Performance of lightness difference formulae

W Chou,[a] H Lin,[a] M R Luo,[a,*] S Westland,[a] B Rigg[a] and J Nobbs[b]

[a]*Colour & Imaging Institute, University of Derby, Kingsway House, Kingsway, Derby DE22 3HL, UK*

[b]*Department of Colour Chemistry, University of Leeds, Leeds LS2 9JT, UK*

Coloration Technology

Society of Dyers and Colourists

There are large variations between different previously published lightness difference experimental data sets. Two hundred and eight pairs of matt and glossy paint samples exhibiting mainly lightness differences were accumulated. Each pair was assessed about twenty times by a panel of fourteen observers using the grey scale method. The results were used to derive a new lightness difference formula (CII), and to a large extent, a new CIE lightness difference formula (CMC99). Both formulae were found to be more accurate than the typical deviation of an individual assessment from the mean of a panel of 20 observers, and outperformed the existing formulae using the present data set. The new CMC99 lightness difference formula is integrated into the new CIE colour difference equation CIEDE2000. The results also showed that special attention should be paid to measuring very dark samples. This is caused by poor instrument repeatability and inter-instrument agreement in this colour region.

Introduction

There are a number of advanced colour difference formulae currently in use that have been extended from CIELAB [1], which include CMC [2], CIE94 [3] and BFD [4,5]. They have been proven to perform much better than CIELAB for predicting the small to medium colour differences typically found in surface colour industries. The CMC and CIE94 colour difference formulae are the current ISO standards for the textile [6] and paint [7] industries, respectively. However, there are large differences between the three advanced formulae (CMC, CIE94 and BFD) in predicting lightness differences [8]. This is mainly caused by a large disagreement between the earlier published data sets [9–12] which were used for deriving these formulae.

Figures 1a–d show the CIELAB colour differences (ΔE) divided by the visual differences (ΔV) plotted against the L^* values for the BFDF–*L* [9], RIT–DuPont–*L* [10], BFDB–*L* [11] and BFDL–*L* [12] data sets, respectively. For each data set, sample pairs were chosen from the main set (BFDF, RIT–DuPont, BFDB and BFDL) to exhibit mainly lightness

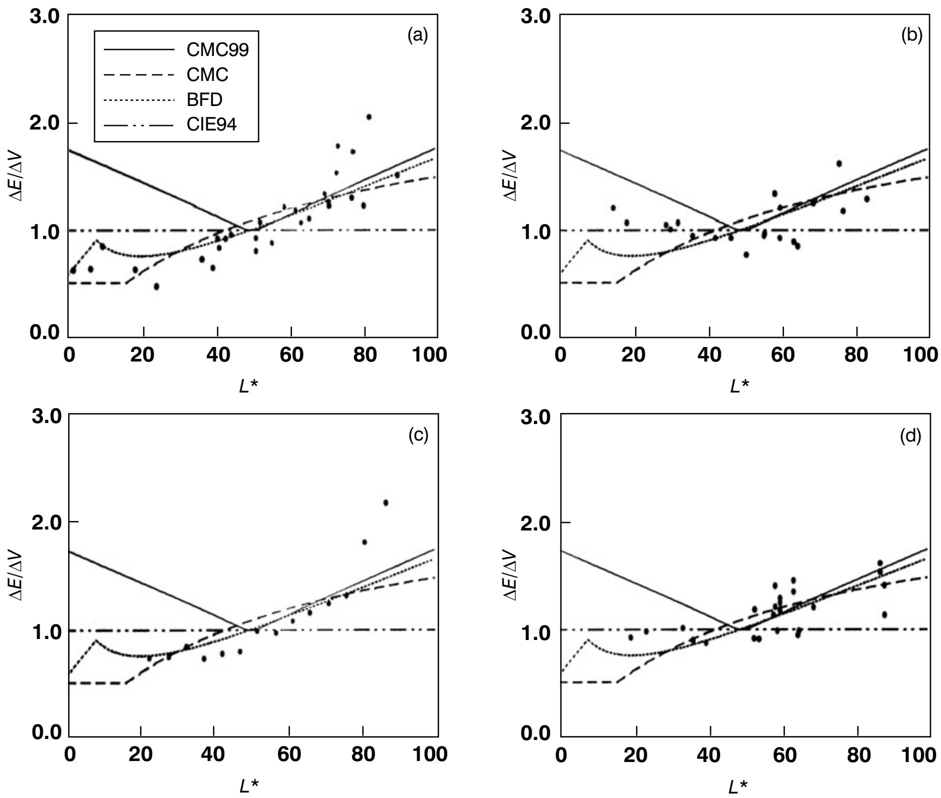

Figure 1 A plot of $\Delta E/\Delta V$ values against L^* scale for (a) BFDF–*L*, (b) RIT–DuPont–*L*, (c) BFDB–*L* and (d) BFDL–*L* data sets; and a plot of the lightness weighting functions of the CMC99, CMC, BFD and CIE94 equations

differences, i.e. $|\Delta L^*|/\Delta E_{ab} > 0.9$. The BFDF–$L$ data set was used to derive the BFD lightness scale and the RIT–DuPont data set was used to derive the CIE94 formula, which has the same lightness scale as the CIELAB L^* scale. The BFDF and RIT–DuPont experiments used glossy paint samples for visual assessments, while textile samples were used for the BFDB and BFDL data sets. The plots in Figure 1 are designated 'trend diagrams'. The data points plotted are not scaled. Hence, it is important to compare their trends, not their magnitudes. In addition, the lightness weighting functions for the CMC99 (see later), CMC, BFD and CIE94 formulae are also plotted in the trend diagrams. It can be seen that the four lightness weighting functions are considerably different. The CMC and BFD functions agree reasonably well except for very dark samples and both imply that the CIELAB L^* scale gives too small ΔL^* values for lightness differences for dark samples and too large ΔL^* values for light samples. The CMC99 equation shows a completely different trend from those of BFD and CMC for dark samples.

For perfect agreement between the visual results and CIELAB ΔE, or ΔL^* values, all data points should lie on the horizontal line with $\Delta E/\Delta V$ equal to a constant. Figures 1a–d show quite different trends, i.e. the BFDF–L and BFDB–L data sets agree well and the trend is reasonably predicted by the CMC and BFD functions. In each case the $\Delta E/\Delta V$ values are lowest for the darkest pairs and gradually increase as L^* increases. However for the BFDL-L and RIT–DuPont–L data sets, there is no evidence to show a decrease of $\Delta E/\Delta V$ value for the darker region ($L^* < 50$). The RIT–DuPont–L data seem to indicate an opposite trend, i.e. an increase of $\Delta E/\Delta V$ values towards the very dark region. This trend is better predicted by the CMC99 function. A series of questions are raised from the above results: which data set represents the true visual results; which formula should be used for industrial applications; and are these discrepancies caused by variations of viewing parameters such as textures (textile and paint), magnitudes of colour difference or methods of assessment? It must be emphasised that the differences between the various functions are large. For a series of pairs which the L^* scale predicts to have the same visual differences, the other scales predict that the visual differences could vary by a factor of two or more.

With this in mind a research project, funded by the Society of the Dyers and Colourists and managed by the University Research Working Group of the Colour Measurement Committee (CMC), was conducted to test different lightness scales. The project was carried out at the University of Keele (Keele) in 1998 [13] and the Colour & Imaging Institute (CII) at the University of Derby in 1999. This article summarises the results obtained from these projects.

A CIE Technical Committee, TC 1-47 on Hue and Lightness Dependent Correction to Industrial Colour Difference Evaluation, has been working since 1998. In fact, the CMC and this CIE TC have been working closely together in order to recommend a new colour difference formula for all industrial applications. It is important for this committee to provide strong evidence to show a significantly better performance for the new formula compared to the existing colour difference formulae. Otherwise, a new equation will be considered to be unnecessary. In the CIE Division 1 Meeting at Warsaw in 1999, evidence was provided showing that there is a need to standardise a new CIE colour difference equation. The basic structure of the formula was determined. It will be again based upon the modification of CIELAB by including the lightness, chroma and hue weighting functions. However, these functions are dissimilar to those used in the CMC, CIE94, BFD and LCD [14] formulae except for the chroma weighting function, in which the CIE94 function was adopted. (The LCD formula was developed by Kim and Nobbs at the University of Leeds.) In addition, an interactive term between chroma and hue differences similar to those in the BFD and LCD formulae will be introduced. The final full equation was finalised in 2000 and was named as CIEDE2000 [15]. The equation is considerably better than either CMC or CIE94, particularly for some areas of colour space, and will lead to one formula being adopted as standard for all surface colour applications.

A new lightness difference formula, designated CMC99 in this article, is given in Eqn 1:

$$\Delta L = \Delta L^* / S_L$$

$$\text{where } S_L = 1 + \frac{0.015(L^* - 50)^2}{\sqrt{20 + (L^* - 50)^2}} \tag{1}$$

where S_L is a lightness weighting function. Nobbs fitted the above formula using a number of experimental data sets including the present one [16]. In fact, the present data set played a large part in the development of the above formula. This formula has been accepted by the CIE TC 1-47 as the lightness equation of the new CIEDE2000.

In this paper, the Keele and CII results are described and discussed in terms of observer repeatability and accuracy, as well as trend diagrams. Different lightness difference formulae were tested using the present data set. A number of lightness weighting functions were also derived to fit the present results from different phases. One of the functions, designed as CII in this article, was selected as giving the most accurate predictions to the combined data set (CII–Keele). In addition, the CMC99 formula also accurately predicted not only the present data, but also the majority of the earlier data sets.

Experimental

Sample preparation

Separate experiments were carried out at Keele and CII. The glossy and matt sets of Munsell neutral paint samples kindly supplied by GretagMacbeth were used. The two sets represent a typical texture difference. The glossy set had a Munsell Value (V) ranging from 0.25 to 9.5 and the matt set from 2.0 to 9.5. Both sets of samples had a 0.25 V interval. Note that the Munsell value corresponds roughly to $L^*/10$. Thus the matt samples had L^* values of approximately 20 to 95. A number of sample pairs were chosen to cover a wide lightness range and were divided into three groups according to the magnitudes of the lightness differences: small, medium and large colour differences corresponding to the CIELAB ΔL^* values averaging 2.7, 5.3 and 10.0, respectively. The $|\Delta L^*|/\Delta E_{ab}$ value for each pair was at least 0.96, showing that the main difference was due to the lightness component. Each experiment was carried out under a different VeriVide viewing cabinet with a neutral background and a D65 simulator. Table 1 summarises the six experimental phases studied.

Table 1 Summary of experimental phases

			CII		Keele	
Phase	Texture	Magnitude	Pairs	Obs	Pairs	Obs
GS	Glossy	Small	20	400	36	720
GM	Glossy	Medium	18	360	17	323
GL	Glossy	Large	16	320	13	247
MS	Matt	Small	14	280	26	520
MM	Matt	Medium	13	260	14	266
ML	Matt	Large	11	220	10	190
Total			92	1840	116	2266

Two separate sets of neutral samples were used in the Keele and CII experiments. The sample pairs selected were also different but followed the same principle. Four and ten observers were used for the Keele and CII experiments, respectively, all of which had passed the Ishihara Colour Vision Test. Nineteen to twenty assessments were made for each pair. Each Keele observer assessed each pair four or five times and each CII observer assessed each pair twice. The sample sizes used were also different, 3 by 3 inches and 4 by 4.5 inches for the CII and Keele experiments, respectively. The L^* of the backgrounds used in two experiments were 40 and 50 for the Keele and CII experiments, respectively.

Colour measurement

A Datacolor SF500 and a GretagMacbeth sphere-based spectrophotometer were used for measuring the Keele and CII samples, respectively. The measuring conditions used were large aperture and specular excluded. It is important to apply the specular excluded condition rather than that of included because it agrees better with visual assessments [17], especially for very dark samples. In a later stage, four spectrophotometers were used to measure the CII samples to investigate the inter-instrument agreement. These are the: Datacolor SF500 (d/8) (S500), GretagMacbeth CE7000A (d/8) (CE7000A), XRite-938 (45/0) (XR-938) and GretagMacbeth CE741 multi-angle (CE741) spectrophotometers. The latter includes four specular angles: 25, 45, 75 and 110°, but the 45° angle was the only one used in this work.

Grey scale method

The grey scale scaling method [12,18,19] was used for both experiments. (The grey scales used here are different from those used for assessing fastness of colour in industry.) The number of samples in the grey scales and their grade numbers were different between the two experiments, but the principle is the same. Only the grey scale used in the CII experiment is described in detail here. Eight neutral matt samples having the same size as the experimental samples were chosen to form a grey scale. This scale was used to assess both the glossy and matt sample pairs. The CIELAB $L^*a^*b^*$ values for each grade and standard under CIE D65 illuminant and 1964 standard colorimetric observer are given in Table 2, together with ΔL^* and ΔE values calculated between each grade and the standard. The results show that the ΔL^* values closely agree with the ΔE values. This indicates that all differences were essentially lightness differences. Figure 2 illustrates the sample arrangement. Each observer was asked to provide the visual results in terms of grade. For example, suppose that a sample pair has a colour difference between two pairs formed by the 'standard' and grade 3 and the 'standard' and grade 4, and is close to the former pair. Hence, the visual result should be in the range from 3.1 to 3.4.

Table 2 CIELAB values of the grey scale used in the CII experiment

No	L^*	a^*	b^*	ΔL^*	ΔE
Standard	38.86	0.06	–0.22		
5	38.94	0.07	–0.23	0.08	0.08
4	41.26	–0.40	–0.34	2.40	2.45
3	43.84	–0.82	–0.32	4.98	5.05
2	47.11	–0.11	–0.64	8.25	8.26
1.5	49.36	0.03	–0.52	10.50	10.50
1	53.59	–0.25	–0.53	14.72	14.73
0	64.19	–0.11	–0.39	25.33	25.33

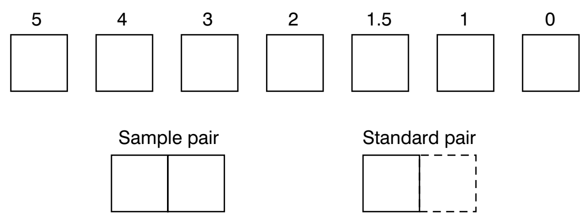

Figure 2 Sample arrangement in the grey scale experiment

For the Keele experiment, each observer was asked to sit in front of the viewing cabinet with a fixed 45/0 (viewing/illuminating) geometry. However, observers were asked to stand in front of the viewing cabinet for the CII experiment. He or she first found a particular viewing angle, at which the sample pair exhibited the largest colour difference, and then made a judgement against the grey scale. This method was used because in the Keele experiment, the adaptation period and viewing angle were found to be important viewing parameters for assessing lightness differences, especially for very dark glossy samples. There was no time constraint for assessing each sample pair.

Data analysis

Performance factor measure

In the present study, many comparisons were made between two sets of data. For example, visual assessments were compared with ΔE values from various colour difference formulae. Earlier studies have shown that applying different statistical measures sometimes can lead to different conclusions. Luo and Rigg [4] devised the performance factor (PF) given in Eqn 2 to overcome this problem by combining four measures (suitably weighted) into one value. The PF also eases the comparison since one value is obtained, and avoids making a decision as to which of the measures is the best. PF is denoted as:

$$\mathrm{PF} = 100 \times [\gamma + V_{\mathrm{AB}} + CV/100 - r] \tag{2}$$

where r represents the correlation coefficient, CV represents the coefficient of variation, γ proposed by Alder *et al*. and V_{AB} derived by Schultz [21], and are each respectively given in Eqns 3 to 6:

$$r = \frac{N\sum(X_iY_i) - \sum X_i \sum Y_i}{\sqrt{\left[N\sum X_i^2 - \left(\sum X_i\right)^2\right]\left[N\sum Y_i^2 - \left(\sum Y_i\right)^2\right]}} \tag{3}$$

$$CV = \frac{\sqrt{\frac{1}{N}\sum(X_i - fY_i)^2}}{\bar{X}} \times 100 \quad \text{where } f = \frac{\sum X_iY_i}{\sum Y_i^2} \tag{4}$$

$$\log(\gamma) = \sqrt{\frac{1}{N}\sum\left(\log\left(\frac{X_i}{Y_i}\right) - \overline{\log\left(\frac{X_i}{Y_i}\right)}\right)^2} \tag{5}$$

$$V_{\mathrm{AB}} = \sqrt{\frac{1}{N}\sum\frac{[X_i - (FY_i)]^2}{X_iFY_i}} \quad \text{where } F = \sqrt{\sum\frac{X_i}{Y_i}\Big/\sum\frac{Y_i}{X_i}} \tag{6}$$

where N is the number of pairs, and X_i and Y_i are values for pair i. When evaluating the goodness of fit of a colour difference formula using the above measures, the X and Y sets contain the visual results and the colour difference values calculated from a particular formula, respectively.

However, previous studies [18,19] found that in some cases the correlation coefficient r was quite inconsistent with the other measures and it was thus decided to use only the remaining three measures. The PF/3 is calculated using Eqn 7.

$$\mathrm{PF}/3 = 100 \times [(\gamma - 1) + V_{\mathrm{AB}} + CV/100]/3 \tag{7}$$

Some idea of the meaning of these measures can be obtained by considering simple hypothetical cases. For no errors, $CV = 0$, $V_{\mathrm{AB}} = 0$, $\gamma = 1$. Suppose that X and Y are the same size [on average, i.e. f (Eqn 4) and F (Eqn 6) are equal to 1], the errors are all 0.5 and the mean of X is 5, with a fairly small range of X (and Y). Then $CV = 10$, $\gamma = 1.1$ and $V_{\mathrm{AB}} = 0.1$. If the errors are doubled, $CV = 20$, $\gamma = 1.2$ and $V_{\mathrm{AB}} = 0.2$. Thus $CV = V_{\mathrm{AB}}$ x 100 and also $(\gamma - 1)$ x 100, and CV corresponds to the percentage error. The correspondence is not exact, particularly when the errors are not constant and the range of X is wide. Combining these three measures as in Eqn 7 weights each correctly and gives a result corresponding in magnitude to CV, i.e. very roughly gives a typical error as a percentage.

Note that it is desirable to consider errors in percentage terms in this type of work. If one colour difference formula is identical to a second, except that it gives ΔE values twice as big, the absolute error in ΔE doubles, but the percentage error remains the same, consistent with the fact that the equations are of equal merit.

Conversion of grey scale ratings to visual differences

For experiments conducted using the grey scale method, the raw data in Grade (G) units were transformed to the visual difference for each pair, ΔV, using Eqn 8 for CII experiment.

$$\Delta V_i = 25.347 - 13.692G_i + 3.116G_i^2 - 0.2781G_i^3 \tag{8}$$

The coefficients in Eqn 8 were obtained by fitting a third-order polynomial between the ΔE and Grade values in Table 2. Figure 3 shows the relationship of grades (G) and CIELAB ΔE (or ΔV) values. It clearly shows that there is good fit of Eqn 8 to the data. Each observer's Grade value was transformed and the arithmetic mean for each pair was calculated. The mean values were used to test various colour difference formulae.

Results

Observer variations

The PF/3 measure is used to indicate the disagreement between observers in terms of repeatability and accuracy. It was computed by using two sets of visual results as X_i and Y_i and setting f and F to one in Eqns 4 and 6, respectively. This assumes that all observers judged colour differences based on the reference grey scale.

Observer repeatability

As mentioned earlier, panels of four (Keele) and ten (CII) observers assessed each pair, with each observer assessing

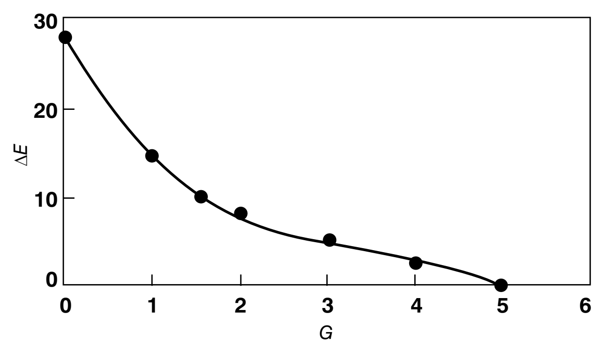

Figure 3 Grades (G) plotted against CIELAB ΔE values

each pair four to five (Keele) or two (CII) times. The PF/3 measure was calculated between the two repeated sessions for each of the observers in the CII experiment. The PF/3 measure was again used to calculate all possible pairs between the four to five sessions for each Keele observer. The mean value was then calculated. Finally, the mean value in each phase was calculated to represent typical repeatability performance. The mean PF/3 value for each phase is listed in Table 3 together with the maximum value.

Table 3 Summary of observer variations for each phase in terms of PF/3 values

	Keele		CII	
Phase	Mean	Maximum	Mean	Maximum
Repeatability				
GS	33	48	47	55
GM	25	34	41	57
GL	23	37	38	54
MS	41	53	53	64
MM	22	28	40	70
ML	19	28	40	67
Mean	27		41	
Accuracy				
GS	24	33	39	54
GM	43	58	33	46
GL	21	29	27	37
MS	29	34	39	67
MM	38	52	36	66
ML	17	20	28	40
Mean	29		35	

From Table 3 it can be seen that the Keele results are more repeatable than the CII results, the mean values being 27 and 41, respectively. Some systematic patterns were found, i.e. the observer repeatability improves with the larger colour difference magnitudes. The worst performance is for the MS set (matt samples with small colour differences). The results for one or two of the CII observers seemed particularly erratic. The mean results of those including all observers and those excluding the least reliable observers were plotted. The two sets still showed a very good agreement. It was decided not to exclude any of the results.

Observer accuracy

The PF/3 measure was also calculated between each individual observer's and the mean visual results to represent observer accuracy. For each CII observer, the mean between two repeated sessions was calculated and the mean of four to five sessions for each Keele observer. The results are also given in Table 3 in terms of the mean and maximum PF/3 values. The mean value of 35 PF/3 units for the CII observers agrees well with those found by Guan and Luo [18] and Kuo and Luo [19]. The overall accuracy performance is considered to be typical for assessments by the grey scale experiment.

The accuracy of 30 to 35 PF/3 units provides a base-line for the development of the lightness difference formula, i.e. a reliable formula should predict the visual results to better than 35 PF/3 units.

Comparing visual results using trend diagrams

The trend diagram is used to compare the visual results from different phases. These are given in Figure 4a–l in the 12 trend diagrams for the different phases/experiments. In each case $\Delta E/\Delta V$ is plotted against L^* where ΔE is calculated from the CIELAB colour difference formula and ΔV is the mean of the ΔV values obtained from Eqn 8. The lightness weighting functions for CMC99, CMC and BFD are also plotted together with that of CIE94 or CIE L^*, which has a constant $\Delta E/\Delta V$ value of one. The data points plotted are not scaled. It is important to compare their trends, not their magnitudes.

In comparing the trends between the different phases, it can be seen clearly that the visual results agree consistently for all phases except for those matt phases in the Keele experiment (see Figure 4h,j and l). The main trend can be fitted with a U or V shape curve indicating that the smallest lightness differences occurred for the darkest and lightest regions and the largest difference around L^* of 50. (For each set, the ΔL^* and ΔE values are roughly constant. Hence, high values for the $\Delta E/\Delta V$ correspond to low values for the visual differences ΔV.) This could be caused by the 'crispening effect' [22], in which a grey background enhances the sensitivity of the observer to lightness differences, between grey samples of about the same lightness as that of background. Alternatively, the observed trend could simply reflect deficiencies in the CIELAB L^* scale.

Comparing the lightness weighting functions in Figure 4, it can be seen that CMC99 gave a close fit for almost all the sub-data sets. The functions of CMC and BFD predicted well for lightness differences in the lighter regions with L^* greater than 50, but give poor predictions in the darker regions.

The trend for the matt/Keele phases shows a different pattern from that of the others, i.e. for lighter lightness differences the perceived differences are smaller than those of darker ones. However, there are larger scatters in these diagrams and a lack of sample pairs covering dark regions due to the nature of matt samples. The lowest L^* value for the matt samples was 20. The difference could be caused by the different experimental techniques used in the Keele and CII experiments. The CII experiment allowed observers to select an angle which exhibited the largest colour difference, while the Keele observers were instructed to make the judgement with a fixed viewing angle. The results also show larger scatters of data points for small-magnitude phases (Figures 4a,b,g and h) than those the other phases, and for Keele phases than CII phases.

Because of the great similarity of trends between the data sets with different magnitudes of colour differences, it was decided to first combine three colour difference magnitude data sets to form four combined data sets, i.e. Keele–glossy (KG), Keele–matt (KM), CII–glossy (CIIG) and CII–matt (CIIM). Subsequently, combined glossy (ALLG) and combined matt (ALLM) data sets were also formed by merging the Keele and CII experimental results. Finally, all data were combined and these were designated as the CII–

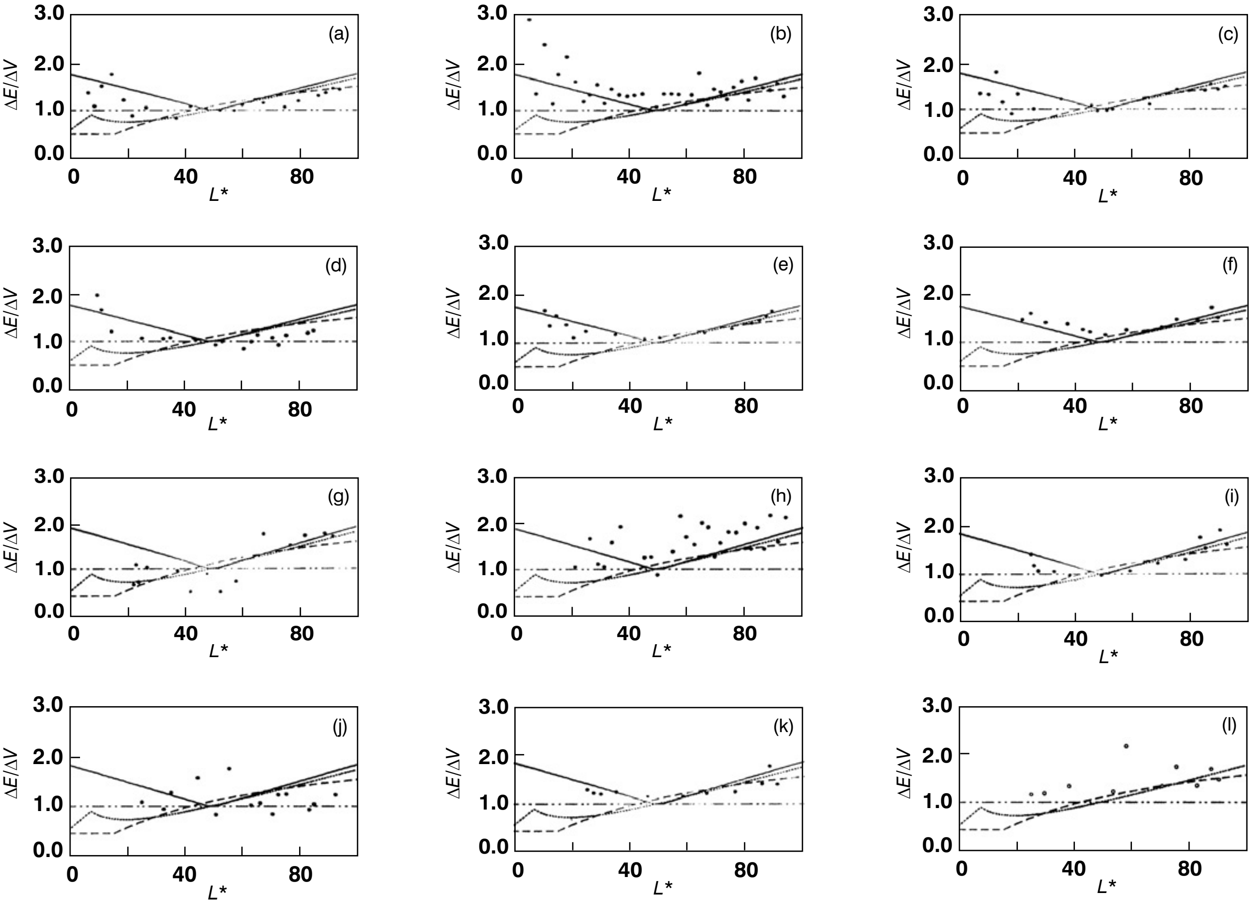

Figure 4 A plot of $\Delta E/\Delta V$ values against L^* scale for (a) CII–GS, (b) Keele–GS, (c) CII–GM, (d) Keele–GM, (e) CII–GL, (f) Keele–GL, (g) CII–MS, (h) Keele–MS, (i) CII–MM, (j) Keele–MM, (k) CII–ML and (l) Keele–ML data sets; and a plot of the lightness weighting functions of the CMC99, CMC, BFD and CIE94 equations; for key see Figure 1

Keele data set. A second-order polynomial curve was fitted to the experimental data in each combined phase. The general equation is as defined in Eqn 9:

$$\Delta L = \Delta L^* / S_L$$
$$\text{where } S_L = a_0 L_m{}^2 + a_1 L_m + a_2 \qquad (9)$$

where L_m is the geometric mean of the L^* values of the two samples considered, a_0 and a_1 are the second- and first-order terms, respectively, and a_2 is the offset term.

Table 4 summarises the coefficients for each combined phase and from this table it can be seen that all coefficients for all data sets are quite similar, except for those for the KM and ALLM data sets. This is consistent with Figure 4, i.e. a different trend can be seen for the matt phases in the Keele experiment. It is encouraging to find that the coefficients for the CII–Keele data set agrees well with those of the other data sets except for KM and ALLM.

Table 4 Coefficients of the second-order polynomials for combined data sets where a_0 and a_1 are the second- and first-order terms, respectively, and a_2 is the offset term

Data set	a_0	a_1	a_2
KG	0.000328	–0.0367	2.17
CIIG	0.000244	–0.0237	1.60
ALLG	0.000258	–0.0271	1.82
KM	–0.000045	0.0099	0.95
CIIM	0.000240	–0.0193	1.35
ALLM	0.000043	0.0020	1.00
CII-Keele	0.000236	–0.0233	1.70

Figures 5a–d are the trend diagrams for the KG, KM, CIIG and CIIM data sets, respectively, and Figure 6 for the CII–Keele data set. For all diagrams, the second-order polynomial fitted to the CII–Keele data set together with a line with $\Delta E/\Delta V$ equal to one are also plotted. These diagrams show that the polynomial and CMC99 functions fitted to the CII–Keele data set fits quite well for almost all data sets. The worst fit is naturally for the KM and ALLM data sets. Even for these sets the fit can be considered to be quite acceptable due to the large scatter and lack of any clear trend. The worst fit can generally be found at low L^* values. This is due to poor instrument repeatability in these regions (see later).

Hence, the equation fitted to the CII–Keele data set can form the basis of a new lightness difference formula, designated CII in the following discussion, and is given as Eqn 9 where the value of S_L is defined as in Eqn 10:

$$S_L = 2.4(L_m^*/100)^2 - 2.4(L_m^*/100) + 1.7 \qquad (10)$$

where L_m^* is the geometric mean of the L^* values for the two samples considered.

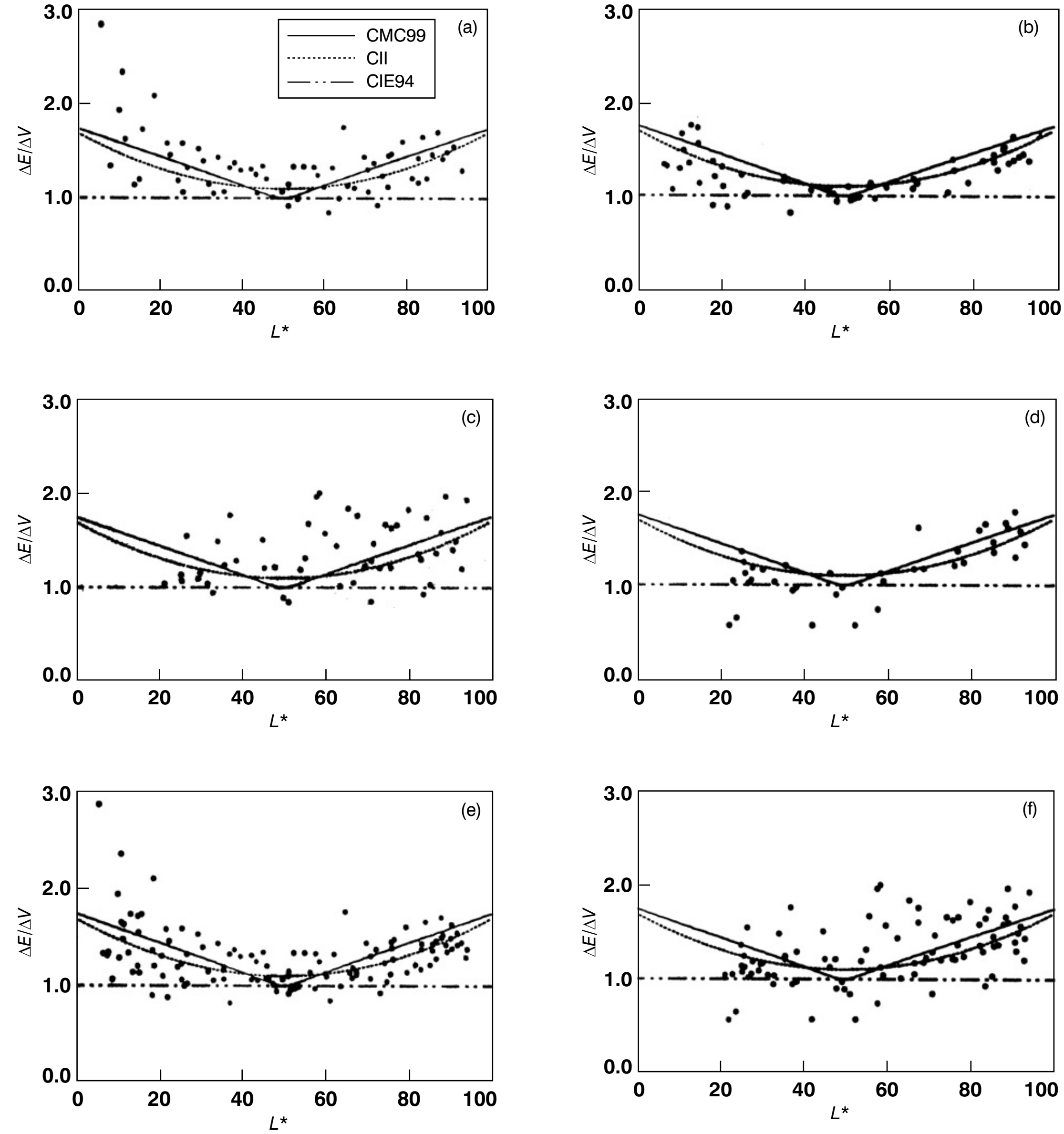

Figure 5 A plot of $\Delta E/\Delta V$ values against L^* scale for (a) KG, (b) CIIG, (c) KM, (d) CIIM, (e) ALLG and (f) ALLM data sets; and a plot of the lightness weighting functions of the CMC99, CII and CIE94 equations

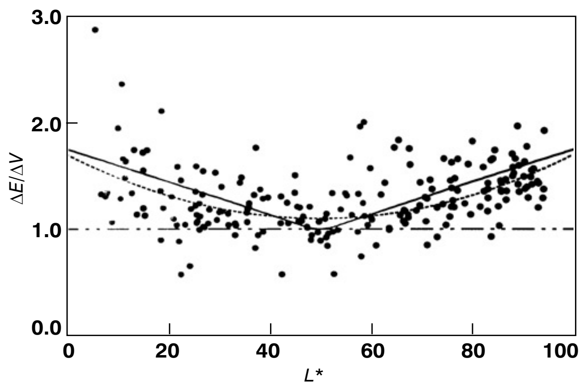

Figure 6 A plot of $\Delta E/\Delta V$ values against L^* scale for CII–Keele data set; and a plot of the lightness weighting functions of the CMC99, CII and CIE94 equations; for key see Figure 5

Note that, except for very large differences, almost identical results are obtained if the L^* value for a standard sample is used instead of L_m^* for the derivation of Eqn 10.

The coefficients in Eqn 10 are slightly different from those of CII–Keele data set in Table 4. It is a simplified equation with a maximum S_L value (1.7) at L^* of 0 and 100 and a minimum value (1.1) at L^* of 50. The CMC99 function is also plotted in Figure 6. It can be seen that this also gives a reasonably good fit to the CII–Keele data set. The CII and CMC99 functions are very similar in shape.

Testing lightness difference scales

The PF/3 measure is used to indicate the disagreement between visual differences (ΔV) and predicted differences (ΔE) calculated from a particular colour difference formula. The f and F in Eqns 4 and 6 were calculated to adjust ΔV values to the same scale as the ΔE values. This is necessary because different colour difference formulae have different magnitudes. The CMC formula for example gives ΔE values which on average are smaller than those from the CIELAB formula.

The Keele and CII experimental results were used to test six colour difference formulae: CMC99, CMC, BFD, LCD,

Table 5 Testing colour difference formulae performance using the results from each individual phases and combined phases[a]

Data set	PF/3 value GS	GM	GL	Mean	MS	MM	ML	Mean
CII Data sets								
CMC	52	49	48	50	36	28	28	31
CIE94	19	19	15	18	41	20	11	24
BFD	30	28	28	29	30	21	22	24
LCD	23	21	20	21	**27**	14	16	19
CMC99	15	14	9	13	33	**10**	9	**17**
CII	**13**	**13**	**7**	**11**	35	**11**	**6**	17
Keele data sets								
CMC	55	58	25	46	22	29	**18**	23
CIE94	22	20	11	18	22	**21**	**18**	**20**
BFD	37	35	25	32	23	33	23	26
LCD	31	28	18	26	**21**	32	24	26
CMC99	**19**	**13**	**7**	**13**	**21**	29	24	25
CII	**19**	**13**	**7**	**13**	**21**	25	21	22

Data set	PF/3 value CIIG	KG	ALLG	CIIM	KM	ALLM	CII-Keele
Combined data sets							
CMC	52	51	53	33	27	30	46
CIE94	19	22	21	28	**25**	27	24
BFD	30	37	34	27	30	34	32
LCD	23	29	27	**22**	29	27	27
CMC99	14	19	18	**22**	29	27	22
CII	**12**	**18**	**17**	**22**	27	**26**	**21**

a The formula giving the best performance is in bold for each data set

CIE94 and the newly derived lightness difference formula (CII). The CIELAB and CIE94 formulae use the same lightness scale (L^*). Hence, only CIE94 was tested. The results are summarised in Table 5. For all the formulae tested except for BFD, their weighting functions were calculated using geometric mean of the L^* values of the two samples considered. This is necessary due to the involvement of the medium (Munsell value difference of 0.5) and large (Munsell value difference of 1.0) lightness differences. If mean values are not used, the performance for medium and large difference data sets will be much worse. (The BFD formula's lightness scale was derived by Coates *et al*. using the BFDF–*L* data set [9]. Because of the way in which the scale is structured, there is no need to use the mean value.) In Table 5, the PF/3 value for the formula giving the best performance is printed in bold for each data set.

In testing the equations' performance using the CII data sets, the CII and CMC99 formulae generally outperformed the others by a large margin and the CMC formula gave the worst prediction for all phases except for the MS phase. The performance of the CMC99 is close to that of the CII, i.e. always within 3 PF/3 units. Almost all formulae gave a similar performance in predicting the CII MS phase results due to the larger scatter (see Figure 4g). When comparing the equations' performance using the Keele data sets, again the CMC99 and CII formulae predicted more accurately than the other formulae except for the MM and ML phases, in which CIE94 performed the best. There is not much difference between all formulae for the results from the matt phases.

For the combined data sets (ALLG and ALLM), the results more or less agree with those found earlier, i.e. the CII and CMC99 formulae performed the best, followed by CIE94, LCD and BFD, with the CMC formula the worst. The trend is more marked for the glossy sets than the matt sets. For the CII–Keele data set, the CII and CMC99 formulae predicted the results more accurately than the third best formula, CIE94, by 2 units. This might seem a small improvement from the CII to CIE L^* formulae, but it must be remembered that the CII and CMC99 formulae outperformed the CIE L^* formula in the majority of the individual and combined phases. They are quite different from the CIE L^* scale. For pairs of samples all with the same difference according to CIE L^*, the CII and CMC99 formulae predict differences varying by up to 50%. Both formulae cannot be correct. The present results consistently support the CII and CMC99 formulae and suggest that the other formulae are not even as good as CIELAB. The main differences between the CII and CMC99 formulae are the predictions for lightness differences where L^* is greater than 100, which frequently occurs for metallic coatings for angles of viewing less than 30° from the normal angle of illumination of the sample. The CII formula could under-estimate the lightness differences in comparison to the CMC99 formula by 40% for a pair of samples around L^* of 150.

Testing lightness formulae using earlier data sets

Five sets of experimental data were accumulated by the CIE TC 1-47 [23] for the development of the new CIE colour difference formula. These were also used to investigate the performance of different lightness formulae. These data sets are RIT–DuPont [10], Witt [24], Bradford Combined Perceptibility (BFD–CP) [12] and the Leeds results obtained using grey scale and pair comparison methods (Leeds–GS and Leeds–PC) [14]. For each data set, the pairs exhibiting mainly lightness differences were selected. The criteria used were: $(\Delta C^*/\Delta E)^2 < 0.25$ and $(\Delta H^*/\Delta E)^2 < 0.25$ (all in CIELAB units), which provide a more strict condition than that used earlier ($|\Delta L^*|/\Delta E > 0.9$). The results in terms of PF/3 are given in Table 6 together with the present

Table 6 Testing colour difference formulae performance using the RIT–DuPont, Witt, Leeds–PC, Leeds–GS, BFD–CP and CII–Keele[a]

Data set	RIT–DuPont	Witt	Leeds–PC	Leeds–GS	BFD–CP	CII–Keele
CMC	34	43	30	28	42	46
CIE94	19	40	33	29	45	24
BFD	22	40	28	29	**41**	32
LCD	18	43	24	24	**41**	30
CMC99	**15**	39	**21**	**20**	45	22
CII	16	**38**	26	23	44	**21**

a The formula giving the best performance is in bold for each data set

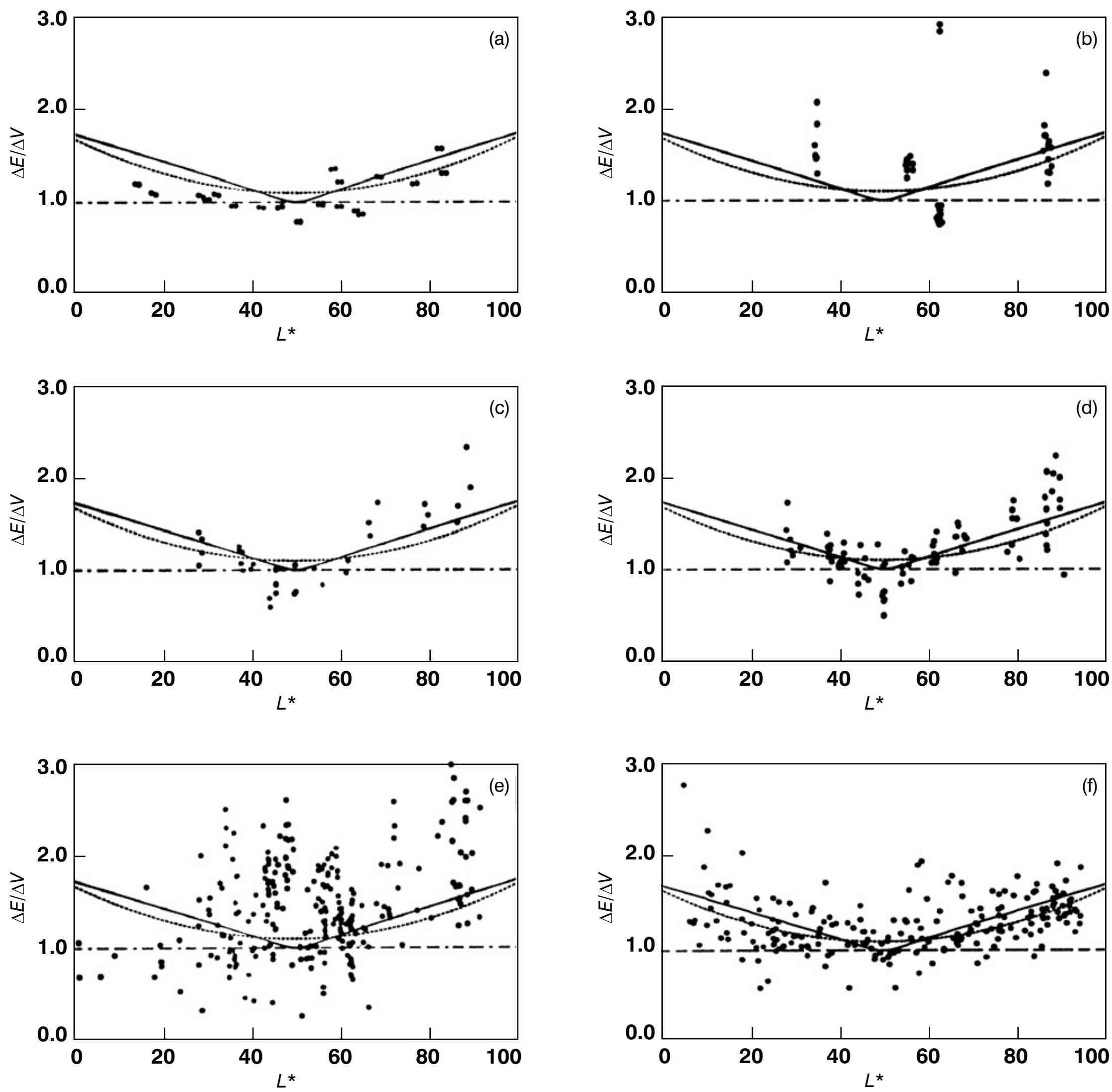

Figure 7 A plot of $\Delta E/\Delta V$ values against L* scale for (a) the RIT–DuPont, (b) Witt, (c) Leeds–PC, (d) Leeds–GS, (e) BFD–CP and (f) the present (CII–Keele); and a plot of the lightness weighting functions of the CMC99, CII and CIE94 equations; for key see Figure 5

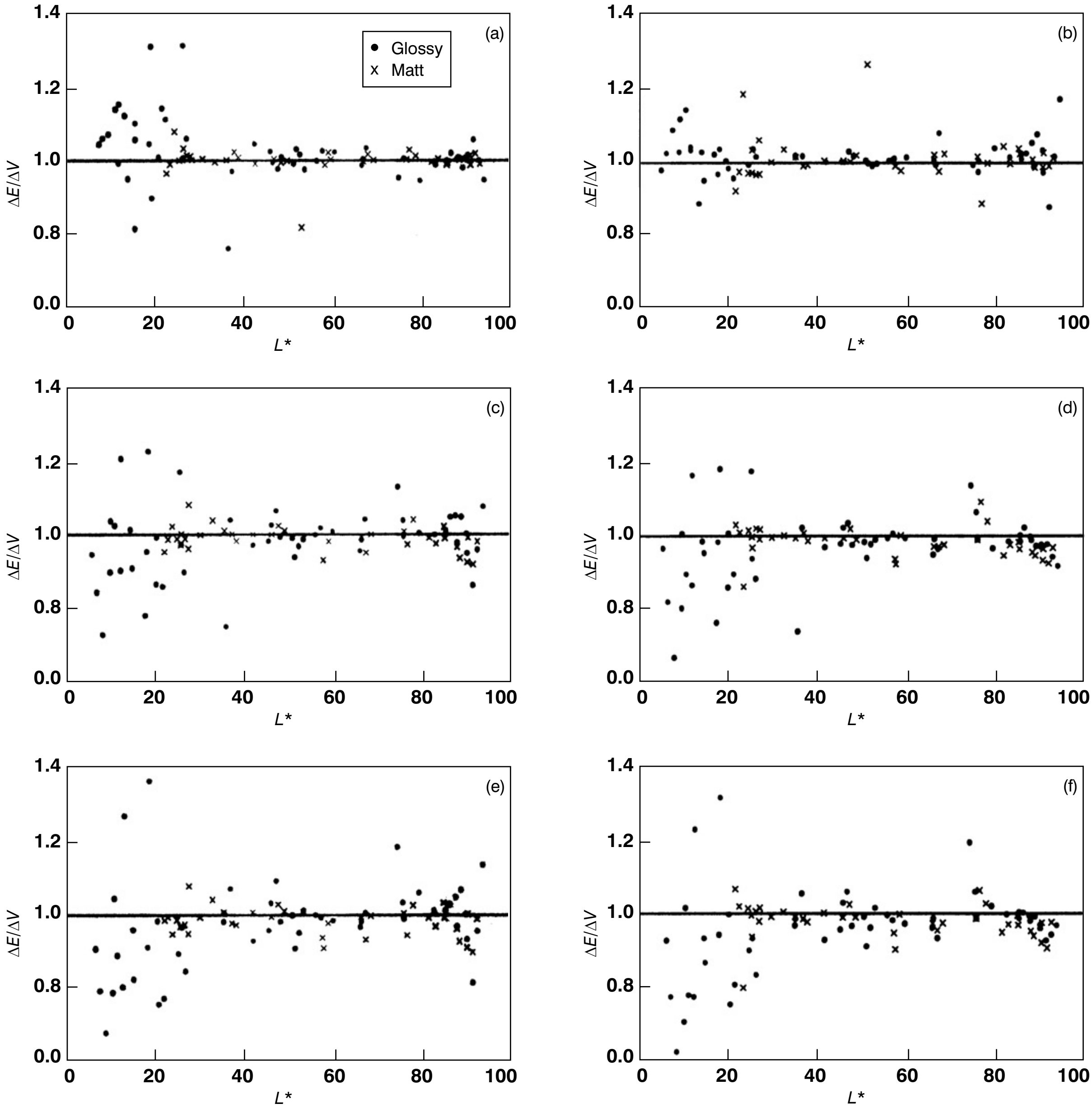

Figure 8 Inter-instrument agreement for measuring lightness differences with (a) SF500 vs CE7000A, (b) XR-938 vs CE741, (c) SF500 vs CE741, (d) SF500 vs XR-938, (e) CE7000A vs CE741 and (f) CE7000A vs XR-938 for glossy and matt samples

combined data set (CII–Keele). In this table the formula giving the best performance is printed in bold for each data set. The trend diagrams also provided in Figure 7 for the (a) RIT–DuPont, (b) Witt, (c) Leeds–PC, (d) Leeds–GS, (e) BFD–CP, and (f) the present CII–Keele data sets, and the lightness weighting functions of the CMC99, CII and CIE94 equations are also plotted.

In Table 6, the results show that the CMC99 and CII formulae gave very similar performance and predicted more accurately than the other formulae for all data sets except for BFD–CP data sets. The BFD–CP data set is the only data set based mainly upon textile materials. Although the results show that the BFD formula gave the best fit to the data, there is a very large scatter in the trend diagram (Figure 7e). It is debatable which formula should be chosen for this data set. Further experiments are being carried out using textile samples to cover the dark areas.

In conclusion, the CMC99 gives the best fit to the majority of data sets. Hence it was adopted by the CIE as the new lightness difference formula [15].

Inter-comparison between different instruments

As mentioned earlier, the CII samples were measured using four different instruments. For the GretagMacbeth multi-angle spectrophotometer (CE741) only the 45/0 geometry results were used in this study. These results were inter-compared. It is well known that it is difficult to measure accurately the dark samples. For each pair of samples, the CIELAB ΔE values were calculated. The ratio between two ΔE values from two instruments for each pair was then calculated. These ratios were plotted against mean L^* value of each pair as illustrated in Figures 8a–f. For a perfect agreement between two instrumental results, these ratios should all equal one. The more deviation from one, the worse the agreement.

For all six comparisons between different instruments, there is a clear trend in that the best agreement occurs for sample pairs with L^* values between 40 and 70, followed by lighter pairs with the darker pairs much the worst. The agreement is better for two instruments having same optical design, i.e. sphere based instruments for SF500 and CE7000A, and 45/0 geometry for CE741 and XR938. The sample pairs were also measured twice using the SF500 to investigate the medium-term repeatability (with a period of 8 h). It was found that the ratios for darker samples ranged between 0.8 and 1.2. This confirms that it is difficult to accurately measure glossy dark samples. These results raise doubt about the reliability of the CII lightness formula derived in the last section, especially for the dark region. Hence, it was decided to use the colour measurement results for each instrument to re-calculate colour differences for all colour difference formulae. Subsequently, these new ΔE values were tested using visual results from each individual and combined phases. The results show that there are very little changes in PF/3 values (± 2 units) in comparison with those in Table 5 and the ranking order for each formulae is also unchanged. In addition, the trend in Figure 6 is still the same.

Conclusions

Two hundred and eight pairs of near-neutral samples exhibiting mainly lightness differences were accumulated. Each pair was assessed about 20 times by a panel of 14 observers using the grey scale method. It was found that the observer accuracy was about 32%, which agreed with previous studies. The trend diagrams revealed that the visual results showed a consistent trend for all phases except those from matt phases of the Keele experiments. The main trend is that a perceived lightness difference is the smallest for the darkest and lightest sample pairs and the largest difference occurs for L^* of 50. The results are consistent with the majority of the earlier work (see Figure 7) in that $\Delta E/\Delta V$ clearly increases as L^* increases from $L^* = 50$. For lower L^* values, the present results clearly show that $\Delta E/\Delta V$ increases again as L^* decreases from $L^* = 50$. All the samples were paint. It is just possible that textile samples might give a different pattern. However, dark textile samples (L^* less than 20) are difficult to prepare and assess. These samples are important for verifying the shape of the lightness weighting function. The CMC is currently conducting new experiments to clarify this.

The combined visual results were used to derive the CII lightness difference formula, and to a large extent, the CMC99 lightness difference formula. Both equations fitted well to all the visual data accumulated in the present and earlier studies. Both new formulae predicted more accurately than the typical deviation of an individual assessment from the mean of a panel of 20 observations. In addition, the CMC99 weighting function also gives reasonable fits to the majority of the previous lightness difference studies. This formula, which was already integrated with the new CIE colour difference equation, outperformed the CMC, CIE94 and BFD equations using the present and the majority of the earlier experimental data sets.

We also found that special attention should be paid to measurements of very dark samples, e.g. to measure these samples more times and take the average, and to clean dirt and finger prints before the colour measurement. The main cause of this is due to poor instrument repeatability and inter-instrument agreement in this colour region. In general, the instruments of the same optical geometry do have better agreement.

* * *

The authors would like to thank the Society of Dyers and Colourists for sponsoring this research work, Mr Heptistall for preparing the experiments at the University of Keele, GretagMacbeth for supplying experimental samples and the many observers for taking part in this study.

References

1. *Colorimetry*, 2nd Edn, CIE Publ. 15.2 (Vienna: CIE Central Bureau, 1986).
2. F J J Clarke, R McDonald and B Rigg, *J.S.D.C.*, **100** (1984) 117.
3. *Industrial colour-difference evaluation*, CIE Publ. 116 (Vienna: CIE Central, 1995).
4. M R Luo and B Rigg, *J.S.D.C.*, **103** (1987) 86.
5. M R Luo and B Rigg, *J.S.D.C.*, **103** (1987) 126.
6. ISO 105-J03:1995 *Textiles – Tests for colour fastness. Part 3: Calculation of colour differences*.
7. ISO/DIS 7724 *Methods of test for paints. Part 6: Determination of colour and colour difference.*
8. B Rigg, *Colour-difference formulae: Recent developments*, Autumn School '95, University of Derby (1995) 87.
9. E Coates, K Y Fong and B Rigg, *J.S.D.C.,* **97** (1981) 179.
10. R S Berns, D H Alman, L Reniff, G D Snyder and M R Balonon-Rosen, *Col. Res. Appl.*, **16** (1991) 297.
11. S A Badu, *Large colour differences between surface colours*, PhD thesis, University of Bradford (1986).
12. M R Luo, *New colour difference formulae for surface colours*, PhD thesis, University of Bradford (1986).
13. D H Kim, *The influence of parametric effects on the appearance of small colour differences*, PhD thesis, University of Leeds (1997).
14. S. Westland, *The perception of achromatic differences*, Colour Science '98, Colour Physics Symposium, Harrogate, UK (1998).
15. M R Luo, G Cui and B Rigg, *Col. Res. Appl.*, in press.
16. J Nobbs, personal communication (1999).
17. B Rigg, *J.S.D.C.,* **97** (1981) 504.
18. S S Guan and M R Luo, *Color Res. Appl.,* **25** (1999) 331.
19. W G Kuo and M R Luo, *J.S.D.C.*, **112** (1996) 312.
20. C Alder, K P Chaing, T F Chong, E Coates, A A Khalili and B Rigg, *J.S.D.C.*, **98** (1982) 14.
21. W Schultz, *Color metrics* (Soesterberg, Netherlands: AIC, 1972) 245.
22. T Kaneko, *Acta Chromatica*, **1** (1964) 51.
23. All experimental data sets accumulated by the CIE TC 1-47 can be obtained by using <ftp//colour.derby.ac.uk/cietc147>.
24. K Witt, *Col. Res. Appl.*, **24** (1999) 78.