COLORIMETRY OF DISPLAYS

Issue 1.00 - January 1997

Bentham Instruments Limited
2 Boulton Road, Reading, Berkshire, RG2 ONH
Tel: +44 (0)118 9751 355   Fax: +44 (0)118 9312 971 Email : sales@bentham.co.uk

Contents

Introduction 1

First we need to be aware of what Colorimetry is about and to know something of its limitations.

Our eyes are our most powerful sense organ with which we obtain detailed knowledge of our environment. They tell us about the shape of objects around us, their size, their disposition, their orientation, their movement. These objects are also mapped in three different wavelength bands so that the brain, by comparing relative signals in these bands, can obtain clues about the material from which the object is made. (Rather like a multi-spectral earth resources satellite can reveal terrain usage.) This data which our eyes extract from the spectral distribution has evolved into an important functional and aesthetic part of our lives known as "colour". It is emphasised that, although the physical stimulus for colour is amenable to precise definition and measurement, the resultant sensation is entirely subjective and colour science cannot (except in a fairly gross sense) predict what "colour" we will sense.

Nevertheless, colour science has provided methods by which very strong predictions can be made about how to match stimuli so that they produce the same colour sensation. This facility has great utility in an industrialised society as two examples illustrate:

i) A garment retailer decides that, for economic reasons, he will get jackets made in Manchester and trousers made in Korea, but he expects get a colour match between jacket and trousers. The traditional solution of cutting both from the same cloth is logistically difficult but if each manufacturer works to an agreed colour specification then jackets and trousers will always match.

ii) An artist uses his skill and creativity to create a painting which evokes a particular mood or idea. A publisher wishes to replicate this picture by the thousand with sufficient fidelity to satisfy the artist and the customers. With a few copies it would be feasible to seek the artists approval on each copy, but for thousands of copies applied colour science is the only economic solution.

There are many more examples from highly diverse fields of activity but they are all the application of colour science to colour matching, and colour measurement is a key element in this process.

Strictly speaking, colour cannot be measured but colorimetry is the accepted shorthand for "measurement of stimulus physical properties in a standardised manner that will enable colour matching".

2.1 Scope

The Commission Internationale d'Eclairage (CIE) has introduced a number of standards and recommendations which have been internationally accepted.

The CIE formulations are, to date, all concerned with colour matching in the particular case of colours viewed in the same visual context. Thus two stimuli with matching CIE parameters will appear to be the same colour provided that the two stimuli are surrounded by the same set of colours and have the same immediate history of surrounding colours. This proviso is acceptable for a wide variety of applications.

The modelling of the effects of the visual context on colour matching is a complex issue which is currently the subject of much research interest under the title of "Colour Appearance" (see Ref. 1) but is outside the scope of this report.

2.2 Origins

Because colour sensation is mediated by the distribution of radiant power in the visible waveband the definition and replication of this distribution should be sufficient to ensure colour matching. However, this would be impractical to the point of impossibility not so much because the definition would comprise a list of some 40 parameters, but matching the reflectance properties of say painted metal and moulded plastics at 40 wavelengths would be a highly elusive goal.

Fortunately, distribution matching is sufficient but not necessary. It is a matter of observation that it is possible to find many different spectral distributions that all yield the same colour sensations. This phenomenon, know as metamerism, is very important in applied colour science and arises from the trichromatic theory of colour vision.

It has been observed that the human eye behaves as if it comprises three types of receptor each having different, but overlapping, wavelength response functions. As a result of this it is found that any given colour sensation can be matched from a mixture of three primary colours so that, if the primaries are defined in reproducible physical terms, a complete colour specification needs just three parameters - the amount of each primary needed for a match.

Historically, the subject of brightness matching (a subset of colour matching) was addressed first.

2.3 Brightness Matching

The shape of the spectral distribution determines the sensation of coloration and the general power level of the stimulus gives rise to the brightness sensation.

However, not all wavelengths contribute equally to the brightness sensation so the CIE adopted a Brightness Matching function which allows stimuli of different wavelength composition to be numerically assessed for brightness matching from knowledge of the wavelength distribution. The assessment is defined by the operation:-

$$\text{Luminance} = 683 \int_{380}^{780} P(\lambda)\, V(\lambda)\, d\lambda \quad cd/m^2 \quad \text{[Equa. 1]}$$

where:

$P(\lambda)$ = Spectral power distribution of the stimulus in $Wm^{-2}\ sr^{-1}\ nm^{-1}$

$V(\lambda)$ = CIE 1924 Brightness Matching Function, sometimes described as the Photopic Repsonse. The shape is shown in Fig. 1 and the CIE definition is in Reference 2.

The resultant luminance parameter serves as a correlate of Brightness.

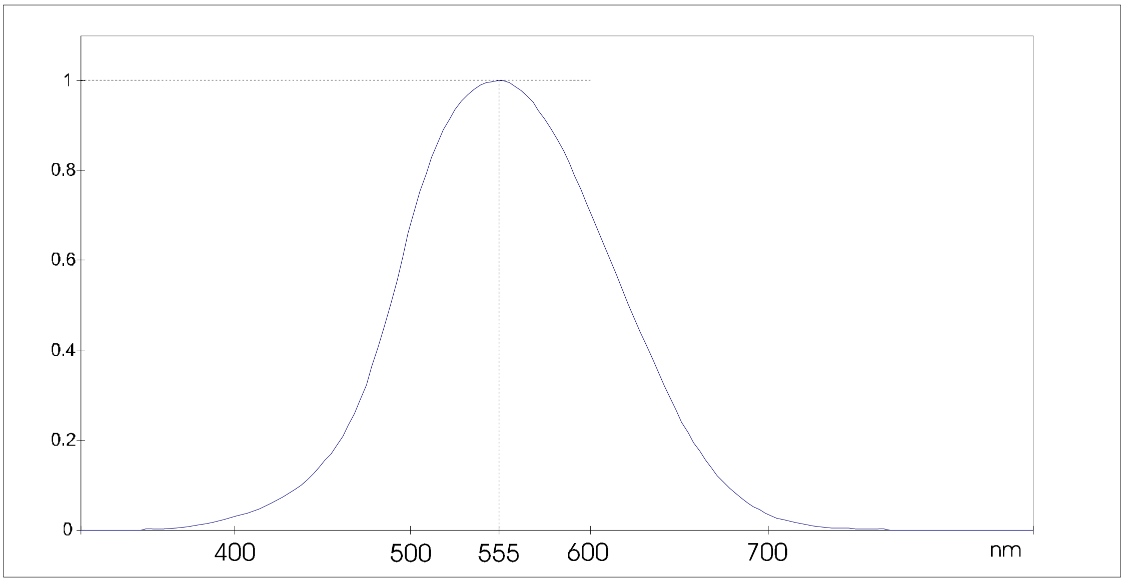

Fig. 1 1924 CIE Brightness Matching Function

The relationship between luminance and brightness is:-

- Luminance is a physical property of the stimulus but brightness is the sensation caused by the stimulus.

- Two stimuli which have the same luminance will give the same brightness sensation.
- The sensation of brightness is not proportioned to luminance. There is not yet any agreed function to relate these two different parameters although many suggestions have been made.

It is unfortunate that many users of photometry and colorimetry insist on using luminance and brightness as interchangeable words.

2.4 Colour Matching

The CIE system, like the human eye, derives from the stimulus spectral distribution a set of three signals. These signals are the Tristimulus Values and are defined by the set of three integrals:-

$$
\left.\begin{aligned}
X &= \int_{380}^{780} P(\lambda)\,\bar{x}(\lambda)\,d\lambda \\
Y &= \int_{380}^{780} P(\lambda)\,\bar{y}(\lambda)\,d\lambda \\
Z &= \int_{380}^{780} P(\lambda)\,\bar{z}(\lambda)\,d\lambda
\end{aligned}\right\} \quad \text{[Equa. 2]}
$$

where:

X, Y, Z are the Tristimulus Values (conventionally represented in upper-case).

$P(\lambda)$ is the spectral power distribution of the stimulus.

$\bar{x}$, $\bar{y}$, $\bar{z}$ are the CIE 1931 Colour matching functions (conventionally represented in lower-case) which are internationally agreed functions. Their shape is shown in Fig 2 and complete definitions can be found in Ref. 3.

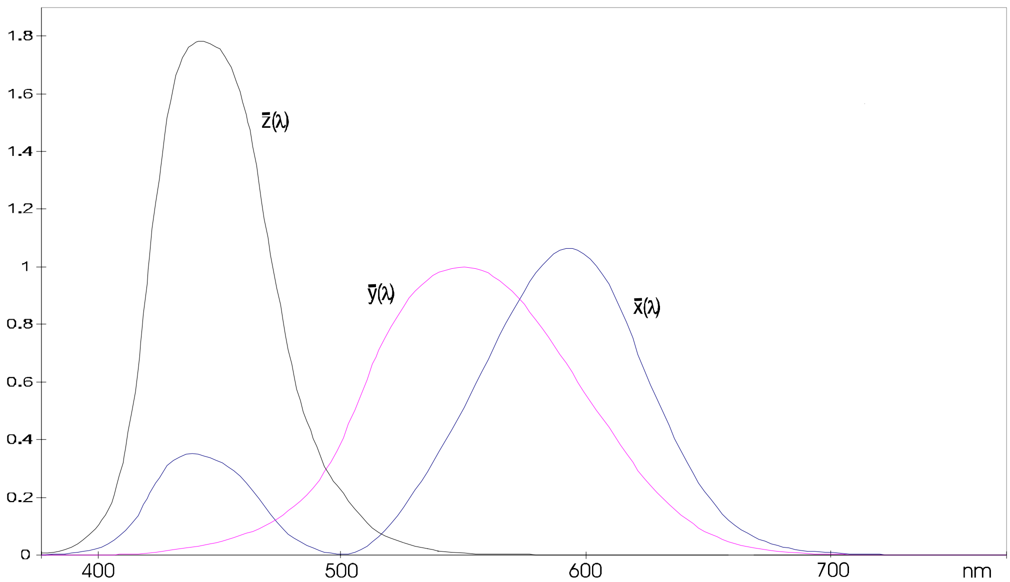

Fig. 2 1931 CIE Colour Matching Functions

The achievement of the CIE 1931 system resides in the fact that colour stimuli having the same Tristimulus Values will yield the same colour sensation (when viewed in the same visual context). Over 60 years later the basic system is used universally and internationally to very great effect.

Notes on the CIE Colour Matching Functions (CMF)

a) The CMFs are related to, but by no means equal to, the actual spectral responses of the three receptor types in the eye.

b) The 1931 CMF have been devised to give some practical conveniences such as equating $\bar{y}$ to V(λ) such as to put all the brightness producing effect into the Y Tristimulus.

c) The 1931 CMFs result from response matching experiments with stimuli that subtend 2 deg. at the observer's eye. However, matching behaviour is known to have some dependence on the stimulus size, so in 1964 the CIE introduced another set of CMFs appropriate to stimuli subtending 10 deg.

However, it is safe to assume that, unless explicitly stated otherwise, the 1931 2 deg. observer is to applied.

2.5 Chromaticity and Luminance

In practice it is convenient to separate the brightness producing effect of a stimulus from the degree and type of coloration.

The Tristimulus Values X, Y, Z defined in Equa. 2 have units related to the units in which P (λ) is measured so a normalisation is introduced:-

$$x = X / (X + Y + Z)$$
$$y = Y / (X + Y + Z) \qquad \text{[Equa. 3]}$$
$$z = Z / (X + Y + Z)$$

Where x, y, z are dimensionless chromaticity values. Because the normalisation yields $x + y + z = 1$ one of the three parameters is redundant and x, y values are usually used.

The general level of the stimulus, a necessary third parameter of a colour matching, is defined by the Y Tristimulus which is treated in one of the two ways depending on the field of application:

a) For surfaces emitting light (for example displays) the Y value is scaled to represent luminance in cd/m2, as indicated in Equa. 1.

b) For reflective surfaces (e.g. paper, textiles, paint) the Y value is expressed as a percentage of the Y value that would obtain with a perfectly reflecting surface under the same illumination.

The description of the coloration of a stimulus as x, y chromaticity values lends itself to graphical conceptualisation of colour relationships in a rectangular co-ordinate plot of x, y resulting in the chromaticity chart. The utility of the chromaticity chart arises mostly from the practice of putting various items of fixed chromaticity on the plot to assist in establishing a quantitative context for the visualisation. The fixed data can be any or all of the following:-

- The chromaticity locus of all monochromatic light stimuli, annotated with wavelength. This lop-sided horseshoe shape, the spectral locus, sets a boundary within which all physically realisable stimuli must fall.

- The chromaticity locus for a material which radiates because it is hot. Planck's law for a Black Body radiator is assumed to apply and many real materials (such as tungsten lamps) are a reasonable approximation to a Black Body, usually annotated with temperature in degrees Kelvin.

- Points representing the chromaticity of standard illuminants associated with the application.

- Points representing the primaries in a particular colour display system. The triangular area defined by these three points defines the gamut of chromaticity available from that display.

- Sub-division of the chart into areas with a colour name associated with each area. (The "Kelly Chart") The colour naming is useful to the tyro but should be treated with great caution because of the subjective and culture-dependant nature of colour naming.

The additions to the 1931 Chromaticity Chart is illustrated in Fig. 3.

However, the display engineer is more likely to encounter a variant of this using the u', v' co-ordinates.

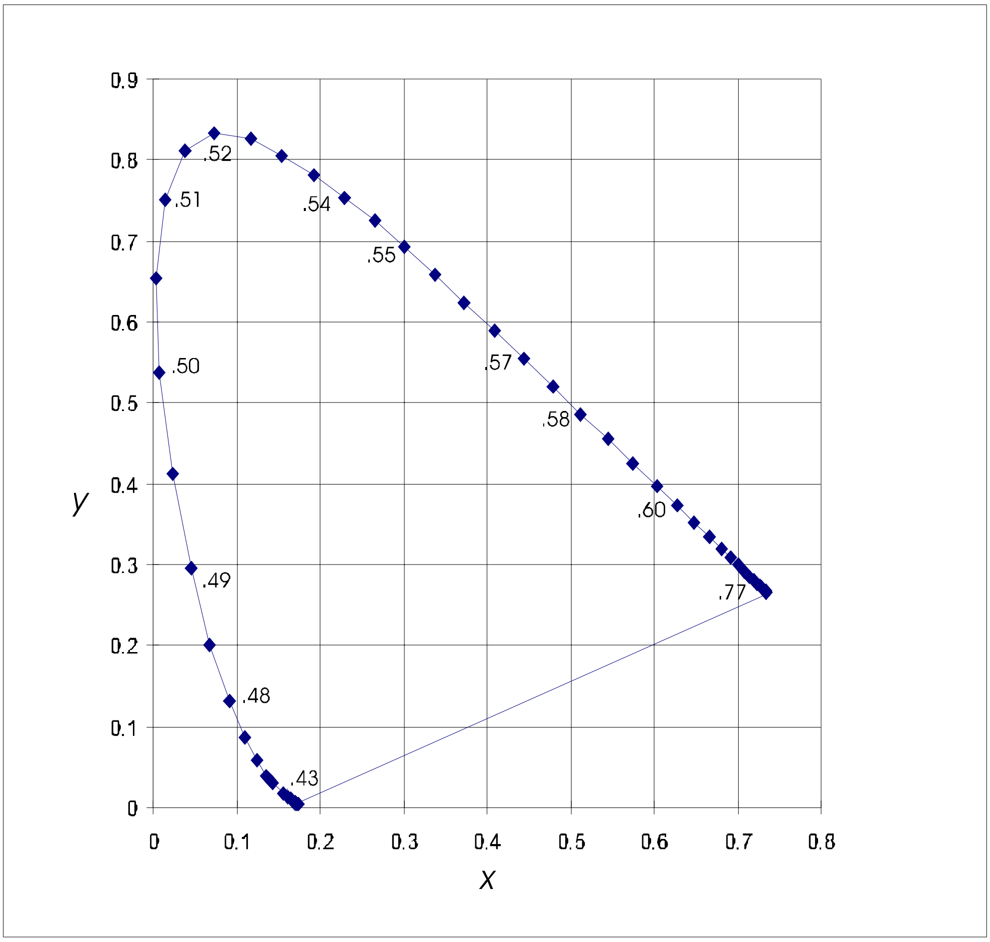

Fig. 3 1931 CIE x, y, Chromaticy

2.6 Uniform Chromaticity Spaces

A very common use of chromaticity co-ordinates is the setting of boundaries and tolerances representing the maximum deviation that the end user can tolerate. The x, y chromaticity space can be, and is, used to set tolerance bounds but it suffers from the disadvantage of being a highly non-uniform space. The just detectable chromaticity difference varies widely over the space (a range of 20:1) so that tolerance boundaries have to have a different size and shape for each colour. Many attempts to find a more uniform space have been pursued leading to promulgation of the CIE 1960 u, v Uniform Colour Space (UCS) and the 1976 u', v' UCS. Neither of these are perfectly uniform but both are a real improvement on the x, y space. These modifications of x, y space are an oblique projection from the x, y plane on to a new plane and the defining transformations are:

1960 UCS

$$u = 4x / (3 - 2x + 12y)$$

[Equa. 4]

$$v = 6y / (3 - 2x + 12y)$$

1976 UCS

$$u' = 4x / (3 - 2x + 12y)$$

[Equa.5]

$$v' = 9y / (3 - 2x + 12y)$$

The 1960 UCS is obsolete and rarely used except for a few cases where it became embodied in other formulations such the definition of Correlated Colour Temperature and the avionics industry parameter Discrimination Index.

The 1976 UCS is now widely used in colorimetry of displays and the chromaticity chart can be annotated with fixed reference points in the same manner as the 1931 chart as illustrated in Fig 4. Because of approximately uniform nature of the 1976 UCS it is reasonable to define tolerance zones as circles in the Chromaticity Chart.

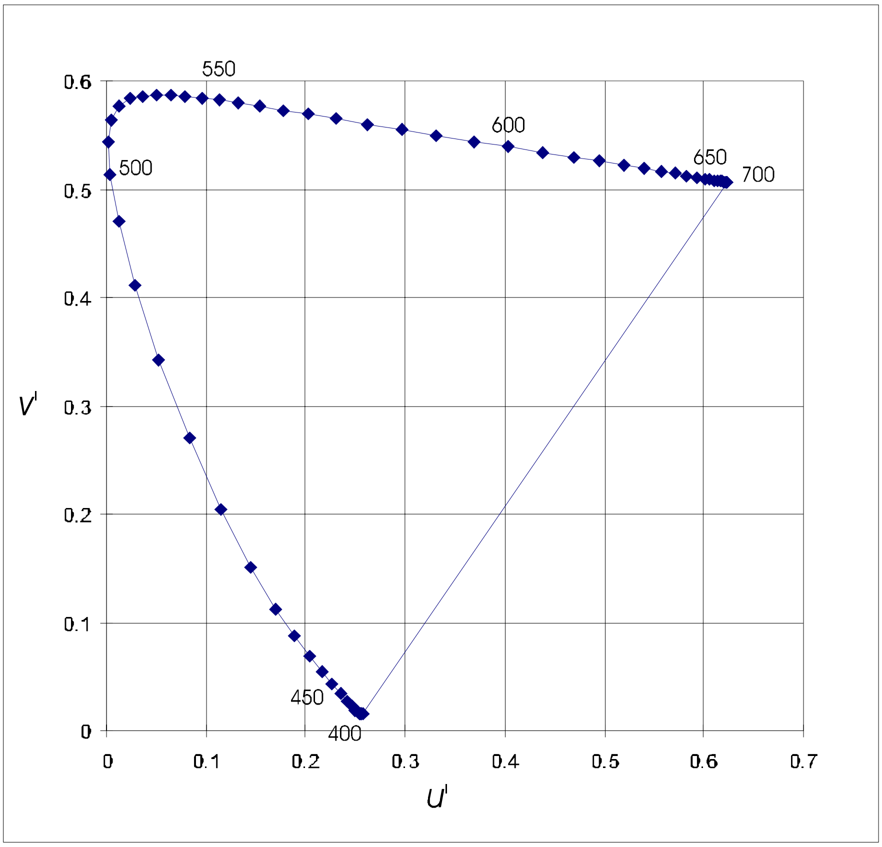

Fig. 4 1976 CIE u', v' Chromaticity

2.7 The CIE L*, u*, v* System

The widespread use of the 1976 UCS has lured some into thinking that a pair of u', v' values is all that is needed to specify a colour, and leads to questions like "where is brown on the Chromaticity chart?" The answer is that brown is likely to have the same Chromaticity as orange but brown differs from orange in the missing third dimension of relative luminance.

The CIE LUV space is a three dimensional space originally formulated for object colours an important feature of which is the postulation of a "surround". The "surround" does not necessarily have to surround the object but is the brightest white object in the visual field which the eye uses to establish a black-grey-white scale under the prevailing illumination.

The CIELUV space is defined by :-

$$L^* = 116\,({}^{Y}/_{Y_n})^{1/3} - 16$$

$$u^* = 13\,L^*\,(u' - u'_n) \qquad \text{[Equa. 6]}$$

$$v^* = 13\,L^*\,(v' - v'_n)$$

Where Y, u', v' relate to the object

Y_n, u'_n, v'_n relate to the surround

Again the formulation is an approximation to a perceptually uniform space and it is reasonable to define tolerance zones as spheres in the space.

The CIELUV system is started being used for display colour specification but there is lack of agreement about how to interpret the "surround" in display parameters. This authors recommendation is that the Y_n, u'_n, v'_n parameters be identified with the display white design centre values for the product.

2.8 Other Derived Parameters

There are two other parameters which can be derived from chromaticity values and which sometimes appear in colorimetric specifications. These are:-

a) Dominant Wavelength

This is a single number parameter which is related to the subjective sensation of hue.

Dominant wavelength is defined as the wavelength of monochromatic light that needs to be added to white light in order to match an unknown colour. As can be seen from the graphical method depicted in Fig.4a the concept is meaningless without an accompanying definition of the white point.

The dominant wavelength is not necessarily the wavelength with maximum spectral radiance.

b) Correlated Colour Temperature (CCT)

This is a single number parameter for chromaticity co-ordinates which are on or are close to the black-body locus.

CCT is defined as the temperature in degrees Kelvin of an ideal black-body radiator whose chromaticity is closest to the unknown and when closeness is measured in CIE 1960 u, v space. The term is believed to have its origin in photography when the only artificial light source was the incandescent lamp. If such a lamp is ran hotter its spectral distribution becomes more flat and colour rendering properties improved but at the expense of reduced lamp life, so lamps were rated by temperature to indicate where the compromise was set.

However, to apply the concept to white light sources whose spectral distribution is nothing like an incandescent lamp (i.e. metameric whites) can be a very misleading indication of colour rendering. Nevertheless it is still sometimes used, inappropriately, to define the white point of TV monitors.

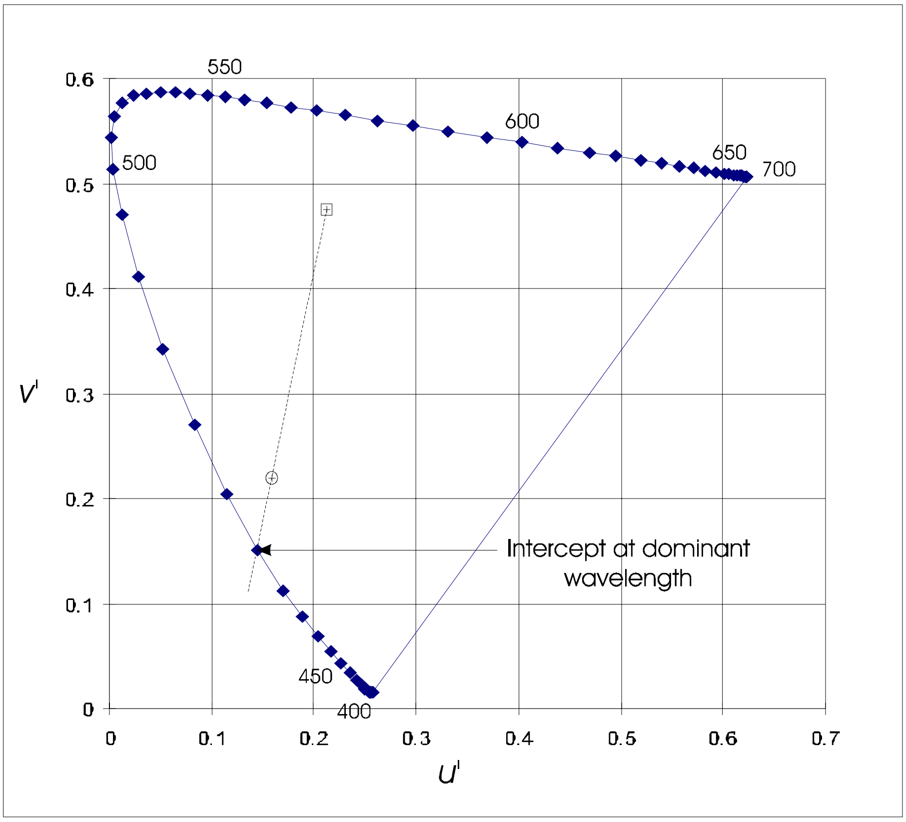

Fig. 4a Graphical Determination of Dominant Wavelength using 1976 Chromaticity Chart

Application Needs 3

Displays have permeated into a large and increasing proportion of workplaces and because most of them offer multi-colour the subject of colorimetry has come to the fore.

Various types of application exist:-

- PC Visual interfaces
- Work stations
- Media transfer
- Airborne displays

3.1 PC Visual Interfaces

The largest application of colour displays is the colour monitor as the prime interface between computer and operator. The wise use of a modest number of colours can enhance the utility of the interface, but provided the colours are chosen to be visually distinct from each other there is no need for anything more than "eyeball colorimetry" on the part of the user.

The monitor manufacturers will use colorimetry in the design process but the monitors are not sold against a colour specification.

3.2 Work Stations

CAD Work-stations used for mechanical design, thermal design, stress analysis etc. use colour monitors with improved resolution and colour rendering performance. Complex situations are portrayed requiring many more discrete colours which will benefit from colorimetry to set-up and maintain the colour palette.

3.3 Media Transfer

The most demanding applications are those CAD systems where colour is significant in its own right (not just as a means of coding some other parameters). This embraces activities such as graphics design, textile design, architectural design etc. where an experienced practitioner uses a monitor to show a design as it evolves and, when satisfied with the design, expects the image to be faithfully rendered in the target media (paper, textile, paint etc.). Such transference between media is an extraordinary difficult task and is beyond the scope of this report but it clearly requires some high - accuracy colorimetry of the monitor display.

3.4 Airborne Displays

Multi-colour displays are now established in cockpit of civil and military aircraft. Colour is mainly used as a coding method to categorise and “chunk” complex display formats to increase speed of information assimilation and minimise ambiguities. Relatively few colours are used in a format so their is no need for high accuracy colorimetry. However, two factors put demands on colorimetry:-

a) The displays are sold against a colour specification which requires comprehensive measurements on each delivered item, putting an emphasis on measurement speed.

b) The colorimetry has to be performed over a wide luminance range from around 1000 cd/m^2down to a fraction of 1 cd/m^2.

3.5 Accuracy Requirements

The colorimetric accuracy required is application dependant with requirements being imposed by customer specifications, by regulatory agencies or by market forces. However, a start can be made by asking how big is a just-noticeable colour difference?

The answer is of the order of 0.001 units in CIE u’, v’ co-ordinates and we have to be a bit vague about it because like all subjective responses it is influenced by many contextual factors. Thus in very demanding applications such as media-transfer (described in para. 3.3) where the aim is to produce a copy which is indistinguishable from the original the product colorimetric tolerance needs to be in the region of a just-noticeable-difference. To get the product to this degree of fidelity will need measurement accuracy to a significantly tighter tolerance.

In an application such as Airborne Displays where colour fidelity is not the issue we find product colorimetric tolerances in the range 0.015 to 0.04 chromaticity units. To support the development and testing of such products it is desirable to work with instruments having ±0.001 u’, v’ accuracy. Such accuracy is well outside the scope of filter colorimeters, can be met under some circumstances with a diode-array spectroradiometer and can readily be met by a scanning spectroradiometer.

The practice has emerged of presenting measured chromaticity to four significant figures (i.e. within 0.0001 units) presumably on the basis of keeping avoidable rounding errors to an order of magnitude less then a just - noticeable difference.

However, a measurement accuracy of ±.0001 u’, v’ units will be elusive for a commercial instrument used in an industrial context, since it would require spectral radiance scale accuracy to within 1%, better than 0.1nm wavelength accuracy together with negligible error due to stray-light, noise etc.

However, this author has observed agreement to within 0.0015 units between instruments of different operating principle and different manufacturer and repeatability of each instrument in the ± .0002 region.

So far we have discussed only chromaticity errors but not the third dimension of luminance.

Because chromaticity is effectively the ratio between energy in different parts of the spectrum (see sect. 2.5) it does not matter if the spectral radiance scale is not absolute.

However, for the luminance term an absolute scale is necessary so we ask again how accurate does it need to be?

Again the answer has to be vague because it is a subjective issue. For a side-by-side comparison of two bright areas a difference of around 2% is detectable. However, a different experiment in which the viewer sees a display which is than taken away and replaced by another display at different luminance the difference will need to be at least 20% if the viewer is to detect it.

Which is fortunate because a well designed and carefully maintained instrument will have an absolute accuracy no better than ±4% in luminance and plenty of lesser performance instruments exist.

Again do not be mislead by readout of luminance to 4 significant figures (which are provided because the resolution is useful for many engineering design purposes to track changes and differences).

The instrument industry provides a reasonable variety of colorimetric products suitable for display measurement. No attempt is made here to evaluate individual products but some indication of the strengths and weaknesses of a few generic types are discussed. First a few words on the topic of light collecting optics which is common to all types.

4.1 Light-Collecting Optics

In most display applications practical issues dictate the use of the “spot-meter” type of optics. Essentially this is a telescope with the eyepiece replaced by a field-stop in the image plane behind which is the radiation sensor. The field-stop defines the area of the display surface which is being measured, see Fig. 5. The reasons for using a spot-meter are:

- It is usually necessary to measure the emission from a specific well defined area of the display.
- Direct masking at the display image plane is confounded because it is usually some distance behind front surface.
- It is often necessary to measure display emission at specific viewing angles, particularly on LCD devices.
- There are situations where the instrument needs to be remote so as not to obstruct access to controls are/or adjustments on the display.

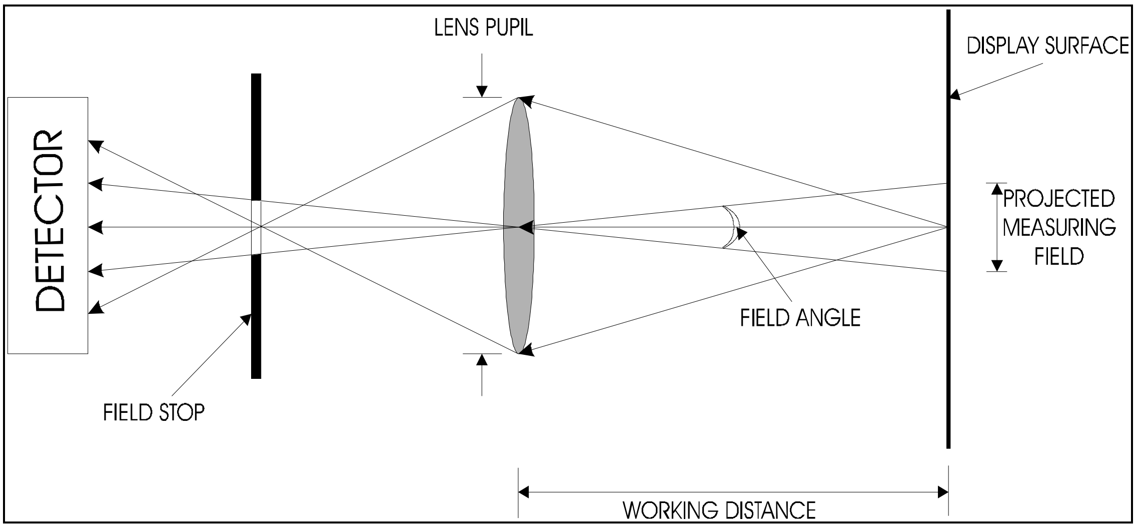

Fig. 5 Spot-Meter Optics

Within the basic spot-meter description there are wide variations in additional facilities, in soundness of design and in quality of manufacture. The following are the major items that should be evaluated on a candidate instrument:-

a) Aiming Facilities

It is vital that the measuring field of the instrument is placed such that the field is filled and is sampling the appropriate area of the display so the instrument must have the facility for accurate aiming. Various configurations can be encountered including the side-mounted aiming telescope and the swing-mirror that flips between aiming and measuring states. Both of them are reliant on the accuracy and stability of the optical alignment and do not give much confidence in knowing where one is measuring.

The aperture-mirror approach, as illustrated in Fig. 6, is strongly recommended because it does not rely on any alignment and it is inherent that the part of the view-finder image that cannot be seen, the black-spot, is the part of the image that is sampled by the detector.

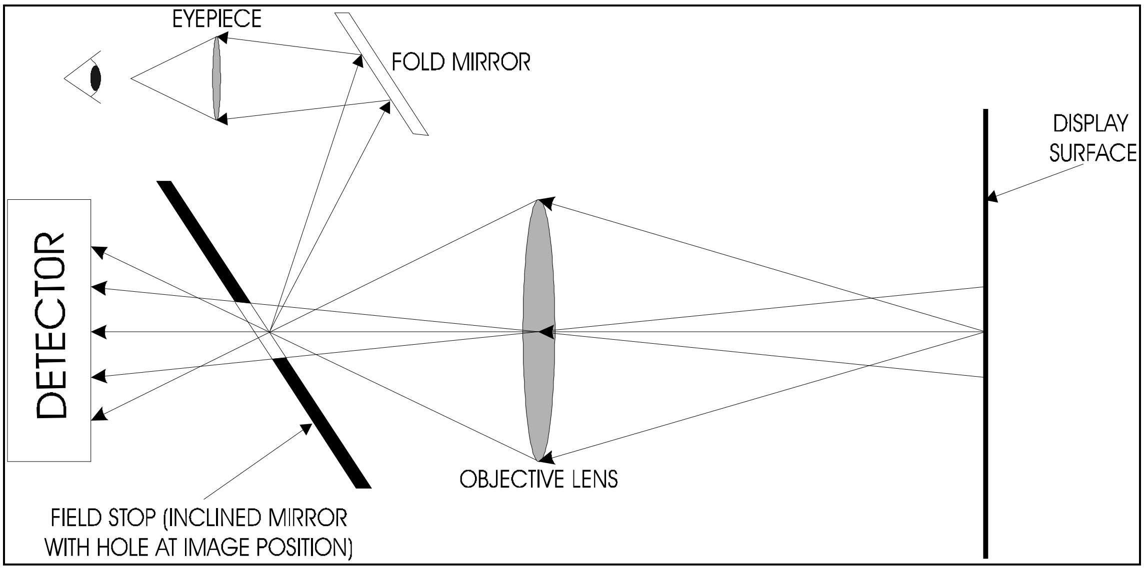

Fig. 6   Aperture-Mirror Self-Aligned Aiming System

b) Field Size

Bearing in mind the fact that the measuring field must be completely filled by the luminous source if the instrument calibration is to be valid some thought is needed to check that the field size suits the application.

A low-cost instrument may have only one field size, probably in the 1 deg. region, but given a reasonable focusing range the field size at the display can be varied from about 9mm at say 0.5 metre working distance to 35mm at say 2 metre working distance.

The more expensive instruments will offer interchangeable aperture-mirrors for a wide range of measuring field size, either manually insertable or on some sort of turret.

c) Numerical Aperture

A large size objective lens with a high numerical - aperture will collect a lot of light and allow measurements down to lower levels. However, it will cost more, will have reduced depth of focus and the large angular acceptance cone may invalidate measurements on displays with strong angle-of-view effects.

d) Working Distance

Check that the instrument will focus at the shortest and longest working distance that is needed.

e) Lens Flare

It is desired that the instrument responds only to light emitted from the area defined by the spot-meter optics. However there will be scattering and internal reflections in the objective lens which will collect some light from the surrounding bright area to influence the measurement.

Measurements on a "black hole" in the middle of a bright field can be quite revealing. See Fig. 7.

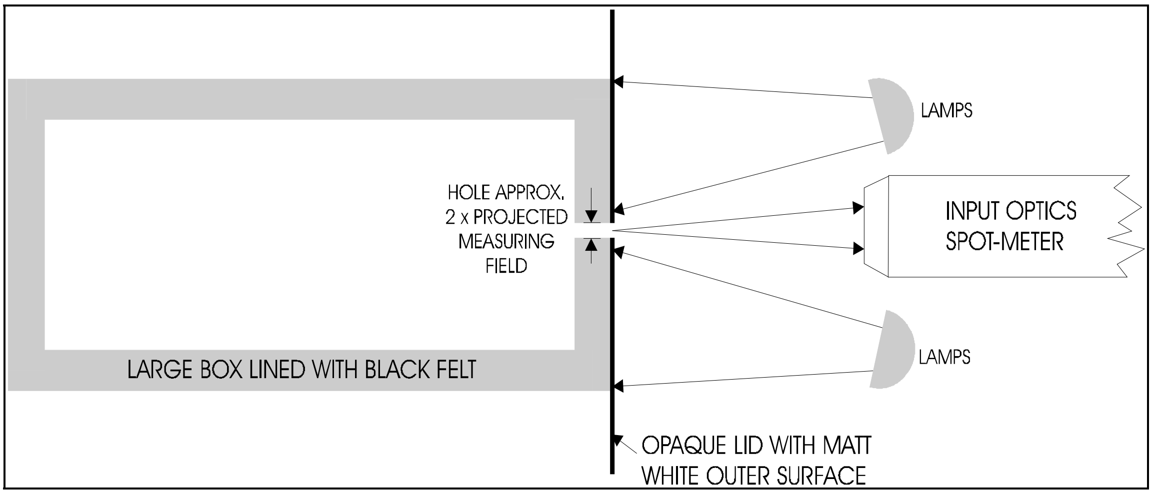

Fig. 7 Lens-Flare Test

4.2 The Photometer

A photometer performs only one part of the colorimeter function - it just measures luminance (the brightness correlate).

The photometer must implement Equation 1 of para 2.3 and it is the weighting with the CIE $V(\lambda)$ function that causes the problems. The instrument manufacturer seeks a filter, or combination of filters, which when used with the spectral response of the detector yields a good match to $V(\lambda)$. Both the design and control over manufacturing consistency costs money particularly if the sample to sample variation in detector spectral response is accommodated by matching each filter to it's detector.

Inevitably there will be mismatches to $V(\lambda)$ but the manufacturer will try to balance out over and under matching to minimise errors on sources which have energy spread over the whole visible range.

However, display emissions often comprise isolated spectral lines so luminance measurement on such sources could pick up the peak errors in $V(\lambda)$ matching.

Beware instrument specifications like "--- response matching to $V(\lambda)$ within ±2%". The 2% sounds good but they actually mean 2% of the $V(\lambda)$ peak value so that a blue emission at 450nm, where $V(\lambda)$ reduces to 0.038, could be subject to more than 50% error!

The more compact photometers use a silicon photo-diode as the detector element which is small, robust, linear and stable over time.

For a more sensitive photometer the photo-multiplier detector has several orders of magnitude more sensitivity but it is more bulky and expensive, is a bit fussy about it's environment and requires frequent calibration.

The more exotic interments will offer a range of extra functions such as:

- A range of measuring field size
- Neutral density filters to accommodate very bright sources.
- Polarisers to separate p and s components of emissions.
- Analogue fast response output for oscilloscope display of light temporal modulation.
- Filter sets to implement the Filter Colorimeter function.

Choose only those extra functions that are really needed and remember the dictum that " a tool which does everything does nothing well"!.

4.3 The Filter Colorimeter

The heart of the Filter Colorimeter is three sensor channels each with it's own spectral response to implement Equa. 2 in para. 2.4. The colorimetric accuracy depends entirely on how well the channel responses approximate to the CIE Colour Matching Functions (CMF), rather like the photometer problem but made more acute. The $\bar{x}$ function causes particular problems because it is double - humped (see Fig 2). Some instruments use a fourth channel to provide the other hump but others omit the short-wavelength hump and approximately compensate by introducing a fraction of the Z channel output into the X channel. Skilful design can do a lot to balance out the mismatch on sources with a continuous spectral distribution.

Thus measurement of reflected light from dyes and pigments (which tend to have smooth reflectance functions) can yield reasonable colorimetric accuracy. These instruments are usually based on silicon diode detectors and can therefore be designed to yield good short and long term precision. The advent of cheap microprocessors means that readout can be offered on all those parameters that can be derived from the basic tristimulers values, plus storage of alternative calibration factors and go/no go indication against stored limits.

However, for display application the following issues must be assessed:

a) <u>CMF Errors</u>

Line spectrum sources will arise in displays which will exaggerate any errors in the CMFs. On the subject of CMF errors the suggestion will be made at some time that such errors be "calibrated-out" for the particular application. Before going down this route consider:-

- It will be possible to devise two light sources which have identical true CIE parameters but which yield different readings on the filter colorimeter, i.e.. what might be termed "instrument metamerism". The correction required is <u>not</u> just a function of the raw reading (unlike say correcting a linearity error on a voltmeter).

- The magnitude and logistics of the task of establishing calibration corrections for each source, promulgating and applying the corrections and updating them for each new source.

b) <u>Light splitting between channels</u>

The flux collected by the input optics has to be consistently split three ways into the tristimulus channels. In some instruments/configurations the split proportions can be corrupted by non-uniformity of the source across the measurement field. Some designs avoid this by inserting the filters in sequence into one time-shared channel.

c) The use of time-shared channels or other internal dynamic processes may have interactions with display dynamic parameters such as refresh rate or CRT phosphor decay times.

4.4) The Spectroradiometer

The spectroradiometer is an instrument which measures the radiant power as a function of wavelength. In the context of colorimetry it explicitly measures the P(λ) parameter in Equa. 2 expressed as a list of numbers in a computer file.

Having measured the P(λ) function the weighting with Colour Matching Functions (CMF) and integration can be implemented in the computer with zero error in the CMFs. The spectroradiometer, compared to a filter colorimeter, is:-

- More bulky
- More expensive
- Slower to measure

However, the advantages of the spectroradiometer are very important:-

- Very high colorimetric accuracy

- Can implement other CMFs such as the 10 deg. observer or the Scotopic Brightness matching function without hardware change and without re-measurement.

- The P(λ) measurement is valuable in it's own right to the display engineer.

- Display design includes analysis and control of reflective surfaces, transmission through filters and other components. Such analysis requires spectral data so the spectroradiometer is used to measure spectral reflectance, spectral transmission, and spectral emission of candidate materials.

The spectroradiometer is likely to include software for manipulation and analysis of data in the spectral domain and therefore will yield colorimetric data on a candidate design before it becomes hardware.

There are many spectroradiometer products on offer but there are two important categories of operating principal that confer their own particular advantages.

All spectroradiometers start with a diffraction grating for which the angle of strong reflection is wavelength dependant in a systematic manner. An incident beam of wide-band light is dispersed in angle so the placement of a light detector at a particular angle will result in a detector that responds to just a narrow band of the incident light.

The conventional (or scanning) spectroradiometer uses a single light detector but mechanically varies the angle to scan the detector sequentially through all wavelengths of interest. The diode-array instrument is also based on the diffraction grating but uses a multiplicity of detectors each at a different angle so that each detector is associated with it's own narrow waveband.

The diode-array type is a relative new comer and brings advantages in respect of compactness (no precision moving parts) and speed of measurement (all wavelengths processed simultaneously rather than sequentially).

Any comparison of the two types is overlaid with differences between instruments suppliers both in respect of design skills and of their perception of market needs. However, some issues that need to be considered are:-

4.4.1) Dedicated or Kit of Components?

At one extreme is a system that the user can assemble from a choice of input optics, monochromators, detectors, electronics and various accessories. It offers flexibility by reconfiguring to suit the particular measurement task and, as free issue, the opportunity get an inappropriate configuration!.

At the other extreme is a dedicated system in integrated form where the supplier has chosen compatible components. It will work well and be easy to use but you have to be sure it will perform the task(s) you want performed!.

4.4.2) Stray-Light

Stray-light performance is the name given to the ability of the instrument to accept at the detector only the energy at the turned wavelength and reject energy at all other wavelengths. Thus parameter is important for all applications but is of particular importance for:-

- Line spectrum emissions having large empty areas in the spectral distribution. If the instrument gives non-zero response at the empty wavelengths the resultant file will look as if some white light has been added.

- Spectroradiometer calibration is usually performed with a reference tungston lamp source. The lamp spectrum will be similar to illuminant A and exhibit 25 times more energy at 780nm than at 380nm so that during calibration stray-light will over estimate the instruments response at 380nm .

- Stray light becomes very significant for NVG compatibility measurements discussed in Reference 4.

Stray light has its origins in grating blemishes, dust on grating and mirror surfaces and subsidiary unintentional reflecting and scattering paths that bounce their way past the selective element. Good detail design can reduce stray light to a reasonable level but for best accuracy further measures are needed. The simplest improvement is a turret of band pass filters that can be introduced progressively as the scan proceeds and which will attenuate spectral components well away from the tuned wavelength.

The best improvement comes from cascading two monochromators to get a net rejection which is the product of the two individual rejection ratios.

The filter turret and the double monochromators measures are applied to scanning type instruments but it is difficulty to see how they could be applied to diode-array instruments. Thus one of the limitations of the diode-array type arise from stray light response.

4.4.3) Higher Order Response

The diffraction grating higher order interference patterns have the effect of giving a response not only at the tuned wavelength λ but also to input radiation at λ/2, λ/3, λ/4 etc. Thus a spectroradiometer when turned to say 700nm will respond to UV energy input at 350nm unless countermeasures are taken. In a scanning instrument it is straightforward to introduce a steep-cutting minus-blue filter when the scan has reached say 500nm. In a diode-array instrument the rejection of 2nd order responses is something of a compromise.

4.4.4) Dynamic Range

For displays measurement there are two dynamic range issues.

First there may be a wide range of luminance levels to accommodate from less than 1 cd/m2 to over 1000 cd/m2. A diode-array instrument copes with this by varying the detector exposure time electronically, perhaps supplemented by a neutral density entrance filter. The scanning instrument copes by electronic range changing and again supplemented by ND filters. Secondly there is the dynamic range within any one spectral distribution between the smallest and largest components. In a scanning instrument the electronic range - changing can be invoked within the scan on a wavelength by wavelength basis yielding some 8 orders of magnitude between noise and quantitising at the bottom end and non-linearity at the top end. However, the diode-array instrument inherently has the same dynamic limits at all wavelengths, typically 3.5 orders of magnitude. It is not difficulty to find display emissions where the major component is a strong narrow spectral plus minor broad band components that are significant at 4 orders of magnitude down on the spectral peak.

4.4.5) Calibration Stability

All spectroradiometric instruments need wavelength and sensitivity calibration by the methods indicated in Section 6.

The scanning instrument probably uses a photomultiplier detector which have poor long term stability so re-calibration is advised for every measurement session.

The diode-array instrument uses silicon-diode detectors which are stable enough to relegate calibration to a 6-monthly event.

Before pressing the button for a colorimetric measurement it is worth asking oneself some questions about the display it self such as:-

- Has the display had time to warm-up and become stable?
- Is there a prescribed procedure for setting the black and white levels?
- When was black and white setting last performed?
- Is the display sensitive to the local environment (e.g. temperature, magnetic)?
- If so, can the environment be repeated?
- Is the proper local area of the display screen being measured?
- Are there features present which allow the measuring position to be accurately revisited?
- Is the appropriate viewing-angle being used and can it be repeated?
- Is the picture-input signal defined?
- Will the measurement be affected by ambient lighting?

Particular display hardware or application will prompt other questions and it is worthwhile addressing these questions up-front rather than puzzling afterwards over non-repeatable measurements.

As far as colorimetric instruments are concerned this author has always found it valuable to get a technical dialogue going with the instrument manufacturer to find out how the instrument works, what are it's strengths and shortcomings, and how to get the best out of it.

The rest of section 5 discusses the major issues of technique which will arise.

5.1 Filling the Measurement Field

The colorimeter or photometer will probably use the spot-meter type of light-collecting optics which will define the small area of the display surface from which light will be collected. It is important this measurement area is completely filled by the luminous patch whose colour to be measured otherwise the luminance will certainly be under-recorded and there may well be chromaticity errors as well.

With a fully addressable display based on say CRT or AMLCD there is usually no difficulty in providing suitable luminous test patches for colorimetry to ensure field-filling but where this proves difficult beware choosing too a small field-size for reasons outlined in 5.2.

There are a few specialised circumstances where under-filling yields valid answers but detail understanding of the instrument is needed first, together with experimental validation so the best advice is always fill the field.

5.2 Display Fine Structure

The most commonly encountered colour display technologies use spatial division, on a fine scale, to enable three independently controllable primary colour channels. The display designer's objective is to make the pattern fine enough to be invisible at normal viewing distances, and they mostly succeed.

If a white patch on a colour monitor is observed under gradually increasing magnification the continuous white will be perceived as having some repetitive structure composed of all white elements. As the magnification is increased further the white elements will separate into red, green and blue elements with discernible shapes. Various shapes and array configurations will be encountered (e.g. see Fig. 8) whose rationale and technology does not concern us here but we do have to take note that the display colour we wish to measure is spatially divided.

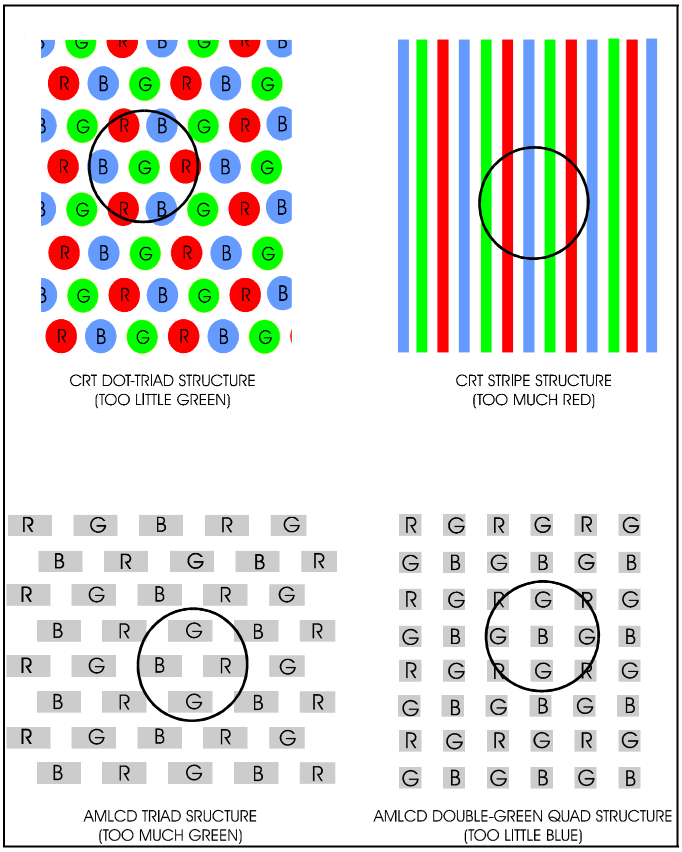

Fig.8 Measuring Field Too Small for Colour Pixel Structure

In this situation the display emission should be averaged over an area big enough to ensure chromatic fusion of the red, green and blue components. In practice this means making the measuring area big enough to ensure that small changes of the measuring area placement with respect to the display structure yield no significant change in averaged value, and a rule of thumb says that the measuring field diameter needs to exceed three times the distance between colour-groups.

In the discussion above we have assumed that the instrument performs a true averaging job on the light within the measuring field and that the instrument response is uniform across the field.

However, the instrument supplier will have assumed that the field is uniformly filled by the display, so there is potential for trouble in applications where neither of the uniformity assumptions is valid. Thus if the display has some fine structure it is necessary to investigate the internal configuration of the measuring instrument to check for non-uniformity of response. Two examples of non-uniform configurations are:-

a) A filter colorimeter which uses a trifurcated fibre-optic bundle to split the incoming light into three channels can yield a different splitting proportion between channels from point to point across the measurement field.

b) A diode-array type of spectroradiometer images the wavelength - dispersed measuring area on to the surface of the diode-array. However, the edges of the measuring area overlapped the edges of the diode active area resulting in non-uniform response across the measuring field.

5.3 Temporal Modulation

The most common subject of colorimetry is the reflecting surface colour for which the reflected light is continuous in time. However, most display devices emit their light in short sharp pulses albeit at a rate which gives the illusion of continuity to the human eye.

Many measuring instruments can have very fast response but we want the instrument reading to relate as close as possible to the human eye response. The Talbot-Plateau law is invoked which says that the visual response is mediated by the temporal average of the light provided that the pulse rate is high enough not to evoke a sensation of flashing.

Colorimetric instruments should respond therefore to the temporal average of pulsed light and it is fortunate that most designs gravitate towards this response. The instrument may well have been designed for use with straightforward reflective samples and not designed to accommodate the extreme situations of a CRT display. Any one phosphor element on a CRT with 25/50 Hz interlace scanning will receive an excitation pulse about 100 nS wide at 40 mS intervals. The phosphor decay characteristic will broaden the pulse to perhaps 40µS for a blue phosphor and up to 10mS for a green phosphor, so the light wave forms can readily exhibit peak: mean ratio of 1000:1. It is important therefore to find out how the instrument responds to temporal light modulation. Examples of what to look for are:-

a) The instrument light detector may have fast response and reproduce the pulsed-light waveform but the following electronic amplification may overload on the pulse peaks (although handling the mean level satisfactorily) so that subsequent extraction of the temporal average is in error.

b) Some instruments (for example the diode-array spectroradiometer) extract the temporal average by integrating the light signal over a known time period and accommodate to a range of display brightness by choosing the integration period. This can give problems when the display is very bright. If, for example, the display is refreshed at 20mS intervals and the instrument integrates over say 70mS the integration will sometimes capture 3 pulses, will sometimes capture 4 pulses resulting in a ±14% variability in reading.

Two remedies are available to ameliorate this sort of problem:-

- Attenuate the light reaching the instrument (using a neutral density filter) to allow longer integration periods.

- Contrive that the integration period is an exact integer-multiple of the display refresh period.

c) All instruments will feature an analogue-to-digital converter somewhere in the signal chain. The combination of pulsed light from the display and a sampling type A/D will yield highly variable readings. The best instrument configuration will comprise an analogue temporal filter to remove the worst of the modulation followed by an integrating type A/D which has a inherently lower response to any remaining modulation.

d) The fact that the red, green and blue components of a display have differing temporal modulation will exacerbate and confuse all the potential problems in a), b) and c) above. Make sure that the fixes work for all three primaries.

5.4 Spectral Distribution

We have already noted that line-spectrum emission often encountered in displays can give rise to inaccuracy due CMF errors in filter colorimeters (para. 4.3). Line-spectra have a potential for trouble in spectroradiometers unless the relationship between monochromator spectral bandwidth and measurement wavelength interval is properly controlled.

If a dedicated integrated instrument is being used it can safely be assumed that the proper relationship has been taken care of by the manufacturer. If, in the interests of flexibility of application, something more kit-like being used than the user has also bought himself the opportunity to get it wrong!.

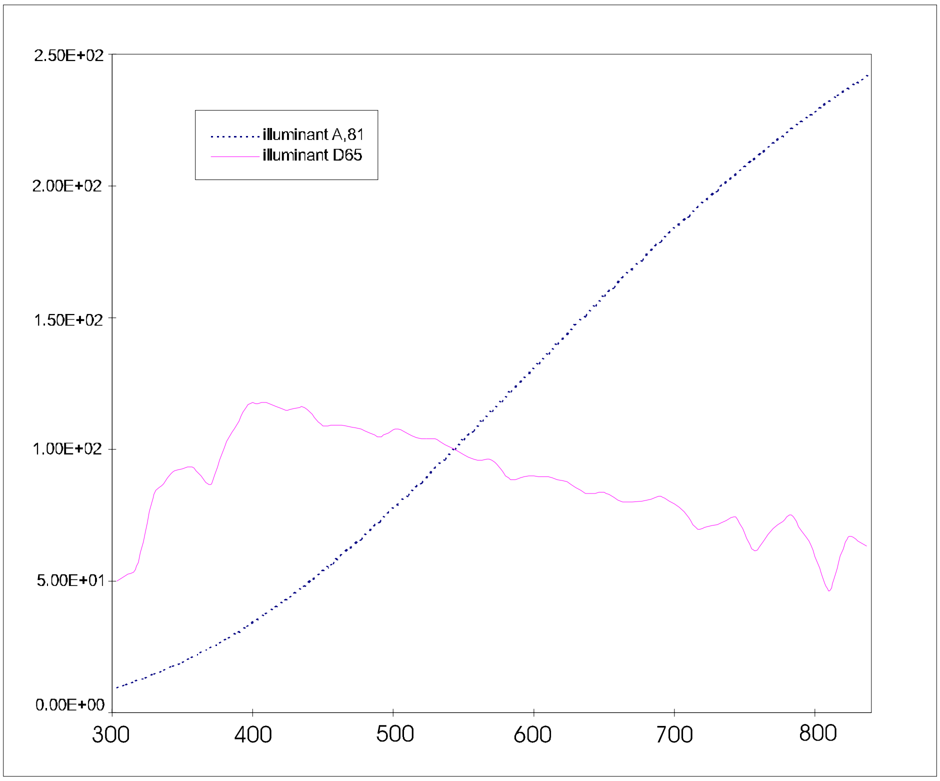

Fig. 9 Spectral Distribution for CIE Illuminants A & D65

However, the proper relationship is straightforward to define and maintain. The rule is: The monochromator spectral bandwidth (width at 50% of peak response) should be an integer multiple of the wavelength interval between measurements.

Thus, for example:-

- 5nm bandwidth with 5nm steps - OK.
- 5nm bandwidth with 3nm steps - Errors.
- 5nm bandwidth with 1nm steps - OK.
- 5nm bandwidth with 10nm steps - Errors.

Failure to observe this relationship will result in failure to record the true power in any spectral lines present, resulting in colorimetric errors.

Strictly speaking the relationship holds for the case of a monochromator with a triangular bandpass function, but most monochromators do yield a good approximation to this when fitted with matching entrance and exit slits.

For the purposes of colour measurement using spectroradiometer methods a spectral bandwidth of 5nm is as narrow as is needed even when line-spectra occur and it does not matter that two adjacent spectral lines get merged into one 5nm wide band.

Finer resolution spectral data may be needed for purposes other than strict colorimetry, for example:-

- Measurement on line-spectrum sources and sharp-edge filters separately with a view to computing the effect of a filter on a source.

- For investigative purposes where a high-resolution spectrum is needed to identify wavelengths and clustered lines with a view to identifying materials or sources of an emission.

Avoid the use of high resolution scanning except where strictly necessary because it will take longer to do the measurement and the measurements will have poorer accuracy.

5.5 Polarisation Effects

The optical systems of some instruments will be sensitive to polarisation of incoming light particularly at mirrors and diffraction gratings. When displays always meant CRTs polarisation was not an issue but the introduction of LCD technology has changed the situation.

Before an instrument is used for LCD measurements it should be evaluated for polarisation sensitivity. This can be accomplished by taking two measurements on the display which have identical test conditions in all respects except for a 90 degree rotation about the instrument optical-axis between the instrument and display. The two readings should be identical (within repeatability of the instrument and stability of the display) so any difference is attributable to polarisation effects.

If the difference is significant for the application then the remedies are:

a) Investigate additions to the instrument which will randomise the polarisation such as an integrating sphere (very effective but accompanied by low throughput of light) or a fibre-optic bundle which give reasonable randomisation, high throughput and geometric flexibility which may be useful in it's own right.

b) Always take dual readings with 90 degree rotation between them and take the arithmetic mean of the two readings. For a spectroradiometer the mean should be taken on wavelength by wavelength basis. Such an operation should be part of the arithmetic functions available in the operating software for the instrument.

5.6 Instrument Numerical Aperture

The numerical aperture (NA) of the light collecting optics is a measure of the light collecting power and defined as the ratio of lens pupil diameter to lens focal length. Clearly a high NA is desirable to avoid signal: noise limitations which eventually plague most measurements but this must be judged in the context of three other effects of the NA:-

a) High NA implies small depth of field and proper focusing becomes more critical.

b) High NA implies measurement over a wide angular range around the intended measurement angle. This will give rise to measurement errors on displays with strong angle of view dependence. Such displays include:

- Those using sharp-cut interference filters or deliberate angle control materials such as “Display Film” made by 3M.
- Liquid Crystal Displays
- Some light emitting diode displays dependant on control of light emitted from the chip edge.

c) For projected image displays such as the Head-Up Display, the Helmet Mounted Display or Virtual Reality Headsets the display optics may be able to fill the users eye-pupil(s) but may not be able to fill the pupil of a high NA instrument resulting in readings which are not representative of what the intended user of the display will see.

Instrument Calibration 6

All colorimetric instruments should be subject to regular periodic calibration to ensure that they are yielding the correct answers. This is desirable to give the user confidence in measurements but is also essential to satisfy the Quality Assurance requirements of regulating agencies.

The complexity and frequency of calibration depends very much on the instrument type so the prime guidance on calibration procedure and interval must come from the instrument supplier. A good approach when faced with a new unfamiliar instrument is to start with a short calibration interval, say 3 months, keeping a record of the extent of any changes in calibration. If this record shows a stable calibration then the interval can be increased and/or the procedure pruned to embrace fewer parameters.

An essential item of hardware for calibration of colorimetric instruments of all types is the standard source.

6.1 Standard Source

In this context a standard source is on a light source which provides a luminous output surface having stable and known spectral radiance (and therefore luminance) which can be used to check or calibrate instruments readings of luminance, chromaticity and, in the case of spectroradiometers, spectral radiance.

Such sources are nearly always based on an incandescent lamps specially designed and manufactured for stability, and run from special purpose stabilised power supplies which include a soft-start to avoid inrush-current lamp stresses. The lamp is mounted in a housing in the form of a baffled integrating sphere, maybe with a diffusing window, to yield an output window of uniform luminance and a good approximation to Lambertian emission.

The lamp is usually powered to yield a colour temperature consistent with CIE Illuminant A (see Appendix) not because it gives a spectral distribution best suited for this calibration function but because it represents a lamp filament temperature conducive to good stability. A standard source as defined here is an essential calibration tool but is useless unless it is carefully nurtured, used only for the minimum time needed to perform the calibration function and itself subject to calibration at regular intervals in a manner traceable to national standards. The standard source can be calibrated at either an accredited test house or at the National Physical Laboratory (NPL), Teddington and will result in a test certificate tabulating spectral radiance at 5nm intervals from 380nm to 780nm. (If the standard source is to be used for purposes other colorimetry than calibration over an extended range is needed).

6.2 Filter Colorimeter Calibration

Essentially the calibration process is exposure to a standard source and adjusting the gain of the three stimulus channels to obtain readings which match the declared values for the standard source. First the Y channel is adjacent for correct luminance reading followed by interactive adjustment of the X and Z channels to obtain correct chromatcity reading. In a modern instrument the hardware adjustments will be replaced by storing new values in a non-volatile memory associated with an internal microprocessor, with the values generated by a resident algorithm.

Having calibrated the colorimeter in this way the only subsequent application that will surely yield accurate readings is measurement of sources with the same spectral distribution as the standard source.

6.3 Spectroradiometer Calibration

Unlike the filter colorimeter the calibration of a spectroradiometer is performed in the spectral radiance domain and does not invoke CIE parameters.

The detail procedure will be different for each instrument design but there will always be the two aspects of establishing the wavelength scale and establishing the radiance scale at each wavelength.

The calibration process will be incorporated into software routines which generate correction files in computers memory and these files are invoked in all subsequent measurements.

a) Wavelength Scale

There will be an approximately linear relationship between array position and tuned wavelength in a diode array type of instrument and in a conventional scanning instrument the sine-bar mechanism will yield a linear relation between lead-screw rotation and tuned wavelength. However, the relation may not be exactly linear and will exhibit some wavelength offset errors.

The wavelength scale is established by scanning a source which exhibits strong line-spectrum emissions at a few suitable wavelengths. Analysis of the wavelength errors at each spectral line provides the coefficients of an error polynomial which are stored and invoked for all subsequent measurements.

Small capillary-tube discharge lamps with appropriate filling provide line spectrum emission by mechanisms which are sufficiently fundamental to be wavelength standards in their own right without resort to a lighter level of calibration.

Both mercury filling and cadmium filling provide spectral lines suitable for spectroradiometer wavelength calibration. It is worth noting that the most common displays include spectral lines of known wavelength so that each measurement provides a check that there have been no major shifts in wavelength scale. Colour CRT phosphors usually incorporate a line spectrum red at 625nm and AMLCDs are usually backlit with fluorescent lamps having a mercury line at 546nm.

b) Spectral Radiance Scale

All the elements in a spectroradiometer exhibit variations in performance with wavelength:-

- The transmission of lenses, fibre-optic bundles etc. is wavelength dependent.
- The throughput of the monochromator is wavelength dependent.
- The detector responsivity is wavelength dependent.

The net result is an instrument whose uncorrected response varies widely with wavelength. The basic approach to this is to establish the detail of the instruments response by using the instrument to measure the apparent spectral radiance of a Standard Source. This process can be formalised as follows:-

A calibration measurement will yield a spectral file $Mc(\lambda) = S(\lambda) \times R(\lambda)$

[Equa. 7]

Where $S(\lambda)$ = Standard Source spectral radiance
$R(\lambda)$ = Instrument response function

With calibration measurement $Mc(\lambda)$ and the known $S(\lambda)$ compute and store a calibration file $C(\lambda)$ using $C(\lambda) = S(\lambda)/Mc(\lambda)$

[Equa. 8]

Subsequent measurement on an unknown source $P(\lambda)$ yields a measurement $M(\lambda)$ where

$$M(\lambda) = P(\lambda) \times R(\lambda) \quad \text{[Equa. 9]}$$

Use Equa. 7 and 8 to replace $R(\lambda)$ and obtain

$$P(\lambda) = M(\lambda) \times C(\lambda) \quad \text{[Equa. 10]}$$

Thus the stored calibration file can be used, with one arithmetic operation, to convert all subsequent raw measurements $M(\lambda)$ to valid $P(\lambda)$ data. For colorimetric data the measured spectral radiance $P(\lambda)$ is plugged into the CIE formulations in section 2.

Clearly the validity of the radiance scale thus generated depends very much on maintaining the same instrument configuration during measurements as was used to generate the calibration data.

The above general description for establishing wavelength and radiance scales apply to both the scanning and diode-array type of spectroradiometer but the practical implementation is likely to be very different. The diode-array instrument, being based on the silicon photodiode detector, will have good long-term stability of calibration. This stability coupled with the fact that the instrument configuration is fixed with few user options allows long intervals between calibration. The calibration process may be performed in a central department or outside test-house so that the process is transparent to the user. However, do not let it become transparent to the point of forgetting to tell the software that you have changed lenses and that a different C(λ) file should be called.

The scanning instrument presents a different situation in that:-

- In the interests of flexibility of application it allows many configuration options from a choice of lenses, of field stops, of spectral bandwidth, of measurement wavelength interval, of wavelength range and of filters inserted in the optical path. Each configuration option will require differing C(λ) calibration files which virtually dictates that the user should perform an in-situ calibration whenever the configuration is changed.

- In the interests of sensitivity the photomultiplier tube (PMT) is the preferred detector. However, the excellent sensitivity of a PMT is effected by many parameters of its local environment such as temperature, magnetic fields, mechanical shock and recent history of light exposure. For this reason it is recommended that a spectral radiance calibration is performed at the start of each day's measurement session. Allow about half an hour after application of power to the PMT before proceeding with calibration, and also allow the standard source about 10 minutes to reach a temperature equilibrium.

For the best confidence in equipment stability perform another calibration at the end of the measurement session and compare the new C(λ) file with the old C(λ) file.

Standard Illuminants

For regular colorimetry of reflecting surfaces the spectral distribution of the illumination of the surface is a central issue, so the CIE have established a few well defined spectral distributions as CIE Standard Illuminants.

The illumination characteristics are a less important issue in display colorimetry but will be encountered in some instrument and display specifications so they are briefly described here.

CIE Illuminant A

Intended to simulate artificial lighting conditions and is implemented by a tungsten filament lamp operated at a correlated colour temperature of 2856K.

The detail spectral distribution is fully defined in Table 1.1. of Ref. 3.

CIE Illuminants B and C

These two illuminants represented two phases of daylight and could be implemented with prescribed liquid filters in conjection with Illuminant A.

However, they have fallen into disuse and CIE have declared Illuminant B to be obsolete.

CIE Illuminant D65

This illuminant, the preferred item in a set of daylight illuminants, has no physical implementation yet and is defined solely in numerical spectral distribution terms.

D65 has CCT of 6500K but the complete set covers a wide range of different daylight phases.

The set is numerically defined in Ref. 3.

Units Conversion

There is a plethora of photometric units but most are now of historic interest only. The important conversions are:

Luminous flux in lumens.

Luminous flux is radiant power which is weighted with wavelength according to brightness producing ability using the V(λ) function. The rate of exchange at V(λ) peak is: 1 Watt at 555nm = 683 lumens.

Luminance

Luminance is the physical correlate of surface brightness and is expressed in Candela/Square meter in the Metric system but the Foot-lambert, is almost universal in the USA.

The rate of exchange is:

1 Foot lambert = 3.426 candela /m2

Illuminance

Illuminance is a measure of luminous flux per unit area falling on a surface. The metric unit is the lux (1 lumen/m^2) but the Foot Candle is widely used in the USA. The rate of exchange is:

1 Foot candle ≡ 10.76 lux

References

1. “Measuring Colour”
R W G Hunt

2. “Light as a trace visual quantity : Principles of Measurement”
CIE Publication No. 41, 1978

3. “Colorimetry”, Second Edition
CIE Publication No. 15.2, 1986

4. “NVG Compatibility (a primer)”
Bentham Instruments, January 1997