Evaluation of Colour Appearance Models using Transmissive Media

Kiran Deshpande and Lindsay MacDonald, London College of Communication, London, United Kingdom

Abstract

A new colour appearance data set for transmissive media was collected through psychophysical experiments. Five experimental phases with different luminance levels and back-ground luminance factors were conducted in dark surround conditions. Three of the phases were performed at high luminance level, i.e. luminance of reference white above 3300 cd/m^2. The magnitude estimation method was used with an experimental set-up similar to the classical LUTCHI experiments. Sixteen observers participated to give a total of 10500 estimations. The lightness, colourfulness and hue of 50 test colours were judged. Observer performance was evaluated in terms of repeatability and accuracy of the observers. Colour appearance changes due to luminance level and background luminance factor were analyzed. Five colour appearance models – CIELAB, Hunt94, CIECAM97s, CIECAM02 and Kwak03 – were tested by using the colour appearance data set. Except for CIELAB, all models performed well in terms of the ability to predict the mean visual data and successfully predicted lightness and colourfulness changes under different luminance levels and backgrounds. CIECAM02 showed best overall performance amongst models.

Psychophysical Experiments

The magnitude estimation technique was used for the psychophysical experiments. The experiment was divided into five phases according to different luminance levels and background luminance factors. The data collected is summarized in Table 1. These five phases can further be classified into two subgroups. The first subgroup includes three phases with different luminance levels (high, medium and low) and the second subgroup comprises three phases with varying background luminance factors (black, grey and white background). All three phases of the first subgroup were conducted using a grey background. The phases in the second subgroup used high luminance level (luminance of reference white ranging from 3320 to 3650 cd/m^2). Thus, the phase with grey background and high luminance level was the common phase between the two subgroups. The lightness, colourfulness and hue of 50 test colours were assessed by a panel of 16 observers to give a total of 10500 estimations.

Test Colours

Fifty colours were selected as the test stimuli for all of the experimental phases. The same test colours were used for all phases of the experiment. The size of each colour sample was 3.5×3.5 cm. Colours were selected to cover a wide colour gamut and range of lightness values. The test stimuli were measured using a JETI telespectroradiometer (TSR). Measured values were plotted in the CIE u' - v' uniform colour space (Fig. 1). It can be seen clearly that the points create a horse-shoe like shape and some of the points lie close to the spectral boundary.

The distribution of test colours is also shown in the CIE a^*-b^* plane in Fig. 2, indicating a good spread across all four quadrants of the a^* - b^* plane. Measurements of test colours were carried out in viewing conditions identical to those of the actual psychophysical experimental phases. The JETI telespectro-radiometer was placed 20 cm from the test colour, positioned normal to the surface of the light table on which the test colours were placed.

Table 1: Summary of the phases used in the experiment

Phase	Light source (CCT)	Luminance of reference white L_w (cd/m^2)	Background	Background luminance factor Y_b	No. of observers	No. of estimations
Black background	5998	3320	Black	0.53	16	2400
White background	5925	3650	White	100	14	2100
Grey background / High luminance	6119	3480	Grey	52.2	16	2400
Medium luminance	6140	298	Grey	53.52	12	1800
Low luminance	6225	27.4	Grey	50.06	12	1800
					Total	10500

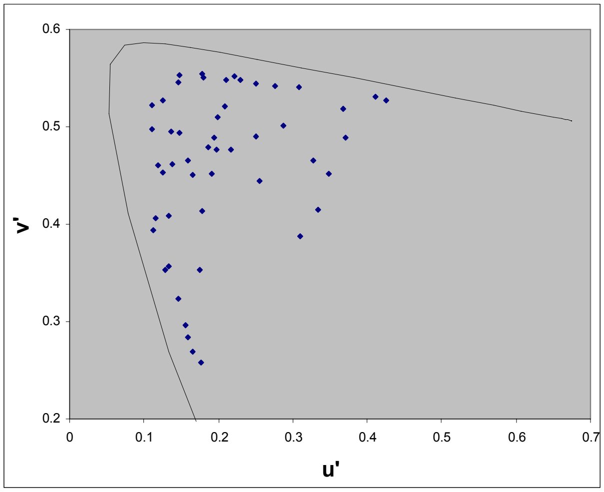

Figure 1*. Distribution of 50 test colours in u' – v' plane*

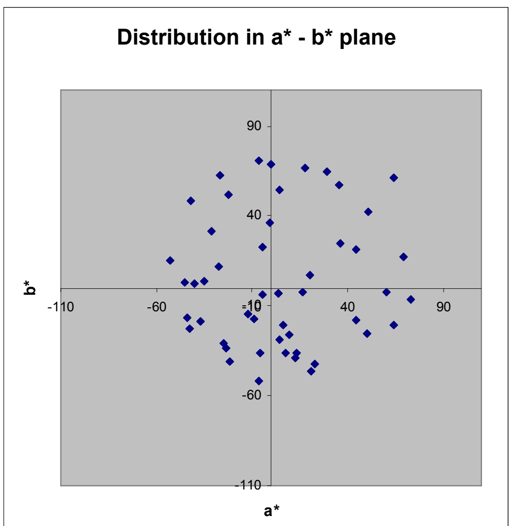

Figure 2*. Distribution of test colours in a*-b* plane*

Design of Viewing Patterns

Viewing patterns were designed for three different backgrounds (i.e. black, grey and white). For three luminance levels (i.e. high, medium and low) the grey background was used. Fig. 3 shows the viewing pattern with the grey background.

The size of the viewing pattern was 115×104 cm. These large viewing patterns, which include the background, were printed on a transparent material (acetate sheet) using an Inca Eagle wide-format ink-jet printer. The viewing pattern consisted of a test stimulus in the centre, with adjacent reference white and reference colourfulness, and a peripheral decorating pattern of random colours. This pattern simulates a complex image and renders the test colours as related colours. For each background, twenty-seven different decorative colours were used.

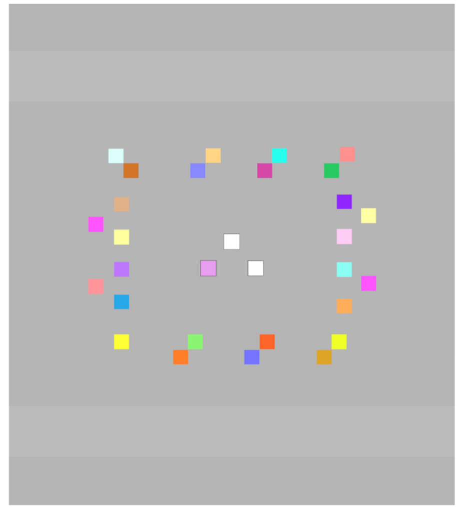

Figure 3. *Viewing pattern for grey background phase*

Experimental Setup

A large light table of size 115×104 cm was used for viewing the test colours. The intensity of the light was controlled by an analogue dimmer. The light sources inside the table for back-illuminating the viewing pattern were ten fluorescent tubes with colour temperature of approximately 6000K.

All the phases were carried out in a dark surround. The distance between the test colour and observer was controlled at 1 m throughout the experiment. The light table was tilted so that its surface was normal to the observer's resting line of sight. Fig. 4 shows the front, side and top views of the experimental set up.

Sixteen observers participated in the psychophysical experiment. All had moderate experience in psychophysical experiments related to colour judgment, but for most it was their first time to perform a magnitude estimation experiment. Of the 16 observers, 11 were male and 5 were female. Most were aged between 20 and 30 years. All observers had normal vision and passed the Ishihara test for colour deficiency. Thirteen observers took part in all of the five phases, one observer finished four phases and two observers finished only three phases.

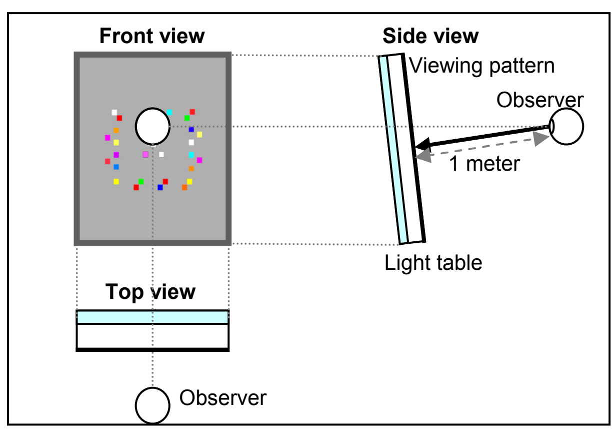

Figure 4. *Schematic diagram of the experimental set up*

Experimental Procedure

The test colour to be judged was presented in the middle of the pattern. Each colour patch was 3.5×3.5 cm, thus subtending an angle of two degrees from the viewing position. Test colours were affixed with the aid of adhesive ‘blue tack’, very small pieces of which were stuck at the four corners of the central area of the viewing pattern.

Each observer was given time to adapt to the surround conditions before starting a session. Adaptation time was changed according to the luminance level. For the high luminance level, a minimum adaptation time of 5 minutes was allowed; for the low luminance, it was 15 minutes. Before starting the experiment, each observer was given training to get acquainted with the task. The time duration for a typical observer session (of one experimental phase) was approximately one hour.

Different reference colourfulness samples were used in the three different viewing patterns. Before starting a new phase, observers were asked to estimate the reference colourfulness sample. This colourfulness value became the anchor point for scaling of test colours throughout the phase. For scaling of lightness, all observers used the same numerical scale between 0 and 100; hence, the arithmetic mean values of lightness were used for further analysis.

For hue scaling, the results were transformed onto a 0-400 scale as follows: Red–Yellow: 0–100, Yellow–Green: 100–200, Green–Blue: 200–300 and Blue–Red: 300–400. The arithmetic mean and standard deviation were calculated from these values. For mixed responses such as 380 and 10, one of the values was moved to the other end of the scale between 0 and 400 e.g. 380 became -20. For calculating the absolute differences for repeatability, accuracy and testing of models, the hue scale of 0–400 was transformed onto 0–100 so that the hue differences could be compared with those of other attributes.

For colourfulness, every individual’s result was first transformed onto the common scale by using the reference u value predicted by that observer. Thus, the colourfulness results of all observers were on the same scale. The geometric mean was used as an averaging method of colourfulness. This geometric mean was used for subsequent analysis.

Observer Performance

Intra-observer and inter-observer variability were examined by evaluating the repeatability and accuracy of observers. In each of the phases, three test colours were repeated randomly for each observer and the difference between the two judgments of the same test colour was calculated. Absolute difference between the two estimations was used as the statistical measure to evaluate the repeatability of the observers. The average of all differences for all observers was calculated for each of the five phases, with results as shown in Fig. 5. It can be seen that colourfulness was most difficult to judge under the black background. The medium luminance (298 cd/m^2) phase, gave the most reliable results.

The accuracy of each observer can be assessed in terms of the closeness of observer’s estimates to the mean visual results. Thus the deviation between the individual’s and the mean visual results was evaluated by using absolute difference and correlation coefficient (CC) as the statistical measures.

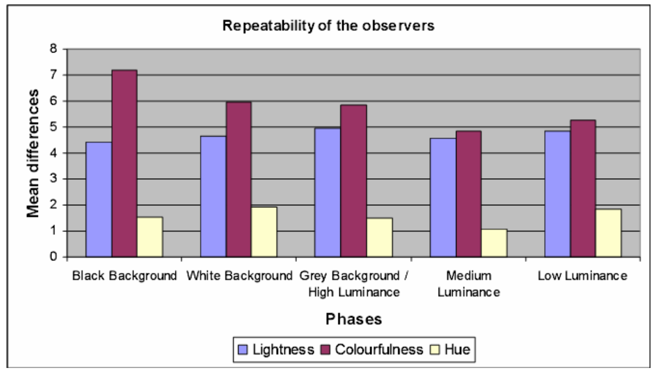

Figure 5. *Repeatability of the observers for all five phases*

For colourfulness, the accuracy was calculated after applying the scaling factor to put all values onto the same scale. Hue values were converted to the 0–100 scale for comparability with lightness. Accuracy results are shown in Fig. 6. Observers found it very difficult to judge the lightness of test colours against the white background.

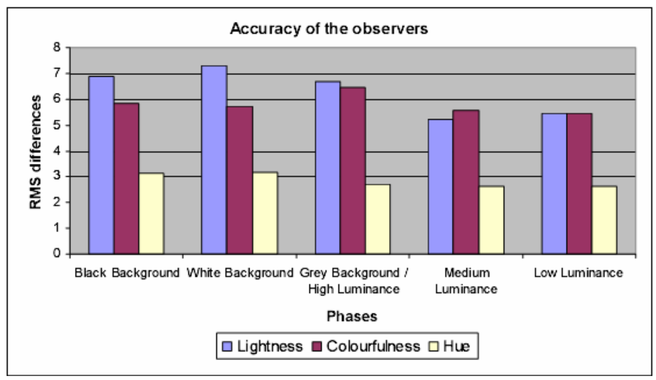

Figure 6. *Accuracy of the observers (absolute difference) for all five phases*

Colour Appearance Phenomena

The effects of luminance levels and background luminance factors on colour appearance were studied through qualitative as well as quantitative comparison between different phases. Mean visual results were used for the comparisons. For quantitative analysis, correlation coefficients and average values of absolute differences were calculated for each attribute.

Effect of Luminance Level

A wide range of luminance levels (luminance of reference white from 0.53 to 3650 cd/m^2) was covered in this study. The visual results of three phases (high luminance, medium luminance and low luminance) were compared to show the effect of luminance level on three colour attributes. All of these phases were carried out under dark surround with grey background. Table 2 gives the absolute differences between the mean visual results of different phases for lightness, colourfulness and hue. The greatest difference was in the judgment of lightness between the high and low luminance conditions.

Table 2: Absolute differences between mean visual values of lightness, colourfulness and hue for different luminance levels

Attribute	High & Low luminance levels	High & medium luminance levels	Medium & low luminance levels
Lightness	5.58	3.03	3.66
Colourfulness	2.91	2.88	2.53
Hue	1.10	1.54	1.55

Effect of Background Luminance Factor

Mean visual results of three phases were compared to study the effect of background luminance factor (Y_b) on colour attributes. The range of background luminance factor was wide i.e. from 0.53 for the black background phase to 100 for the white background phase. Reference white luminance was high, i.e. 3320 to 3650 cd/m^2. All the phases were conducted with dark surround condition. Table 3 shows the effect of background luminance factor on lightness, colourfulness and hue in terms of absolute differences between the mean visual results of different phases. The greatest difference was in judgments of colourfulness between the white and black backgrounds.

Table 3: Absolute differences between mean visual values of lightness, colourfulness and hue between two phases with different Y_b

Attribute	High & Low luminance levels	High & medium luminance levels	Medium & low luminance levels
Lightness	5.58	3.03	3.66
Colourfulness	2.91	2.88	2.53
Hue	1.10	1.54	1.55

Testing Colour Appearance Models

Colour appearance data obtained from the psychophysical experiments were used to test the accuracy of different colour appearance models. The performance of the models was evaluated by testing their ability to predict the mean visual results and the changes in colour appearance under the various viewing conditions. The models tested were CIELAB, Hunt94, CIECAM97s, CIECAM02 and Kwak03. Each of the models requires tristimulus values (*XYZ*) of the test colours and the reference white ($X_wY_wZ_w$) as input parameters. All models except CIELAB need input information of luminance of reference white (L_w), background luminance factor (Y_b) and surround. Apart from these input data, the Hunt94 and Kwak03 models require rod contribution information.

Method

The models were tested for the following parameters: lightness (*J*), chroma (*C*), hue (*h*) and colourfulness (*M*). As the CIELAB model does not include colourfulness, the chroma predictor was tested. Although chroma was not judged by the observers, colourfulness and chroma are related to each other by the brightness of the reference white. Thus, they can be assumed to be similar for an isolated experimental phase. In all experimental phases, observers were allowed to adapt fully to the viewing condition and the chromaticity of reference white. Hence the degree of chromatic adaptation, *D*, was set equal to 1.0 for all models. Absolute and RMS differences between mean visual data and model-predicted data were calculated for each attribute.

Performance of Lightness Predictor

Fig. 7 shows the performance of lightness predictors in terms of average colour difference. It can be seen that all the models except CIECAM02 showed poor performance for the black background phase. The CIECAM02 model showed consistent performance throughout all phases.

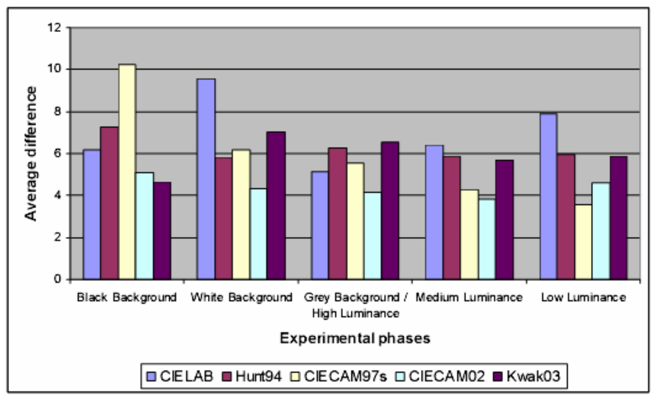

Figure 7. *Performance of lightness predictor in terms of average difference*

Performance of Hue Predictor

As expected, all models performed better for prediction of hue than for lightness and chroma (Fig. 8). The Kwak03 model showed slightly worse results for the hue predicted for all phases.

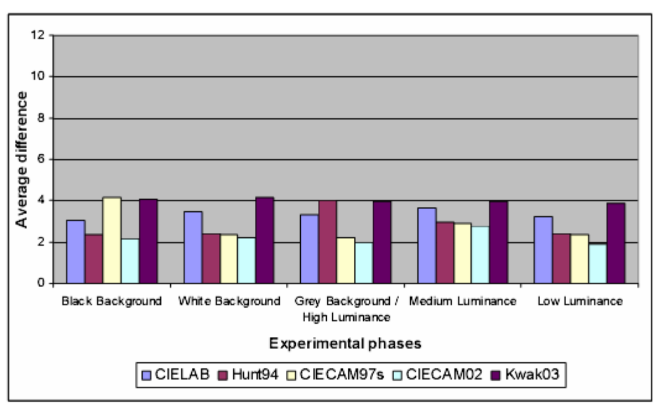

Figure 8. *Performance of hue predictor in terms of average difference*

Performance of Colourfulness Predictor

Because scaling factors were applied to both chroma and colourfulness predictors by linear fitting with the mean visual colourfulness, the average differences for chroma and colourfulness predictors were almost identical. Performance of all models was found to be equally good for all phases except Hunt94 and CIECAM97s for the black background (Fig. 9).

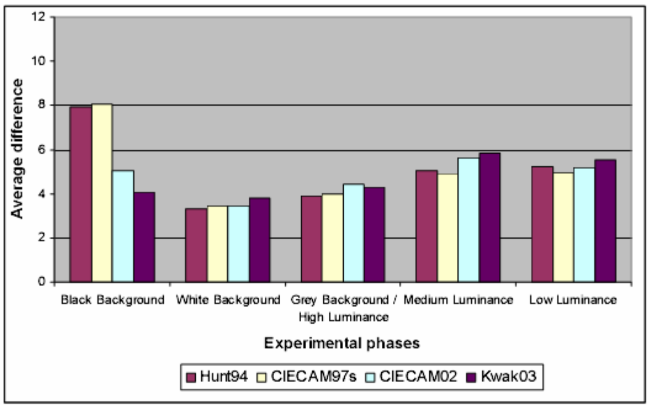

Figure 9. *Performance of colourfulness predictor by average difference*

RMS Differences

Root mean square (RMS) differences were calculated from average differences of lightness, chroma and hue. These can be regarded as the overall colour differences (comparable to ΔE^* for CIELAB) between the mean visual data and model predictions. Fig. 10 shows the RMS differences for each model. The values varied between approximately 6 and 12 units, suggesting that none of the models was particularly good overall in predicting the colour appearance reported by the observers. The worst result was for CIECAM97s with the black background. The most consistent model overall was CIECAM02.

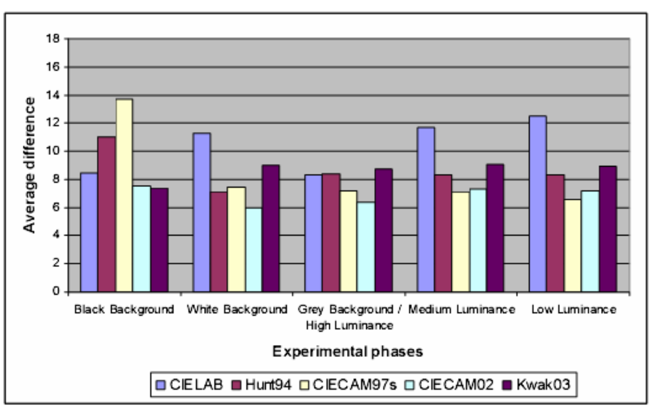

Figure 10. *RMS differences between mean visual data and model predictions*

Qualitative Performance of Colour Appearance Models

For qualitative comparison between the model prediction and mean visual data for the five different phases, scatter diagrams for lightness, chroma, colourfulness and hue were used. Mean visual data were plotted on the *y*-axis and model predictions were plotted on the *x*-axis. The scatter was lowest for hue, indicating that all models predicted hue fairly well. The largest scatter, indicating the greatest differences between the predictions of the models, was for lightness (Fig. 11).

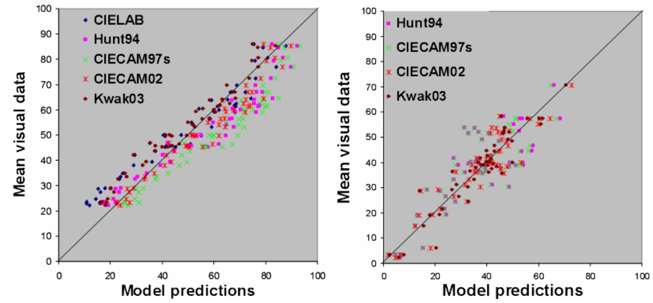

Figure 11. *Qualitative performance of lightness predictors (left) and colourfulness predictors (right) for the Black Background phase*

Prediction of the Effects of Luminance Level and Background

Model predictions of the effects of luminance level and background were compared with the visual data, again by the use of scatter diagrams between phases (Figs. 15 to 18). Luminance level had very little effect on perceived hue.

Prediction of lightness change and colourfulness change by luminance level

Luminance level did not have a strong effect on the visual data, except that dark colours were generally judged lighter at high luminance than at low luminance, perhaps because of flare effects (Fig. 12). The Kwak03 model overall gave the best prediction, followed by CIECAM02. The Hunt94 and CIECAM97s models tended to over-predict the lightness of medium and light colours at higher levels of luminance.

Colourfulness change by luminance level for all models is shown in Fig. 12. The CIELAB model was excluded as it does not predict colourfulness. All models gave moderately good predictions of colourfulness change.

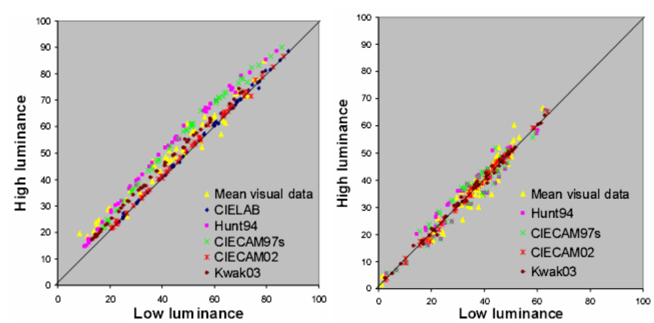

Figure 12. *Prediction of lightness change (left) and colourfulness change (right) by luminance level.*

Prediction of lightness change and colourfulness change by background luminance factor

Lightness decreased with increasing background luminance factor (Fig. 13). This effect was predicted by all models except CIELAB. CIECAM02 generally fitted the visual data best, followed by Kwak03. The CIECAM97s and Hunt94 models tended to over-predict lightness for the black background (a simultaneous contrast effect).

The mean visual data showed a strong increase in colourfulness with darker background. All four models successfully predicted this effect although with some differences between models (Fig. 13). There was a surprising degree of scatter in the predictions of the models against the black background, especially for CIECAM97s and Hunt94.

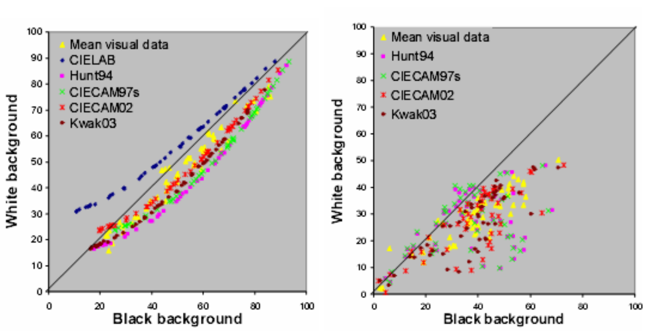

Figure 13. *Prediction of lightness change (left) and colourfulness change (right) by background luminance factor*

Conclusion

This experiment extended the colour appearance data set by the use of backlit transmissive samples at very high levels of luminance (up to 3600 cd/m^2). The visual data gathered in the experiments was found to be reliable, with satisfactory repeatability and accuracy of observers. At the high luminance level observers found it difficult to judge the colour appearance, which resulted in poorer repeatability and accuracy than at medium and low luminance levels.

The results of the white background phase showed the poorest accuracy for lightness. This indicates that observers found the lightness of colour samples hard to judge under a white background with high luminance, probably because of intraocular flare. Hue predictions were the most consistent among all appearance attributes in terms of both repeatability and accuracy. At high luminance levels, both lightness and colourfulness of the colour samples increased as compared to lower luminance levels. The results thus confirmed the Hunt effect, which states that colourfulness increases with luminance level. A darker (lower luminance factor) background induced higher lightness and colourfulness compared with a lighter background. Most of the colour samples therefore appeared lighter and more colourful against darker backgrounds. Lightness differences were found to be less than colourfulness differences. The perceived colourfulness was most affected by a change of background.

The performances of all colour appearance models, except CIELAB, were found to be good in terms of both average colour differences and RMS differences. CIECAM02 showed the best performance for all phases whereas CIELAB was found to be the worst. Hunt94 and CIECAM97s showed poor performance for chroma and colourfulness predictors with black background. Performance of the Kwak03 model was somewhat worse than that of CIECAM02 although the chroma and colourfulness predictions of both models were similar. The Kwak03 model had previously been claimed to give the best results [8] for the luminance level range 0.1–250 cd/m^2 but had not been tested for the very high luminance level used in this experiment. The performance of lightness predictors of all models at the high luminance level was found to be worse than under low and medium levels. The hue predictors of all models were best among all the colour attributes. The changes in colour appearance due to luminance level and background luminance factor were successfully predicted by all models, except CIELAB. The CIECAM02 and Kwak03 models performed better than others in predicting lightness and colourfulness changes. The CIELAB model failed to predict most of the colour appearance changes.

References

[1] M. R. Luo, A. Clarke, P. Rhodes, A. Schappo, S. Scrivener and C. Tait, "Quantifying colour appearance. Part I. LUTCHI colour appearance data", Color Res. Appl. 16, pp. 166-180 (1991a).

[2] M. R. Luo, A. Clarke, P. Rhodes, A. Schappo, S. Scrivener and C. Tait, "Quantifying colour appearance. Part II. Testing colour models performance using LUTCHI colour appearance data", Color Res. Appl. 16, pp. 166-180 (1991b).

[3] M. R. Luo, X. Gao, P. Rhodes, H. Xin, A. Clarke and S. Scrivener, "Quantifying colour appearance. Part IV. Transmissive media", Color Res. Appl. 18, pp. 191-209, (1993).

[4] M. R. Luo, X. Gao and S. Scrivener, "Quantifying colour appearance. Part V. Simultaneous contrast", Color Res. Appl. 20, pp. 18-28, (1995).

[5] R. W. G. Hunt, "An improved predictor of colourfulness in a model of colour vision", Color Res. Appl. 19, pp. 23-26, (1994)

[6] CIE, "The CIE 1997 interim colour appearance model (simple version)", CIE Publ., 131 pp. (1998)

[7] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo and T. Newman, "The CIECAM02 color appearance model", Proc. 10th Color Imaging Conference, Scottsdale (2002)

[8] Y. Kwak, "Quantifying the Colour Appearance of Displays", Doctor of Philosophy Thesis, University of Derby (2003).

Author Biographies

Kiran Deshpande received his B.E. in printing and graphics communication from the University of Pune, India (2002) and his MSc in digital colour imaging from London College of Communication (2006). His research interests include colour appearance and colour management solutions.

Lindsay MacDonald is Professor of Digital Media at the London College of Communication and is a Fellow of IS&T.