Issues of Colour Measurement and Assessment

Tony Johnson and Phil Green*

Keywords: Colorimetry, Spectrophotometry, Differences

Abstract: Colorimetry in Graphic Arts is faced with a number of uncertainties. Many of these, such as prediction of appearance under varying viewing conditions and measurement precision, have been (partially) resolved by standards defined by CIE and ISO. However, other uncertainties still remain and two of them are addressed in this paper:

- When making colour measurements of samples that differ in gloss the geometry of the instruments used for the colour measurement can produce very misleading predictions of colours that match.
- Since the introduction of acceptability parameters into colour difference assessment - with the introduction of the CMC and CIE94 colour difference formulae - few attempts have been made to determine what, if any, weightings are appropriate for Graphic Arts.

An experiment is being set up to evaluate the acceptability parameters appropriate for the use of CIE94 in Graphic Arts, and provide additional data on any residual non-uniformity from this formula which may be significant for our industry. In addition we will correlate gloss and colour measurements in order to produce a model that will provide measurements that improve the prediction of matching pairs with different gloss, for typical Graphic Arts viewing conditions. Some of the background to this work, together with some preliminary findings, will be discussed.

Introduction

Uncertainty in Colorimetry - Colorimetry is not as unambiguous as many people seem to believe. The reasons for this lie in the history of its development and the variables it is attempting to contend with. However, to understand the issues we are addressing in this paper it is important that the reasons for this possible ambiguity are understood. For this reason we will start with a brief summary of the main characteristics of the popular standard used for colorimetry.

*London College of Printing

The colour matching functions of the standard colorimetric observer (essentially the average of the results of colour matching experiments obtained from a number of observers) are the basis of the CIE system of colour measurement. The original set were defined in 1931 and specify the amount of each of three imaginary stimuli required to produce a colour match to each wavelength of the light within the visible spectrum by a 'typical' observer. Because the mixing of coloured lights is mathematically linear it follows that if the amount of each wavelength of light emitted by a sample is known, the total amount of each stimulus required by a 'typical' observer to match the sample can be predicted quite easily. This is achieved by summing the product of the colour matching functions and sample emission over the range of visible wavelengths. The three values that result are known as the tristimulus values.

In 1964 a supplementary set of colour matching functions were developed by the CIE for measurement in situations where the application requires assessment of large areas of colour. This was necessary because the original data was only correct for small fields of view, of less than 2° angular subtense. The resultant colour matching functions define the 1964 (10°) standard colorimetric observer. Whilst these are widely used in the paint and printing ink industries they are not generally used for measurements produced to support the reproduction of complex coloured images, for which the 2° observer is generally preferred. It should be noted that it is not generally possible to convert data obtained from one observer calculation to that of the other.

Two instrument geometries are specified by CIE for reflectance measurement. The first of these is usually abbreviated to 0°/45° (or 45°/0°) and specifies that the sample shall be illuminated with a narrow beam of light at 0 or 45 degrees to the sample. The reflected light is then measured at the alternate angle. The second geometry is 0°/diffuse (or diffuse/0°) and this specifies that the sample shall be illuminated with a narrow beam of light at 0 degrees, and the diffusely reflected light integrated by a highly reflecting sphere so that the total reflected light may be measured. Alternatively, the incident light may be diffuse and the reflected light measured at 0 degrees. (In practice it is normal to offset the 0 degrees to 8 degrees, but for most samples this makes no significant effect). A 'gloss trap' may be inserted or removed for most instruments of this sort so that a significant proportion of the specularly reflected light may be removed from the measurement if desired. Again, it should be noted that it is not generally possible to convert data obtained from one instrument geometry to that of the other.

Using spectrophotometers (of either geometry) to provide relative spectral reflectance and transmittance data, and then computing the tristimulus values for any particular illuminant overcomes many of the problems associated with tristimulus filter colorimetry. As the price of spectrophotometers has fallen, they have become the most common type of instrument used for the colorimetry of reflecting and transmitting samples. However, since CIE does not provide observer weighting functions for the 10nm and 20nm bandpass spectrophotometers widely used in the graphic arts industry (only 1nm and 5nm are recommended and provided by CIE), the weighting functions used are normally interpolated from the CIE data. Since there are many ways of undertaking this interpolation it would be possible to find instruments that would report different tristimulus values even if the reflectance data obtained were identical.

The standardisation by the CIE of the spectral power distribution of various illuminants (of which illuminant A and the daylight series are the most important) has helped in producing reasonable consistency between tristimulus data obtained from such instruments. It should be noted that illuminants A and D65 were standardised in 1991 as ISO/CIE 10526 (Standard Colorimetric Illuminants). However, there are a wide range of illuminants and care has to be taken in selecting that appropriate for the purpose; D50 is that most commonly used in Graphic Arts calculations.

To ease the confusion that can arise in colorimetry because of the various measurement options (2 measurement geometries, 2 standard observers and a wide range of possible illuminants) an International Standard (ISO 13655) has been issued that specifies the conditions to be used for colour measurement in Graphic Arts. The standard also specifies other parameters needed to make unambiguous measurement. Thus it was agreed that Graphic Arts should use the 2° observer, D50 illumination, a triangular band-pass when calculating tristimulus data from spectral data obtained from measurements made with a 0°/45° or 45°/0° instrument geometry and a black backing to the print during measurement. Weighting functions have also been specified, based on those proposed in the US standard test method ASTM E-308 (1985).

Despite the high degree of standardisation implied by the above summary, variations in results obtained from different measurement procedures, but particularly different instruments, are quite common. Poor tolerance specification by CIE inevitably means that different instruments will produce somewhat different results, even if they are nominally of the same geometry and other instrument design issues are not significant. When these other sources of variability such as stray light; wavelength error; polarisation; linearity; bandwidth and accuracy of the white reference are considered it is clear that absolute accuracy must be limited. When issues associated with the sample itself - such as thermochromism; fluorescence; flatness; directionality and uniformity are also considered it indicates how much care is required in measurement to achieve reliable data.

For many aspects of colorimetry some of these issues are of limited importance. Applications such as the control of paper and ink manufacturing are usually limited to internal control within companies, using a restricted range of instruments. Such measurement is primarily concerned with the specification of differences from some approved sample and in such situations issues of inter-instrument agreement are not important (apart from between those in use within the company, which are often all of the same type). So long as sufficient care is taken to calibrate the instrument and sufficient measurements are made to minimise sample effects the result obtained is adequate, even if the absolute accuracy achieved is fairly low. It is only when data is required to be exchanged that absolute accuracy (or at least high levels of inter-instrument agreement) is required.

However, when we consider issues associated with cross-media matching we are precisely in this situation – even if all the data exchange is still internal to one company. Measurements need to be made of monitors, prints and photographic transparencies, with the objective of defining a colour match between the media. Usually such measurements will be made with different instruments, and even where it is possible to use a common instrument it will require different illumination and viewing geometries. This introduces a new problem, which is that the geometries specified in all of the standards are essentially measurements of a flareless condition. But most practical viewing situations include some degree of flare. Furthermore, in some situations (particularly for the viewing of reflecting materials) the geometry used for measurement will not be at all similar to that used for viewing. So, as the gloss of any samples that have to match varies from one to another this will introduce yet another source of error in the prediction of a match.

Issues of Colour Difference Measurement - The non-uniformity of the colour space defined by XYZ means that the system is not useful for defining colour differences (or tolerances). Thus in 1960/63 a "uniform" transformation of XYZ was introduced by CIE. This was superseded in 1976 by the transformations now known as CIELUV and CIELAB. However, neither of these spaces is particularly uniform. Two pairs of colours with a just noticeable difference between them will have colour difference ratios of at least 3:1 or 4:1 when selected from different regions of colour space. However, this is substantially better than the ratio of at least 10:1 obtained in the 1931 system. In fact, there is actually little to choose between the two 1976 uniform colour spaces in overall terms of uniformity. The major difference is the way that the non-uniformity is distributed. Therefore, the choice of the CIELAB formula, commonly used by the dye and pigment colorist, comes down either to prejudice or to the importance of certain regions of colour space to the users.

The search for a more uniform space has continued since 1976 although it is generally believed that any really major improvements are unlikely without considerable complexity. An alternative approach is to use one of the above 1976 colour spaces and modify the colour difference formula associated with them. The UK Colour Matching Committee of the Society of Dyers and Colorists proposed such a procedure, in which new parameters were derived from CIELAB values, Clarke et al (1984). The resultant formulae became colloquially known as CMC (l:c) and are widely used in the textile industry. They are derived from considerable experimental data concerning the acceptability of colour matches in the textile industry and appear to give substantially better uniformity for small differences in colour. In 1995 a somewhat simplified proposal, based on the RIT/Du Pont tolerance data derived from experiments with automotive paints, Berns et al (1991), was accepted by CIE and published as CIE_{94} (1995). It is defined as follows:

$$\Delta E_{94} = \sqrt{\left(\frac{\Delta L^*}{k_l S_l}\right)^2 + \left(\frac{\Delta C^*}{k_c S_c}\right)^2 + \left(\frac{\Delta H^*}{k_h S_h}\right)^2}$$

where $S_l=1$, $S_c=1+0.045C^*_{ab}$ and $S_h=1+0.015C^*_{ab}$; and k_L, k_C and k_h are constants that can be set to improve acceptability (rather than perceptibility) tolerances.

Essentially CIE_{94} corrects for a major uniformity error in that pairs of colours of high chroma produce far larger numerical differences than their appearance difference justifies. Thus, pairs of very similar high chroma colours produce very large numerical differences with CIELAB that the correction in CIE_{94} reduces. However, it is known that this correction alone is still not adequate to provide a reasonably uniform difference space. Certain hues, particularly yellows, still produce larger numerical differences than other hues and it is anticipated that a correction for this will be introduced at some time, when sufficient data is available to fully quantify the error.

It is worth noting that CIE_{94} is non-Euclidean. The way in which the formula is defined leads to a difference in ΔE according to which of the two samples is taken as the reference. For some high chroma sample pairs, which may have an acceptable difference with chroma differences as high as 12, this can lead to a difference of about 0.5ΔE depending upon which sample is taken as the reference. We assume that this is why the CIE recommendation is that CIE_{94} should only be used if the CIELAB ΔE is less than 5. However, it would be an easy problem to fix, by a simple re-definition of the formula, if it proves to be a significant issue for acceptability of the equation.

It should also be noted that any such space (even were it perfectly uniform) can only be strictly correct for one condition of viewing. Thus as visual adaptation effects take place (in which the visual mechanism 'adjusts' as the level or colour of the of illumination changes), or simultaneous contrast effects are introduced (in which the sample is seen in context with other colours surrounding it) the uniformity of the space must deteriorate. We are fortunate that colour constancy prevails which means that many changes are small; nevertheless they exist. For many practical purposes such problems are not relevant; the changes due to adaptation are small compared to the non-uniformity of the spaces themselves.

Nevertheless, it should be noted that the uniform colour spaces are normally assumed to be 'uniform' for illuminant D65. It is important to ensure that the 'uniformity' of this space also applies to any differences specified for Graphic Arts viewing conditions, using D50. In future it seems likely that colour differences will be based on colour appearance measurements (rather than simple measures of the stimulus) but there is much work to be done to get to that point. In the meantime data is required to establish whether the existing formulae are satisfactory for our needs, and the extent to which the weighting of the various parametric constants in the models will improve them for Graphic Arts applications. That same data can also be used as a means of evaluating any future models, possibly including those based on appearance modelling if the experiments are sufficiently comprehensive.

Proposed Work Programme

The discussion above raises a number of issues that we hope to address in our study. Our objective is to determine the following:

1) The magnitude of perceptible and acceptable differences in colour for typical Graphic Arts viewing conditions, as well as for a limited range of alternate conditions.
2) The influence of gloss on the measurement of colour appearance.

To this end we have printed, by lithography, a large number of colours (approximately 60 in total) on a variety of substrates. These have been selected to sample colour space reasonably uniformly, but the set also contains some colours that are fairly close to each other because of the method used for production. Each of these samples is approximately 2.5cm square and has been produced together with a number of similar samples that differ slightly from the 'aim' colour. We used a polynomial model, obtained by regression analysis from the 'SWOP' characterisation data published by ANSI, to compute colours with a difference of approximately 1, 2, 3, 4 and 5 units of difference in each of hue angle, chroma and lightness from the aim colours. We then added additional colours, with a larger difference, to ensure we had samples that were clearly visually different from the aim colours. In all each 'aim' colour has approximately 24 of these 'adjacent' samples, leading to a total in excess of 1400 colour patches. Additional variation was obtained by the natural fluctuation of ink weight over the run of about 500 copies.

For Phase 1 of the experiment we have selected 25 'aim' colours from these samples, printed on a gloss paper. Figures 1 and 2 show how these aim colours are distributed. Figure 1 shows the samples plotted in an a^* vs b^* diagram, and figure 2 shows them when L^* is plotted against C^* for all hues together.

Figure 1 – Distribution of the colours to be used for assessment (a*/b* plane)

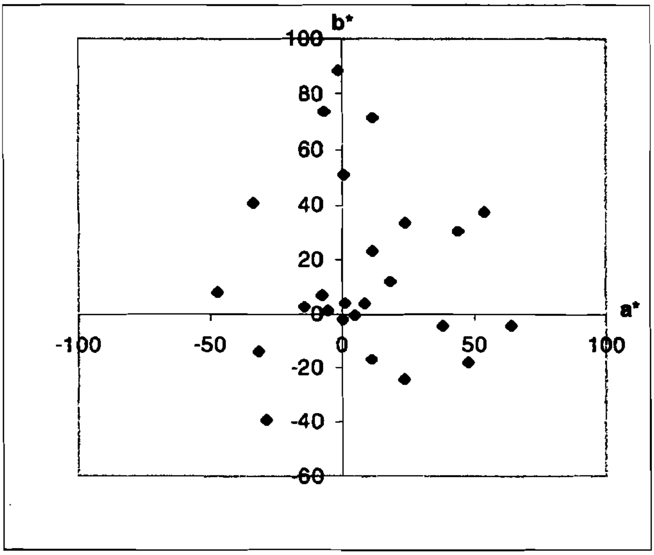
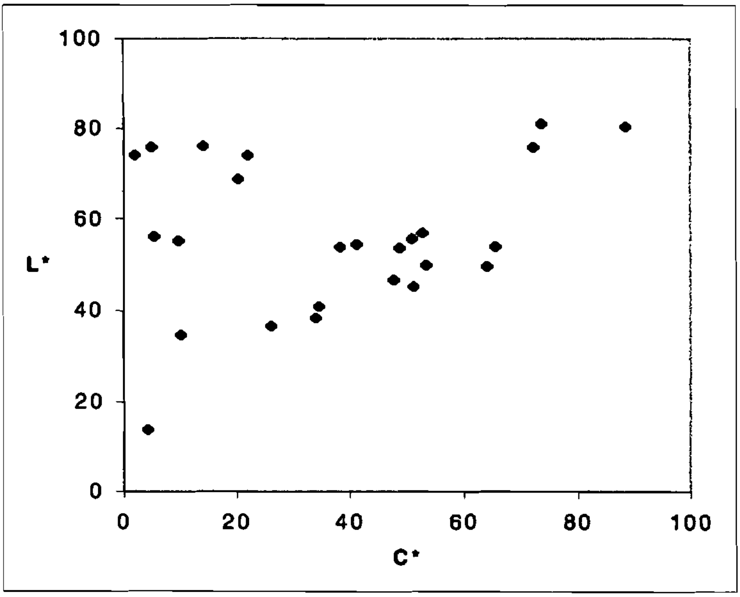


Figure 2 – Distribution of the colours to be used for assessment (L*/C* plane)



For each set of samples observers are presented with the 'aim' colour, together with each sample in turn. They are asked to specify:

1) Is there a perceptible difference in colour between the colour pairs, and
2) If there is a difference would it be acceptable as a proof to production print deviation, or as variation within a run, for a critical production process such as packaging consistency or matching a house colour?

For this study we are using a viewing booth simulating the D50 illuminant, at an illumination level close to that defined for critical appraisal in ISO 3664 (2000 lux). However, for a limited number of samples we will use more than one level (the second will be close to the lower level of illumination proposed in ISO 3664), and other illuminants; D65 and A.

The samples are viewed on a black background at an angle designed to minimise specular reflection. The viewing distance of 0.4m ensures that the visual effect of screen angle interference is not distracting and corresponds to an angular subtense of 1.8 degrees. Samples have a 1.5mm unprinted surround, and are positioned adjacent to each other. Observers are permitted to exchange sample locations during the experiment.

CIE guidelines for research on colour difference were published by Robertson (1978) and updated by Witt (1995). These describe the parameters in need of investigation and recommend experimental procedures and methods of analysis. Since we are particularly interested in determining parameters for use in the graphic arts we have found it necessary to depart from certain of the CIE guidelines (particularly the choice of illuminant and illumination level). However, we will be including conforming conditions in some of the experimental work in order to allow comparisons to be made to other studies.

Phase 2 of the experiment, which is to attempt to introduce parameters for gloss and flare into the measurement of colour appearance, is distinctly more difficult. A small number of colours will be selected for comparison across the substrates. These vary in gloss significantly as the substrates range from Newsprint, through a matt-coated paper to a high gloss coated paper. As well as repeating the assessment above for the other substrates observers will undertake magnitude estimation to evaluate the hue, chroma and lightness of the samples presented.

Colour measurements of all the samples will be made using telespectroradiometry and both of the standard CIE geometries. Spectrogoniophotometric measurements will also be made on the samples to determine various gloss attributes. This combination of measurements and subjective judgements will permit us to investigate models that 'correct' colorimetric data to compensate for both viewing flare and differences in gloss between media. Such information will be useful in the definition of colour appearance when combined with the corrections specified in CIECAM97s. Of course those corrections do partially correct for the effects of viewing flare (as well as simultaneous contrast and luminance level) and so the data obtained will also enable us to test the effectiveness of this correction for typical Graphic Arts viewing conditions.

The data from Phase 1 will be analysed to investigate the uniformity of the colour difference equations in current use, as well as the establishment of parametric constants for acceptability in CIE 94. As stated earlier, the data from Phase 2 will be used to determine a model for colour appearance based on both gloss and colorimetry.

We hope that phase 3 of the experiment will deal with complex images. As a result of the models developed from the data collected in phases 1 and 2 we should have a good indication of the variations required to achieve a colour appearance match, for proximal patches of colour, across different substrates. However, we anticipate that colour tolerances can be much larger as we deal with complex images, owing to the greater spatial separation of individual elements and the assimilation effect that arises from the very small angular subtense of individual pixels. Our intention is to set up an experiment to validate this hypothesis.

Description of study to date

As we write this paper we are concentrating on Phase 1 of the study. We have produced the printed samples and measured the colour of those produced on a gloss-coated paper using the standard Graphics Arts geometry (i.e. 45°/0°). We have made a study of the accuracy and consistency of the spectrophotometers to be used for this work, evaluated the uniformity of the samples which have been produced and have undertaken a pilot study of the visual experiment. This pilot study is the work we are reporting in this paper. The remaining work outlined above will be continued over the next 18 months and we hope to report on it at the next 2 TAGA conferences.

Summary of the literature review

There is a substantial body of literature dealing with the uniformity of the CIELAB, CIELUV, CMC and CIE_{94} colour differences discussed above, as well as proposals for alternative colour difference formulae. We hope to report on this in substantially more detail elsewhere. However, a brief summary will be presented here. For simplicity we have separated discussion of studies that are specifically directed towards Graphic Arts applications toward the end of this section. The initial discussion focuses on more general studies, or those using samples more appropriate to other industries.

Since the first CIE Guidelines for research in this area were published, the following types of study have been completed:

A. Determination of colour difference tolerances, either by defining tolerance ellipses or by defining the median tolerance of a distribution. [Rich & Billmeyer (1983), Luo & Rigg (1986), Alman et al (1989), Berns et al (1991), Indow & Morrison (1991), Berns (1996), Melgosa et al (1997), Ebner & Fairchild (1998b) and Qiao et al (1998)].

B. Development of new colour difference metrics or more uniform colour spaces based on experimental results. [Luo and Rigg (1987) and Ebner & Fairchild (1998a)].

C. Analysis of the performance of colour difference equations and their weighting functions. [Kuehni (1982 and 1998), Alman et al (1989), Mahy et al (1994), Melgosa et al (1994, 1995 and 1996), Witt (1994 and 1999) and Berns (1996)].

D. Parametric analysis designed to improve understanding of the effects of surround, illuminant, surface texture etc. [Witt (1990)].

E. Studies of other aspects of the colour difference problem, such as observer variability, the scaling of colour differences between large and small differences, or the relationship between perceptibility and acceptability. [Kuehni (1982), Stokes et al (1992a and b), Fairchild & Alfvin (1995), Pointer & Attridge (1997), Witt (1987, 1995 and 1999)].

Samples - All studies involved a single type of sample, and between one and 17 colour centres. Glossy acrylic paints were the most commonly used colorant, although a number of studies used colour prints on glossy paper for ease of sample creation. Witt (1994) investigated a tolerance database of textile judgements. A small number of studies used CRT samples or images. No studies used prints made by commercial printing processes, or more than one type of sample with difference reflectance properties.

Psychovisual experiments - Pair comparison and magnitude estimation experiments were used in many studies, most commonly using a reference or anchor image or pair of samples. The anchor pair were commonly achromatic samples with an L^* lightness of approximately 50, differing by around 1 ΔE. Although this makes it possible to compare the results from different experiments, it is difficult to see how observers could extend a near-threshold lightness difference to the judgement of supra-threshold chromatic difference.

Some authors reported observer difficulty, fatigue or loss of motivation, often associated with long sessions.

Colour difference equations and their weighting parameters - There is general agreement on the improved performance of CIE_{94} over other difference equations. Other equations (CMC, BFD etc) have historically given good results with individual data sets, but when applied to other data sets their weighting parameters cannot be adjusted to give significantly better results than CIE_{94}, and in most cases the performance is worse.

Some studies have reported an element of interdependence between weighting functions and values of lightness, hue, and saturation or chroma. However, these interdependencies were generally weak or not significant, and not robust when applied to other data sets.

Studies investigating the weighting for lightness in glossy paint samples agreed that it should be set to unity, with no significant improvement by using a lightness-dependent function ($S_L = K_1 + K_2L^*$, where K_1 and K_2 are constants). Witt (1994) found that a weighting for lightness close to 2 gave a good prediction of the data for judgements of a non-glossy textile material, and this result was supported by Berns (1996).

The results of studies investigating the tolerances for hue and chroma were less clear, with considerable variation in results between different studies. Hue non-uniformity in CIELAB makes the interpretation of results for hue weighting difficult to interpret.

Acceptability and perceptibility - The main study on acceptability and perceptibility in complex images, by Stokes et al (1992a and b), concluded that scene content affects acceptability but not perceptibility. A mean perceptibility tolerance of 3ΔE was found; while for acceptability the mean tolerance was 6ΔE. A simple linear scaling from perceptibility did not model the acceptability results well. Acceptability results showed greater variance, possibly indicating that the task was not well defined or that observers applied their own criteria. However, this study was on CRT images and its applicability to hard copy is not clear.

Observer variability - Studies that reported inter and intra observer variation generally did so with a single observer for intra-observer variation. As might be expected, the former was usually found to be higher than the latter, typically by a factor of 2 (both for homogenous samples and complex images). Alfvin & Fairchild reported mean inter-observer variability for homogenous samples of 2.5 ΔE, with small differences in different types of media.

Graphic Arts studies - Because of the nature of the samples selected, and the viewing conditions used, it is uncertain how much of the above is directly relevant to Graphic Arts. The two most relevant studies for this application are those by Bassimer et al (1995) and Paul (1998). Bassimer and his colleagues prepared 26 pairs of matte, non-metameric colours that were visually fairly similar. There were 6 high chroma colours plus a brown and a grey. The sample pairs varied in colour difference from approximately 0.3 to 3 for many pairs but the differences for 4 pairs of yellow samples were much higher. A panel of observers was asked to judge the perceptibility of any differences, and also to evaluate the characteristics of any differences that they found. The important conclusions from this study were that:

- perceptible differences were approximately 0.3 for grey up to 5 for yellow (a ratio of 16:1),
- for most colours there was reasonable correlation between perceptibility of differences and ΔE, but this was not the case for acceptability,
- the visual rankings for deviations in chroma, hue and lightness correlated well with those determined by measurement,
- there was little difference in perceptibility as the viewing illuminant was altered from D50 to A,
- there were significant differences between observers (and possibly a gender difference also).

Paul asked observers to compare 34 pairs of colours with varying colour differences between them. They were asked to categorise the difference into 6 categories and these results were processed, in an unspecified manner, to produce a single value for visual assessment of the difference for each colour. He found that the correlation between visual ranking and ΔE_{ab} was low, but improved significantly when the visual ranking was compared to ΔE_{94}. It is difficult to compare the improvement in terms of a uniformity ratio but it would appear to show that for ΔE_{ab} it was at least as bad (16:1) as Bassimer and his colleagues found and certainly improved by at least a factor of 2 for ΔE_{94}. We understand from the principal author that the data of Bassimer et al also showed a major improvement in uniformity when ΔE_{CMC} was subsequently used to define the colour differences for results of their study.

Thus, both studies show that there is a significant problem for the Graphic Arts industry with the use of ΔE_{ab} and both suggest that it can be reduced by the use of CIE_{94}. However, the degree of improvement, whilst significant, is difficult to quantify from the Paul paper, and Bassimer and his colleagues have not published their data in this regard. Neither is there much indication in either study as to whether the acceptability weighting factors that CIE_{94} introduces could be useful in our industry.

Although the observers in the study by Bassimer et al were asked to state whether samples were different in lightness, chroma or hue, no attempt seems to have been made to correlate this information with batch acceptance. This is presumably because there were really too few samples to permit this. It seems clear from one of their samples (a yellow) that hue differences were considered particularly critical in this colour region. The smallest of 4 ΔE_{ab} values was deemed by their observers to be the most visible and unanimously unacceptable difference, probably because the colour shift for that pair of colours was predominantly in hue. However, that is all the information pertaining to the reason for acceptability that can be obtained from the published data. Thus it is not possible to determine whether the poor correlation between ΔE and acceptability could be improved by adjustment of the weighting factors in the CIE_{94} difference formula.

This is an area where we hope to extend the important work reported by these two authors, as well as adding to the database of information that should enable the industry to eventually produce firm recommendations on the determination of acceptability by numerical difference.

Results to date

At the time of writing this paper we have only completed a pilot study which has been performed to validate the testing procedure. Thus the data gathered is for only 12 of the 'aim' colours we intend to gather data for, and for only 4 observers. The colorimetric data for the 12 colours assessed to date, calculated for illuminant D50, is given in table 1.

Table 1 – CIELAB values for the 'aim' colours used in this study

Colour	*L**	*a**	*b**
Pink	74.00	18.33	12.21
Light Grey	74.17	0.37	-1.99
Purple	38.37	23.67	-24.44
Black	13.91	1.22	4.14
Dark Orange	49.89	43.81	30.52
Dark Green	41.07	-31.68	-13.98
Greenish-Yellow	80.38	-1.35	88.54
Mid-Grey	56.11	-5.25	1.31
Brown	54.48	23.78	33.66
Dark Grey	34.84	-7.40	6.90
Mid-Green	56.87	-33.58	40.73
Light Purple	68.86	11.40	-16.88
Reddish-Yellow	75.90	11.69	71.31
Cyan	53.67	-28.66	-39.35
Magenta	49.65	63.99	-4.44

Because of the limited number of observers used to date the data has not yet been subjected to any systematic statistical analysis. The results in table 2 are based on largely on inspection, and show the upper limits for both perceptible and acceptable difference obtained when 3 observers, or more, produced consistent assessments. The difference is expressed in terms of both CIELAB and CIE_{94} with no weighting factors applied.

Table 2 – Upper limits of colour difference for both perceptibility and acceptability

Colour	Perceptible difference		Acceptable difference	
	CIELAB	CIE_{94}	CIELAB	CIE_{94}
Pink	2	2	4	4
Light Grey	1	1	3	3
Purple*	2	2	5	5
Black	1.5	1.5	4	4
Dark Orange**	1	0.75	4	4
Dark Green	2	1.75	4.5	3.5
Greenish-Yellow*	1	0.5	5	4
Mid-Grey*	1.5	1.5	3	3
Brown	2	1.5	4	4
Dark Grey	1.5	1.5	5	4.5
Mid-Green***	2	1.5	5	4
Light Purple	1	1	4	3.5
Reddish-Yellow***	1	0.6	7	3.5
Cyan	2	1.5	6	6
Magenta	1	0.75	6	4

* The assessments of these samples indicated that the hue weighting should be reduced.
** The assessments of these samples indicated that the lightness weighting should be reduced.
*** The assessments of these samples indicated that the chroma weighting should be reduced significantly.

Analysis and Conclusions

From the literature review we conclude that colour difference assessment can be difficult, and that the values for both perceptibility and acceptability are very dependent upon the media, the position in colour space and the nature of the difference between the samples.

Our results show that perceptible differences are obtained at around 1 to 2 ΔE_{ab} units. What surprised us is that this limit did not seem to change significantly when high chroma yellows were compared to grey samples and for this reason CIE_{94} actually performed worse, with respect to producing a single value tolerance, than CIELAB. This is in contradiction to almost all the published data and requires further investigation to determine the reason.

However, for acceptability CIE_{94} proved far more useful. The formula produces greater uniformity for the magnitude of acceptable difference across different colour regions. It was particularly noticeable for 2 high chroma samples – the mid-green and reddish yellow - where observers were accepting differences of 9 and 12 respectively for ΔE_{ab}. These differences were reduced to 6 or less for CIE_{94}. If we are looking for a single value for acceptability $\Delta E_{94} = 4$ seems a good value to choose! This is lower than the value of 6 obtained by Stokes (1992) but that is not unexpected since those results were obtained with complex images.

It is interesting to note that that the limits of acceptability are approximately 2 or 3 times those for perceptibility. This is similar to the findings of Stokes (1992). However, our results show an exception for high chroma colours, where this could increase to 5 to 7 times perceptibility regardless of which difference formula was used.

There is some evidence that H* errors need to be weighted to make them more significant (by reducing k_h) and that C* can be reduced in significance for high chroma colours (by increasing k_c) even beyond the reduction achieved by CIE_{94}. We also note that for dark, high chroma colours L* errors may be more significant. However, if this is so it means that k_l would not be a simple constant but a function of both L* and C*. Such interdependencies have been found before, but have generally not been confirmed by other studies.

Work to be done

Much more work needs to be done to verify the results above. This is in hand. The number of observers will be around 20, to increase the statistical significance of the data, and we anticipate that most of these will have experience in the field of colour matching. We intend to investigate whether there is any correlation between colour matching experience and discrimination thresholds. However, we also intend to use print buyers as well as those involved in production. It will be interesting to see if they draw different conclusions!

The second phase of the work will then be the main focus of our future study; to investigate the effects of surface gloss on colour acceptance. This will be a complex task that we hope to report upon in future TAGA conferences.

References

Alfvin, R. L. and Fairchild, M. D. (1997) *Observer variability in metameric color matches* Col Res App **22** 174-188

Alman, D. H., Berns, R. S., Snyder, G. D., and Larsen, W. A. (1989) *Performance testing of color difference metrics using a color appearance data set.* Col Res App **14** 139-151

Bassimer, R., Costello, G., DiBernardo, A., DiPiazza, J., Kuna, D., Paulius, K., Rybny, C. and Zawacki, W. (1995) *A comparison of visual and spectrophotometric evaluations of paired color prints* TAGA proc. 558-578

Berns, R. S. (1996) *Deriving instrumental tolerances from pass-fail and colorimetric data* Col Res App **21** 459-472

Berns, R. S., Alman, D. H., Reniff, L., Snyder, G. D., and Balonen-Rosen, M. R. (1991) *Visual determination of suprathreshold color-difference tolerances using probit analysis* Col Res App **16** 297-316/14 139-151

CIE (1995) *Industrial Colour Difference Evaluation* CIE Publication 116, Vienna

Clarke, F. J. J., McDonald, R. and Rigg, B (1984) J. Soc Dye Col **100** 128

Ebner, F. and Fairchild, M. D. (1998) *Development and testing of a color space (IPT) with improved hue uniformity* Proc IS&T/SID Colour Imaging Conf **6** 8-13

Ebner, F. and Fairchild, M. D. (1998) *Finding constant hue surfaces in color space* Proc Col Img: Device Independent Color, Hard Copy and Graphic Arts 107-117

Fairchild, M. D. and Alfvin, R. L. (1995) *Precision of color matches and accuracy of color-matching functions in cross-media color reproduction* Proc IS&T/SID Colour Imaging Conf **3** 18-22

Indow, T. and Morrison, M. L. (1991) *Construction of discrimination ellipsoids for surface colors by the method of constant stimuli* Col Res App **16** 42-56

Kuehni, R. G. (1982) *Advances in color difference formulas* Col Res App **7** 19-23

Kuehni, R. G. (1998) *Hue uniformity and the CIELAB space and color difference formula* Col Res App **23** 314-322

Luo, M. R and Rigg, B. (1986) *Chromaticity-discrimination ellipses for surface colours* Col Res App **11** 25-42

Luo, M. R and Rigg, B. (1987) *Colour difference formula BFD* J. Soc Dye Col **103** 86-94

Mahy, M., Van Eycken, L. and Osterlink, A. (1994) *Evaluation of uniform colour spaces developed after the adoption of CIELAB and CIELUV* Col Res App **19** 105-121

Melgosa, M., Hita, E. and Perez, M. M. (1995) *Sensitivity differences in chroma, hue and lightness from several classical threshold datasets.* Col Res App **20** 220-225

Melgosa, M., Hita, E., Poza, A. J. and Perez, M. M. (1996) *The weighting function for lightness in the CIE94 color-difference model* Col Res App **21** 347-352

Melgosa, M., Hita, E., Poza, A. J., Alman, D. H. and Berns, R. S. (1997) *Suprathreshold color difference ellipsoids for surface colors* Col Res App **22** 148-155

Melgosa, M., Quesada, J. J. and Hita, E. (1994) *Uniformity of some recent color metrics tested with an accurate color-difference tolerance dataset.* Appl Opt (Optical Technology) **33** 8069-8077

Pobboravsky, I. (1988) *Effect of small color differences in color vision on the matching of soft and hard proofs.* Proc TAGA Conf 62-79

Pointer, M. R. and Attridge, G. G. (1997) *Some aspects of the visual scaling of large colour differences* Col Res App **22** 298-307

Qiao, Y., Berns, R. S., Reniff, L and Montag, E. (1998) *Visual determination of hue suprathreshold color-difference tolerances* Col Res App **23** 302-313

Paul, A. *Colour fluctuations in printing* J. Prepress & Print Tech **2** 18-28

Rich, D. C. and Billmeyer, F. W. (1983) *Small and moderate color difference IV Color difference perceptibility* Col Res App **8** 31-39

Robertson, A. (1978) *CIE guidelines for co-ordinated research on colour difference evaluation* CRA **3** 149-151

Stokes, M., Fairchild, M. D. and Berns, R. S. (1992) *Colorimetrically quantified visual tolerances for pictorial images* Proc TAGA Conf 757-778

Stokes, M., Fairchild, M. D. and Berns, R. S. (1992) *Precision requirements for digital color reproduction* ACM Transactions on graphics **11** 406-422

Witt, K. (1987) *Three-dimensional threshold of color difference perceptibility in painted samples: variability of observers in four CIE color regions* Col Res App **12** 128-134

Witt, K. (1990) *Parametric effects on surface color-difference evaluation at threshold* Col Res App **15** 185-195

Witt, K. (1994) *Modified CIELAB formula tested using a textile pass-fail data set.* Col Res App **19** 4 273-276

Witt, K. (1995) *Linearity and additivity of small color differences* Col Res App **20** 36-43

Witt, K. (1995) *CIE Guidelines for co-ordinated future work on industrial colour-difference evaluation* CRA **20** 399-403

Witt, K. (1999) *Geometric relations between scales of small colour differences* Col Res App **24** 78-92