R + GY_G + BY_B$$
$$Z = RZ_R + GZ_G + BZ_B$$

Note that the R,G,B co-ordinates are so defined that (1,0,0) denotes the light emitted at maximum strength by the red phosphor, (0,1,0) denotes the light of the green phosphor and (0,0,1) the light of the blue phosphor. In order to span the whole space of possible chromatic valencies R, G and B must sometimes take on negative values, as well as values greater than one. (This is the case for those chromatic valencies that are not possible to represent on the screen as real physical light.)

In our case, we actually want to transform from (X,Y,Z) to (R,G,B) so we have to use the inverted matrix:

$$T = \begin{pmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{pmatrix}^{-1} \qquad \text{so that} \qquad \begin{pmatrix} R \\ G \\ B \end{pmatrix} = T \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

T is supplied as the data-file "matris.dat", calculated by using measured values of (X,Y,Z) for the phosphors of a monitor, which we assume to be representative, see Appendix.

In the program the calculation is repeated each time you increase or decrease any bin in the graph of the spectral distribution by one step.

Now, as mentioned above, only those (X,Y,Z) can be reproduced as real lights on the screen which give an (R,G,B) that fulfils the condition that R, G and B should all be positive and ≤ 1.

On the other hand, in the visual situation model we delimit ourselves to those spectral distributions $f(\lambda)$ which are generated by selective absorption from an initially constant distribution $f(\lambda) \equiv f_{max}$. Hence $(X,Y,Z)_{max}$ was provisionally set to (100,100,100).

The luminance range of an ordinary monitor is relatively limited. To use it optimally for our purpose, we should reproduce our chromatic valencies with as high luminance as possible. This can be accomplished in many ways; in the present program two alternatives have been implemented. (It is essentially a question of adjusting f_{max} to an appropriate level.)

Alternative A: The maximum value that any of R, G or B can attain in transformations from (X,Y,Z) to (R,G,B) is used to normalise all other R,G,B values, i.e. values obtained with various other shapes of $f(\lambda)$. From studies, where ideal distributions were used to find the maximum of R, G and B respectively, it was found that R is the one that may reach the highest value, call it R_{max}.[10] Thus for all other cases we put $R = R/R_{max}$, $G = G/R_{max}$ and $B = B/R_{max}$.

Alternative B: For each individual colour, i.e. each momentary shape of $f(\lambda)$, the R,G,B-values are normalised, so that one of them is always at maximum, i.e. is = 1. In other words, we put $C = \max(R,G,B)$ and $R = R/C$, $G = G/C$, $B = B/C$.

In the program, here described, you toggle between these two alternatives by just pushing key X. Alternative A is the one to use if you wish to compare and match spectral distributions, produced by selective absorption out of one and the same f_{max}, when the lightness relations are important. If only chromaticness and hue is of interest to compare, then you may use alternative B; especially if one of (or both) distributions are the result of strong absorption, as when it consists of only two or three spectral lines. Alternative B can be interpreted as using different scales on the vertical axis in the two diagrams.

The so determined coefficients should be used to produce a light, with the spectral distribution:

$$C(\lambda) = R \cdot r_f(\lambda) + G \cdot g_f(\lambda) + B \cdot b_f(\lambda)$$

chromatically equivalent to the light represented by the spectral distribution $f(\lambda)$.

STEP 3: To accomplish the emission of this light on the monitor screen, the signals from the computer program to the three electron guns must be such that the three phosphors do really emit their respective lights in the particular proportion given i the last mentioned formula. To guarantee this we have to know the *excitation functions* of the hardware involved.

The colour code (r,g,b) in the program, for each channel selects one out of a number of levels of voltage in the signal sent to the monitor (usually 64 or 256). The excitation function tells what fraction of maximum emission that this level in reality means. You need not to know the details of the hardware, since the excitation function can be determined by measuring the emitted light directly on the screen by help of a photometer. It is usually not linear, a typical example is shown in figure 10.

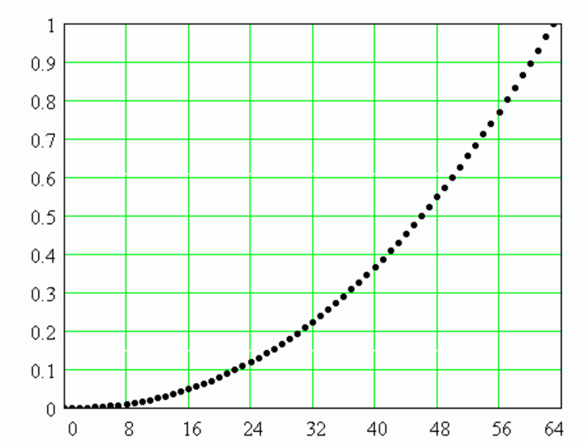


FIGURE 10 The excitation function γ

The program is designed for 18 bit DAC, which means that each of r,g,b denotes one out of 64 levels. Hence, the excitation function is supplied as a list of 64 float numbers (fixed at 0.0 at level 0 and 1.0 at level 63). The data-file is "c.dat". Remember that the R,G and B coefficients are normalised, as previously described. They are all less than or equal to one. Given, for instance, R, the instruction "r=digit(R)" calls a small routine which scans the list, finds the nearest value and returns the index for that position in the list. If any one of R,G or B is negative the colour cannot be reproduced on the screen and hence R, G and B are all three put to zero, and a warning text is printed on the screen.

STEP 4. Having determined the three integers r,g and b these are sent to the palette as the momentary setting of the colour assigned to the area on the screen where the light representing $f(\lambda)$ is displayed. This light should then have the spectral distribution

$$C_{r,g,b}(\lambda) = \gamma(r) \cdot r_f(\lambda) + \gamma(g) \cdot g_f(\lambda) + \gamma(b) \cdot b_f(\lambda)$$

where γ denotes the excitation function, as described above.

Limitations

1. The model of the colour world, we have settled to delimit ourselves to, is the chromatic valencies (i.e. equivalence classes of spectral distributions) that are generated by all kinds of functions that fulfil the condition of having all its values confined in between zero and a fixed upper limit. This condition implies that the (X,Y,Z) values obtained are confined within a certain subspace of the total XYZ-space, a "colour solid" called the Luther-Nyberg solid.[11]

Likewise, on the other hand, the possible spectral distributions $C(\lambda)$ that can be realised on a computer monitor are limited, and thus the corresponding colour valencies are confined within a subspace, in RGB-terms called the "RGB-cube", with (0,0,0) in one corner, and (1,1,1) in the diagonally opposite corner of the cube.
Our ambition is to reproduce, as far as possible, the Luther-Nyberg solid by help of the RGB-cube. With a suitable normalisation, the RGB-cube can be covered within the

Luther-Nyberg solid, but the reverse situation is not attainable. By far not all f(λ) of our model can be represented by an equivalent C(λ). This concerns for instance light with greater spectral purity than the phosphors emit. For example a deep red - the red from a Helium laser emitting mono-energetic red light. But, in fact, the same is true for some yellow, green, blue, violet and purple colours as well. Computationally these cases show up by one or more of the calculated R,G,B-values being negative. Then the box, where C(λ) is displayed, is turned black.[11]

2. Observe that what is displayed in the small box on the computer screen is a light which is designed to be equivalent to a light having the spectral energy distribution shown in the diagram. Equivalent means that the two lights should look the same under the same conditions of presentation, that is (in our case) as this particular square luminous area on the monitor. Which perceived colour - in terms of hue, lightness and chromaticness - this area would be seen to have is not predicted by colorimetry. It depends on the chromaticity and luminance of surrounding areas as well as on the current adaptation of the visual system.

3. The luminance range of an ordinary monitor is in practice quite limited, among other things due to stray light incident on the screen, if you do not work in complete darkness. The colour gamut area is severely limited for dark colours on the screen. If r=g=b=0 does not imply that the screen is completely black, a certain amount of "background" should be added to C(λ). This background is usually neutral or almost neutral and tends to diminish the saturation of the colour. The background is mainly due to the fact that the monitor is used in an illuminated room, but the brightness control set at a too high level will also have this effect of diluting the colours.

A short description of the computer program "SPECTRA"

To get the program contact the author pehr.sallstrom@swipnet.se

The program consists of five settings among which you choose in the main menu, which comes up when you start running the program. The settings are captioned:

1. Spectral distribution and colour
2. Matching two spectral distributions
3. Complementary distributions
4. Variously illuminated colour samples
5. Illustrating subtractive colour mixing

I will now in turn describe each setting, its contents and suggested use.

SETTING 1: Spectral distribution and colour

This is the basic setting, demonstrating the relationship between the spectral energy distribution of a light flux and the colour this light is perceptually associated with.

In the big box the spectral distribution is shown, in the small box the chromatically equivalent light, as composed by R,G,B, is displayed. By pushing one of the *keys 1 to 8* you bring in a default example. Any one of them can be taken as the starting point for constructing a modified distribution, using the *arrow keys*. The distribution is reproduced as 31 bins, each bin corresponding to a sampling interval of 10 nm, at sampling points from 400 nm to 700 nm. For each bin 31 levels of intensity are at disposal (including zero). This makes a total of 31^{31} possible different distributions!

You will find that for some distributions, you construct, there will not appear a corresponding colour in the small box. Instead black i shown, and you get the message: "Not possible to reproduce". This means: a light equivalent to the one having this spectral distribution cannot be reached by any combination of the emissions from the phosphors R, G and B. (More about this below!)

If you arrive at a distribution that you find interesting and want to save, just push *key S* and the distribution will be saved as a file, named "spek1.dat". Later on you may recall this distribution by pushing *key 0*.

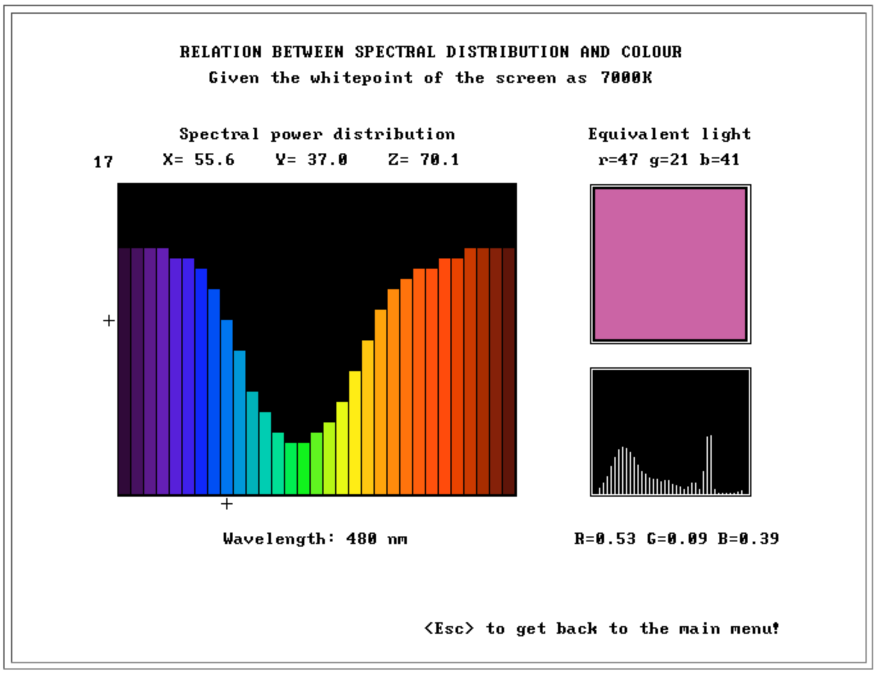
The default distributions are as follows:

1. straw	a distribution, essentially monotonously increasing with wavelength, corresponds to a yellow-making disposition
2. light blue	a distribution, essentially monotonously decreasing with wavelength, corresponds to a blue-making disposition
3. grey	a mainly constant distribution corresponds to neutral colour

4. green	a distribution where central parts dominate gives a green-making disposition to the light
5. magenta	conversely, if the short- and/or long-wave sides of the distribution dominate, then this gives a magenta-making disposition to the light
6. yellow	a stronger yellow is obtained if the distribution is characterised by having low values at the short wavelength end and high values elsewhere.
7. blue	a strong blue is obtained if the distribution is strongly accentuated towards the short-wave side.
8. sunset	here you can see the sequence of colours of the setting sun. You push 8 several times. Each time the distribution you get corresponds to a thicker layer of turbid medium. The formula for the distribution is:

$$f(\lambda) = 30 \cdot e^{-[(700/\lambda)^4 - 1] \cdot d} \qquad d = 0.0 \;\; 0.15 \;\; 0.30 \;\; 0.60 \;\; 1.0 \;\; 1.5 \;\; 3.0$$

Suggestions for using this setting: You can start, for instance, with example number 1 and pull down the staples on the short wave side of the distribution, thereby making the yellow colour more saturated. In the same way, starting with number 2, you can intensify and deepen the blue colour by pulling down the staples on the long wavelength side.



RELATION BETWEEN SPECTRAL DISTRIBUTION AND COLOUR
Given the whitepoint of the screen as 7000K

Spectral power distribution
17 X= 55.6 Y= 37.0 Z= 70.1

Equivalent light
r=47 g=21 b=41

Wavelength: 480 nm

R=0.53 G=0.09 B=0.39

<Esc> to get back to the main menu!

FIGURE 11 - Setting 1. In the left box the spectral power distribution is shown in arbitrary units (running from 0 to 30) together with the corresponding tristimulus values X,Y,Z. Also indicated is the wavelength at which the function is at the moment modifiable (450 nm) together with the current function value (14). To the right, the equivalent light, as produced by the phosphors, is shown together with a small graph showing the spectral distribution of this light. The R,G,B values are displayed, as well as the corresponding r,g,b code. The equivalent light is presented as "Optimised!" which means that it has been given the highest luminance possible. (In this case R= 1.00) The white-point of the screen, used in calculating these values, is indicated as 7000 K.

Starting with number 3 you can study what influence small deviations from constant distribution may have on the colour. Starting with number 6 or 7 you can find out how far the colour can be intensified before you reach the limit beyond which the electronic screen cannot reproduce it.

However, these and other interesting demonstrations can as well be made in the settings to follow.

SETTING 2: Equivalent distributions

Two boxes are shown. In each of them a distribution can be set up by help of any of the keys 1 to 8. It can then be modified according to your wish by help of the arrow keys. The two distributions are independent and *you shift between them by help of the space bar*.

Above the distributions a box with two adjacent fields is shown. There the two lights are displayed which are equivalent to lights with the shown spectral distributions. You may adjust these until they become metameric. (This happens when the border between the two adjacent areas disappears.)

The following distributions are default, and called by pushing 1 - 8 :

Left box: straw, light blue, neutral, green, blue, grey, D65, variable level
Right box: yellow, blue, neutral, magenta, blue, grey, three-band, variable level

Compare 1 in the left and right box and you see what makes a yellow colour intense in its hue. The same for a blue colour is shown in 2. Then you can study how distributions of different shape may match chromatically. For instance, start by 3 in the left box, then search the distribution (among the given ones in the right box) that matches that one. They are metameric grey. Go on investigating which other pairs of distributions are metameric!

Setting 2 is the adequate one to use when you want to construct an arbitrary pair of distributions, for instance a pair of metameric samples that you want to study under various illuminants (in setting 4), or transmission functions for a pair of colour filters that you want to study in their capacity to create a new colour by putting them together, as in subtractive mixing, setting 5.

If you push key 8 repeatedly, you set up a constant distribution of increasing level. It starts at zero and increases in thirty steps, then begins at zero again. Key 8 has the same function on both diagrams, independently.

With *key X* you toggle between the two alternative calibrations mentioned on page 16. Normally the lightness of the colour shown is relative to the maximum full spectrum. But when the signal "Optimised!" is shown, the current distribution is shown as a colour with as high luminance as possible. If you establish a metameric pair of distributions, both optimised, then the scale is different in the two diagrams. In the non-optimised position the scale is the same in both diagrams, so the distributions can be directly compared. For instance, at key 7 you have standard daylight to the left; to the right you have so called "three-band light". They are shown to be metameric, after pushing key X.

A particularly worthwhile pair can be saved as a file by pushing key S. Actually you get two files called "spek2v.dat" and "spek2h.dat". At a later occasion you can recall these distributions by pushing key 0. The two files can be read and used in settings 3,4, 5.
Pushing key 9 you set up a pair of distributions previously saved from setting 3.

SETTING 3: Complementary distributions

This setting is very much like setting 2, but the two distributions are not independently variable. What you take away from one of them is added to the other one.

Keys 1 - 5 give some default pair of distributions to start from. Keys 9 and 0 gives you the possibility of using pairs you have earlier made and saved.

The default examples are as follows:

1. To the left yellow, to the right a very dark, neutral colour.
This setting can be used for the following important, thought provoking demonstration.
From the distribution shown to the left you take away energy in the yellow part of the spectrum, i.e. at 580 nm and 570 nm. These are then automatically added to the distribution to the right. The colour corresponding to this is dark brown. By pushing key X you can however optimise the luminance of both lights, which makes it easier to see that they both have the same hue, and furthermore the same hue as the one you started with on the left side. Yet one of them is dominated by photons with wavelengths at 570 and 580 nm, the other one has no photons in this (yellow!) part of the spectrum.
This shows that the colour is not determined by the presence or absence of any specific wavelengths in the distribution, but rather by the distribution as a whole. The characteristic feature of a spectral distribution of a light flux, perceived as yellow (as compared to daylight, perceived as white, i.e. colourless), is that it is low at short wavelengths, i.e. the violet end of the spectrum.

2. To the left orange-yellow, to the right a dark, neutral colour.
This example is similar to the previous. From the left one you take away energy in the yellow part of the spectrum. You start at 580 nm, then 570 and 590 and partly 560 and 590 nm. This time you end up with two *metameric* distributions, both of them yellow - or, since they are relatively dark, brown. (By pushing key X you can increase the luminance of both lights, which makes it easier to see that they have the same hue as the one you started with on the left side.) Still, one of them is dominated by photons in the region around 580 nm, the other one has no photons in this part of the spectrum! (See figure 12)

3. An example of two complementary colours, related to spectral distributions that add to constant distribution: Yellow and blue.

4. The same, but based on idealised spectral distributions.

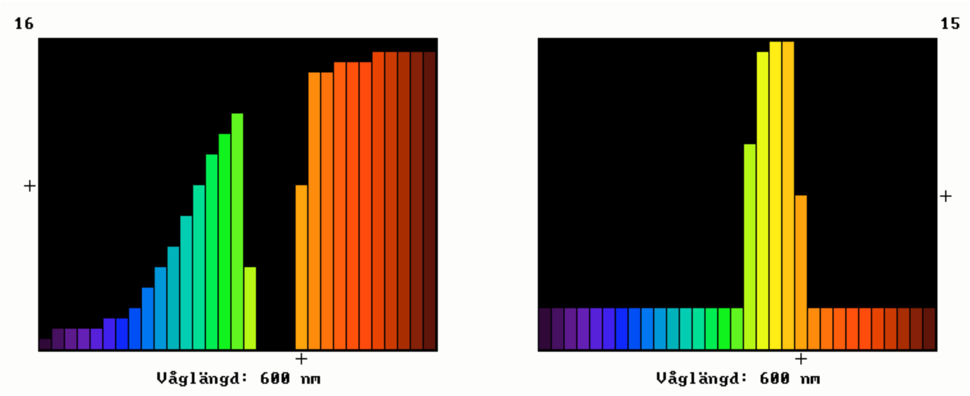


FIGURE 12 - Metameric yellow lights. Both distributions represent one and the same yellow colour. The one to the right is dominated by photons with wavelengths in the yellow region of the spectrum, whereas the distribution to the left has no photons belonging to that part of the spectrum.

5. Another example, magenta and green as ideal distributions. A certain amount of neutral has been added - otherwise the colours could not be shown on the screen.

6. Gives you a further means for studying complementary colours. To the left you have a full spectrum, to the right essentially darkness. What you take away from the left side is added to the right side, and the corresponding colour grows from a dark one to a successively brighter one. The two colours are always complementary in the sense that the corresponding spectral distributions add up to white. You can construct pairs of strong complementary colours as well as weak pastel colours.
Suppose you carry over all light at 540 nm to the right side. There you get a distribution that i essentially mono-energetic light, in this case dark green (dark because it is relatively little energy). To the left you have a distribution which is full-spectrum with a "dark line" in it. This distribution corresponds to a light pink colour. In principle: to each mono-energetic (or narrow-band) light there exist a complementary light with a spectral distribution consisting of a full-spectrum *minus* a mono-energetic (narrow-band) component. Try for instance with taking away 450, 460 and 470 nm, creating a quite broad dark space in the full spectrum. This results in a bright golden yellow colour and a deep blue. Push key X to see the example in bright position.

By pushing key 9 you can set up a pair of distributions, previously constructed in setting 2. And, as usual, by pushing key S you can save the current pair of distributions, and call on them later with key 0.

SETTING 4: Illuminated metameric colour sample

In this setting the degree of complication has increased, in that now you study simultaneously three spectral distributions. The first one to the left represents an illumination, falling onto a pair of colour samples, the reflectances of which are shown in the other two boxes. The box above shows the illumination and (as two smaller boxes inside) the illuminated samples.

You can select various illuminations. In some of them the two samples will look identical, in other illuminations they will look different. So this makes it possible to demonstrate the practical consequences of metamerism.[13]

You may choose between 4 given illuminations (key 1-4):

E	constant distribution ("equal energy spectrum")
D65	standard daylight at 6500 K correlated colour temperature
T65	three-band fluorescent light, matched to D65
A	standard incandescent light

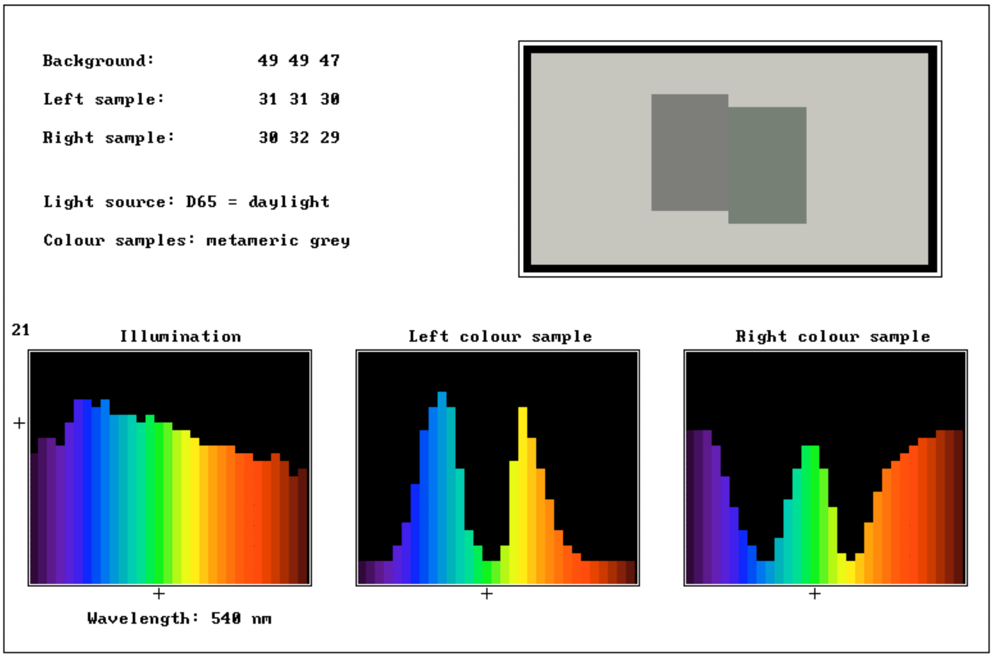
The distribution can be modified by help of the arrow keys, saved with S and recalled with 0. The default examples of metameric colour samples are:

Key 5	Metameric grey (cp. number 6 in setting 2)
Key 6	Metameric yellow
Key 7	Metameric green
Key 8	Metameric blue

For each of these - study them in the various illuminations 1-4 ! Construct your own examples in setting 2 and test them in this setting! (You save them in setting 2 by pushing S, then push 9, in the present setting, to recall them.)

Note that, since it is not very rewarding to try to reproduce absolute luminance on an electronic screen, the illuminants 1-4 are in relative terms, being automatically normalised so as to take best care of the limited luminance of the screen.

You will find it highly curious that incandescent light (key 4), as displayed on the monitor, looks so strongly coloured. It is hardly to believe that this colour should be chromatically equivalent to the colour of a white paper, illuminated by an ordinary electric bulb! However, this is actually the case, as you can convince yourself about by matching the two (by a procedure described in my previous report, USIP 98-09).



Background: 49 49 47

Left sample: 31 31 30

Right sample: 30 32 29

Light source: D65 = daylight

Colour samples: metameric grey

Wavelength: 540 nm

FIGURE 13. Setting 4 with illuminant D65 and metameric greys

SETTING 5: Illustrating subtractive colour mixing

Another situation where three distributions are shown is so called subtractive colour mixing. A flux of light passes through two colour filters, one after the other. The spectral power distribution of the light flux is modified in two steps.

You may for instance have a yellow filter and a light blue filter. When the light, initially white, has passed these two filters its spectral distribution has changed in the same way as it would have been changed by a green colour filter. The rule of thumb says: yellow and blue, subtractively mixed, give green.
Mathematically subtractive colour mixing corresponds to multiplying (wavelength by wavelength) three spectral distributions: that of the illumination, that of the first filter, that of the second filter.

In this setting three boxes are shown, displaying the spectral distribution of the first filter, that of the second filter and the product of them. The illumination is assumed to be the equal energy case. The box above shows the initial as well as the resultant colour, if possible to display.

Using this setting you see clearly why yellow together with blue gives green - and what shape of distributions will give the best green - and moreover, why certain filters give just black subtractively. You will also see what happens when two identical filters are put together. Usually the colour becomes more saturated and shifts towards red. Interestingly enough it may also happen that two filters, of the same colour, will give a completely new colour when put together. For instance two green filters, which together make a red filter. This is called *dichroism*.

The default examples are as follows:

1. Straw and light blue give a faint green
2. Yellow and turquoise give green
3. Yellow and blue give dark green
4. Yellow and magenta give red
5. Green and magenta neutralise each other to grey
6. Two orange filters make a deeper and redder orange
7. Two moss green filters that make a dark red (dichroic filters)
8. Two medium grey filters give a darker grey.

As usual you can select any one of the cases as starting point and modify the distributions, by help of the arrow keys, according to your wish. You shift between the two filters with the space-bar. You may save a couple of distributions with S and recall them with 0. Pushing 9 introduces the two distributions you have previously set up in setting 2. You may for instance study how metameric distributions behave when modifying light in succession. What about complementary distributions?

Note that this setting can also be interpreted as displaying the case when a coloured light (distribution 1) illuminates a colour sample (distribution 2). This pertains to the situation on a theatre, when coloured floodlighting is used together with painted scenes.

Some further demonstrations

The empirical fact that human colour vision is essentially trichromatic means that the totality of chromatic valencies can be described with reference to three independently variable lights. As far as colour is concerned, this means that the spectral distributions with 31 independently variable intensity values is heavily over-determined. In fact you need only three of them. If you select for instance the bins at wavelengths 620 nm, 540 nm and 460 nm and put f(λ) = 0 elsewhere, *by varying these three you can generate all colours possible to display on the monitor.*

To study this, use setting 2. In the left diagram, set up green (by pushing key 5) . Push X. Go to the right diagram (space bar!), set up zero background by pushing key 8. Then start building a function with non-zero values at 460, 540 and 620. Push X to see the result at full luminance. Adjust the bins until you get perfect match between the two colours!

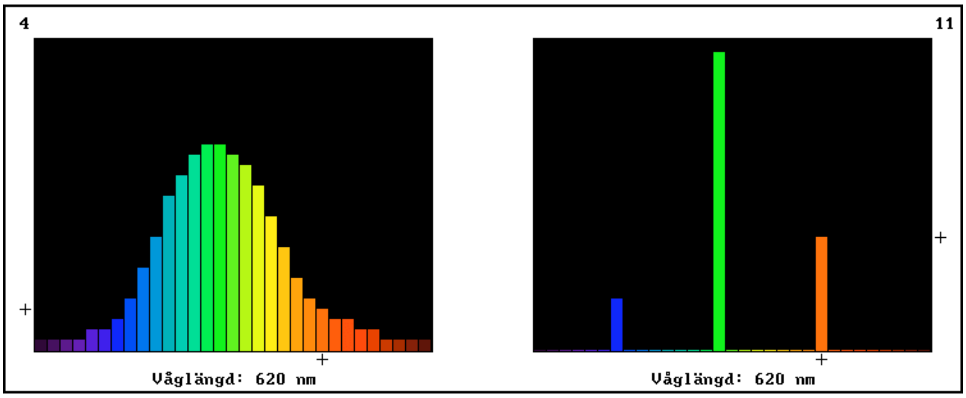


FIGURE 14 - Two distributions belonging to the same equivalence class.
(The vertical scale is not the same in the two diagrams)

Try to match some other colours, including grey (key 4). Do also try to find matches using some other triple of wavelengths. You can verify that to match a neutral grey it may suffice with just *two* monoenergetic (monochromatic) lights, if the wavelengths are chosen appropriately. For instance, try with 580 nm and 480 nm. (Push X to get the result as "white"). In a weak sense, such wavelengths can be called "complementary". How is it possible that two single wave-lengths add up to white light?

Evidently, complementarity can be defined in various ways. I have preferred to define complementary spectral distributions as those which add up to a full constant distribution at the level of the white-point of the colour system. (One may speak of complementarity in a strict sense by adding the condition that the two distributions should not have any wavelength component in common. That is, if realised as two filters, they should give just black, if put together. Complementary idealised distributions behave that way.) So far this concerns complementarity in a physical sense. One may also speak of complementarity in a perceptual sense. Any two members of the equivalent classes to which a pair of complementary distributions belong would add up to a light belonging to the same equivalence class as the constant distribution, i.e. the class of the white-point. So even chromatic valencies can be said to be complementary (not only spectral distributions). Next, spectral lights that add up to form a distribution that happens to be equivalent to a constant distribution (thus perceptually neutral, or achromatic) may be called complementary, in this weak sense. Of course "quasi-white" light composed of only two spectral lines is useless as illumination, since the colour rendering is very bad.

Some questions

1. *Is the colour displayed in the small box on the screen reliable?*
This is a tricky issue, in fact. The perceived colour of light of a given spectral distribution depends on the situation, for instance surrounding areas and the state of adaptation of the eye. The relation between spectral distribution and perceived colour is not fixed but variable.
What the program does, is to present - in the small box on the screen - a light that, from the point of view of the so called "standard observer", is chromatically equivalent with light having the spectral energy distribution shown in the graph below. Equivalence means that the two lights, under exactly the same circumstances, to a standard observer would give rise to exactly the *same* colour perception - but *which* colour (hue, saturation, lightness) depends on further conditions.
When looking at a computer screen you are adapted to a "white" which is rather cool (correlated colour temperature, say, 7000 K) which means that the colour you see in the little box on the screen is probably somewhat warmer than this light would seem under usual conditions (for instance indoor daylight or incandescent light). Make the following test: Take a piece of white cardboard, make a small hole in it, and hold it so that you see the little colour box on the screen through it, but does not see any other parts of the monitor screen!

2. *How does one calibrate the program against the computer display?*
Calibration is usually a heavy task. If the display and the program are not calibrated towards each other, the colour shown in the small box will not be "true", that is to say, the light presented there will not belong to the same equivalent class as light having the spectral energy distribution shown in the graph. A strict calibration is only possible if you are in a position to be able to measure your display spectro-photometrically. What you need is the excitation function(s), and either the X,Y,Z values for the three phosphors at maximum excitation, or the chromaticity coordinates of the three phosphors and the white-point. From these you construct the transformation matrix "matris.d", used by the program.
If you cannot do these measurements, you have to rely on the default matrix and excitation function - for educational purposes this will certainly be satisfactory. The most problematic aspect is the white-point, since that varies quite a lot from screen to screen. The default is intended for a correlated colour temperature or 7000 K.

3. *Why not include the effect of chromatic adaptation?*
If a certain light on the screen, representing a colour sample, shall give you the same impression of hue, chromaticness and lightness as that colour sample regarded in ordinary illumination would do, it might have to be shifted somewhat away from red. Since, if you have adapted to a screen-white of 7000 K, this adaptation tends to neutralise the "blue look" of samples, regarded in such light. The program does not include any feature to take care of chromatic adaptation of the visual system. But there is a possibility to use the transformation "matris.equ" (instead of "matris.d") which is constructed so as to make an equal energy spectrum being reproduced as R=G=B, i.e. look white if you are

adapted to the screen's white-point. For educational purposes this would do as a first approximation to a correction for chromatic adaptation.

4. *What about colours that lie outside the RGB triangle?*
The only colours (i.e. equivalent classes of spectral distributions) that can be represented on the display are those which contain some linear combination of the lights from the three phosphors, the orange-red, the yellow-green and the blue.[14] You cannot represent a deeper red than the orange-red phosphor, and not a green, more saturated than the pale one, obtained by mixing G with an appropriate amount of B, to get a pure green hue. As a matter of fact laser light, as well as the light from light diodes, cannot be matched against any colour on the screen. The same is true for some very strong colour samples produced with dyes and paints. Interestingly enough, if you have, say, a yellow diode at some distance from the screen, you may find that it looks identical to a yellow you produce on the screen by a certain mixture of R and G. But if you arrange it so as to be able to compare side by side the two yellow lights, then it suddenly becomes apparent - a quite striking effect - that the yellow of the screen is pale. We seem to have quite limited capacity of visually judging saturation, or chromaticness, in absolute terms. Alternatively, you can address it as a kind of contrast effect.

Anyway, it quite often happens, when you design a spectral curve on the diagram in any of the settings of the program, that you arrive at a distribution, the colour of which cannot be displayed. In that case, the small box is set to black, and you get the message "Cannot be reproduced!" (Keep an eye on the r,g,b -values displayed, so that you know which one needs to be compensated, to get back to a colour possible to display.)

5. *Can you really display a pure spectrum?*
Certainly not! The colours shown in the diagrams where the spectral distributions are displayed are illustrative, to make the picture more vivid. Monoenergetic light belongs to equivalent classes that lie far outside the subspace of RGB that can be reached by superimposing light from the three phosphors. But the hue can be reproduced, if not the saturation. In the default palette, given as two files "high.dat" and "low.dat" use has been made of the known position of the unique (or logically defined) hues in a spectrum. Yellow, defined as the point where the hue is neither green-yellow nor orange-yellow, is at 580 nm. Orange, as the colour that is as yellow as it is red, is at 600 nm. Blue, defined as the point where red-blue changes over to green-blue is found at about 470 nm and the hue that is as blue as it is green is at 500 nm. The whole range between 510 and 570 looks essentially green. From 620 nm upwards the spectrum looks essentially red. The brightness of the spectrum is greatest at 555 nm, decreasing relatively symmetrical towards both ends. (See figure 9 !) Since the data-files are in ASCII-code, you may easily adjust the colours of the "spectrum", as displayed on your screen, so that they really fit these criteria, according to your own visual judgement.

Acknowledgements

The facilities put at my disposal within the Science Laboratory at the Department of Physics, Stockholm University, are gratefully acknowledged. The study was supported by the FLIP project (Flexible Learning in Physics).

Appendix

Data used in the program "spektral.exe":

Data-files used for computation:

"cie.d"	tristimulus functions for calculation of X,Y,Z
"matris.d"	3x3 matrix for transformation from X,Y,Z to R,G,B
"c.dat"	excitation function from r,g,b to phosphor radiation
"red.d"	radiance of the red phosphor
"green.d"	radiance of the green phosphor
"blue.d"	radiance of the blue phosphor

The default matrix, "matris.d", is based on the following measured three-stimulus coordinates of the three phosphors RGB and white-point of a typical monitor, formally adjusted to a correlated colour temperature of 7000 K:

	R	G	B	white-point	
X	44.68	34.08	20.12	98.88	
Y	24.81	73.16	8.864	106.83	cd/m^2
Z	2.586	13.02	105.5	121.11	
x	0.620	0.283	0.150	0.303	
y	0.344	0.608	0.066	0.327	

The matrix in "matris.d" is the inverted matrix, thus:

	X	Y	Z
R	+0.02994	-0.01313	-0.004606
G	-0.01022	+0.01836	+0.0004061
B	+0.000527	-0.001944	+0.009541

In addition "matris.d" contains an index giving the corresponding colour temperature, in this case 7000.

As an alternative, the matrix "matris.equ" is supplied. It is based on the same x,y values for the phosphor, but with the white point at (0.333 0.333) i.e. the point for an equal energy spectrum, having X=Y=Z. By backing up "matris.d" as "matris.d70" and then copying "matris.equ" to "matris.d" you can use this instead in the program, which for instance makes the background in setting 5 white. (Then, of course, if your monitor is really at 7000 K, the light displayed in the small box is no true equivalent to the graph shown, but is on the other hand more suitable for educational demonstrations.)

The excitation function (data file: "c.dat") is assumed to be the same for the three channels, and given as a power function between 0 and 1, with gamma = 2.204, which means 50% excitation at level 46, if maximum excitation is at level 63.

Take care that the brightness control on the monitor is adjusted such that you really have black at the lowest excitation, and avoid getting stray light from windows or luminaries onto the screen!

The software

The program was written in the C language in the form of six subprograms, compiled and linked together into a MS-DOS exe-file by Borland's Turbo C compiler. To be able to use advanced features of SuperVGA graphics, the library SVGACC, developed and licensed by Zephyr software, Texas, was used together with the standard libraries. (For more information on this graphics library see "www.zephyrsoftware.com")

The program is down-loaded in form of an archive file "spektral.exe" which is self-expanding, when started. After that, you click on (or order run of) "spekt.exe". All other files, packed in the archive, are data-files in ASCII format.

This means that you can identify the data files and rename them and make changes inside them, should you wish. In "readsp.txt" the current data-files are listed.

Notes

1. What is the colour of a single photon? (Well in practice you will not see it, but if you could ...) There is a certain probability of this single photon being absorbed by anyone of the three (or four) receptor types (cones and rods) in the retina. It could trigger a long-wave sensitive cone, but could alternatively trigger a medium-wave cone or a short-wave cone. So it could evoke a red, a green or a blue sensation, as the case may be. In short: "the Rays are not coloured" in the strict sense that they do not even have the *disposition* to stir up the sensation of this or that specific colour. When the receptors are stimulated with a flux of photons of a certain wavelength the rate of photon catches is however systematically different for the three types of cones. Say that the wavelength is 500nm, then the rate of catches in the long-wave cones is half that of in the medium-wave cones, and the catches in the short-wave cones is about one tenth of that in the medium-wave. At 633nm (red laser light) the catches in medium-wave cones are about a tenth of that in the long-wave cones.

2. In other words, it is not enough to study the *working* causality - the chain of local interactions - that underlies the process of colour vision; you also have to study the *formal* causality governing these processes. Ernst Cassierer says, in his excellent little book "Logik der Kulturwissenschaften" (1942):
"Form-Analysen und kausale Analysen erscheinen nunmehr als Richtungen, die einander nicht widerstreiten, sondern die einander ergänzen, und die sich in allem Wissen miteinander verbinden müssen."

3. One may partition a full spectrum into two parts in several ways - for instance: slice it up in narrow wavelength intervals (as the bins in the diagrams in the program) and collect every other slice to one distribution and let the remaining slices constitutes the other one. The narrower the interval, the more similar the two distributions will be - both will do as white light - but nevertheless they will have no single wavelength in common, i.e. be strictly complementary! In conclusion: Complementary distributions will not necessarily show complementary colours; but if you divide the full spectrum of white light in two parts and one of the parts turns out to be coloured, then the other part is also coloured, and the two colours (hues) may reasonably be called "complementary colours".

4. The spectral distributions produced by putting double-refractive material between polaroids can all be described as parts of a sine function. It may by constituted by only a few degrees along such a curve (appearing as variations between light and dark, slightly bluish or yellowish tones) or it may comprise about one period (which gives the most brilliant colours) or it may comprise several periods, resulting i more dull an finally achromatic colours. From the point of view of colormetry, all in all it turns out to be roughly the same colour gamut as the one obtained with the RGB-system of an electronic screen. Which means e.g. no deep red hues. (For a theoretical treatment see Sällström 1996)

5. In fact these colour samples where especially prepared long ago by the research group of Karl Miescher in Basel, trying to replicate as fairly as possible the colour wheel of Wilhelm Ostwald's "Vollfarben".

6. The procedure can be modelled from an equal energy spectrum by cutting out a region of wave-lengths of appropriate width, and letting this "absorption band" pass through the spectral range between 400 and 700 nm. To generate the full circle one has to use two consecutive absorption bands, not to far apart. As far as I have found, by help of computer simulations using standard CIE 1931 colorimetry, the optimal choice seems to be two absorption bands of 150 nm width, 100 nm apart. This filter generates a series of chromaticities of "full-colours" as it passes through the spectrum.
Generalising, any kind of comb-filtering of white light generates a colour circle around the white point of the chromaticity diagram. According to the parameters, it may be more or less saturated and more or less symmetrical or accentuated in the yellow/blue or red/green directions. (See Holtsmark and Sällström 1999)

7. You may wonder: Why plot energy versus wavelength in the diagrams? Since the light detection at the cones in the retina is most likely a photo-chemical process, and thus based on absorption of photons with a probability which is a function of the photon energy (and thus the frequency of the light), it would seem

most adequate to represent the spectral distribution graphically as photon density versus frequency. If one does this, a constant function will better represent neutral or "white" light than in our case.
The intensity of the flux in terms of number of photons per unit area and second is not the same as intensity measured in terms of energy per unit area and second. A photon carries an amount of energy which is a function of its frequency, ν. Furthermore, since frequency is inversely proportional to wavelength, the scale on the horizontal axis will also be somewhat different if you choose to plot photon density versus ν instead of versus λ. Sometimes the wave number, λ^{-1} is used.
However, the established praxis is to show energy (actually power) versus wavelength, since energy and wavelength is what one measures. In a program used for educational purposes I did not find it justified to try to introduce a new praxis.

8. Intuitively, the principle is not self evident, since it means that light does not take up space (as material entities do). If we look at each other and I see your eye and you see my eye, must not the rays of light collide on its way between us? Furthermore, how can all these rays of light from the environment in front of me enter into my eye through the narrow entrance pupil without there being severe crowding, destroying my visual impression of the world? Already Alhazen, in his famous book on vision, translated into latin some time around 1200, discusses this issue in a surprisingly "modern" way (see for instance D.Park: The Fire within the Eye - A Historical Essay on the Nature and Meaning of Light, Princeton univ. Press 1997, p. 83 f) and Roger Bacon, the "doctor admirabilis", in the thirteenth century, discussed the nature of the *species* that were thought to be in some way responsible for the communication between the external object and the visual organ. Species cannot be three-dimensional, Bacon points out, since if they were, it would be impossible for two rays of light to intersect and afterwards go on following their original separate directions. (See G. ten Doesschate: Oxford and the Revival of Optics in the Thirteenth Century. Vision Res. Vol.1, p.332, 1962).
What I here call "the superposition principle" has many forms. The superposition of intensities of light is the oldest one. It is valid under the conditions I characterise as "the visual situation". Next step in the history of ideas was to describe light as a kind of periodic motion (a "wave"), and then the fundamental superposition is that of amplitudes, not intensities, which opens up the possibility of interference phenomena. In modern physic, "the superposition principle" is usually referred to as the basic postulate of quantum mechanics; in that case it again means something different, since it applies to quantum mechanical states (or "probability amplitudes"), which is a much generalised concept, as compared with the concepts used under the heading of "waves" in classical physics.

9. The so called colour rendering index (Ra-index) of artificial light sources is based on this idea. It has however serious shortcomings. One is that it only compares metameric sources. How should sources with different colour temperature be compared? As is well-known the spectral power distribution of incandescent light differs appreciably from that of daylight, giving a relative prominence to the long-wave end of the spectrum. However, the colorimetric shifts obtained in going from daylight to incandescent light are aesthetically acceptable and in most cases not even noted. On the other hand, the shifts obtained when we go from daylight to fluorescent tube light, are sometimes not acceptable. The Danish architect and designer Poul Henningsen pointed out the difference between "distortions" and "transpositions". He says: "If a blue tie turns a bit darker in candle light, the other colours do also follow to various degree, thus it is a transposition, not a distortion." (P.Henningsen: Om Lys, Kobenhavn 1974) Going from one kind of illumination to another, the colorimetric shifts of colour samples may be appreciable, but a long as all of the significant colour relations within a scene are consistently like those of the original scene, the colours will actually seem to be the same in both. This is what Ralph Evans called "the consistency principle" (see R.Evans: The Perception of Color, 1974, p. 205).
Another aspect of importance for assessing the colour rendering capacity of light sources is how colour scales transform under various illuminants. Suppose we start with a fine-graded scale, equally spaced according to some visually assessed quality such as lightness, hue or chromaticness. Under a new illuminant, the individual steps in the scale may be more or less influenced some of them may degenerate into one; at other places there will appear gaps and clusters. It may even happen that the order between neighbouring steps is exchanged. Gaps may lead to harsh contrast, glare, veiling. Degeneration and clustering means a reduced discrimination in certain regions of colour space. Visual texture may be influenced: it can get course and exaggerated or alternatively washed out, according to circumstances. Such effects are typical of three-band sources. (See Sällström 1977 or Valberg, Seim and Sällström: Colour Rendering and the Three-Band Fluorescent Lamp, Proc. of the 19th Session of CIE, Kyoto 1979.)

10. From the T-matrix (see the Appendix!) you can see that R gets its highest value if X is big and Y and Z are small. In principle, if X=100 and Y=Z=0 R would be 3.0, but this is never the case, because the matching functions x,y,z overlap to a great extent. Trying to find a distribution which gives a high value of X necessarily implies that you also get contributions to Y and Z. The result is that R will at most be 1.58 in case of this particular T-matrix. The value of R for X=Y=Z=100 (i.e. the white-point) is 1.22, hence a strategy that works with an arbitrarily normalised matrix is to put $R_{max} = 1.32*R$ for the white-point, as calculated from the elements of the matrix. This is the way it is done in the program.

11. So called after two of the pioneers on this field, R.Luther (1927) and Nikolaus Nyberg(1928). In his beautiful article, "Zum Aufbau des Farbenkörpers im Raume aller Lichtempfindungen", in Zeitschrift für Physik **52,** 406-419, Nyberg introduces complementary colours and discusses the symmetry properties of the colour solid; he shows that the solid can be completely constituted by ideal colours (i.e. colour materials having idealised reflectances; se page 7 above) with Ostwald's "Vollfarben" as a ring on the surface. For further information on the importance of ideal colours, optimal colours and "Vollfarben", see for instance P.J.Bouma: Physical Aspects of Colour, Eindhoven 1971 (published by MacMillan)

12. Actually, in many practical cases neither the chromaticness (saturation) nor the lightness can be reproduced. To give you a concrete example. My colleagues at the Physical Institute wanted to display on a computer screen (and even print out on a colour printer) how the colour of ice changes with the thickness of the layer. They had made measurements of the spectral absorption of ice at the Antarctic, where they had found up to kilometres of very clear ice. In principle (according to my calculations based on colorimetry) the ice, being at first colourless, then turquoise, should deepen into blue an even violet, with increasing thickness. However, simultaneously it should darken by several exponents of 10. So, faithfully displayed on the screen, the colour of the ice would turn black before you could see the change of hue from turquoise to deep blue. To be able to follow the hue change, you had to increase the luminance, when the thickness of ice was increased - for instance assuming that you made use of a successively stronger light source, to shine through the ice. Moreover, the deep blue and violet tones in reality were much more saturated than could be represented on a computer monitor.

13. Concerning the choice of sources. D65 is the most usual choice to represent daylight. That we also want a light, representing incandescent light, is self-evident. But what about "three-band"? It is meant to represent fluorescent tube lighting. During the sixties it was thought that to get good colour rendering with fluorescent light one has to try to make the spectral energy distribution relatively continuous, containing all wavelengths over the range to which the eye is sensitive. In other words, one tried to imitate ordinary daylight and sunlight. This is however somewhat against the philosophy of fluorescent tubes, which are based on electrical excitation of mercury vapour, since this emits its energy at certain specific wavelengths, mainly in the region of UV, blue and yellow-green. The UV component is taken care of in the coating of the tube, where it is converted into a red component.
In the seventies it was found by extensive studies, by among others W.A.Thornton at Westinghouse, New Jersey, that good colour rendering properties could be obtained with light having the energy concentrated at three relatively narrow bands, if these were accurately positioned. The most demanding one is the red band that must be at or near 610 nm. If at lower wavelengths, red materials will be reproduced as dull, brownish colours. On the other hand, if it is positioned at higher wavelengths, the red materials will look self-luminous and very unnatural. (See Sällström 1997)
The spectrum used here is based on measurements on an ordinary so called "low-energy-lamp", but adjusted to a much cooler composition of the three bands, so as to match D65.

14. In view of the fact that the phosphors in use are actually far from perceptually pure red, green and blue, Klaus Richter at BAM in Berlin suggests that instead of R G B one should use O L V to designate the primaries of electronic displays. O stands for "orange-red", L for "leaf-green" and V for "violet-blue". Then R G and B are reserved for elementary colours (according to Hering's definitions). See Backhaus et al (ed.) 1998.

References

1. For a general overview of current progress in colour vision research, see for instance:

Backhaus, Kliegl & Werner (ed.) : Color Vision - perspectives from different disciplines
W. de Gruyter, Berlin, New York 1998

Byrne and Hilbert (ed.) : Readings in color - II. The science of color. MIT Press 1997

2. Complementary essays by the author:

P. Sällström: Colour and Physics - Some remarks concerning the physical aspects of human colour vision. USIP Report 73-09 (May 1973)

P. Sällström: Puzzling Questions about Colours. (Lecture draft with demonstrations) Inst. of Physics, Stockholm university (april 1998).

P. Sällström: Using a Personal Computer as a Visual Colorimeter.
USIP Report 98-09 (Dec. 1998)

T. Holtsmark and P. Sällström: On Boundary Colours Fusing into Grey.
University of Oslo and University of Stockholm (to be published)

3. For Swedish readers:

P. Sällström: Belysning och färgseende. USIP Report 77-21 (Sept. 1977)

P. Sällström: Liten kurs i färgfysik. (Kompendium)
Fysikum, Stockholms universitet, Jan. 1995.

P. Sällström: Angående vissa färgfenomen i samband med polariserat ljus. (Utkast) Fysikum, Stockholms universitet, April 1996.

P. Sällström: Färglära på datorn. Kosmos förlag 1998 (ISBN 91-86066-43-9)

G. Tonnquist: Färgsystemanalys. Byggforskningsrådet, Stockholm 1995