The Uncertainty of Colour-Matching Data

Boris Oicherman, Ronnier Luo and Arthur Tarrant, Colour and Imaging Group, Department of Colour and Polymer Chemistry, University of Leeds, Leeds, United Kingdom
Alan Robertson, National Research Council, Montreal Road, Ottawa, Ontario, Canada (retired)

Abstract

A colour matching experiment of the maximum-saturation type has been conducted on a group of five observers. The sources of uncertainty in its outcomes were identified and analysed. The variability introduced by the instrumentation was found to make a significant contribution to the total data variation for a single observer. Variability for a single observer, in its turn, contributes significantly to the overall observer-related variability. The CIE Standard Deviate Observer significantly underestimates the variability of colour matching according our study. The choice of the primary lights has an effect on the variability of colour matching. The results generally agree with previous findings by other researchers, although there are some differences in detail.

Introduction

The two CIE standard colorimetric observers represent the colour - matching characteristics of an average observer with normal colour vision. The properties of individual "real" observers differ in varying degrees from the standard, and from one another. Moreover, the colour matching results of the same observer vary due to the natural sensitivity thresholds of the colour vision mechanism. The result of these two types of variability is that there is some degree of uncertainty associated with every colour match: a pair of colour stimuli which are perceived as identical by one observer may be perceived as different by another. (Note that physical metrologists generally use the term "uncertainty" to characterize the dispersion of values that could reasonably be attributed to a measurement result. Colour scientists use the term "variability" to denote the dispersion of colour matches by a group of observers. In this paper we have used the two terms more or less interchangeably to denote the dispersion of the measured values that characterize a colour match.)

Understanding all sources and magnitudes of uncertainty of colour matching is a task of high scientific and industrial importance. The scientific interest lies in a better understanding of the mechanisms of colour vision, achieved by the analysis of the variations in colour matching properties as an indicator of variations in colour perception. The industrial interest is in a tool, which would allow colour matching practitioners in painting, automotive, graphic and other industries to make a reliable estimation of the uncertainty of colour matching; an improved Standard Deviate Observer which would, as the previous study[2] suggested: *"…better quantify the variability associated with observer metamerism in practical cross-media color matching"*.

Reported here are the results of a colour matching experiment of the maximum-saturation type. The sources of uncertainty in the experiment outcomes were identified. Particular care was taken to include sources of instrumental uncertainty in the analysis. The following topics were investigated:

1. Magnitude of the physical (instrumental) variability
2. Variability of colour matching within a single observer (intra-observer)
3. Variability of colour matching data between observers (inter-observer)
4. Performance of the CIE Standard Deviate Observer (SDO) in predicting our experimental results
5. Dependence of the variability of the colour matching results on the spectral location of the primary lights.

Uncertainty of Colour Matching

The Colour Matching Functions (CMF) are the relative amounts of three monochromatic primary stimuli required to match in colour every monochromatic test stimulus in the visible range of wavelengths. Two standard sets of CMFs, which represent the average observer with normal colour vision, are defined by the CIE[3] as the 1931 Standard Colorimetric Observer and the 1964 Standard Colorimetric Observer - for the field sizes of 2º and 10º, respectively.

The CMFs are the result of a visual colour measurement, which is performed by means of a colour matching experiment[4] (p. 121). Result of the visual colour measurement, just like any measurement, has some degree of uncertainty associated with it. With the aim of basic colorimetry defined as the prediction of metameric matches, the uncertainty of visual colour measurement can also be termed "The Uncertainty of Colour Matching": what is the probability that, if a given pair of stimuli is metameric with respect to the standard observer, it will match to the real one?

The sources of uncertainty of colour matching can be divided into four general categories:

1. Variability related to the instruments. These include visual colorimeters used to measure the colour matching functions, the measurement equipment used to measure the stimuli in the experiment, and the reference equipment used for calibration.
2. The variability in the colour matching data of a single observer, or *intra-observer variability*. This uncertainty is the result of natural thresholds of sensitivity to colour differences of the colour perception mechanism, as well as of the observer's experience in performing colour matching task, fatigue and, perhaps, motivation.
3. The variability in the colour matching properties of different observers. This variability is the result of the differences in the colour vision mechanisms of different but otherwise colour-

normal humans, and is referred to as the *inter-observer* variability. The result of the inter-observer variability in colour matching is the phenomenon of *observer metamerism*: a pair of colour stimuli with different spectral power distribution can be judged to match by one observer (being a *metameric pair*), and to mismatch by another.

4. The validity of the principles which underlie the mathematical construct of colorimetry, with the most important one being the Trichromatic Generalisation[4] (pp. 118).

To differentiate between the variability introduced by the observers and by the instruments, the terms "psychophysical variability" and "physical variability" are used here, respectively.

In the colour matching experiment, the physical and the psychophysical types of variability are difficult to separate from each other. The stimulus is altered by the observer and is measured by a tele-spectroradiometer (TSR). The same TSR is used to evaluate the temporal and the spatial variability of the stimulus. Thus, the fluctuations of the stimulus, the observer judgments and the TSR have some degree of correlation between them, and are all reflected in the final variability of the tristimulus values. In this study, detailed analysis of the correlation between all the elements of the colour matching experiment setup was not performed. Rather, the variability of each of the elements is reported, and the variability of the tristimulus values as measured by the TSR is assumed to represent the combined variability of the whole system.

The term "tristimulus values" in the following text refers to the values derived directly by means of the colour matching experiment (RGB values), rather than the CIE XYZ tristimulus values - unless otherwise stated.

The reported experiment is the first stage in a project which aims to investigate the whole complex of uncertainty factors in colour matching. Here, results are reported for the variability of colour matching in an experiment of the maximum saturation type, with low illuminance, large field and narrow-band stimuli. The results of an analysis of the failure of the assumption of additivity for the same conditions are reported elsewhere.[1]

Methods

The Instrument

Figure 1 illustrates the Tarrant Visual Colorimeter,[5] which was initially developed for teaching and demonstration purposes, and was adapted for research and used in this study. This instrument provides a vertically-divided bipartite field of 6° in size and allows making matches of the maximum saturation type[4] (p. 379). The test and the primary stimuli are created by means of filtered tungsten lamps projected onto a white diffusing surface on the back wall of the instrument. Four projection units are mounted rigidly, while a system of apertures allows each channel (three primaries and the test) to be switched to either side of the field, thus allowing one set of primary filters to be used for projection on both sides of the bipartite field. The viewing is free and binocular, through the aperture located at the front wall, at the distance of approximately 1500 mm. The luminance at each channel is controlled by varying the electrical current fed to the lamp; change of the light chromaticity as the result of this adjustment is avoided by use of narrow-band interference filters for stimulus generation.

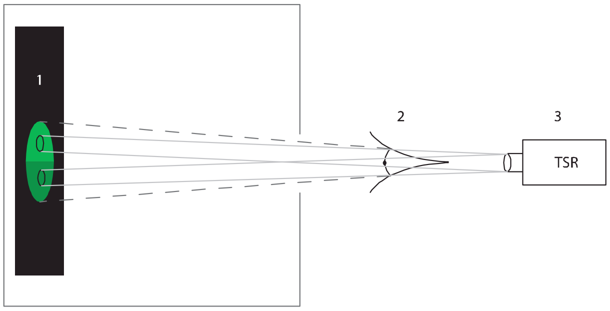

Figure 1. Schematic illustration of the experimental setup (top view). The observer (2) views binocularly the bipartite field on the screen (1). Immediately after the match has been performed, the observer moves aside and a radiometric measurement of both fields is taken by the TSR (3) located just behind the observer's head.

The Stimuli

One set of primaries was chosen to be similar to that of the final primaries in the Stiles and Burch experiment[6] leading to the establishment of the CIE 1964 observer. This set was labeled as "traditional" ("T") and included stimuli at 441 nm, 521 nm and 641 nm. Another set was similar to the so-called prime colours[7] ("PC"), and included stimuli at 451 nm, 530 nm and 603 nm. The test stimuli were at 461 nm, 500 nm, 541 nm, 584 nm, 650 nm and 661 nm. All the primary and the test stimuli were generated using narrow-band interference filters with a bandwidth of 10 nm at half-height. The relative spectral power distributions (SPD) of the stimuli are illustrated in Figure 2.

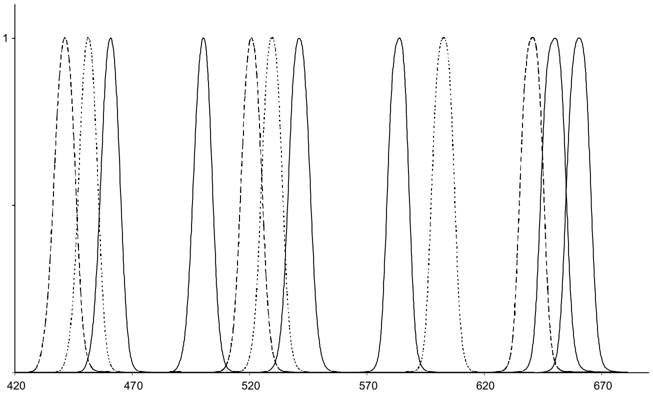

Figure 2. SPD of the experimental stimuli, normalised to have the value of unity at the peak: Long dashes - T primary set; short dashes - PC primary set, solid lines - test stimuli.

The filters at 500 nm, 584 nm and 650 nm – close to those known as "anti-prime" (AP) colours[7] – were originally planned to be used as an additional set of primaries. However, they were abandoned due to complaints by the observers that they are "not intuitive" and are extremely difficult to use in the maximum saturation type of colour matching. Thornton's reported experiment,[7] which used the

AP lights as primaries, was of the Maxwell type and not of the maximum saturation type.

The experiment was conducted at low levels of illumination: the photopic luminance values of the test stimuli were between 0.07 and 3.3 cd/m^2; which corresponds to 3.2 to 125 photopic trolands calculated with the Trezona model of pupil size.[8]

Experimental Procedures

Five observers made repeated trichromatic matches of nine test stimuli with two different sets of primary stimuli. The nine test stimuli included the six listed above, and the three primary stimuli of the other primary set. Four observes repeated each colour match three times, and one observer ("B") performed ten repetitions of every colour match. All the repetitions were made on different days. The data of observer "B" were used for analysing the intra-observer variability, and mean values of all the observers were used to analyse the inter-observer variability.

Radiometric measurements were taken from the test and the matching fields immediately after the observer pronounced a match. These SPD measurements then were used to calculate the tristimulus values, using a procedure similar to the one described by Thornton.[7]

All four CIE SDO deviate functions[9] were used to evaluate the variability of the standard deviate observer. The experimental variances and covariances were converted to CIE 1964 XYZ values by applying a standard model of error propagation.[10] The converted values were used to plot 95% confidence ellipses in the CIE 1964 chromaticity diagram, and to plot the variability results versus spectrum position of the test stimuli. Thus, the experimental results derived with different primary sets and the SDO were all compared in the same primary system.

Experimental Results

Physical Variability

The following sources of physical uncertainty were identified and evaluated:

1. Combined long-term temporal fluctuations of the visual colorimeter, including the optical and the electrical systems and the TSR. This was measured as variations in spectral transmittance of a neutral filter projected on the test side of the bipartite field
2. Bipartite field spatial non-uniformity – separately for each channel and each field side
3. Cross-talk between the colorimeter channels
4. NPL white calibration gauge, as published in the gauge documentation. The NPL gauge was used to calculate the correction curve for TSR measurements.

Table 1 summarises the physical variability in terms of coefficient of variation (CV) which is calculated by means of Equation (1):

$$CV = \frac{s(q)}{\overline{q}} \times 100\% \tag{1}$$

where $s(q)$ is the combined standard deviation and $\overline{q}$ is the mean value.

Table 1: Summary of the variability introduced by the instruments: All the evaluated factors with exception of field uniformity (a); field uniformity (b)

TSR + Visual Colorimeter	1.44%
Cross talk between the colorimeter channels (mean)	1.10%
NPL calibration gauge	1.45%
Combined	2.32%

(a)

R(left)	R(right)	G(left)	G(right)	B(left)	B(right)	Test
0.4%	0.9%	0.9%	0.3%	6.9%	2.5%	1.2%

(b)

Psychophysical Variability

As discussed above, the variability of the tristimulus values is the result of both physical and psychophysical types of variability. In this study, no attempt was made to separate the variability of the two kinds from one another (although their relative magnitude is discussed below). Hence, it might be worthwhile to differentiate between the *observer variability* (psychophysical only) and the *variability of tristimulus values* (psychophysical and physical) in colour matching experiments. However, to keep to the accepted terminology, we use the common notation of intra- and inter-observer variability, while keeping in mind that the physical variability is included with both. The graphical representation of the results in CV terms is can be found in Figure 3, where the intra- and the inter-observer types of variability are compared with each other and with the Standard Deviate Observer in terms of CIE 1964 tristimulus values. Figure 4 illustrates the relation between the intra-observer variability and the SDO in the form of 95% confidence ellipses in the CIE 1964 chromaticity diagram.

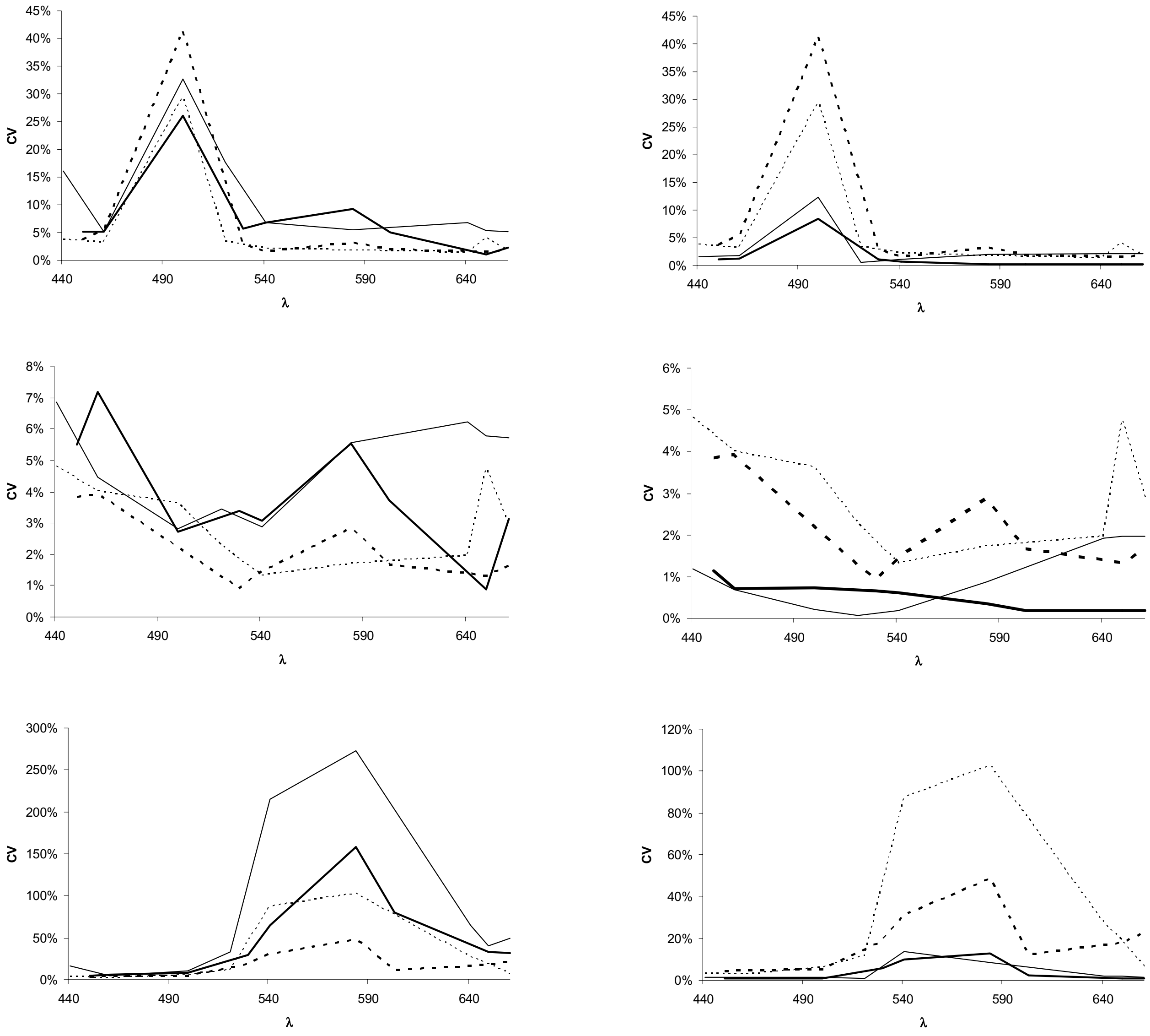

Figure 3. Coefficient of Variation (CV) values plotted against wavelength of the test stimulus; the discrete data points are connected by lines for illustrative purposes only. Left column: intra-observer variations (dashed lines) and inter-observer variations (solid lines). Right column: intra-observer variations (dashed lines) and the Standard Deviate Observer (solid lines). The thick lines stand for the T set, and the thin lines - for the PC set. The plots are for X_{10} (top), Y_{10} (middle) and Z_{10} (bottom) tristimulus values.

Discussion

In the analysis of the variability of the colour matching data collected in our experiment, the following questions were considered:

1. What are, in quantitative terms, the physical and the psychophysical types of variability?
2. What is the relation of the physical and the psychophysical types of variability?
3. What is the relation of the intra- and inter-observer variability?
4. What is the performance of the CIE Standard Deviate Observer[9] in predicting our experimental results?
5. Do the results depend on the choice of the primaries?
6. How does the variability of the tristimulus values of a stimulus depends on its spectral position?

In order to adequately compare colour matching results obtained with different primaries, they need to be transformed to one common primary system; we used system of the CIE 1964 observer. The CIE SDO[9] defines four deviate observers, and also provides a method to estimate the inter-observer colour matching variability for a sample having given spectral power distribution. However, it does not define a method to estimate the variability of the tristimulus values derived in a colour matching experiment. In order to compare the present results from the two primary sets with each other and with the SDO, we needed to evaluate the SDO values for each test stimulus matched with each primary set. This was done using the following procedure:

1. Deriving a transformation from each of the primary sets to the CIE XYZ and to each of the four Standard Deviate Observers.
2. Transforming the mean tristimulus values of all observations to 5 sets of $X_iY_iZ_i$ tristimulus values, where *i* corresponds to one of five observers: the CIE 1964 observer and the four Deviate Observers . The variability within the set of five values was used as the representative of the variability of the SDO.

Instrumental and the Intra-Observer Variability

Comparison of the physical and the psychophysical intra-observer variability was performed on the tristimulus values in the original primary systems (T and PC) without transformation to a common primary system; this is to avoid additional uncertainty being introduced by the transformation. The combined physical variability was found to account on average for about 74% of the total intra-observer variability value. Given the rather low relative value of the combined physical variability - about 3% on average - it can be concluded that the high ratio of the physical to the intra-observer variability is not due to the instruments' poor performance or inappropriate design of the colorimeter, but rather due to the low intra-observer variations. It can be noted that the intra-observer results were generated by a motivated observer who is experienced in colour matching. Hence, the variability introduced by the instruments can be considered to be the most significant part in the variability of the results of a single observer in our experiment. However, as mentioned above, the relation of the physical and psychophysical intra-observer variations is a complex one; it was not analysed in the course of this study and requires further investigation. An example of the complexity of the subject is that, for some tristimulus values, the relative physical variability exceeds the intra-observer variability.

The high figure of 74% is strongly influenced by those matches where the total relative variability was low. In matches where the total relative variability was high, the physical component is much smaller than the psychophysical one; however, these high relative variability values are mainly in the areas of the spectrum where the absolute tristimulus values approach zero – such as X_{10} tristimulus values in the region of 500 nm, or Z_{10} values above 540 nm.

As can be seen in Figure 3, the variability values are different in the two primary sets. The inter-observer variability in the PC set is higher than in the T set in X_{10} values around 500 nm and both intra- and inter-observer variability in Z_{10} values are higher in the PC set in the range of approximately 500-640 nm. For Y_{10}, however, the CV values are rather similar and generally low compared to the other two, with the mean CV value for intra-observer variability being around 2-3% - values which are comparable with the combined instrumental variability, and probably are the lowest obtainable in our experimental conditions.

The SDO defines[9] *"...a method for evaluating the degree of color mismatch for a metameric color pair (…) when an actual observer with normal color vision is substituted for the standard colorimetric observer."* Hence, it aims to represent the variability in colour matching properties between real human observers. It is clear from Figures 3 and 4 that the SDO tends to significantly under-estimate even the variations within single observer; this is true for both primary sets and for all the test stimuli - except in the blue region in PC primary set, where the sizes of the ellipses are rather similar.

The relation of the experimental intra-observer variability with the SDO is somewhat different in the two primary sets. In the PC set both the size and the orientation of the experimental and SDO ellipses differ for all the stimuli with exception of 441 and 461 nm, for which the sizes (but not the orientation) of the ellipses are fairly similar (Figure 4). The sizes of the ellipses in T set are significantly different as well, however, the orientation of the corresponding ellipses in the two sets is remarkably similar, especially in the green-yellow region; a clue for the explanation of this can lie in the similarity of the T primary set with the one of Stiles and Burch set – on which the SDO is based.

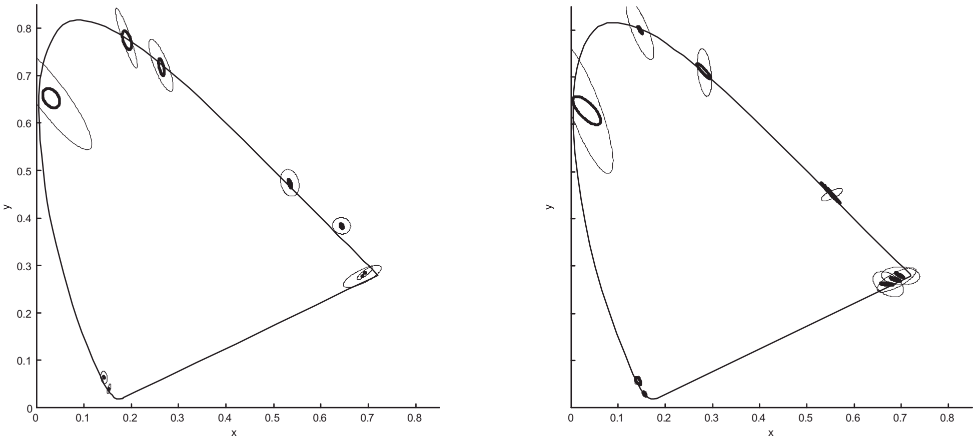

Figure 4. 95% confidence ellipses (scaled by the ratio of 10) of intra-observer colour matching variations in the CIE 1964 chromaticity diagram. The light lines ellipses represent the experimental data, and heavy line ellipses represent the SDO. The position of the ellipses was adjusted so the centres of each pair would coincide. Left: T primary set; Right: PC primary set. As can be seen, some values fall outside of the locus of the monochromatic stimuli. This could be the result of the uncertainty introduced by the process of the transformation of the primary space, or the additivity failures in the experimental results. [1]

On average, the SDO accounts for about 20% of the intra-observer experimental variability in the T primary set, and about 40% in the PC primary set. Alfvin and Fairchild[2] reported the results of comparing the experimental variability with the SDO prediction for matching a broadband cyan transparency with CRT primaries. The data were presented as 95% confidence ellipses in the CIELAB a*b* plane. Although the experimental conditions were different, the ratio between the SDO and intra-observer ellipses (as evaluated from the plots) is of the same order as presented here - approximately 25%. On the other hand, our results contradict those reported by North and Fairchild.[11] They reported that the variability of the SDO and that of single observer are of the same order. Note however that in all cases, as discussed above, the reported variability includes the physical variability.

Inter-Observer Variability

As expected, the variability between different observers is higher than the variability within a single observer. On average, the ratio between the intra-observer and inter-observer variations is about 50% for the PC primary set and 30% for the T set. In one case, however, the intra-observer variability exceeds the inter-observer (in CV terms) – for X_{10} at 500 nm for the T primary set. Thus, the intra-observer variability is a significant part of the total variability of colour matching in the reported results.

The differences between the intra- and the inter-observer variability are illustrated for the 541 nm stimulus using the PC primaries in Figure 5, where the 95% confidence ellipses are plotted in the CIE 1964 chromaticity diagram. For this particular stimulus matched by the PC primaries - the ratios of the intra-observer to the inter-observer variation are 0.30, 0.46 and 0.40 respectively for the X_{10}, Y_{10} and Z_{10} tristimulus values. Figures of the same order were reported by Alfvin and Fairchild[2]: intra-observer variability in their data was approximately 50% of the inter-observer variability. Nimeroff[12] analysed the Stiles and Burch large field colour matching data and reported significantly larger differences: i.e. the inter-observer variability is approximately 5.7 times larger than the intra-observer one.

It was found that the SDO under-predicts the inter-observer experimental variability to a significant degree. On average, it accounts for only 16% of the experimental inter-observer variability in the PC set, and 12% in the T set. North and Fairchild[11] did not report numerical data in their paper, however they also stated that the SDO significantly under-predicts the observed variations in colour-matching data in their experiment.

Conclusions

For the given experimental conditions, the following conclusions are drawn:

1. The uncertainty introduced by physical sources (instruments, stimuli, etc.) accounts on average for about 74% of the intra-observer variability in our colour-matching data.
2. Inter-observer variability was found on average to be twice as large as intra-observer variability, which agrees with the publication by Alfvin and Fairchild.[2]
3. The variability of colour matching was found to be dependant on the primary set, with generally larger – in relative terms – variability for the prime colours set than for the traditional (Stiles and Burch[6]) one.
4. The CIE SDO was found to significantly under-predict the observer variations of colour matching data in our study, accounting on average for 30% of intra-observer variability and 15% of inter-observer variability

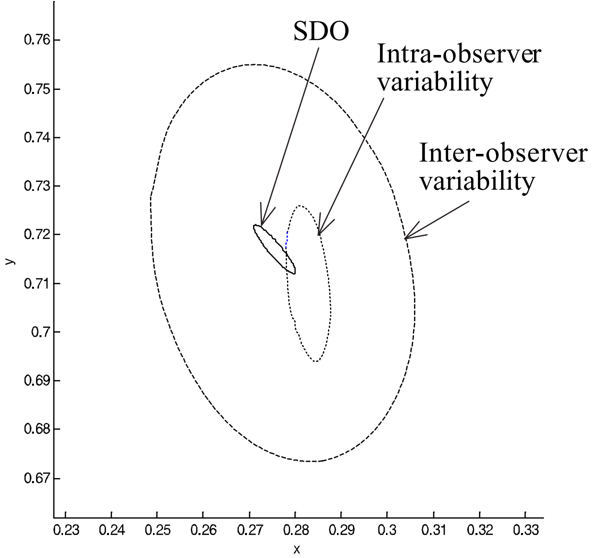

Figure 5. 95% confidence ellipses of the 541 nm stimulus matched in the PC primary set, plotted in the $x_{10}y_{10}$ chromaticity diagram. The plot shows the relationship between the CIE Standard Deviate Observer and the intra and inter-observer variability.

References

1. Oicherman B., Luo R. M., Robertson A., Tarrant A. Experimental Verification of Colorimetric Additivity Assumption. 10th Congress of the International Colour Association; 2005; Granada, Spain.
2. Alfvin R. A., Fairchild M. D. Observer variability in metameric color matches using color reproduction media. Color Research and Application 1997;22(3):174-188.
3. ISO/CIE, Colorimetric observers, joint ISO/CIE standard 10527(E). Vienna: ISO/CIE; 1991.
4. Wyszecki G., Stiles W. S. Color Science, Concepts and Methods, Quantitative Data and Formulae: John Wiley and Sons, Inc.; 1982.
5. Tarrant A. W. S. Visual colour matching equipment for teaching and research. 10th Congress of the International Colour Accotiation, Proceedings of SPIE; 2002. p 985-988.
6. Stiles W. S., Burch J. M. NPL colour-matching investigation:final report (1958). Optica Acta 1959;6:1.
7. Thornton W. A. Toward a more accurate and extensible colorimetry, Part I: Introduction. The visual colorimeter-spectroradiometer. Experimental results. Color Research and Application 1992;17:79-122.
8. Trezona P. W. Luminance level conversions to assist lighting engineers to use fundamental visual data. Lighting Research and Technology 1983;15(Light. Res. Technol. 15, 83, 1983):83.
9. CIE, Special metamerism index: change in observer. Central Bureau of the CIE, Vienna, Austria; 1989.
10. Wyszecki G. Propagation of errors in colorimetric transformations. Journal of the Optical Society of America A 1959;49:389.
11. North A. D., Fairchild M. D75. Measuring Color-Matching Functions, Part II: New data for accessing observer metamerism. Color Research and Application 1993;18:163-170.

12. Nimeroff I., Rosenblatt J. R., Dannemiller M. C. Variability of spectral tristimulus values. Journal of Research of the National Bureau of Standards - A 1961;65(6):475-483.

Author Biographies

Boris Oicherman is currently a PhD student in the Colour & Imaging Group in the Department of Colour and Polymer Chemistry, University of Leeds; doing his research into the subject of "Uncertainty of Colour Matching". He studied Printing Technologies, Computer Programming, and completed his Masters degree in Digital Colour Imaging with distinction in 2003. Prior to his Master's studies, Boris worked in the R&D department of the Indigo Electronic Printing Systems company (now a part of the Hewlett Packard company), working on the implementation of colour management systems for digital printing presses.

Ronnier Luo is Professor of Colour and Imaging Science, and the head of the Colour & Imaging Group at the Department of Colour and Polymer Chemistry, University of Leeds. He is the author of numerous publications in the field of colour appearance modelling, colour difference evaluation, colour management and digital imaging. He chairs a Technical Committee of CIE TC 8-02. He is a a recipient of the Centenary Medal from the Society of Dyers and Colourists in 2004, Royal Photographic Society's Davies Medal in 2003, Bartleson Research Award in 1994 and the Silver Medal of the Worshipful Company of Dyers in 1987.

Arthur Tarrant studied physics at the University of London (Chelsea) and first worked at the National Physical Laboratory, working with precise measurements in spectrophotometry, spectroradiometry, and colorimetry. He moved to the University of Surrey in 1958 as a lecturer in spectroscopy, and was awarded his PhD in 1967 for his work on daylight. His main researches have covered the spectral composition of daylight, colour in visual displays, scattered light in optical instruments, lighting, colour names and visual colour matching instruments. Although now supposedly retired, he continues both teaching and research in colour science and optics

Alan Robertson retired in 2000 after spending 35 years at the National Research Council of Canada working on colour science and metrology. He is a former President of the International Color Association (AIC) and Vice President of the International Commission on Illumination (CIE). He received his Ph.D. in 1965 from the University of London where he studied under David Wright. He currently acts as a consultant to the Colour & Imaging Group of the University of Leeds. E-mail: ccdbo@leeds.ac.uk