i-PERCEPTION a Pion publication
i-Perception (2014) volume 5, pages 132–142
dx.doi.org/10.1068/i0626 ISSN 2041-6695 perceptionweb.com/i-perception

Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder

Anja Moos*
School of Critical Studies, University of Glasgow, Glasgow, Scotland, UK; e-mail: anja.moos@gmail.com

Rachel Smith
School of Critical Studies, University of Glasgow, Glasgow, Scotland, UK; e-mail: rachel.smith@glasgow.ac.uk

Sam R. Miller
Glasgow, Scotland, UK; e-mail: sam.miller.r@gmail.com

David R. Simmons
School of Psychology, University of Glasgow, Glasgow, Scotland, UK; e-mail: david.simmons@glasgow.ac.uk

Received 17 September 2013, in revised form 1 May 2014; published 27 May 2014

Abstract. Human speech conveys many forms of information, but for some exceptional individuals (synaesthetes), listening to speech sounds can automatically induce visual percepts such as colours. In this experiment, grapheme–colour synaesthetes and controls were asked to assign colours, or shades of grey, to different vowel sounds. We then investigated whether the acoustic content of these vowel sounds influenced participants' colour and grey-shade choices. We found that both colour and grey-shade associations varied systematically with vowel changes. The colour effect was significant for both participant groups, but significantly stronger and more consistent for synaesthetes. Because not all vowel sounds that we used are "translatable" into graphemes, we conclude that acoustic–phonetic influences co-exist with established graphemic influences in the cross-modal correspondences of both synaesthetes and non-synaesthetes.

Keywords: coloured vowels, cross-modal perception, colour vision, synaesthesia, vowel sounds.

1 Introduction

What do you see when you hear a speech sound? The automatic and involuntary experience of a perception in a modality different from the one stimulated, such as seeing colours when listening to speech sounds, is called synaesthesia (Simner, 2007). For many "grapheme–colour" synaesthetes, both reading and listening to words can induce vivid and consistent concurrent perceptions of colour. Surprisingly little, however, is known about how their experiences might be affected by the acoustic–phonetic properties of speech. Instead, such cases are generally considered to be mediated by written language (e.g. Baron-Cohen, Harrison, Goldstein, & Wyke, 1993; Simner et al., 2006). The reasoning is that abstract letter units (graphemes) predict synaesthetes' perceived colours better than abstract sound units (phonemes) do: For example, many synaesthetes report the same colour for the differently-pronounced <c>s in cat and cease, but different colours for the differently-spelled /k/ sounds in cat and koala (cf. Simner, 2007). This paper investigates whether the vowel sounds of speech can be found to systematically influence colour percepts, when the focus is on vowels' concrete acoustic–phonetic properties, rather than their phonemic identity.

Although vowel sound to colour associations have been proposed previously, no systematic, rigorous tests have compared synaesthetes and controls' behaviour. Jakobson (1962) and Marks (1975) both direct attention to the main acoustic features of vowel quality: vowel formants. Formants are peaks of the sound spectrum, i.e. accumulations of acoustic energy at certain frequencies. Vowels are usually distinguishable by their first two formants, F1 and F2. Figure 1 illustrates how eight reference vowels, the *primary cardinal vowels* (International Phonetic Association, 1999), are distributed in acoustic F1–F2 space. Cardinal vowel C1 or [i], similar to the vowel in *cheese*, is characterised by a low F1 and a high F2, i.e. the two energy peaks are far apart. C5, or [ɑ], similar to the vowel in the first syllable of *father,* has a high F1 but low F2. C1–C4 are termed front vowels, C4–C5 open vowels and C5–C8 back vowels, according to the articulatory settings of the mouth and tongue. So, the *cheese*

*Corresponding author.

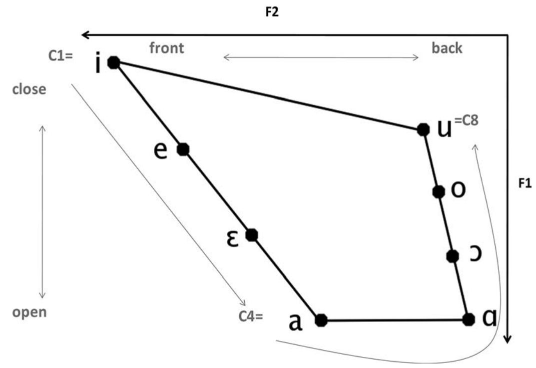

Figure 1. Schematic vowel quadrilateral of the eight primary cardinal vowels C1–C8, positioned in the frequency space of the first two formants F1 and F2 and the articulatory space from close to open mouth setting and front to back tongue body setting.

vowel is produced with a nearly closed mouth and fronted tongue, and the *father* vowel with an open mouth and retracted tongue.

Vowels with both high F1 and F2, like the [a] in *cat,* have been associated with perceived colourfulness (Jakobson, 1962) or redness (Marks, 1975). Increasing F2 frequency has, moreover, been linked to a vowel's perceived lightness, so that [i] should be perceived as light but [ɔ] as in *caught* as dark (Marks, 1975). However, these proposals are based on case reports that lack the methodological control characteristic of synaesthesia research today. Wrembel (2009) reports more stringent tests using articulatorily defined vowel phonemes, which indicate lighter colour associations with front vowels, but only non-synaesthetes were tested. What is thus missing is a direct comparison of synaesthetes' and controls' behaviour, using both acoustic–phonetic and advanced colorimetric analyses to substantiate previous ideas and findings. A detailed analysis and comparison across groups may help to identify cognitive mechanisms underlying the associations, similar to the proposals of Ward, Huckstep, and Tsakanikos (2006) who assume shared mechanisms in musical pitch—lightness mappings by synaesthetes and controls. Potentially, scales such as high–low and light–dark are treated in a similar fashion across modalities by the brain (Maurer & Mondloch, 2005; Spence, 2011), which could also play a role in sound symbolism.

Understanding the interplay of acoustic–phonetic and graphemic influences in synaesthesia may also help in explaining its neural basis, and the cross-modal interactions underlying perceptual awareness and feature binding more generally. One prominent theory of synaesthesia assumes cross-activation between brain regions, particularly adjacent ones (Hubbard, Brang, & Ramachandran, 2011). On this view, acoustic–phonetic structure should influence grapheme–colour synaesthesia only minimally, since the commonly implicated brain areas (the "visual word form area" in fusiform gyrus and "colour area" V4) are not usually associated with auditory processing. The notion that synaesthesia originates early in development (Maurer & Mondloch, 2005) accords well with a role for acoustic–phonetic structure: Infants first encounter sounds phonetically, gradually systematising their experiences into phonological knowledge, which, in turn, forms a basis for literacy. Thus, linguistic synaesthesiae might initially be driven by concrete phonetic attributes of speech, with phonemic and ultimately graphemic influences supervening later in development (cf. Simner, 2007).

To test whether the acoustic–phonetic structure of vowels affects synaesthetic perception, we presented self-reported grapheme–colour synaesthetes and control participants with vowel sounds, and they chose the best corresponding colour. If acoustic–phonetic structure influences concurrent synaesthetic percepts, we would expect systematic shifts in chromaticity of colour choices as F1 increases (hypothesis 1), and in the lightness of grey-shade choices as F2 increases (hypothesis 2). If, in contrast, graphemic influences dominate to the exclusion of acoustic–phonetic ones, we would see no influence of formant frequencies on colour choices. As observed for musical sounds (Ward et al., 2006) and graphemes (Simner et al., 2005), we expect broadly similar response patterns among synaesthetes and controls, with synaesthetes behaving more consistently (hypothesis 3).

2 Methods

2.1 Participants

Eleven synaesthetes (two male, mean age 26, range 18–69) and 20 controls (nine male, mean age 24, range 19–37) participated, all recruited locally. All reported normal hearing, normal or corrected-to-normal eyesight and normal colour vision, except one synaesthete who reported a slight colour deficiency. Tests of colour vision (Ishihara, 1978) confirmed the participants' self-reports. All had English as their native language. More specifically, all controls and six synaesthetes were Scottish; the other five synaesthetes had other varieties of English (including two bilinguals with French). Synaesthetes underwent a short directed interview before the experiment. All reported having grapheme–colour synaesthesia. Some synaesthetes, but not all, reported having additional synaesthetic experiences (e.g. shapes and/or colours in response to music and/or voices). However, as we have shown elsewhere (Moos, Simmons, Simner, & Smith, 2013), the presentation of synaesthesia is often multi-faceted: something which may only become apparent after rigorous testing.

The study was approved by the ethics committee of the Faculty of Arts at the University of Glasgow, and participants provided informed consent before testing.

2.2 Auditory stimuli

Stimuli were 16 different vowel sounds, selected to represent the acoustic vowel space rather than to instantiate the specific vowels of English (AuditoryStimuli.wav in the supporting information available on-line). Recordings of the eight primary cardinal vowels (CVs; Figure 1) were made in a high-specification sound studio. They were spoken in isolation on a level pitch by an English male trained phonetician (65 years old).

To create a richer vowel continuum, eight intermediate vowels were made by morphing each neighbouring pair of CVs using the STRAIGHT modification and synthesis procedure (Kawahara, 2003). Interpolation between each pair of vowel recordings used five vectors: f0, spectral amplitude, aperiodicity, time and formant frequencies. Morphed vowels are referred to as C1.5 (between [i] and [e], C2.5…C8.5 (between [u] and [i]).

The 16 vowels were adjusted in intensity (to 80 dB_{SPL}) and duration (to 1049 ms, the mean duration of the original stimuli) using Praat's PSOLA function (Boersma & Weenink, 2011). F0 varied minimally, from 120 to 124 Hertz, and was not equalised. The frequencies of F1 and F2 in Hertz were measured manually from LPC (linear predictive coding) and fft (fast Fourier transform) spectra in Praat, for subsequent analyses.

2.3 Visual response display

To separately examine the contributions of chromaticity (hypothesis 1) and lightness (hypothesis 2), we created two visual response displays, "colour" and "grey-shade." The response space comprised 16 choices. We chose not to offer a larger set of colours because of time restrictions. As auditory memory decays rapidly (Pisoni, 1973), fast responding was required; this is only possible with a limited set of colours. Nonetheless, 16 choices expand the range used in previous work without unduly lengthening the number of trials required (e.g. Marks, 1975; Wrembel, 2009). Each display consisted of 16 circles on a mid-grey background in a 4 × 4 matrix, see Figure S1 in the supporting information on-line.

The coloured display consisted of the 11 focal colours (white, black, red, green, yellow, blue, brown, grey, orange, pink and purple; Berlin & Kay, 1999), plus five further colours (dark green, light green, pale pink, cyan and dark blue) to fill the gaps in colour space. The grey-shades were visually equidistant.

The focal colours were displayed by converting the Munsell codes defined by Rosch Heider (1972) into CIE coordinates. The CIE coordinates of these and other display colours were measured with a chromameter (Minolta, CS-100, Konica Minolta Sensing Europe) and converted to CIELUV space for analysis (Westland & Ripamonti, 2004; see supporting information Table S1).

2.4 Procedure

Participants were tested individually in a high-specification studio, in darkness to avoid room lighting affecting their colour perception. They wore high-quality headphones and viewed the LCD monitor from a distance of 65 cm.

First, participants carried out the colour task, followed after a break of 5–15 min by the grey-shade task. They were instructed to click on the colour or grey-shade that best matched the sound. On each trial, the sound and visual response display were presented simultaneously. After a response had been chosen, a grey screen was displayed for 2 s to allow the after-image to dissipate. The order of the colours on the screen was randomised on each trial, whereas the grey-shades were ordered from white (top left) to black (bottom right).

The presentation order used a type 1 index 1 sequence (Nonyane & Theobald, 2007). Here, every stimulus is followed by itself as often as by any other stimulus. This resulted in 16*16+1 = 257 stimuli for each of the colour and grey-shade tasks (total 514 stimuli), providing a good test of within-subject reliability. The experiment lasted approximately 1.5 hr including self-paced breaks after each block of 32 stimuli.

2.5 Statistical analysis

For the colour experiment, L*, u* and v* values (representing luminance, the red–green axis and the yellow–blue axis, respectively) were the dependent variables in a repeated-measures analysis of covariance model (ANCOVA). For the grey experiment, L* was the sole dependent variable. In both cases, fixed predictors were the first two formant frequencies of the vowel stimuli (F1, F2) as covariates[1], participant group and the interactions of F1 × Group and F2 × Group. Between-subject random effects were fitted for intercept, F1 and F2. To aid interpretation of the formant frequency and interaction terms, estimates of the “slope” are presented with standard errors, where the slope is the change in the dependent variable associated with a change of 100 Hz in the relevant formant frequency in the relevant participant group. The between- and within-subject variances were estimated separately for each group. Equality of variances between groups was assessed using an *F*-test. Goodness of fit of the models was demonstrated by comparison of -2 times the difference in log likelihood to a χ^2 distribution (χ^2>2,304, df’s = 87, *p*’s <0.001).

To analyse consistency, we employed two complementary approaches. First, we identified each participant’s modal (i.e., most frequent) colour responses to each vowel and then used non-parametric techniques (Mann–Whitney U test) to compare the frequency of these modal responses across groups. This determined how frequently participants chose *exactly* the same colour (or grey shade) across multiple repetitions of a given vowel. Second, we used parametric techniques on the continuous CIELUV variables to look at how close these choices were to each other, using comparisons of the within- and between-subject variances as estimated by the ANCOVA.

Note that results are treated as statistically significant when $p < 0.05$ and as trends when $0.05 \leq p < 0.1$.

3 Results and discussion

3.1 Colour associations

According to self-report, for most synaesthetes, all vowel sounds induced synaesthetic colours. For the rest of the synaesthetes, most vowel sounds induced synaesthetic colours. As the design of the experiment was a forced choice task, colours for those vowel sounds which did not induce automatic colour associations were associated freely, similar to the associations made by non-synaesthetes.

Figure 2 and Table 1 show that u* of the colour choices, which represents the red–green axis, increased significantly with rising F1 for both groups, but significantly more strongly for synaesthetes (Table 1, Colours: u*: F1, F1 × Group; synaesthete slope 14.2 (SE 2.0), control slope 4.4 (2.3)). There was also a trend for u* to decrease slightly with rising F2, independent of participant group (u*: F2; overall slope -1.9 (1.0)). In other words, both synaesthetes and controls tended to judge open vowels such as [a] (with high F1) as containing a high proportion of red, and front vowels such as [i] (with high F2) as containing a high proportion of green, confirming hypothesis 1 with respect to F1 and extending it by finding regularities in F2 for green as well. A trend was also found for front vowels (with high F2) to be judged as more yellow (with a higher v*) than back vowels, with a trend indicating this is specific to synaesthetes (v*: F2, F2 × Group; synaesthete slope 3.7 (1.9), control slope −0.1 (0.7)). These colour associations broadly resemble the results of Marks (1975) who found that open vowels were associated with red (or blue) and front vowels with yellow (or white).

[1] Because a relationship between formants and colour choices was to be expected a priori according to the findings of previous research such as Marks (1975) and Wrembel (2009), including the formants as covariates allows us to more accurately assess the effect of colour associations.

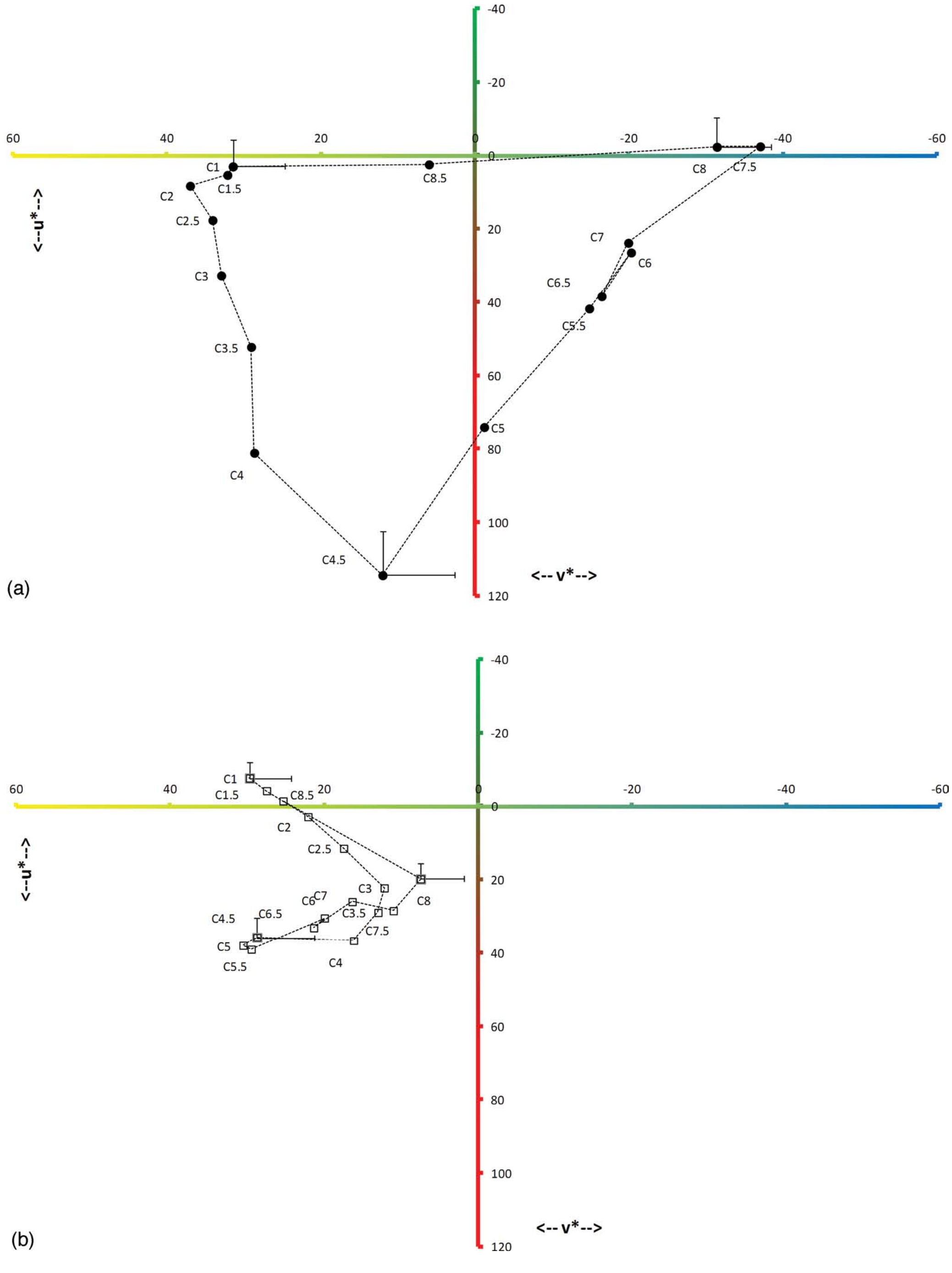

Figure 2. Colour choices of synaesthetes (a) (filled circles) and controls (b) (blank squares) projected onto the CIELUV colour space. Data points show the averaged u* and v* values per group per vowel. Note that the two axes have been inverted and the u* axis is mirrored to resemble how the synaesthetes' responses are similar to the vowel space shown in Figure 1. Whiskers of one standard error (SE) were inserted for the vowels C1, C4.5 and C8 to show the response variability for these corner vowels.

The relation between F1 (or openness) and redness–greenness allows for interpretation at various levels. Many species make themselves appear "larger" both visually and auditorily to threaten adversaries (Harris, Fitch, Goldstein, & Fashing, 2006). Mouth and jaw opening, which contribute to perceived threat, increase both frequency and intensity of F1, as well as overall vowel intensity (Fairbanks, House, & Stevens, 1950). The association of open vowels (high F1) with red might thus reflect a shared semantic association with threat, dominance or warning (Humphrey, 1976); or, at a psychophysical level, with the tendency for red to appear closer in space to the perceiver than

Table 1. Effects of F1, F2, participant group and interactions on colour/lightness choices, from repeated-measures ANCOVA analyses.

Response	Measure	Effect	*F*	df	*p*
Colours	L*	group	3.08	1,17.7	.097
		F1	**7.28**	**1,17**	**.015**
		F1 × Group	0.48	1,17	.499
		F2	**4.61**	**1,19**	**.045**
		F2 × Group	1.46	1,19	.241
	u*	group	0.41	1,16	.531
		F1	**35.94**	**1,27.9**	**<.001**
		F1 × Group	**9.83**	**1,27.9**	**.004**
		F2	3.73	1,17.5	.070
		F2 × Group	0.68	1,17.5	.420
	v*	group	3.00	1,24.5	.096
		F1	0.23	1,18.4	.639
		F1 × Group	2.05	1,18.4	.169
		F2	3.17	1,13.6	.097
		F2 × Group	3.66	1,13.6	.077
Grey-Shades	L*	group	2.04	1,20.7	.169
		F1	**16.46**	**1,23.7**	**.001**
		F1 × Group	0.77	1,23.7	.389
		F2	**25.61**	**1,22**	**<.001**
		F2 × Group	**5.42**	**1,22**	**.030**

Note. Significant differences are in bold.

non-red stimuli (Simmons, 2011). Alternatively, [a] is one of the commonest vowels cross-linguistically (Liljencrants & Lindblom, 1972) and red one of the most prototypical colours (Berlin & Kay, 1999); this statistical correspondence might underlie their association, as for grapheme frequency (Simner et al., 2005).

3.2 Luminance associations

In both the grey-shade and the colour task, L* (luminance) of responses increased significantly both with rising F1 and F2 (Table 1, Colours: L*: F1, F2; overall F1 slope 2.2 (0.8), overall F2 slope 0.7 (0.3); Grey-Shades: L*, F1, F2; overall F1 slope 3.3 (0.8), overall F2 slope 1.5 (0.3)). That is, both synaesthetes and controls chose lighter shades for open (high F1) and front (high F2) vowels, confirming hypothesis 2 with respect to F2, and extending it to F1. The significant interaction of F2 with group in the grey-shade task (L*: F2 × Group) indicates that F2 influences synaesthetes' luminance associations more strongly than those of controls (synaesthete slope 2.2 (0.5), control slope 0.8 (0.4)).

The relations of vowel formants with lightness accord with findings that lightness associates with higher musical pitch and "clearer" timbre (Marks, 1975; Ward et al., 2006). Pitch—lightness correspondences have been frequently documented, in humans and recently in chimpanzees (Ludwig, Adachi, & Matsuzawa, 2011), but it remains unclear whether they result from structural similarities between neural processing of different stimulus dimensions, and/or from environmental statistical regularities and/or (in humans) semantically mediated associations (Spence, 2011).

3.3 Consistency

Figure 3 illustrates differences in response consistency between a representative synaesthete and control. In the colour task, synaesthetes responded more consistently, as predicted (hypothesis 3), choosing their most frequent colour for each vowel significantly more often than controls (12.0 vs. 7.8 times, respectively; Mann–Whitney $U = 20.5$, $z = -3.697$, $p < 0.001$). No such difference was found in the grey-shade task (synaesthetes 7.8 vs. controls 7.4; $U = 78.5$, $z = -1.301$, $p = 0.197$). Apparently,

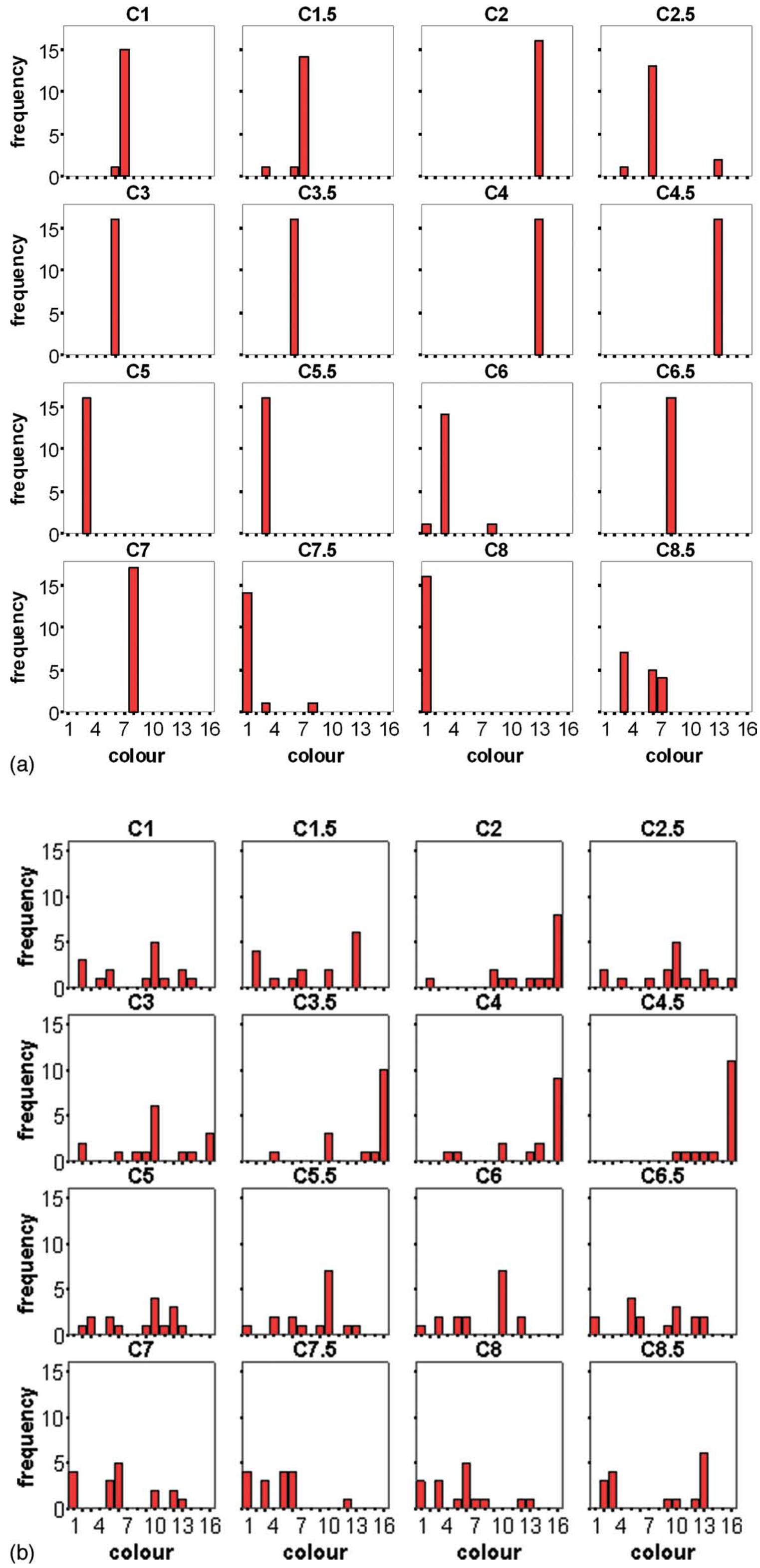

Figure 3. Consistency of colour choice across 16 repetitions per vowel C1–C8.5 for a typical synaesthete (a) and control (b). Bar height indicates how often a colour was chosen per vowel (each box shows responses for one vowel). Colours are arranged in alphabetical order along the x-axis: 1 black, 2 blue, 3 brown, 4 cyan, 5 dark blue, 6 dark green, 7 green, 8 grey, 9 light green, 10 orange, 11 pink, 12 purple, 13 red, 14 rosa (pale pink), 15 white and 16 yellow.

synaesthetes' consistent behaviour relies on the availability of colour information; they are not more systematic than controls under achromatic conditions.

The random effects structure of the ANCOVA results supports this picture: In the colour task (Table 2, Colours), synaesthetes showed significantly greater within-subject consistency than controls in L* and v*. Between-subject agreement was significantly smaller for synaesthetes than controls with respect to F2 on the v* scale. For grey-shade choices (Table 2, Grey-Shades), within-subject consistency was again significantly larger in synaesthetes than controls, but participant groups did not differ in between-subject agreement.

Thus, synaesthetes' individual colour choices were much more consistent than controls' (hypothesis 3), as also observed for grapheme–colour associations (e.g., Simner et al., 2005). The evidence for synaesthetes being more consistent than controls is weaker for the grey shade task, and significant only on some measures. This diminished contrast between synaesthetes' and controls' consistency performance in the grey shade task could be related to the observation that some synaesthetes reported that the task of assigning grey-shades to vowels was difficult or disturbing, presumably because a key dimension of their inherent associations was unavailable, whereas controls seemed to find it easier to use the light–dark dimension in their associations.

4 General discussion

Our results showed that shifts in the chromaticity of colour associations were systematically influenced by vowel acoustics, confirming hypothesis 1. Likewise, hypothesis 2 was confirmed with results of the achromatic part of the experiment: Front vowels with a high second formant were associated with lighter shades than back vowels. These results were found for both participant groups; however, synaesthetes' colour associations were more consistent than those of non-synaesthetes, i.e. the same vowel sounds induced mostly the same or similar colour associations (consistent with hypothesis 3).

Table 2. Results of consistency analyses comparing synaesthetes and controls using the random effects obtained in repeated-measures ANCOVA analyses assessing relationships of F1 and F2 with L*, u* and v*.

Response	Measure	Variance parameter	*F*	df	*p*
Colours	L*	Between-subject mean	0.47	16.2,10.8	.161
		Between-subject F1	0.40	14.4,10.4	.110
		Between-subject F2	0.57	15.9,10.7	.297
		Within-subject	**1.27**	**4313,2790**	**<.0001**
	u*	Between-subject mean	0.34	16,10.8	.053
		Between-subject F1	2.19	15.4,8.9	.241
		Between-subject F2	0.46	16.1,10.7	.149
		Within-subject	0.94	4313,2790	.081
	v*	Between-subject mean	1.11	15.6,10.4	.889
		Between-subject F1	0.49	14.4,10.5	.213
		Between-subject F2	**0.17**	**14,10.8**	**.003**
		Within-subject	**1.84**	**4313,2790**	**<.0001**
Grey-Shades	L*	Between-subject mean	0.71	16.6,10.9	.519
		Between-subject F1	1.02	16.1,10.6	.994
		Between-subject F2	0.83	16.4,10.8	.714
		Within-subject	**1.35**	**4314,2791.1**	**<.0001**

Note. The "*F*" column gives *F* ratios for the differences in variance between synaesthete and control groups, for the intercept ("between-subject mean"), for specific fixed effects ("between-subject F1, F2") and residually ("within-subject"). An *F* ratio of 1 implies equality of variance between participant groups; $F > 1$ indicates that controls are more variable than synaesthetes; $F < 1$ indicates that synaesthetes are more variable than controls. Significant differences between participant groups are in bold.

Our synaesthete participants considered themselves to have letter–colour associations, yet showed systematic, fine-grained influences from acoustic–phonetic structure. Is grapheme–colour synaesthesia therefore fundamentally a phonetic phenomenon? We do not take this view, but rather propose that acoustic–phonetic influences co-exist with established effects such as graphemes' frequency of occurrence (Simner et al., 2005). Suggestive supporting evidence comes from participants' responses to different sounds that can correspond to the grapheme <a>. In Scottish English, the commonest pronunciation of <a> in words (e.g. *hat*) resembles C4.5; but the *name* of the letter <a>, and its pronunciation in words like *hate,* resembles C2. For several Scottish synaesthetes in our sample, C2 had the same colour as C4.5, and formed an "island" distinct from the more finely graded colours of phonetically-neighbouring vowels, suggesting its colour was influenced by the pronunciation of <a>.

If phonetic and graphemic influences do co-exist, the presentation modality in specific experiments may be key to understanding the balance among them. Our method, with its multiple presentations of subtly-varying sound qualities, was better able to detect acoustic–phonetic influences than are written surveys eliciting responses to letters from a small number of visual presentations.

In terms of their potential neural basis, our data cannot be accounted for solely by hyper-connection between adjacent brain regions dedicated exclusively to processing visual graphemes and colour. They could be explained in several other ways, e.g. if visual grapheme regions also perform more multimodal or integrative functions (cf. Sharp, Scott, & Wise, 2004), or in terms of other routes, or alternative interactional mechanisms, connecting auditory and colour areas (cf. Grossenbacher & Lovelace, 2001; Hänggi, Beeli, Oechslin, & Jäncke, 2008; Smilek, Dixon, Cudahy, & Merikle, 2001). Some grapheme–colour synaesthetes also show hyper-connectivity in the parietal regions implicated in binding, suggesting that their unusual abilities might relate to letter–sound integration in general, rather than visual letter processing specifically (see Hubbard et al., 2011, and Rouw, Scholte, & Colizoli, 2011, for further discussion). Furthermore, given the developmental precedence of language learning over reading, the foundations of synaesthesia could be laid even earlier in development than previously thought (Maurer & Mondloch, 2005). As for the non-synaesthetes, they seem to show the same trends in responses as the synaesthetes but to a lesser degree and with significantly less consistency. Whether this provides evidence for a "synaesthesia continuum" or simply "cross-modal integration" will depend on theoretical standpoint (Spence, 2011).

5 Conclusions

We found that synaesthetes and controls show regular patterns in their vowel sound—colour associations, while synaesthetes' associations are stronger and more consistent. The reduced difference between the two groups in association consistency with grey shades emphasises that synaesthetes use a dimension in their associations which is not obvious to controls, such as the scale of grey shades from white to black might be. As Spence (2011) has observed, the relationship between acoustic stimulation and concurrent visual perceptions in both synaesthetes and non-synaesthetes is hard to classify in terms of other cross-modal correspondences. Our data add a further dimension to this conundrum by suggesting that the acoustic–phonetic structure must be considered alongside graphemic structure when trying to explain the commonest forms of synaesthesia.

References

Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception, 22*, 419–426. doi:10.1068/p220419

Berlin, B., & Kay, P. (1999). *Basic color terms. Their universality and evolution.* Stanford, CA: CSLI Publications. doi:10.1002/col.5080170514

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [Computer program]. Version 5.1.15, retrieved 31 August 2009 from http://www.praat.org/

Fairbanks, G., House, A. S., & Stevens, E. L. (1950). An experimental study on vowel intensities. *Journal of the Acoustical Society of America, 22*, 457–459. doi:10.1121/1.1906627

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: Cognitive and physiological constraints. *Trends in Cognitive Sciences, 5*, 36–41. doi:10.1016/S1364-6613(00)01571-0

Hänggi, J., Beeli, G., Oechslin, M. S., & Jäncke, L. (2008). The multiple synaesthete E.S.—Neuroanatomical basis of interval–taste and tone–colour synaesthesia. *NeuroImage, 43*, 192–203. doi:10.1016/j.neuroimage.2008.07.018

Harris, T. R., Fitch, W. T., Goldstein, L. M., & Fashing, P. J. (2006). Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. *Ethology, 112*, 911–920. doi:10.1111/j.1439-0310.2006.01247.x
Hubbard, E. M., Brang, D., & Ramachandran, V. (2011). The cross-activation theory at 10. *Journal of Neuropsychology, 5*, 152–177. doi:10.1111/j.1748-6653.2011.02014.x
Humphrey, N. (1976). The colour currency of nature. In T. Porter & B. Mikellides (Eds.), *Colour for Architecture* (pp. 95–98). London: Studio Vista.
International Phonetic Association. (1999). *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet*. Cambridge: Cambridge University Press.
Ishihara, S. (1978). *Tests for Colour-Blindness*. Kanehara Shuppan & Co., Ltd.
Jakobson, R. (1962). *Selected writings I. Phonological studies.* 's-Gravenhage: Mouton & Co.
Kawahara, H. (2003). Exemplar-based voice quality analysis and control using a high quality auditory morphing procedure based on STRAIGHT. In C. d'Alessandro & K. Scherer (Eds.), *ISCA workshop VOQUAL* (pp. 109–114).
Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839–862. doi:10.2307/411991
Ludwig, V., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proceedings of the National Academy of Sciences of the USA, 108,* 20661–20665. doi:10.1073/pnas.1112605108
Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin, 82*, 303–331. doi:10.1037/0033-2909.82.3.303
Maurer, D., & Mondloch, C. J. (2005). Neonatal synaesthesia: A reevaluation. In L. C. Robertson & N. Sagiv (Eds.), *Synaesthesia: Perspectives from cognitive neuroscience* (pp. 193–213). Oxford: Oxford University Press.
Moos, A., Simmons, D., Simner, J., & Smith, R. (2013). Color and texture associations in voice-induced synaesthesia. *Frontiers in Psychology, 4*, 568. doi:10.3389/fpsyg.2013.00568
Nonyane, B. A. S., & Theobald, C. M. (2007). Design sequences for sensory studies: Achieving balance for carry-over and position effects. *British Journal of Mathematical and Statistical Psychology, 60*, 339–349. doi:10.1348/000711006X114568
Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13*, 253–260. doi:10.3758/BF03214136
Rosch Heider, E. (1972). Universials in color naming and memory. *Journal of Experimental Psychology, 93*, 10–20. doi:10.1037/h0032606
Rouw, R., Scholte, H.S., & Colizoli, O. (2011). Brain areas involved in synaesthesia: A review. *Journal of Neuropsychology, 5*, 214–242. doi:10.1111/j.1748-6653.2011.02006.x
Sharp, D. J., Scott, S. K., & Wise, R. J. S. (2004). Retrieving meaning after temporal lobe infarction: The role of the Basal Language area. *Annals of Neurology, 56*, 836–846. doi:10.1002/ana.20294
Simmons, D. R. (2011). Colour and emotion. In C. P. Biggam, C. Kay, C. Hough & D. R. Simmons (Eds.), *New Directions in Colour Studies*. Benjamins: Amsterdam.
Simner, J. (2007). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Science, 11*, 23–29. doi:10.1016/j.tics.2006.10.010
Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K., & Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception, 35*, 1024–1033. doi:10.1068/p5469
Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology, 22*, 1069–1085. doi:10.1080/02643290500200122
Smilek, D., Dixon, M.J., Cudahy, C., & Merikle, P.M. (2001). Synaesthetic photisms influence visual perception. *Journal of Cognitive Neuroscience*, 13, 930–936. doi:10.1162/089892901753165845
Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception and Psychophysics, 73*, 971–995. doi:10.3758/s13414-010-0073-7
Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex, 42*, 264–280. doi:10.1016/S0010-9452(08)70352-6
Westland, S., & Ripamonti, C. (2004). *Computational colour science using MatLab.* Hoboken, NJ: Wiley. doi:10.1002/0470020326.ch4
Wrembel, M. (2009). On hearing colours—Cross-modal associations in vowel perception in a non-synaesthetic population. *Poznań Studies in Contemporary Linguistics, 45*, 581–598. doi:10.2478/v10010-009-0028-0

Anja Moos studied phonetics, psychology, and educational science in Saarbrücken, Germany, and has worked in the field of forensic phonetics. Later, she discovered synaesthesia as a research topic for her PhD. Now she evaluates multi-sensory perception of design and material of cars for Ford with the aim of introducing products that are in cross-sensory harmony.

Rachel Smith studied languages and linguistics at the University of Cambridge, and has worked since 2006 as a Lecturer in Phonetics at the University of Glasgow. As well as speech and synaesthesia, her research investigates how speech perception is affected by familiarity with voices and accents, and speech rhythm and timing.

David R. Simmons D.Phil. (Oxon), B.Sc. (Hons), A.R.C.S.: Originally a physicist, David's research journey has passed from 3D vision through texture to colour vision and aesthetics. However, in recent years, his interests have broadened to include individual differences in perception, especially autism and synaesthesia. David is currently a member of the joint editorial board of Perception/i-Perception.

a Pion publication