Journal of Neuroscience Methods 215 (2013) 156–160

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Basic Neuroscience
Short communication

Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity

Nicolas Rothen[a,b,*], Anil K. Seth[a,c], Christoph Witzel[b], Jamie Ward[a,b]

[a] *Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK*
[b] *Department of Psychology, University of Sussex, Brighton, UK*
[c] *Department of Informatics, University of Sussex, Brighton, UK*

HIGHLIGHTS

- ▶ We report an optimised method for diagnosing synaesthesia.
- ▶ We provide several measures superior to those commonly used.
- ▶ We provide reliable cut-off values for the diagnosis of synaesthesia.

ARTICLE INFO

Article history:
Received 19 December 2012
Received in revised form 7 February 2013
Accepted 12 February 2013

Keywords:
Synaesthesia
Consistency
Diagnosis
Sensitivity
Specificity
Binary classification
ROC

ABSTRACT

The most commonly used method for formally assessing grapheme-colour synaesthesia (i.e., experiencing colours in response to letter and/or number stimuli) involves selecting colours from a large colour palette on several occasions and measuring consistency of the colours selected. However, the ability to diagnose synaesthesia using this method depends on several factors that have not been directly contrasted. These include the type of colour space used (e.g., RGB, HSV, CIELUV, CIELAB) and different measures of consistency (e.g., city block and Euclidean distance in colour space). This study aims to find the most reliable way of diagnosing grapheme-colour synaesthesia based on maximising sensitivity (i.e., ability of a test to identify true synaesthetes) and specificity (i.e., ability of a test to identify true non-synaesthetes). We show, applying ROC (Receiver Operating Characteristics) to binary classification of a large sample of self-declared synaesthetes and non-synaesthetes, that the consistency criterion (i.e., cut-off value) for diagnosing synaesthesia is considerably higher than the current standard in the field. We also show that methods based on perceptual CIELUV and CIELAB colour models (rather than RGB and HSV colour representations) and Euclidean distances offer an even greater sensitivity and specificity than most currently used measures. Together, these findings offer improved heuristics for the behavioural assessment of grapheme-colour synaesthesia.

1. Introduction

Synaesthesia is a phenomenon characterised by involuntary and unusual associations between and within different modalities (e.g., hearing colours, tasting words). The developmental form of the condition is associated with structural and functional differences in the brain (Rouw et al., 2011) and a specific cognitive profile (e.g., Rothen et al., 2012). A hallmark of synaesthesia is that an inducing stimulus (i.e., synaesthetic inducer) is consistently associated with a particular secondary experience (i.e., synaesthetic concurrent). Thus, a common standard in synaesthesia research is to use a measure of consistency as an objective diagnostic criterion for synaesthesia.

However, there are various ways of measuring consistency. In grapheme-colour synaesthesia, measures of consistency can be based on different colour spaces and can utilise different distance metrics within these spaces. To date, direct comparisons among these methods have been lacking. Here, we determine the optimal way of measuring consistency for grapheme-colour synaesthesia to jointly maximise sensitivity (i.e., ability to identify true synaesthetes) and specificity (i.e., ability to identify non-synaesthetes).

Baron-Cohen et al. (1987) first introduced the so-called test of "genuineness", also referred to as the test of consistency (cf. Asher et al., 2006 for a revised version of the test). In the original version of the test, the participant is presented with several instances of potential synaesthetic inducers and has to report a detailed

* Corresponding author at: School of Psychology, Sackler Centre for Consciousness Science, University of Sussex, Falmer, Brighton, BN1 9QH, UK. Tel.: +44 1273 876649.
E-mail address: nicolas.rothen@gmail.com (N. Rothen).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.02.009

description of the resulting (synaesthetic) associations. The test is repeated after a substantial gap, occurring several days to weeks later, and at the initial test, the participant is not informed that they will be asked to repeat the test. The consistency of inducer-concurrent pairings is assessed by independent raters and expressed as percentage of consistent pairings. While synaesthetes can rely on their perceptual-like concurrent experiences when performing such a test, non-synaesthetes are entirely dependent on their memory or other strategies. As such, synaesthetes generally outperform non-synaesthetes on tests of consistency.

Contemporary research has tended to use computer-based colour pickers rather than verbal colour descriptions, and the test-retest interval is (for convenience) performed within a single session. Consistency is then calculated according to a predefined algorithm. Given that most forms of synaesthesia involve colour photisms as synaesthetic concurrents, the algorithm is usually based on distances (e.g., Nikolić et al., 2011) or correlations (e.g., Witthoft and Winawer, 2006) in colour space providing a value of overall consistency for a set of trials.

To date, there is only one published attempt to create a standardised method to present stimuli and quantify consistency (Eagleman et al., 2007). In this study, 15 self-reported synaesthetes and 15 non-synaesthete controls were presented with the graphemes A–Z and 0–9 in random order three times each. Participants had to pick one colour represented in HSV (Hue, Saturation, Value) colour space for each grapheme.

Consistency was then calculated based on the average of 'city block' (e.g., Krause, 1987) distances in RGB (Red, Green, Blue; each from 0 to 1) space (Eq. (1)). Consistency values of the synaesthetes and controls were continuously distributed between approximately 0.2 and 3.2 (with smaller numbers denoting greater test-retest consistency). Even though RGB values are device dependent and city block distances in RGB space do not represent perceptual distances, this approach was powerful enough to detect the strong differences between synaesthetes and non-synaesthetes, with all synaesthetes performing below a consistency value of 1 and all controls above a consistency value of 1. Thus, 1 was defined as the cut-off value. This test is freely available online to synaesthetes and researchers (www.synesthete.org) and is widely used in publications (e.g., Ward et al., 2010; Brang et al., 2011). Although this approach provides a straightforward and simple method of assessing potential synaesthetic experiences online, it is desirable to optimise classification accuracy. For this reason, we compared different colour specifications (in terms of colour spaces and distance metrics) to optimise the diagnosis of grapheme-colour synaesthesia via the above specified standardised internet-based test. We hypothesised that colour specifications that better represent perceptual colour differences should provide higher diagnostic accuracy in terms of sensitivity and specificity. It has been previously suggested that Euclidean distances in CIELUV and CIELAB colour space provide approximate estimations of perceptual colour differences (e.g., Fairchild, 1998). Euclidean distances refer to the linear distances in these colour spaces. Notably, CIELUV is generally used for emitted colours (i.e., monitors, etc.) and CIELAB for printed colours. Hence, we expected that colour consistency measurements based on these colour specifications should be more accurate in classifying synaesthetes and non-synaesthetes than measurements based on other colour spaces, such as RGB and HSV, and other difference measures, such as city block distances, that have little psychological meaning.

High accuracy implies high sensitivity and high specificity. For this reason, it was our aim to evaluate the discriminative performance of the specified colour spaces and distance measures through the application of ROC curve analysis in binary classification of synaesthetes and non-synaesthetes and hence, to provide more reliable consistency criteria for the diagnosis of grapheme-colour synaesthesia. It is important to note that we did not attempt to specify colour representation as precisely as possible. Rather, we were looking for the best and most efficient way to use the internet-based standardised grapheme-colour consistency test.

2. Method

2.1. Participants

We gathered the data provided by 154 self-declared synaesthetes who gave us access to their performance on the standardised grapheme-colour consistency test (letters A–Z and numbers 0–9) of Eagleman et al. (2007). Only self-declared synaesthetes who picked a colour for all three trials of at least five different graphemes were included. Moreover, if there was more than one dataset for a specific synaesthete, only the first was included. A total of 144 datasets met our inclusion criteria. Potential control participants were interrogated about whether they experienced grapheme-colour synaesthesia. Only people who did not report any instances of grapheme-colour synaesthesia were recruited as controls. We tested a total of 96 controls with the standardised grapheme-colour consistency test. Note that the battery does not ask participants to enter demographic details; however, with the exception of four controls, all participants were aged 18 years or over. The youngest of the controls was aged 11 years. All participants aged less than 18 years were accompanied by one of their parents.

2.2. Materials

We used the standardised grapheme-colour consistency test, which is accessible via the Internet (www.synesthete.org). This test uses the letters from A to Z and numbers from 0 to 9 as synaesthetic inducers and possible synaesthetic colour concurrents in HSV space (that is, colours were represented on a plane varying in lightness along the vertical axis and in saturation along the horizontal axis with a separate bar to adjust hue).

2.3. Procedure

Participants conducted the standardised grapheme-colour consistency test via the Internet. Each participant was presented with the graphemes A–Z and 0–9 three times in randomised order (i.e., 108 trials). Self-declared synaesthetes were tested remotely and followed the original instructions provided by the test to choose the colour which most closely resembles the synaesthetic colour associated with the presented grapheme. Controls were tested individually, in person. To obtain a broader range of performance, controls were randomly assigned to one of three conditions. They were instructed by an experimenter either (1) to try to memorise and always choose the same colour for one particular grapheme, (2) to choose the colour which they think goes best with a presented grapheme or (3) to follow a mix of both instructions. All three instructions led to very similar results[1].

The data for each of the different instruction groups can be found in Supplementary Table 1. To acknowledge that synaesthetes were tested with different computer setups, each control was randomly assigned to one of six different computer setups (including different monitors and graphics cards) for this task.

[1] Note that binary classification failed to provide reliable discrimination between the different control groups (i.e., different instructions). That is, DP was always lower than 1, with the exception of CIELUV squared Euclidean mean, when comparing the memory instruction to the intuition instruction (DP = 1.0104 for both), which suggests that there are serious limitations to reliable classification between the various control groups.

2.4. Analysis

The analysis was based on 4 representations of colour (RGB, HSV, CIELUV, and CIELAB), 3 measures of distance (city block, Euclidean, and Euclidean-squared), and 2 summary statistics (grand mean and grand median). Thus, 24 conditions were tested in total.

2.4.1. Colour representations/models

We used RGB and HSV colour representations and two models of human colour perception, CIELUV and CIELAB, as basis to calculate consistency scores. CIELUV and CIELAB each consist of 3 dimensions (L*, u*, and v* and L*, a*, and b*, respectively). The L* axis represents perceived lightness and is the same in both colour spaces. The u* and the a* axis contrast green (negative values) against red (positive values). The v* and the b* axis contrast blue (negative) against yellow (positive).

Colour values for RGB space were provided as part of the output of the standardised consistency test (these values were rescaled so that each dimension lies between 0 and 1; see also Eagleman et al., 2007). RGB values were converted to corresponding values in HSV (MATLAB R2011b), leading to values from 0 to 1 for each of the dimensions H, S, and V. Next, we converted RGB values into CIELUV and CIELAB values. First, we linearised RGB values by applying inverted gamma functions (gamma compression), and converted these linear RGBs into tristimulus values (XYZ; Brainard et al., 2002). Based on these XYZ values, we calculated CIELUV and CIELAB (Hunt and Pointer, 2011, p. 55).

Obviously, we could not measure xyY primary weights and gamma distributions of the specific monitors used by the participants because data collection was via the Internet. To obtain monitor specifications that are most representative of a range of random monitors, we used standard RGB (sRGB; Stokes et al., 1996). For this standard CRT monitor, the white-point corresponds to standard illuminant D65 with the chromaticity coordinates and luminance (xyY) of [0.3127, 0.3290, 80] cd/m^2. The xyY of the primaries are red = [0.640, 0.330, 17.0], green = [0.300, 0.600, 57.2], and blue = [0.150, 0.060, 5.8]. The standard gamma function for all three RGB values is:

(i) RGBl = 1/(1+0.055) × (RGBg + 0.055)^2.4 for RGBg ≥ 0.04045 i.e., for discrete RGBg > 0, and
(ii) RGBl = 0 for discrete RGBg = 0;

where RGBl = linear RGB, RGBg = gamma distributed RGB, and RGB corresponds to R, G, and B. Because the background of the standardised synaesthesia test was set to RGB [255,255,255], we used the monitor-white-point as the adaptation point for the CIELUV/CIELAB conversion.

2.4.2. Consistency measures

Consistency for a grapheme was calculated only if a colour was chosen for all three trials. Following Eagleman et al. (2007), we calculated city block distances in RGB, HSV, CIELUV, and CIELAB colour space for individual graphemes according to Eq. (1) (illustrated here using RGB colour dimensions). We next calculated Euclidean distances in RGB, HSV, CIELUV, and CIELAB colour space for individual graphemes according to Eq. (2) (again illustrated using RGB colour dimensions). This procedure was repeated to calculate squared Euclidean distances (equivalent to Eq. (3) without the square root term) in order to place increasing weight on larger distances. To obtain a consistency measure representing the entire grapheme set (i.e., letters and numbers), we calculated the grand average and the grand median of the ensemble of consistency values.

$$d = \sum_{i=\{1,2,3\}} \left|R_1 - R_2\right| + \left|G_1 - G_2\right| + \left|B_1 - B_2\right| \tag{1}$$

$$d = \sum_{i=\{1,2,3\}} \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2} \tag{2}$$

where i refers to the number running over all combinations between the 3 trials per grapheme (i.e., trial 1 and trial 2, trial 2 and trial 3, trial 3 and trial 1).

2.4.3. Binary classification

We applied ROC curve analysis to binary classification of self-declared synaesthetes and non-synaesthete controls to determine, in each condition, the cut-off value maximising sensitivity and specificity for the given samples (Cardillo, 2006, 2008). For each condition, sensitivity and specificity rates were calculated for all unique consistency values, allowing identification of an optimal cut-off value: the point with the highest true positive rate and lowest false positive rate (see Fig. 1 top left). It is noteworthy that with this approach, sensitivity and specificity are affected by the specific distribution of performance in the performance range of each group rather than unequal sample sizes.

Crucially, the method enables a quantitative comparison of the discriminatory performance of different colour models in combination with different ways of calculating consistency. Discriminatory performance in each condition can be expressed as a single value, referred to as Discriminant Power (DP; also known as test effectiveness; Eq. (3)) associated with the corresponding optimal cut-off for that condition, and which can be interpreted as the standardised distance between the means of two populations. DP values around 1 are regarded as not effective in discriminating between two samples. DP values around 3 are regarded as effective in discriminating between two samples (Cardillo, 2006; cf. also Sokolova et al., 2006).

As an additional measure, the more commonly used Area Under the Curve (AUC) is also provided. AUC is the probability of a given consistency measure and its corresponding cut-off correctly classifying a randomly drawn pair of a synaesthete and a control.

$$DP = \frac{\sqrt{3}}{\pi}(\log X + \log Y) \tag{3}$$

where: X = sensitivity/(1−sensitivity) and Y = specificity/(1−specificity).

3. Results

Table 1 summarises the mean, SD, sensitivity, specificity, cut-off value, discriminant power, and area under the curve for the 24 conditions tested to discriminate between synaesthetes and controls. It is important to mention that the numerical values of the different measures are not directly comparable because they are based on different calculations (i.e., city block, Euclidean, and squared Euclidean) and different colour spaces (i.e., RGB, HSV, CIELUV, and CIELAB). The results in Table 1 are first sorted on the basis of DP and then on the basis of AUC, with higher values indicating better discrimination abilities for the associated measure. The widely used method of Eagleman et al. (2007), means of RGB city block distances, was ranked 13th. Mean CIELUV and CIELAB distances were generally ranked best among the different colour-space/dimension alternatives. The best overall DP was obtained using the mean of Euclidean distances in CIELUV colour space, specifying a cut-off value of 135. Fig. 1 shows the distribution of scores for the three best-performing measures and the original method (Eagleman et al., 2007). Note that specificity is often higher than sensitivity because more synaesthetes performed similarly to the controls than controls performed similarly to the synaesthetes. It is notable that HSV conditions performed particularly poorly. Split-half reliability testing led to very similar results (Supplementary Table 2).

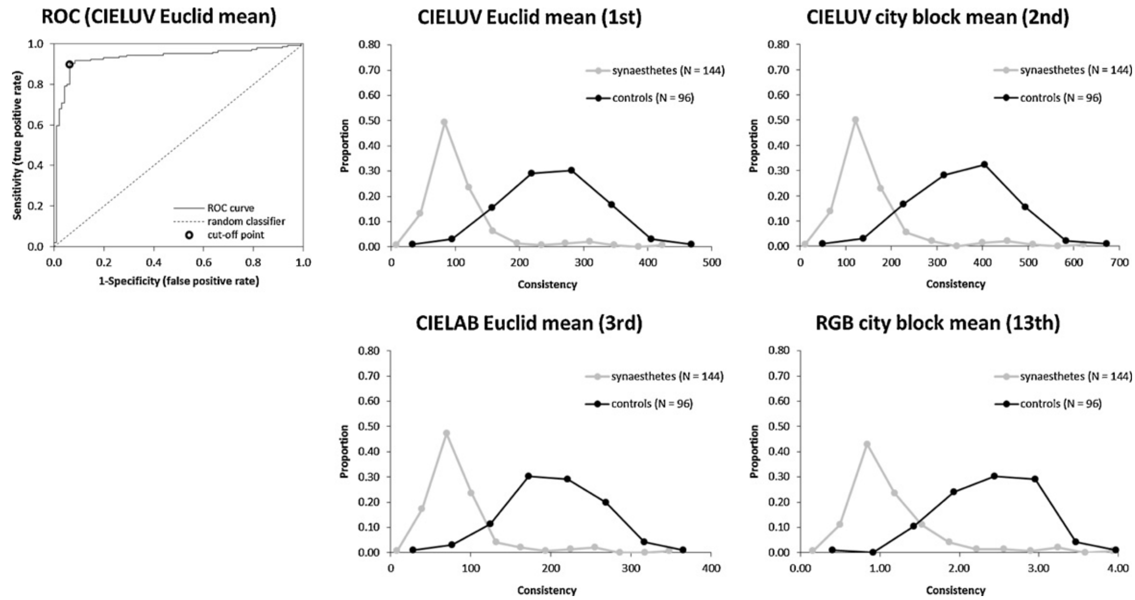

Fig. 1. Top left corner: An example ROC curve. The optimal cut-off value is defined as the point that results in the highest true positive rate (i.e., is highest on the vertical axis) and the lowest false positive rate (i.e., furthest to the left on the horizontal axis). The remaining subfigures depict the distribution of consistency scores of the synaesthetes and the controls for the top three effective discriminant measures – CIELUV Euclidean mean, CIELUV city block mean, CIELAB Euclidean mean – and for the original method (Eagleman et al., 2007), RGB city block mean, which ranked 13th. The cut-off value as defined in Table 1 for synaesthesia versus control is taken as the intersection of these distributions.

To validate the conversions from RGB to CIELUV and CIELAB described above, eight synaesthetes and eleven controls were tested on a calibrated Cathode Ray Tube (CRT) monitor under optimal lighting conditions. The RGB data from these participants were then converted to CIELUV/CIELAB using either the specifications of the calibrated monitor or the previously used standard specifications (sRGB). On the calibrated set-up, we measured monitor specifications with a ColorCal (Cambridge Research Systems, LTD, http://www.crsltd.com/tools-for-vision-science/light-measurement-display-calibration) colorimeter for a typical CRT monitor. The chromaticity coordinates and luminances of this monitor were [0.626, 0.337, 13.01], [0.283, 0.612, 47.84], and [0.151, 0.071, 7.96] for the R, G, B primaries, respectively. The measured white-point of this monitor (RGB = [255 255 255]) was xyY = [0.278, 0.302, 69.75]. The ROC procedure was applied to both datasets in all 24 conditions. Both setups led to nearly identical results (Supplementary Table 3). Hence, accurate classification does not depend critically on whether the monitor has been calibrated.

4. Discussion

We have described a comparative analysis of a variety of colour-consistency tests for grapheme-colour synaesthesia, examining both different colour spaces and different distance metrics within these spaces and taking advantage of ROC curve analyses to identify those tests that maximised sensitivity and specificity by determining an optimal cut-off value for each condition. Our study further extends previous work by considering a comparatively large sample of self-declared synaesthetes. One recent suggestion of Simner (2011) is that high consistency may be a self-fulfilling prophecy (i.e., simply because our selection criterion for synaesthesia specifies high consistency), but our current research, considering self-declared synaesthetes, speaks against this view. There are very few self-declared adult synaesthetes who appear to be inconsistent.

A first result is that the popular method of Eagleman et al. (2007) was not the best among the measures we tested, ranking 13th overall. With this condition, our ROC procedure on a large sample generated a cut-off value of 1.43 for separating synaesthetes from non-synaesthetes, which is considerably higher than the value of 1.0 recommended by Eagleman et al. (2007). Of course, the cut-off value is not fixed but depends on whether the researcher is concerned with specificity, sensitivity, or both (there are instances in which a more conservative cut-off could be justified).

However, the cut-off value of 1.43 could be usefully adopted by future researchers who use the current online battery to diagnose grapheme-colour synaesthesia. In general, the most reliable ways of discriminating between synaesthetes and controls on the basis of consistency were based on colour differences in CIELUV and CIELAB space, respectively. The HSV colour space did not fare well. The superiority of CIELUV/CIELAB space may be partially explained by the observation that these colour spaces are coarsely perceptually uniform whereas RGB (e.g., Glasbey et al., 2007) and HSV colour spaces are not perceptual at all. Note that this argument also lends support to the perceptual quality of synaesthetic colour experiences. Overall, Euclidean distances (which are perceptually relevant) seemed to discriminate slightly better than city block and squared Euclidean distances (which are not perceptually relevant).

Consistency measures based on the mean across items fared much better than consistency measures based on the median in discriminating synaesthetes from controls. Extreme inconsistencies appear to be helpful in discriminating the groups and, as expected, these extremes tend to affect the mean more than the median.

In summary, our recommendations for researchers when diagnosing grapheme-colour synaesthesia via colour consistency tests under less controlled conditions are (1) to transform RGB values to representations in CIELUV space, (2) to calculate the Euclidean distance between the colours selected for each specific grapheme in that colour space, and (3) to compute the mean of the distances to determine whether this value falls below the suggested diagnostic cut-off value of 135. To standardise instructions for controls on the standardised grapheme-colour consistency test, we recommend using the following wording "Always choose the colour that you think goes best with a particular letter or number, memorise it, and then choose the same colour again when the grapheme reappears." No restrictions should be made regarding the "no colour" option,

Table 1
Synaesthesia versus *control*. Summary of measures used to discriminate between synaesthetes and controls. Measures are sorted first according to DP (4th column) and then according to the AUC (5th column). Higher values indicate better discrimination abilities. Sensitivity and specificity are given in per cent. Mean, SD, and cut-off values are dependent on the specific calculation method and colour space and thus, are not directly comparable. Note, the original measure ranked 13th and is highlighted.

Synaesthesia vs control												
Space	Distance	Descriptive	DP	AUC	Ranking	Mean (syn)	Mean (con)	SD (syn)	SD (con)	Sensitivity	Specificity	Cut-off
CIELUV	Euclid	mean	2.7217	0.9308	1	85.51	219.38	58.27	68.87	90	94	135.30
CIELUV	city block	mean	2.7217	0.9306	2	123.04	313.80	83.79	97.51	90	94	192.96
CIELAB	Euclid	mean	2.6756	0.9304	3	69.46	177.05	46.52	53.60	91	93	109.20
CIELUV	sq Euclid	mean	2.6305	0.9350	4	6418.21	27965.69	9117.78	12745.00	90	93	12229.15
CIELAB	city block	mean	2.5957	0.9296	5	104.18	263.35	69.82	79.18	91	92	168.02
CIELAB	sq Euclid	mean	2.5605	0.9361	6	4152.54	18211.02	6004.18	8014.19	92	90	9282.87
RGB	sq Euclid	mean	2.4747	0.9296	7	0.30	1.07	0.36	0.40	88	93	0.55
CIELUV	Euclid	median	2.4078	0.9117	8	70.24	206.96	60.74	91.29	86	93	89.76
CIELUV	sq Euclid	median	2.4078	0.9114	9	3312.02	20249.66	8309.16	15841.39	86	93	3085.03
RGB	Euclid	mean	2.3595	0.9236	10	0.64	1.43	0.37	0.39	88	91	0.96
CIELUV	city block	median	2.2894	0.9099	11	101.73	298.31	88.16	131.12	87	91	137.09
CIELAB	Euclid	median	2.2518	0.9084	12	57.08	165.45	47.91	69.65	92	84	87.69
RGB	city block	mean	2.2360	0.9232	13	0.95	2.13	0.56	0.59	88	89	1.43
CIELAB	sq Euclid	median	2.1815	0.9085	14	2134.11	12653.50	5400.26	9066.18	88	88	2237.83
CIELAB	city block	median	2.1686	0.9058	15	85.92	242.35	72.79	101.40	89	86	118.21
RGB	sq Euclid	median	2.0607	0.9005	16	0.17	0.85	0.35	0.54	87	86	0.23
RGB	Euclid	median	2.0147	0.9006	17	0.55	1.39	0.39	0.52	83	89	0.71
HSV	Euclid	mean	1.9196	0.8796	18	0.74	1.23	0.31	0.28	77	91	0.93
RGB	city block	median	1.9190	0.8962	19	0.80	2.02	0.58	0.78	85	85	1.08
HSV	city block	mean	1.8981	0.8866	20	0.98	1.74	0.45	0.45	76	91	1.21
HSV	Euclid	median	1.8171	0.8681	21	0.57	1.16	0.35	0.39	83	84	0.82
HSV	sq Euclid	median	1.8171	0.8675	22	0.17	0.57	0.24	0.32	83	84	0.27
HSV	city block	median	1.7623	0.8700	23	0.78	1.63	0.47	0.58	84	82	1.13
HSV	sq Euclid	mean	1.7330	0.8634	24	0.42	0.81	0.28	0.25	75	89	0.54

which involves a button to indicate that the presented grapheme does not have a colour association.

Although our research focused specifically on grapheme-colour synaesthesia, we expect that similar principles (i.e., based on a mean of Euclidean distances in CIELUV/CIELAB space) should apply to all other types of synaesthesia involving colour. However, additional empirical research is needed to establish where the optimal cut-off value lies between synaesthetes and controls in other domains. Different inducers (e.g., musical notes, numbers, or months of the year) may require different cut-off values because controls may differ in their ability to generate high levels of consistency depending on the type of material. Similarly, long-term consistency (with longer time intervals between test and retest) might result in different (possibly higher) cut-off values, although we would not expect the various methods to be different in their effectiveness. There is also a trend for children to be less consistent on such tests (e.g., Simner et al., 2009). The method that we have adopted based on ROC analyses will, however, be applicable in these other domains.

Acknowledgements

We would like to thank Aimée Davenport, Sabine Oligschläger, Kimberley Warne, and Daniel Coolbear for their help with data collection.

We gratefully acknowledge support from the Swiss National Science Foundation (Grant PBBEP1_133498) and the Holcim Foundation for the Advancement of Scientific Research to NR, the Dr. Mortimer and Theresa Sackler Foundation to AKS and NR, and the Postdoc-Programme of the German Academic Exchange Service (DAAD) to CW.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2013.02.009.

References

Asher JE, Aitken MRF, Farooqi N, Kurmani S, Baron-Cohen S. Diagnosing and phenotyping visual synaesthesia: a preliminary evaluation of the Revised Test of Genuineness (TOG-R). Cortex 2006;42:137–46.

Baron-Cohen S, Wyke MA, Binnie C. Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. Perception 1987;16: 761–7.

Brainard DH, Pelli DG, Robson T. Display characterization. In: Hornak J, editor. The encyclopedia of imaging science and technology. New York: Wiley; 2002. p. 172–88.

Brang D, Rouw R, Ramachandran VS, Coulson S. Similarly shaped letters evoke similar colors in grapheme–color synesthesia. Neuropsychologia 2011;49:1355–8.

Cardillo, G. Clinical test performance: the performance of a clinical test based on the Bayes theorem. Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/12705, 2006 [14.03.12].

Cardillo, G. ROC curve: compute a receiver operating characteristics curve. Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/19950, 2008 [14.03.12].

Eagleman DM, Kagan AD, Nelson SS, Sagaram D, Sarma AK. A standardized test battery for the study of synesthesia. J Neurosci Methods 2007;159: 139–45.

Fairchild MD. Color Appearance Models. Reading. Massachusetts: Addison-Wesley; 1998.

Glasbey C, van der Heijden G, Toh VFK, Gray A. Colour displays for categorical images. Color Res Appl 2007;32:304–9.

Hunt RWG, Pointer MR. Measuring colour. 4th ed. Chichester, UK: John Wiley & Sons; 2011.

Krause EF. Taxicab geometry: an adventure in non-euclidean geometry. New York: Dover Publications; 1987.

Nikolić D, Jürgens UM, Rothen N, Meier B, Mroczko A. Swimming-style synesthesia. Cortex 2011;47:874–9.

Rothen N, Meier B, Ward J. Enhanced memory ability: insights from synaesthesia. Neurosci Biobehav Rev 2012;36:1952–63.

Rouw R, Scholte HS, Colizoli O. Brain areas involved in synaesthesia: a review. J Neuropsychol 2011;5:214–42.

Simner J. Defining synaesthesia. Br J Psychol 2011;103:1–15.

Simner J, Harrold J, Creed H, Monro L, Foulkes L. Early detection of markers for synaesthesia in childhood populations. Brain 2009;132:57–64.

Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and roc: a family of discriminant measures for performance evaluation. In: Sattar A, Kang B, editors. Adv artif intell. Berlin, Heidelberg: Springer; 2006. p. 1015–21, 2006.

Stokes M, Anderson M, Chandrasekar S. A standard default color space for the internet – sRGB. In: International color consortium (ICC); 1996.

Ward J, Jonas C, Dienes Z, Seth A. Grapheme-colour synaesthesia improves detection of embedded shapes, but without pre-attentive "pop-out" of synaesthetic colour. Proc Biol Sci 2010;277:1021–6.

Witthoft N, Winawer J. Synesthetic colors determined by having colored refrigerator magnets in childhood. Cortex 2006;42:175–83.