Which color differencing equation should be used?

Martin Habekost

Keywords: Colour differencing, Color, CIE, Colorimetry, Differencing, Delta E

Color differencing equations have been in used for quite some time. In 1976 the CIE released the ΔEab-formula, which is still widely used in industry and in research. This formula has its drawback and a number of other color differencing formulas have been issued that try to accommodate how the human observer perceives color differencing in different areas of the color space. Trained and untrained observers in regards to judging color differences were asked to rank color differences of test colors. In both cases the ΔE2000 formula corresponded best with the way both groups of observers perceived these color differences. When industry experts were asked to rank perceived color differences without having a standard to compare to the CMC1:1 formula corresponds well with their observations. Although the ΔE2000 is mathematically more complicated than the ΔEab formula the TC130 is mentioning guideline values (not official standard values) for evaluating process colors in the ISO 12646-2 procedure in its latest draft version. This is an indication that the ΔE2000 formula will soon become the standard color differencing equation and replace the ΔEab formula.

1. Introduction

Color differencing equations have been used for quite some time. In 1976 the CIE published the first internationally endorsed color differencing equation. This formula called ΔEab or ΔE76 deemed a difference or ΔE of 1.0 to be the smallest difference perceivable by the human eye. This formula has been used in many ISO procedures such as 12647-2 for process control in the production of halftone color separations, proof and production prints. This color differencing equation made it possible to better communicate color differences under standard illuminants and observers. The color notation used for this equation was the L*a*b*-color space.

It was soon discovered that this equation had its shortcomings. These shortcomings were, that it was not taken into consideration that the human eye is more sensitive to small colour differences in some regions of the color wheel and less sensitive in others. This means that a ΔE of 1.0 could be a small visible difference in one area of the visible spectrum (i.e. dark blue colors) and a large visible difference in another area (i.e. light pastel type colors).

The introduction of the L*a*b*-color notation by the CIE was done to bring order to the various color notations and color differencing equations that were used (CIE, 1986). The ΔEab equations looks as follows:

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \qquad (1.1)$$

The CIE revised the formula by introducing the ΔE94 formula in 1994. This formula uses the L*C*h*-notation for calculating color differences. The ΔE94-formula looks as follows:

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*}{k_C S_C}\right)^2 + \left(\frac{\Delta h^*}{k_H S_H}\right)^2} \qquad (1.2)$$

This equation has two sets of coefficients. The *k*-coefficients are also known as parametric factors and refer to effects influencing color-difference judgment. The *S*-coefficients account for CIELab's lack of visual uniformity (Billmeyer, 2000).

Although this formula matched closer to the color difference perception of the human eye it lacked some accuracy in the blue-violet region of the color space, which lead to the release of the ΔE2000 formula in 2000. This formula contains a so-called rotational term for the blue-violet region to address the shortcomings of the ΔE94 formula. Since this equation has a deficiency in the blue-violet region a correctional or rotational factor was added. This corrected formula is known as the ΔE2000 equation (CIE, 2001):

$$\Delta E^*_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)} \qquad (1.3)$$

It can be seen from this equation that the LCh has been transformed into L', C' and H'. How the LCh values are transformed into this new notation has been explained in detail (Sharma G. W., 2005). The ΔE2000 formula has five corrections to CIELab: A lightness weighting function ($k_L S_L$), a chroma weighting function ($k_C S_C$), a hue weighting function ($k_H S_H$), an interactive term between chroma and hue differences for improving the performance for blue colors and a factor (R_T) for rescaling the CIELab a*-axis for improving performance for grey colours.

The CIE was not the only body that released a color differencing equation to address the shortcomings of the ΔEab-formula. In 1984 the CMC (Color Measurement Committee of the Society of Dyes and Colorists of Great Britain) (Clarke, 1984) also developed an equation that is based on the L*C*h*-notation of colors.

This equation takes the various color sensitivities of the human visual system into consideration and a ΔE of 1.0 under CMC gives the same visual difference in all regions of the color-wheel.
The formula looks as follows:

$$\Delta E^*_{CMC} = \sqrt{\left(\frac{\Delta L^*}{lS_L}\right)^2 + \left(\frac{\Delta C^*}{cS_C}\right)^2 + \left(\frac{\Delta h^*}{S_H}\right)^2} \qquad (1.4)$$

The S_L, S_C and S_H are the main weighting factors for lightness, chroma and hue. The two factors *l* and *c* are constant and are defined by the user and weight the importance of lightness and chroma relative to the hue of the measured color.

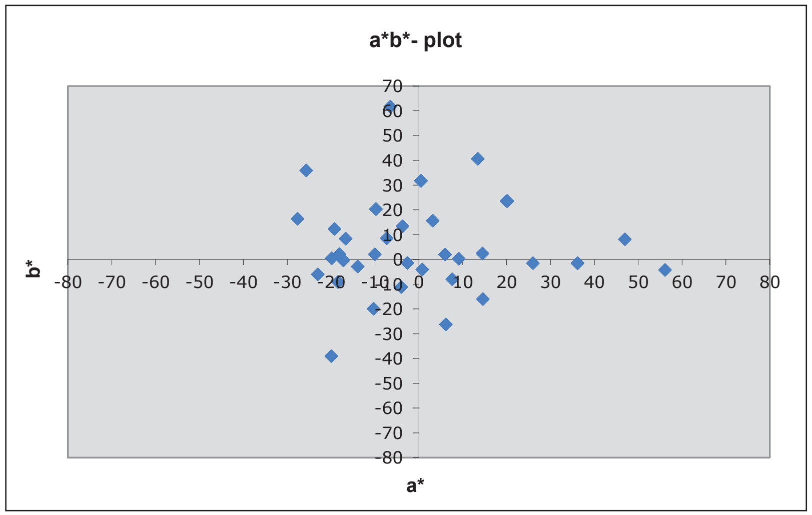

Figure 1: a*,b*-plot of the tested colors.

All these equations try to overcome the drawbacks of the L*a*b-color space by introducing correction terms to the non-uniform L*a*b*-color space.

2. Application of various ΔE-equations

In this paper three individual studies have been combined to achieve a better understanding of the correlation between the ΔE-values that the different color differencing equations produce for the differences between two colors and how the human perception of color differences correlates with these ΔE-values.

2.1 Observations of color differences by untrained observers

The ΔEab equation has been widely used in industry and research, but are the other equations being used or is one more dominantly used and how do the measured color difference correspond with how observers perceive the color differences.

The author of this paper conducted a study (Habekost, 2007) with untrained observers. The observers had to look at 34 different colors and their variations. The variations had a ΔEab of 2, 5, 5.5 and 7 to the tested color. The color patches were 2x2 cm in size. The distribution of the tested colors in an a*,b*-plot can be seen in Figure 1.

The observers viewed the color patches in viewing booth with 5000K lighting. The difference between the standard and the sample patch could be rated into match, slightly different, different, more different and very different. The rankings were translated into numbers from 5 (= match) to 1 (= very different) and a ranking scheme was applied to weight the responses. A typical ranking can be seen in Table 1.

The ΔEab-values were plotted against the total number and the R^2-value obtained. This was done for the ΔE-values from all four equations and all color samples.

A typical plot can be seen in Figure 2.

The R^2-values for these samples for the other equations were 0.96 for ΔE94, 0.94 for ΔE2000 and 0.96 for ΔECMC.

The obtained R^2-values from all color samples and the various ΔE-equations were then plotted against the L, C and h values of all the evaluated colors. This was done to conclude which color differencing equation performed better.

In regards to lightness and chroma ΔE2000 and ΔECMC performed almost in a similar fashion, while in regards to the hue of the tested colors the ΔECMC-equation performed slightly better in the blue region. Basically it can be said that in this study the visual ranking by the untrained observers versus the color difference numbers resulting from the tested ΔE-equations the numbers from the ΔECMC-equation correlate better than those from the ΔE2000 equation. (More details about this can be found in Habekost, 2007.)

2.2. Color patches and trained observers

A similar experiment as mentioned in 3.1 was carried out with industry professionals. This time slightly different colors were chosen to achieve a more uniform distribution of the test colors throughout the color space. The test colors contained neutral as well as intense colors and covered the main color centers. A graphical representation of this can be seen in Figure 3.

Table 1: Ratings and rankings of a color, ΔEab-values

	ΔE: 1.86	Ranking	ΔE: 5.85	Ranking	ΔE: 4.49	Ranking	ΔE: 8.27	Ranking
Match	4	20	0	0	1	5	0	0
Slightly different	11	44	1	4	1	4	0	0
Different	2	6	5	15	8	24	2	6
More different	0	0	5	10	3	6	4	8
Very different	0	0	6	6	4	4	11	11
Total:	17	**70**	17	**35**	17	**43**	17	**25**

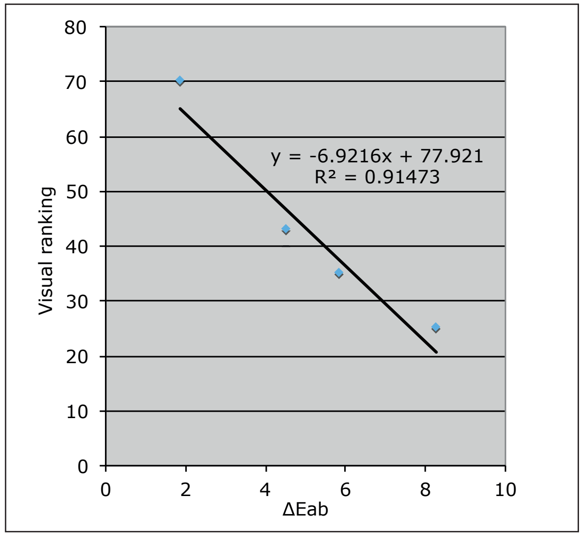

Figure 2: Example of a correlation between color difference ΔEab and rating.

The L*a*b*-values of the test colours were entered into a MatLab worksheet and transformed into L*C*h*-values. These L*C*h*-values were modified in the way that color differences using ΔEab between 2 and 7 resulted. Modifications were applied to each of the three variables individually. For each standard color six variations were generated. This means that there were two variations in regards to Lightness, two in regards to Chroma and two in regards to hue.

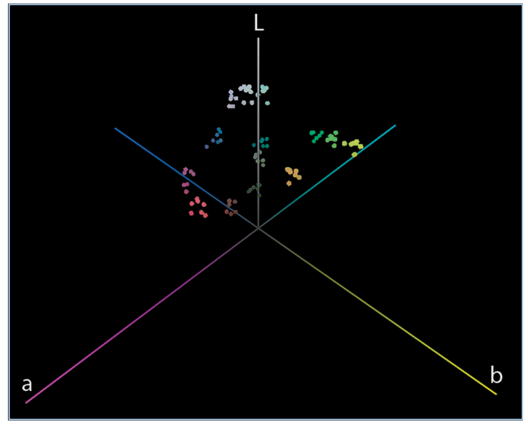

Figure 3: Test colors used for the viewing of the test color chips

2.3 Observations by experienced observers

A set of 14 test colors including neutral colors was chosen for this test. The results from the test have been grouped by Lightness, Chroma and Hue.

From Figure 4 it can be seen that the CMC 2:1 equation and ΔE2000 have the strongest correlation between the perceived and calculated color differences for the darker colors. However this correlation is reduced as the lightness of the samples increases until a lightness values of about 45. All equations show in drop in correlation between observed differences and numerical color differences in the lightness area between 40 and 60. This is the area were most of the colors are present in CMYK color gamut of an offset press. After this the correlation increases again, but the ΔE2000 equation shows truly a better correlation than all the other equations. The correlation drops down to about 60%, which is not a very strong correlation. Only the ΔE2000 equation shows a better correlation after a lightness value of 45. It is also interesting to observe that the DIN99 equation did not perform better than the ΔE2000 equation, although one could be under the impression that that could be the case based on the logic that went into the creation of the DIN99 equation.

Figure 5 shows the performance versus the Chroma of the test colors. The interesting part of this Figure is,

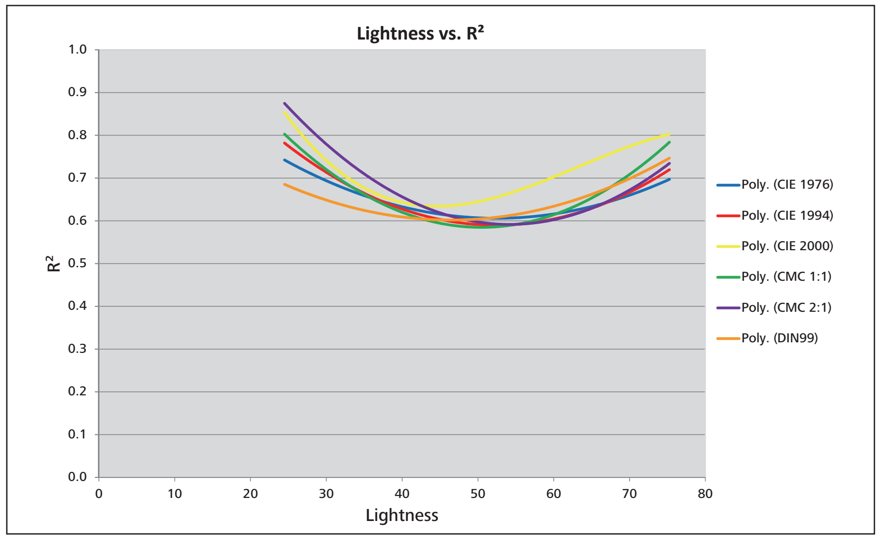

Figure 4: Correlation of calculated vs. observed differences in relation to the Lightness of values of the samples. All lines are 2nd degree polynomial.

that the ΔE2000 equation correlates well between the perceived differences and the calculated differences in the low chroma area from 5 to 15 (low intensity colors) and also well in the area of the high intensity colors. It is only the ΔE2000 equation that shows a small increase in correlation after a chroma value of about 30. As could also be seen in the Figure 4 is that the CMC-equations also show a very good correlation between the perceived and the calculated color differences, but the ΔE2000 equation performs slightly better.

The correlation between observed and calculated differences in relation to the hue of the test colors is

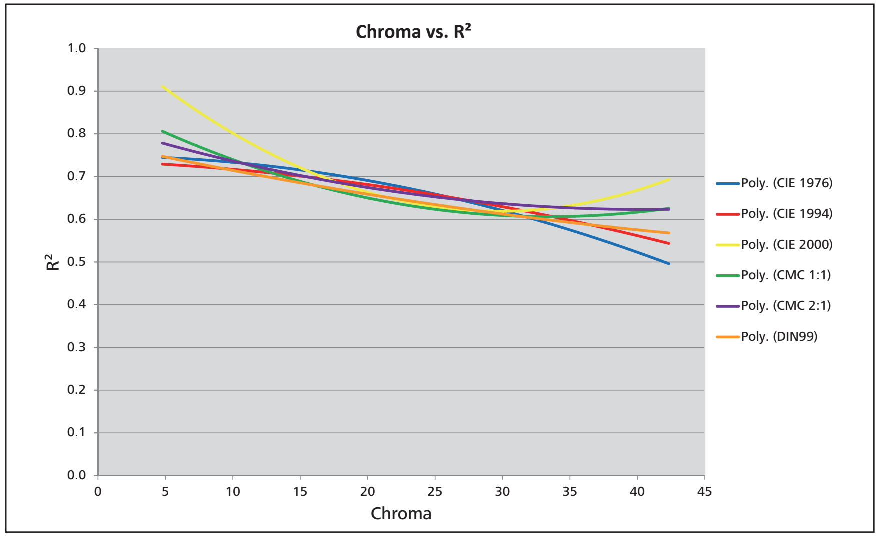

Figure 5: Correlation of calculated vs. observed differences in relation to the Chroma of values of the samples. All lines are 2nd degree polynomial.

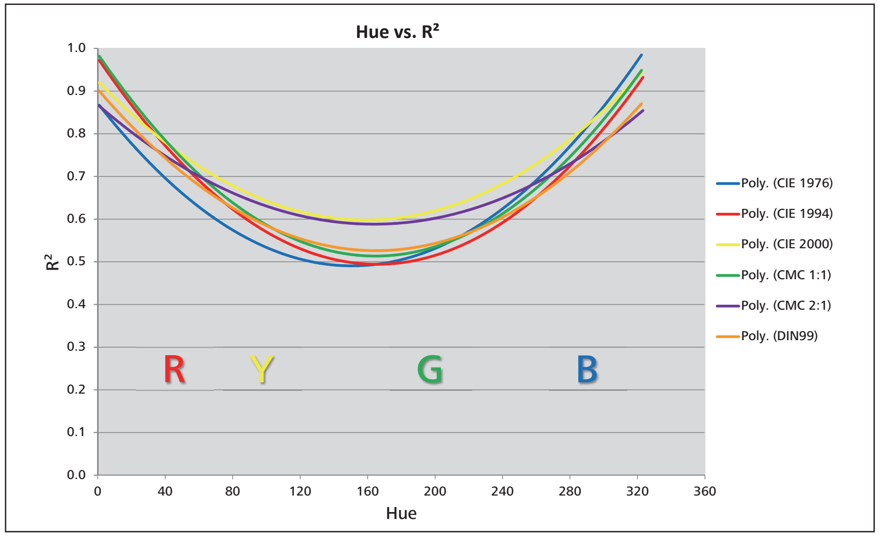

Figure 6: Correlation of calculated vs. observed differences in relation to the Hue of values of the samples. All lines are 2nd degree polynomial.

shown in Figure 6. It gives an interesting insight on how the perceived and the numerical differences correlate. The CMC equations surpasses the correlation between the perceived and the calculated differences of the ΔE2000 equation in the reddish area of the tested colors, but for the majority of the tested colors the ΔE2000 equation shows definitely a better correlation. It needs to be noted that around a hue angle of 180 degrees (green) the correlation of all investigated color difference equations is quite poor. A correlation of 50% is not really a correlation. Only the ΔE2000 and the ΔECMC (2:1) equation show a slight correlation between observed and numerical differences.

From these three figures it can be observed that the ΔE2000 equation performs decent in regards to the Lightness, Chroma and Hue of the tested colors and their perceived differences. This is an important finding of this test, especially if the connection between this test and the test with the inexperienced users is drawn. In both cases it was the ΔE2000 equation that gave the best results in regards present and perceived color differences.

For the evaluation of the possible correlations between perceived and numerical differences using the various color differencing equations the 2nd degree polynomial trend curves where chosen, since no direct or straight line correlation exists between datasets. Having an R^2-value of less than 0.5 for Figure 4 to 6 indicates that there is very little correlation between the observed and calculated differences. The only color differencing equation that shows acceptable correlation values is the ΔE2000 color differencing equation.

In an extension of the work done with the inexperienced users it was also tried to optimize the weighting factors of the ΔE2000 equation (Habekost/Rohlf, 2008). The result of this work was that it was best to leave the weighting factors at their default values of 1.

3. Observations of color differences in images by trained observers

During one of IPA technical conferences a proofing RoundUP test was conducted were vendors and users of proofing systems were invited to submit proofs of a test form. This test form was provided by IDEAlliance and contained several SCID images and the IT8.7/4 test target. The to be generated proofs had to match colorimetrically the GRACoL reference printing conditions represented in the “GRACoL2006_Coated1.txt” file*.

Figure 7 shows the 2-page version of the test form. There was also a 3-page version for smaller format proofing devices available.

For accurate color reproduction it is beneficial to use ICC profiles. A source and a destination profile is required to correctly proof the test form. Participants could use an appropriate ICC profile provided by IDEAlliance or generate their own profile from the GRACoL2006_Coated1.txt file. These two possible routes will provide the source ICC profile. It is beneficial to have also a destination profile that characterizes the chosen proofing device. The principle of source and destination ICC profiles for accurate color reproduction is well documented (Sharma A., 2004).

* Available at: http://www.idealliance.org/downloads/idealliance-2006-swop-gracol-characterization-data-sets

Figure 7: The IDEAlliance CMYK test form consisting of a technical page (left) and a visual page (right).

There were no good or "OK" prints or proofs supplied that had to be matched. The evaluation of the submitted proofs was done solely by measurements of the IT8.7/4 test target. The measurements were compared to a set of established criteria.

If a supplied proof would be outside of one the established tolerances the submission would be classified as failed. A typical measurement set is shown in Table 2 below, which approximately represent the IDEAlliance hardcopy proofing certification tolerances.

Altogether there were 22 submissions from vendors and 64 end-user submissions. The average ΔEab from all vendor submissions was ΔEab = 1.01, while the average ΔEab from end-users was 2.21. This is quite a remarkable result. Despite the fact of this result it was also necessary to see how a visual judging of the supplied proofs corresponds to these ΔEab-numbers and any of the newer color differencing equations like ΔE94, ΔE2000, ΔECMC and DIN99.

Table 2: Evaluation criteria with a set of typcial data

		ΔE_{ab}	Pass/Fail	Tolerance
IT8.7/4 (all patches)		1.12	Pass	Average $\Delta E_{ab} \leq 1.50$
IT8.7/4 (95th percentile)		2.30	Pass	$\Delta E_{ab} \leq 6.00$
Solids	Cyan	3.85	Pass	$\Delta E_{ab} \leq 5.00$
	Magenta	0.90	Pass	$\Delta E_{ab} \leq 5.00$
	Yellow	1.03	Pass	$\Delta E_{ab} \leq 5.00$
	Black	1.32	Pass	$\Delta E_{ab} \leq 5.00$
Overprints	Red	0.63	Pass	$\Delta E_{ab} \leq 5.00$
	Green	3.17	Pass	$\Delta E_{ab} \leq 5.00$
	Blue	0.87	Pass	$\Delta E_{ab} \leq 5.00$
Neutral Gray	(50/40/40)	1.02	Pass	$\Delta E_{ab} \leq 1.50$
Paper White	Delta L*(95.0)	0.43	Pass	$\Delta E_{ab} \leq 2.00$
	Delta a* (-0.02)	1.06	Fail	$\Delta E_{ab} \leq 1.00$
	Delta b* (-1.96)	0.75	Pass	$\Delta E_{ab} \leq 2.00$
Ugra/FOGRA Media Wedge		1.36	Pass	Average $\Delta E_{ab} \leq 1.50$
Sheet to Sheet Variation		0.85	Pass	Max $\Delta E_{ab} \leq 1.50$

3.1 Experimental

During the IPA Technical Conference participants were asked to visually rate the proofs supplied by vendors and suppliers. This was done on two separate days. On the first day the proofs from the vendors were displayed and on the second day the proofs submitted by users. The proofs were displayed in color viewing booths supplied by GTI (GIT EVS-2450/FS). The fluorescent light tubes had a color-rendering index (CRI) of 93 – 95 towards the D50 illuminant.

Participants were given a ranking sheet, as can be seen in Appendix 1. Rankings had to be given for the reproduction of quarter-, mid- and three-quarter tones, as well as for flesh tones and neutral colors (gray).

These areas are encompassing all the critical elements in a printed product. Highlight and shadow areas should give detail reproduction, whilst the mid tone areas are most sensitive to possible dot gain issues. Neutral colors and flesh tones are most perceptive to possible color imbalances.

Judges could use rankings from 1, for the lowest ranking, and 10, for the highest ranking. In order to give some kind of a guideline the rankings were split as follows:

9 – 10 points:	Excellent reproduction / Excellent rendering of flesh tones
7 – 8 points:	Slight shift / very good rendering of flesh tones
5 – 6 points:	Visible shift / good rendering of flesh tones
3 – 4 points:	Visible shift / Questionable rendering of flesh tones
1 – 2 points:	Large shift / Poor rendering of flesh tones

Each sheet was evaluated by an average of 5 people. Although all vendor supplied sheets were visually evaluated this was not possible for all user supplied sheets.

Out of the 64 user submitted proofs 42 proofs (~ 66%) were evaluated by conference participants. This was due to the large number of submitted proofs and the limited amount of space in the three viewing booths. Conference participants work in the Graphic Arts industry and are most likely hands-on color experts. The initial question was, how could the various color differencing numbers be compared with the visual ratings given by the conference participants? Each color differencing equations gives a different average ΔE and the values of all the 1617 patches show then a different standard deviation. A method for comparing different averages or means is the coefficient of variation. The coefficient of variation calculates the ratio of the standard deviation to the mean and is a useful measure for comparing the degree of variation from one data set to another, even if they have different means. The coefficient of variation is defined as:

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Mean}}$$

This allows comparing the data with greatly varying means as they are generated by each entry and the color differencing equations. Also it is important to know that the lower the value of the coefficient of variation is, the better the overall data approximate to the mean.

Entries from the judging sheets were collected and averaged. These results were grouped by vendor and user submissions. These results were further divided into the five categories:

- Quartertones
- Mid tones
- Three quarter tones
- Flesh tones
- Neutrals (Gray)

This was done to see whether one color differencing equation correlates better with visual judging results. Although many submissions had a low average ΔEab some showed a quite high maximum ΔEab-value. A ΔEab above 5 was considered as high and a list was compiled that contained all these patches. These patches were plotted in Chromix® ColorThink software against the reference data from the GRACoL2006_Coated1.txt file. It was also tried to determine if a certain combination of software and proofing device is more bound to cause these outliers than other combinations.

In a last step of the evaluation of the visual rankings a new set of tolerances for each of the equations was set up to see if this results in fewer or more pass/fail results.

3.2 Visual Rating versus ΔE equation for vendor submitted proofs

In the first step of the evaluation the visual ratings from vendor-supplied proofs were grouped by the ranking they received and the coefficients of variation, derived from the average ΔE-value and the standard deviation of the ΔE-values of the 1617 color patches were plotted against the color differencing equations that were used. Vendor submissions have been coded with H and a number to anonymise the supplier entries. A typical plot of this can be seen in Figure 8.

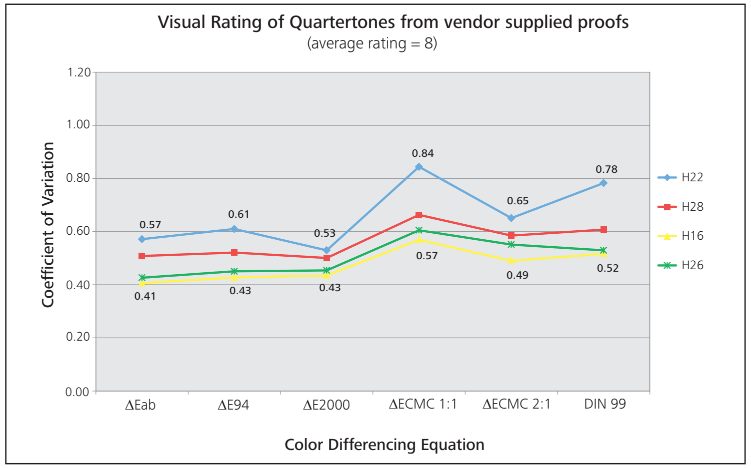

Figure 8: Visual ratings for vendor-supplied proofs in regards to the color differencing equations used.

The four vendor supplied proofs had received similar visual ratings. The plots for similar visual ratings of other vendor supplied proofs look similar. It is interesting to see that the ΔE2000 equation creates a distribution profile in which the proofing systems look as though they have a similar error spread. The ΔECMC (1:1) creates a much different profile in which proofing systems look as thought they have a very high error spread.

In the midtones, three quarter tones and flesh tones the picture changes. It seems that ΔEab results in a lower error spread in regards to the midtones, whilst for the Three Quarter tones and flesh tones it seems that it is ΔE94. For the neutral colors ΔE2000 creates the smallest error spread. For all five visual test criteria ΔECMC (1:1) gives the largest error spread. It would seem that having a smaller error spread is more desirable, since all data gets normalized, but, as will be shown later on, this is not a good representation of the perceived visual differences.

A color differencing equation should give a good numerically representation of the differences that are present. It should not exaggerate or minimize the perceived differences. The majority of the vendor supplied proofs (65%) passed the certification with an average ΔEab of 1.01. The vendor submitted proofs that did not pass the quality assurance evaluation did so due to a failure in only one category. A list of all the categories can be found in Table 2.

3.3 Visual Rating versus ΔE equation user-submitted proofs

What was done for the vendor submitted proofs was repeated for the user submitted proofs. The majority of the user submitted proofs did not pass the verification and had an average ΔEab of 2.21. Although numerically this is a discouraging outcome it needs to be said that there could be many factors contributing to this result. Users might operate the equipment in less than ideal conditions, the ICC profiles that were being used might not be ideal, generic ICC profiles or no profile at all were being used. Nevertheless 21 out of the 64 (~ 33%) user submission achieved on average a ΔEab of ≤ 1.50. This was also quite a remarkable result. Judges had the same rating categories as with the vendor submitted proofs and used the same judging sheet that can be seen in appendix 1. A typical plot of this can be seen in Figure 9.

The user-supplied proofs were given a similar rating by the judges. In the example given below the proofs had received a very good, almost excellent rating.

As seen before with the vendor submitted proofs, the ΔE2000 equation gives the lowest statistical error of all 5 color-differencing equations.

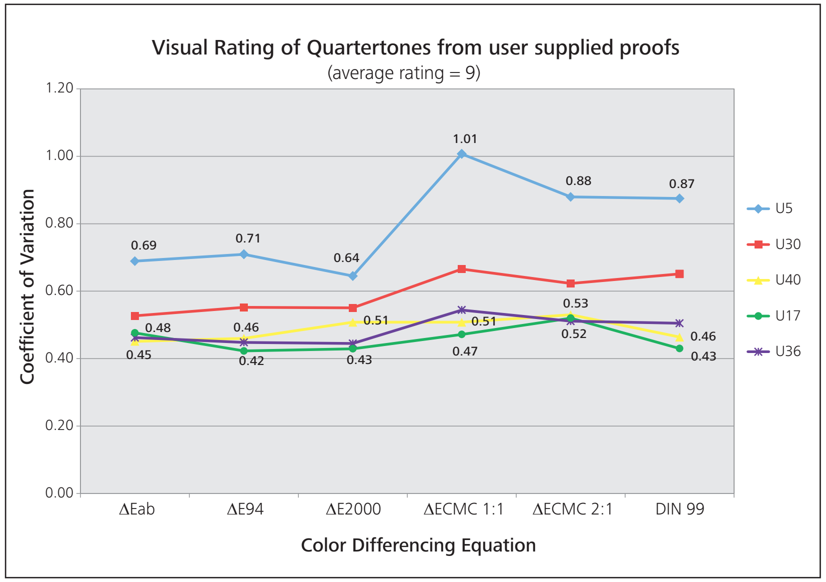

Figure 9: Visual ratings for user-supplied proofs in regards to the color differencing equations used.

The larger spread between the lowest and highest coefficient of variation for ΔE2000 can be attributed to overall larger spread of the 1617 color differencing values. The comparison of the mid-tone and three-quarter tone ratings reveal that the ΔE2000 gives the lowest statistical error. The same applies for the flesh-tones and the visual ratings given for the reproduction of the neutral colors.

3.4 Overall Visual Rating versus the coefficient of variation

In the previous paragraph it was attempted to see which color differencing equation gives the lowest error spread, but is this really giving a true representation of the visual ranking given by the judges in relation to the spread of the ΔE-values under the five color differencing equations under investigation here.

In order to determine a correlation between the visual ratings and the color differencing equations the visual ratings given by the judges were grouped into similar values and the coefficients of variation derived from the average ΔE-value of the 1617 color patches and the standard deviation of these ΔE-values were plotted against each other. A typical plot can be seen in Figure 10. The trend lines used in this plot are 4th order polynomial. From the Figure it can be seen that the data points are not on a straight line, so a straight trend line should not be used. A 4th order polynomial curve followed the data points. Based on the definition given above in regards to the coefficient of variation a smaller number means a lower spread of the data in comparison to the average. Figure 10 shows that although the vendor submitted proofs received quite a high visual (7 – 7.5) rating the error spread from the measured data is quite large. This in turn means that although the measured results were not that good, the judges gave quite a good visual rating. A complete list of the R^2-values listed by category can be seen in Table 3.

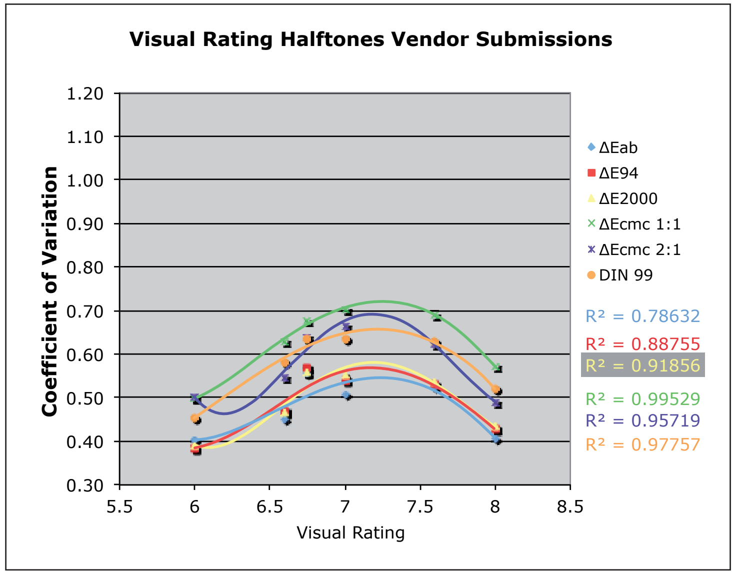

Figure 10: Visual rating for mid-tones from vendor submissions vs. coefficient of variation.

Table 3: R^2-values from vendor submitted proofs in relation to the coefficient of variation. Maximum values per category are highlighted.

	ΔE_{ab}	ΔE94	ΔE2000	ΔECMC (1:1)	ΔECMC (2:1)	DIN99
Quartertones	0.709	0.660	0.710	0.456	0.593	0.505
Mid tones	0.786	0.888	0.919	0.995	0.957	0.978
Three quarter tones	0.195	0.232	0.201	0.367	0.321	0.292
Flesh tones	0.695	0.834	0.843	0.978	0.807	0.969
Neutrals	0.631	0.774	0.762	0.683	0.785	0.667

From this Table it can be seen that in 3 out 5 cases the ΔECMC (1:1) equation gives a better correlation between the ratings given by the judges and coefficient of variation, which relates to the spread of the numerical color differences given by one of the color differencing equations. This means also, that the ΔECMC (1:1) equation gives a better reflection of how the human observers perceived differences present in the submitted proofs. For an unknown reason the correlation in regards to the three quarter tones is low or better non-existent, since the R^2-values indicate only a 37% correlation, which is not really a correlation. Only in the quartertones does ΔE2000 give a better correlation to the perceived differences. This might have to do with the fact that there was no "OK" sheet to compare to. Judges were giving the rankings based on their daily work experience and what, in their mind, is a good reproduction of the images shown in Figure 3. The above-mentioned procedure was also carried out for the user submitted proofs. The results in regards to user submitted proofs are little bit different compared to the vendor submitted proofs. All R^2-values of the 4th order polynomial trend lines shown in Figure 11 are listed in Table 4.

The color differences present in the user submitted proofs were greater than in the vendor submitted proofs. In this case only in 2 out 5 cases did the ΔECMC 1:1 color differencing equation correspond with the differences perceived by the judges.

The ΔE2000 equation does not seem to correlate well with a larger spread of color differences, as they were present in the user submitted proofs. Interestingly enough the DIN99 method for expressing color differences seems to relate better than the other equations with the judged color differences.

From these results it looks like only one color differencing equation seems to stand out. This is the ΔECMC (1:1) equation. Since the data gathered for the visual judging stems from the perceived accuracy or inaccuracy in relation to the five categories mentioned above, it makes sense the ΔECMC equation with a lightness weight of 1 and a chroma weight of 1 gives the best correlation.

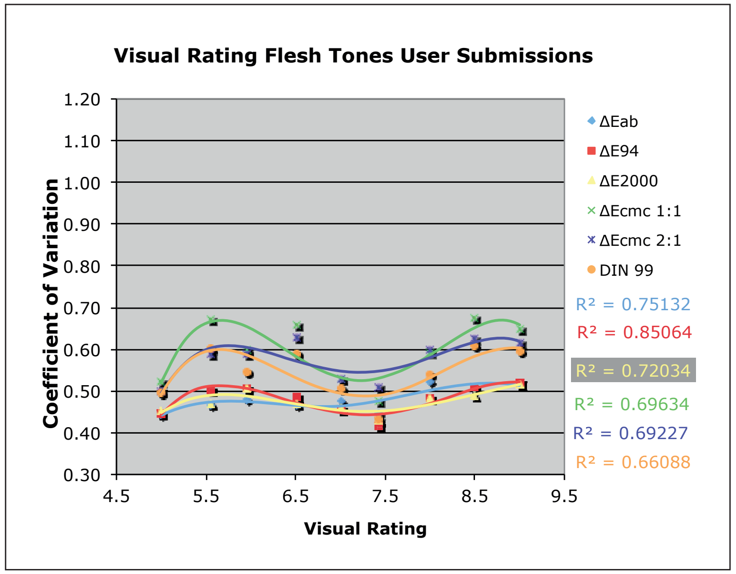

Figure 11: Visual rating for flesh tones from user submissions vs. coefficient of variation.

Table 4: R^2-values from user submitted proofs in relation to the coefficient of variation. Maximum values per category are highlighted.

	ΔE_{ab}	ΔE94	ΔE2000	ΔECMC (1:1)	ΔECMC (2:1)	DIN99
Quartertones	0.501	0.517	0.415	0.638	0.437	0.652
Mid tones	0.514	0.887	0.881	0.954	0.581	0.782
Three quarter tones	0.891	0.827	0.611	0.980	0.855	0.940
Flesh tones	0.751	0.851	0.720	0.696	0.692	0.661
Neutrals	0.851	0.560	0.618	0.570	0.763	0.485

3.5 New tolerances

Based on the results obtained so far it is necessary to reevaluate the data gathered during the IPA conference. Previous paragraphs showed that ΔECMC (1:1) is the color differencing equation of choice when it comes to perceptual color differences. Many studies showed (Luo, 2004; Johnson, 2006; Habekost, 2007) that the ΔE2000 equation gives a quite true numerical representation of a small color difference that is visible between a standard and a sample.

The three equation ΔECMC (1:1), ΔE2000 and DIN99 were used to re-evaluate the present data to see whether the main results were greatly changed or just some minor changes in regards to a pass or fail of the 86 entrants would take place.

Before this re-evaluation can take place it is necessary to set-up tolerance by which the data can be measured in regards to a pass or fail rating. In Table 5 these values are shown. The new tolerances were obtained by looking up the corresponding patches in all 86 submissions. In some submissions the ΔEab-values were in close proximity to the ΔEab-values listed in Table 5. At least three entries per color patch were used and their corresponding ΔE-values under ΔECMC (1:1), ΔE2000 and DIN99 averaged. These averages were now used as the new tolerances as they are listed in Table 5.

As a result of these new tolerances the ratings for some of the submissions changed from "Fail" to "Pass" but also some that passed before under the set of ΔEab tolerances received a failing grade under the new tolerances. This was also looked at in comparison to the overall visual rating some of the proofs received. The submissions that received a failing grade under the new set of tolerances received also low visual grades by the judges. The same applies also for submissions whose grade changed to a passing grade. They usually received high ratings from the judges. This shows that it is beneficial to use the newer color differencing equations, since they reflect better how human observers perceive color differences. Examples of this can be seen in Table 6.

Table 5: Table of new tolerances used for the evaluation of all submitted entries

Criteria	ΔE_{ab}	ΔECMC (1:1)	ΔE2000	DIN99
IT 8.7/4 (All)	1.5	1.15	1.00	0.90
IT 8.7/4 (95 percentile)	6	5.50	4.00	4.10
Cyan	5	2.40	2.10	2.10
Magenta	5	3.10	2.90	2.50
Yellow	5	2.00	1.70	1.60
Black	5	10.00	4.50	6.50
Red	5	3.00	2.50	2.30
Green	5	2.00	1.70	1.60
Blue	5	4.00	3.60	3.50
Gray	1.5	1.70	1.50	1.20
Fogra Wegde (average)	1.5	1.05	0.90	0.95
Paper White (Fogra)	3	3.65	2.60	2.10
Fogra Max	10	8	7	7

Table 6: Examples of the changed ratings under the new set of tolerances

Vendor/User ID	ΔE_{ab}	ΔECMC (1:1)	ΔE2000	DIN99	Visual rating
U50	Pass	Pass	Fail	Fail	6.3
H12	Pass	Pass	Fail	Fail	6.7
H29	Pass	Pass	Fail	Fail	6.8
U35	Fail	Pass	Pass	Pass	7.6
U54	Fail	Pass	Pass	Pass	8.0
U49	Fail	Pass	Pass	Pass	8.4
U32	Fail	Pass	Pass	Pass	8.7

This Table shows that a low visual rating translates in most cases to a "Fail" rating. The opposite applies to a submission that received a high visual rating from the judges. The initial "Fail" under ΔEab transforms into a "Pass" using any of the three newer color-differencing equations listed in Table 5. Therefore it is advisable that any of these three equations should be used to get a better correlation between visual perception and numerical data presentation.

4. Conclusions

From the experiments carried out in the various years it becomes clear that ΔE2000 corresponds better with the way human observers perceive small color differences. All the "newer" color differencing equations try to compensate for the non-linearity of the L*a*b*-color space through the use of weighting factors and correctional terms. The only exception to the rule is the DIN99 equation, which transforms the L*a*b*-color space in such a way, that it becomes linear and the straight Euclidian distance between two colors can be used.

An interesting aspect of the various projects is, that the ΔE2000 seems to correspond well when trained and untrained observers look at color chips and determine if there is a difference between them. When industry professionals looked at images the ΔECMC (1:1) equation correlates quite well with the perceived color differences. This is probably due to the circumstance that there was no good proof to compare too. It was more of a relative comparison than an absolute comparison.

The ISO technical committee TC130 is currently revising the ISO procedure 12647 (ISO TC130, 2012) in its various iterations and is listing ΔE2000 tolerance values for the process colors in offset printing (ISO TC130, 2012). Although these values are given out only "for information only" it is an indication that an adoption of ΔE2000 as the new standard for the determination of color differences will take place in the near future.

Visual Judging Sheet

Rankings:
1 for lowest ranking
10 for highest ranking

NAME: ____________________ DATE: _________________

Proof ID#	Quarter Tones	Mid Tones	Three quarter tones	Flesh tones	Neutrals (Grey)	Comment
Example: #H6	*3*	*6*	*7*	*5*	*5*	*Mid tones too red*

Ranking Scale:
9 - 10 points: Excellent reproduction/Excellent rendering of flesh tones
7 - 8 points: Slight shift/very good rendering of flesh tones
5 - 6 points: Visible shift/good rendering of flesh tones
3 - 4 points: Visible shift/Questionable rendering of flesh tones
1 - 2 points: Large shift/Poor rendering of flesh tones

Appendix 1: Visual judging sheet

5. References

- Billmeyer, S. (2000): „Principles of Color Technology", 3rd ed. New York: Wiley & Sons.
- CIE. (1986): „Colorimetry", 2nd ed. Vienna: CIE Publications No. 15.2.
- CIE. (2001): „Improvement to industrial color-difference evaluation", CIE technical report 142-2001. Vienna: CIE Central Bureau.
- Clarke, F.J.J./McDonald, R./Rigg B. (1984): „Modification to the JPC79 colour-difference formula", Journal of the Society of Dyers and Colourists vol. 100 (4), 128–132.
- DIN, 6. (2001): Colorimetric determination of color differences of surface colors using the DIN99 formula. Beuth Verlag.
- Habekost, M. (2007): „Color Difference Equations and the Human Eye". TAGA Annual Technical conference. Pittsburgh: TAGA.
- Habekost, M. (2009): „The Applicability of Modern Color Differencing Equations in the Graphic Arts Industry". TAGA 2009 annual technical conference. Pittsburgh: TAGA.
- Habekost, M./Rohlf, K. (2008): „The Evaluation of Color Difference Equations and Optimization of DE2000", Journal of Graphic Technology , 4 (3), 149-164.
- ISO TC130. (2012, February 28): ISO/DIS 12647-2. Geneva, Switzerland.
- Johnson, T./Green, P. (2006): „The Colour Difference Formula CIEDE2000 and its Performance with a Graphic Arts Data Set", TAGA Journal of Graphic Technology, vol. 2, 59-71.
- Luo, M.R./Minchew, C./Kenyon, P./Cui, G. (2004): „Verification of CIEDE2000 using Industrial Data". AIC Color and Paints, Interim Meeting of the International Color Association, Proceedings, 97-102.
- Sharma, A./Collins, T./Cheydleur, R./Smiley, S./Suessel, F. (2006): „Visual and colorimetric press to proof matching using the new GRACoL reference printing condition". TAGA Journal of Graphic Technology, vol. 3, 32-49.
- Sharma, A. (2004): „Understanding Color Management", Thomson Delmar, New York.
- Sharma, G./Wu, W./Dalal, E.N. (2005): „The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations", Color Research & Application, vol. 30 (1), 21-30.

(received for the first time: 06-03-12)

Martin Habekost

Dr. rer. nat., Ass. Prof.,
School of GCM,
Ryerson University Toronto,
Canada

mhabekos@ryerson.ca