



Cross-modal associations and synesthesia: Categorical perception and structure in vowel–color mappings in a large online sample

Christine Cuskley¹ · Mark Dingemanse^{2,3} · Simon Kirby¹ · Tessa M. van Leeuwen⁴

© The Author(s) 2019

Abstract

We report associations between vowel sounds, graphemes, and colors collected online from over 1,000 Dutch speakers. We also provide open materials, including a Python implementation of the structure measure and code for a single-page web application to run simple cross-modal tasks. We also provide a full dataset of color–vowel associations from 1,164 participants, including over 200 synesthetes identified using consistency measures. Our analysis reveals salient patterns in the cross-modal associations and introduces a novel measure of isomorphism in cross-modal mappings. We found that, while the acoustic features of vowels significantly predict certain mappings (replicating prior work), both vowel phoneme category and grapheme category are even better predictors of color choice. Phoneme category is the best predictor of color choice overall, pointing to the importance of phonological representations in addition to acoustic cues. Generally, high/front vowels are lighter, more green, and more yellow than low/back vowels. Synesthetes respond more strongly on some dimensions, choosing lighter and more yellow colors for high and mid front vowels than do nonsynesthetes. We also present a novel measure of cross-modal mappings adapted from ecology, which uses a simulated distribution of mappings to measure the extent to which participants' actual mappings are structured isomorphically across modalities. Synesthetes have mappings that tend to be more structured than nonsynesthetes', and more consistent color choices across trials correlate with higher structure scores. Nevertheless, the large majority (~ 70%) of participants produce structured mappings, indicating that the capacity to make isomorphically structured mappings across distinct modalities is shared to a large extent, even if the exact nature of the mappings varies across individuals. Overall, this novel structure measure suggests a distribution of structured cross-modal association in the population, with synesthetes at one extreme and participants with unstructured associations at the other.

Keywords Synesthesia · Cross-modal correspondences · Language · Sound symbolism

Cross-modal associations have been proposed as a key mechanism underlying sound-symbolic phenomena relevant to

Christine Cuskley and Mark Dingemanse contributed equally to this work.

✉ Christine Cuskley
ccuskley@gmail.com

✉ Mark Dingemanse
mark.dingemanse@mpi.nl

¹ Centre for Language Evolution, University of Edinburgh, Edinburgh, UK

² Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

³ Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

⁴ Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

language learning and development (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Imai & Kita, 2014) as well as to language evolution (Cuskley & Kirby, 2013). Yet, the relationship between cross-modal associations and the much rarer phenomenon of *synesthesia* is still an open area of study. Understanding the relationship between synesthesia and cross-modal associations is a key step in understanding the mechanisms underlying both phenomena, and in particular, their strong relationships to learned categories and language (Cuskley & Kirby, 2013; Simner, 2007). Synesthesia is a relatively rare phenomenon, occurring in approximately 5% of the population (Simner et al., 2006), whereas cross-modal associations are much more widespread. Therefore, capturing the relationship between these phenomena requires examining large samples of participants. This article reports a large-scale study examining the associations between vowel sounds, graphemes, and colors, collected online from over 1,000 participants, simultaneously revealing salient patterns in cross-modal associations and identifying a

couple hundred synesthetes in the sample. We begin with an overview of previous studies that have contrasted the associations of synesthetes and nonsynesthetes, with particular relevance to linguistic cross-modal phenomena.

Synesthesia and cross-modality

Synesthesia, from the Greek *syn-* (together) and *-aisthes* (feeling), is a phenomenon wherein a stimulus in one sensory modality (known as the *inducer*; e.g., a sound) elicits a response not only from that modality (e.g., hearing the sound) but also in another (e.g., *seeing* the sound, known as the *concurrent*). Synesthesia is an involuntary, automatic sensory experience and can occur in many forms, ranging from shaped tastes to smelled colors. Although about a hundred distinct forms of synesthesia have been documented, over 88% of attested forms are linguistic in nature (Simner et al., 2006), with the most widely reported and well-studied form being colored graphemes. In recent years, the study of forms of synesthesia with linguistic inducers has shed light on language processing more generally, showing distinct semantic and phonological effects (Asano & Yokosawa, 2011, 2012; Simner, 2007).

Unlike synesthesia, cross-modal associations are not involuntary or automatic, but nonetheless demonstrate strong relationships between sensory modalities. For example, people reliably associate brighter colors with higher pitch and darker colors with lower pitch (Martino & Marks, 2001). Cross-modal associations are often elicited explicitly (e.g., by asking participants to match a sound to a taste; Simner, Cuskley, & Kirby, 2010), though they have also been demonstrated in implicit contexts (e.g., Ward, Huckstep, & Tsakanikos, 2006). Cross-modal associations are widespread and often shared across a population, though some may vary cross-culturally (Dolscheid, Shayan, Majid, & Casasanto, 2013; Styles & Gawne, 2017).

Despite the commonalities between synesthesia and cross-modal associations, relatively few studies have directly contrasted the two phenomena. In large part, these studies have shown that cross-modal associations in nonsynesthetes show commonalities with inducer–concurrent patterns in synesthetes. Here we focus on a few studies especially relevant to linguistic forms of synesthesia.

An important early study tackling the similarities between synesthesia and cross-modal associations was Simner et al. (2005), which tested associations between colors and graphemes (A–Z as well as 0–9) among English- and German-speaking synesthetes and nonsynesthete controls. Using both forced choice (choosing colors from an array) and open answer (name the color that best matches the grapheme) methodologies, they showed that synesthetes and nonsynesthetes shared many trends in their choices. For instance, both

overwhelmingly identified the grapheme *A* as red (cf. Root et al., 2018; Rouw, Case, Gosavi, & Ramachandran, 2014, for recent cross-linguistic confirmations). Using temporally spaced testing (one to three weeks for controls, and two to four months for synesthetes), they also showed that synesthetic participants were far more temporally consistent in their color choices (92%) than controls (32%), despite the longer time interval for testing synesthetes. This study marked the first large-scale demonstration that synesthetes and nonsynesthetes share trends in cross-modal mappings, and it reinforced the idea that temporal consistency is an important feature of synesthesia.

Although graphemes may be the best-known linguistic inducer of synesthesia, it has long been known that phonemes can also play a role, though this remains underexplored (Simner, 2007). In a study of cross-modal associations between vowel sounds and colors, Moos, Smith, Miller, and Simmons (2014) showed that the acoustic properties of F1 and F2 in vowels were significant predictors of color choices in both synesthetes and nonsynesthetes, although more extremely in the former group. F1 and F2 are vowel formants independent of voice pitch (although they are also measured in Hertz), which vary as a result of changing how the vocal tract filters the source sound provided by the vibrating vocal folds. Although not entirely deterministic, changes in F1 and F2 vary predominantly with tongue position in vowel production. The value of F1 is lower when the tongue is higher in the mouth (i.e., high vowels; e.g., the /i/ in *beet*), and higher when the tongue is lower in the mouth (i.e., low vowels; e.g., the /a/ in *bot*). The value of F2, on the other hand, is lower when the tongue is farther back in the mouth, and higher when the tongue is farther forward.¹ Moos et al. used 16 synthesized vowels that spanned the F1–F2 space and tested 11 English-speaking synesthetes and 20 English-speaking controls. Over hundreds of trials, participants responded to each vowel sound multiple times by choosing one of 16 color swatches, which varied in the dimensions of lightness, green–red, or yellow–blue. They found generally that lower values of F1 (i.e., higher vowels) were greener and yellower, but lower values of F2 (i.e., more back vowels) were more red and blue, and that these trends were much stronger for synesthetes than for controls. Moos et al. concluded that acoustic factors play a privileged role in vowel–color associations over graphemic factors. Similar acoustic-led trends in color choices have been found among nonsynesthetic English–Korean bilinguals (Kim, Nam, & Kim, 2018).

¹ Note the slightly confusing terminology here: High vowels have a lower value of F1, whereas low vowels have a higher value of F1. This is because the high/low distinction as applied to vowel quality refers to the position of the tongue in articulation (rather than to the F1 value): A high tongue makes for a lower value of F1, whereas a low tongue makes for a higher value of F1. For a detailed overview of the articulatory acoustics of vowel production, see Johnson (2011).

The goal of the present study was to examine the relationship between cross-modal associations and synesthesia in the domain of vowel–color associations among a large online sample of Dutch speakers. Using the same auditory stimuli as Moos et al. (2014), we make several novel contributions to our understanding of cross-modal associations and synesthesia. We aimed to (i) replicate prior results in a larger sample using more fine-grained color measures, (ii) test the novel question of whether categorical perception plays a role in shaping vowel–color associations, and (iii) contribute a new measure of structural isomorphism in cross-modal associations.

In the first instance, we aimed to replicate and extend the results reported in Moos et al. (2014). We used a larger sample ($n = 1,164$) of Dutch speakers and a more fine-grained color response space. Within our sample, we used consistency measures adapted from Rothen, Seth, Witzel, and Ward (2013) to identify synesthetes. Large-scale, online approaches are essential for the study of cross-modality and synesthesia, for several reasons. The large sample allowed us to identify synesthetes behaviorally and to sample more randomly from the population (as opposed to targeted recruitment of synesthetic volunteers, as in, e.g., Baron-Cohen, Burt, Smith-Laittan, Harrison, & Bolton, 1996). This, in turn, provided us with a larger cohort of synesthetes and nonsynesthetes to compare, a distinct advantage for understanding how synesthesia and cross-modality function across populations.

Large sample sizes have been particularly important in earlier studies that have upended our understanding of both the overall prevalence of synesthesia and the skewed sex ratios in the phenomenon. Studies such as Simner et al. (2006) showed that most earlier estimates of the prevalence of synesthesia were off by several orders of magnitude. Early volunteer-recruited studies with small samples had such high F:M ratios that some have suggested that synesthesia might even be X-linked (Baron-Cohen et al., 1996; Smilek et al., 2002). However, recent work with larger, random samples has shown this to be largely an artifact of the volunteer recruitment strategies in earlier studies (Simner & Carmichael, 2015). In short, since the *variation* across the population is a key issue in the study of cross-modality and synesthesia, a large-scale online approach is essential for a more complete understanding of the phenomena.

The large-scale online approach also presents some challenges. Tasks generally need to be shorter in order to maximize the number of participants completing the task, and the circumstances of participation are less controlled than in a lab setting. Unlike in Moos et al. (2014), the task reported here consisted of only three trials per vowel item, and we expected that our data might be noisier as a result of less controlled participation conditions. To the extent that the methods introduced noise, this should stack the deck against our hypotheses of finding acoustic, categorical, and cross-modal structure in

vowel–color associations, rendering any findings of structure more robust. Also, as in other online experiments in behavioral science, we expected increased noise in the data to be offset by the considerable advantages of having such a large sample (Crump, McDonnell, & Gureckis, 2013).

A second aim of our study was to look beyond acoustic or graphemic factors and examine how vowel phoneme category might influence associations in both synesthetes and nonsynesthetes. While we expected to replicate Moos et al.'s (2014) results regarding acoustic factors, a key focus of our study was the role of *categorical perception*. We hypothesized that since vowel perception is predominantly categorical (Rosner & Pickering, 1994), vowel category may be a better predictor of color choices than are acoustic factors. We predicted that acoustic factors would be a good predictor of color choices insofar as changes in acoustic structure correlated with changes in vowel category. Given the prevalence of grapheme–color synesthesia, the best-approximated vowel grapheme might also form a strong predictor, especially for participants who are grapheme–color synesthetes. However, grapheme category might also play an influential role for nonsynesthetes, since auditory information can automatically invoke graphemic form (Cuskley, Simner, & Kirby, 2017; Ziegler & Ferrand, 1998).

Finally, we presented a novel measure of isomorphic structure in cross-modal associations using the Mantel test (Mantel, 1967). We argue that this measure forms a useful complement to existing consistency tests used to identify synesthetes, especially in large samples. This measure quantifies the extent to which an individual's associations are isomorphically structured (i.e., similar sounds are matched with similar colors) or unstructured (i.e., similar sounds are matched with dissimilar colors). We used this measure to show that while the specific nature of cross-modal mappings may exhibit considerable individual variation, the *capacity* for making structured mappings is shared to a much broader extent in the population. This new measure of structure in cross-modal associations provides a promising way to probe structural isomorphism across domains or sensory modalities generally.

Method

Participants

A total of 1,164 adult participants volunteered to take part in an online vowel–color association task as part of a larger survey on cross-modal associations advertised in the Dutch national press, popular media, and Dutch national TV as the “Groot Nationaal Onderzoek” (“Large National Survey”; van Leeuwen & Dingemans, 2016). This is an annual public engagement initiative of the national broadcaster NTR, aiming to actively involve the general public in research. The funding

attached to this initiative enabled us to design and develop a dedicated web application, and widespread advertisement in the national media facilitated the large sample size.

Participants provided informed consent prior to participation and were not required to provide any personal information. About 85% of participants provided some information in a voluntary pretask survey that could be answered anonymously. The reported age range was 18–88 years old (median = 46, $SD = 16$); reported gender was 667 female, 206 male, and 282 who did not select a gender.² A subset of 398 participants additionally carried out a grapheme–color association task.

Materials

The stimuli in the vowel–color association task were 16 vowel sounds selected to represent points spread through acoustic vowel space. Moos et al. (2014) reported the method for creating these stimuli as follows:

Recordings of the eight primary cardinal vowels were made in a high-specification sound studio. . . . To create a richer vowel continuum, eight intermediate vowels were made by morphing each neighbouring pair. . . . The 16 vowels were adjusted in intensity (to 80 dBSPeL) and duration (to 1049 ms, the mean duration of the original stimuli) using Praat's PSOLA function Boersma and Weenink (2011). F0 [which corresponds to voice pitch] varied minimally, from 120 to 124 Hertz, and was not equalised. (Moos et al., 2014, p. 134)

The stimuli in the grapheme–color association task were the characters A–Z (capital) and 0–9 presented in black sans-serif capitals on a gray background. In each task, items were presented in randomized order and each item occurred three times, making for a total of 48 trials in the vowel task and 108 trials in the grapheme task. Consistency scores across trials per item and task were calculated in order to identify synesthetes in the sample (described in further detail below).

Color responses were recorded using an RGB color picker following widely used methods in online test batteries of synesthesia (Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007; Rothen et al., 2013). Even though RGB values are

² The survey also included the question “Do you have synesthesia?” which was answered by 875 participants, with 73 saying they did have synesthesia, 290 saying they did not, and 512 saying they didn't know. Since this question did not specify what particular type of synesthesia a participant had, these responses were not used in the present analysis.

This questionnaire did not screen for color blindness, but this is unlikely to have drastically affected our participant pool, since color blindness is X-linked recessive, and our participants were predominantly female. On the basis of the reported gender for a large subset of the sample and rates of color blindness (8% in males and 0.05% in females; Chan, Goh, & Tan, 2014), our sample likely had approximately 25–30 colorblind participants, which is a small percentage of our dataset (~2%), and one that was likely to add noise to the results rather than to inadvertently support our hypotheses.

device-specific, data collected using RGB color pickers are robust enough for the detection of synesthesia, especially when converted to the more perception-veridical CIE color space (Rothen et al., 2013) and analyzed in terms of relative distances. For the model analyses below, we converted the RGB color responses to CIELuv space (as in Moos et al., 2014) using the standard illuminant D65, particularly in order to use the CIELuv-based consistency measures from Rothen et al.'s study.

Procedure

Both association tests were available online. To maximize participation while assuring high-quality data, we piloted the tasks across multiple platforms, making sure they would run smoothly across devices without loss of functionality or data. The instructions were kept as clear and concise as possible and were similar to those provided for similar tasks in the lab and on other synesthesia-screening websites. Full code for the web application used to collect the data is available at <http://github.com/mdingemans/colouredvowels>.

For the vowel–color task, the instructions were as follows: “In this test you associate colors to sounds. You hear a sound, then choose a color that you feel fits best. Synesthetes always see the same color for the same sound: is that the same for you? Try to respond as intuitively as possible.” For the grapheme–color task, the instructions were as follows: “In this test you associate colors to letters in a precise way. You see a letter or digit, then choose a color that you feel fits best. Synesthetes always see the same color for a letter or digit. Does this hold for you, too? Try to respond as intuitively as possible.”³

In the vowel–color task, participants first were presented with a brief sound test to check their audio volume. Next, the 16 vowel items were presented in a random order three times, for a total of 48 trials. Each trial began with the audio file autoplating. After the file had played once, participants could relisten to each audio file as many times as they wanted before choosing a color. In the grapheme–color task, the 36 items (the letters A–Z and digits 0–9) were presented in a random order three times, for a total of 108 trials. In both tasks, participants clicked on the “Next” button after choosing their desired color in order to continue to the next trial. The interface for the experiment during a vowel trial is shown in Fig. 1.

In both tasks, the RGB spectrum of the color picker was randomly shifted on the horizontal axis after each trial,

³ In the original Dutch: *Vowel–color task*: “In deze test koppel je kleuren aan taalklanken. Je hoort een geluid, en dan kies je de kleur die je daar het beste bij vindt passen. Synestheten zien altijd dezelfde kleur voor dezelfde klank: is dat bij jou ook zo? Probeer zo intuïtief mogelijk te reageren.” *Grapheme–color task*: “In deze test koppel je wederom kleuren aan letters, maar dan iets preciezer. Je ziet een letter of cijfer, en dan kies je de kleur die je daar het beste bij vindt passen. Probeer zo intuïtief mogelijk te reageren.”

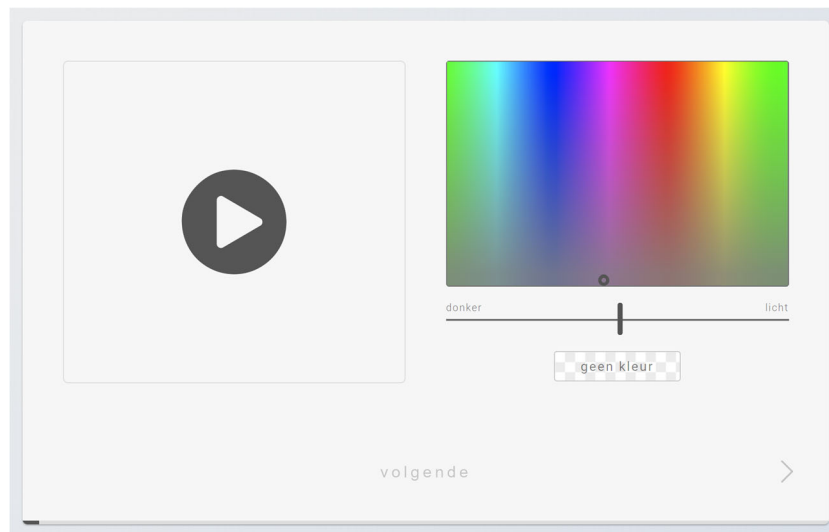


Fig. 1 Screenshot of the interface for the vowel–color association task

making it difficult to achieve high consistency just by clicking in the same region of the color picker (i.e., making a spatial association). A separate bar allowed participants to adjust the lightness of the chosen color. On each trial, it was possible to choose “no color.” This option was used in less than 5% of trials in the vowel color task, and in around 11% of trials in the grapheme–color task.

Data preparation

The test data were logged locally in the participant’s browser and transmitted to an SQL database on a secure server upon task completion. The results were postprocessed offline in R. To safeguard against possible duplicate data entries from the same individuals, responses from the same email address or same name, or from the same IP address without differential name or email address information, were removed. Overall, 11 completed tests were detected as duplicates and removed. Immediately after these quality control steps, the data were anonymized, keeping a random identifier as the sole link between the anonymized demographic metadata and test results.

The color responses in both tasks were converted to CIELuv values as in Moos et al. (2014), using Python’s `colormath` package (Taylor, 2017). CIE spaces are generally preferable to RGB space because they are based on how people perceive color (rather than, e.g., how a computer should render it). CIELuv distances in particular have been shown to be more accurate in detecting synesthetes (Rothen et al., 2013) than RGB distances (Eagleman et al., 2007). In CIELuv space, L corresponds to lightness, u corresponds to a green–red continuum, and v corresponds to a blue–yellow continuum.

Following Rothen et al. (2013), we calculate a consistency score for each participant as follows. First, we calculated d (Eq. 1) as the sum of the paired Euclidean distances in

CIELuv space between the three color choice trials for each vowel or grapheme (see the equation below, adapted from Eq. 2 in Rothen et al., 2013, p. 158). The d values for each item were used to create a mean c across items for each participant, which served as their CIELuv consistency score. Following Rothen et al., we defined synesthetes as those with consistency $c < 135.3$.

$$d = \sum_{(i=1,2,3)} \sqrt{(L_1 - L_2)^2 + (u_1 - u_2)^2 + (v_1 - v_2)^2}, \quad (1)$$

A total of 34 participants were removed from the sample because they chose “No color” for more than half of the items in the vowel association task, making it impossible to calculate a valid vowel consistency score. In the remaining participants, we identified 365 vowel–color synesthetes, with a mean $c = 95.60$ ($SD = 28.00$, 95% CI = 2.88). The nonsynesthetes had a mean of $c = 224.33$ ($SD = 63.47$, 95% CI = 4.5). Figure 2 shows the overall distribution of consistency scores, with the dashed red line indicating the consistency cutoff used in Rothen et al. (2013).

Since we lacked grapheme consistency data for much of the sample, the participants who completed the grapheme association task but did not provide data sufficient for consistency calculations were retained in the sample and classified as UNKNOWN for grapheme color synesthesia, along with participants who did not complete the grapheme task. The grapheme synesthetes had a mean $c = 82.45$ ($SD = 38.91$, 95% CI = 9.21) on the grapheme task, a value only slightly lower (higher consistency) than that reported by Rothen et al. (2013, p. 160; $c = 85.51$, $SD = 58.27$, CI not reported). Nonsynesthetes had a mean $c = 243.19$ ($SD = 64.83$, 95% CI = 10.03), a value slightly higher than the one reported by Rothen et al. (p. 160; $c = 219.38$, $SD = 68.87$, CI not reported).

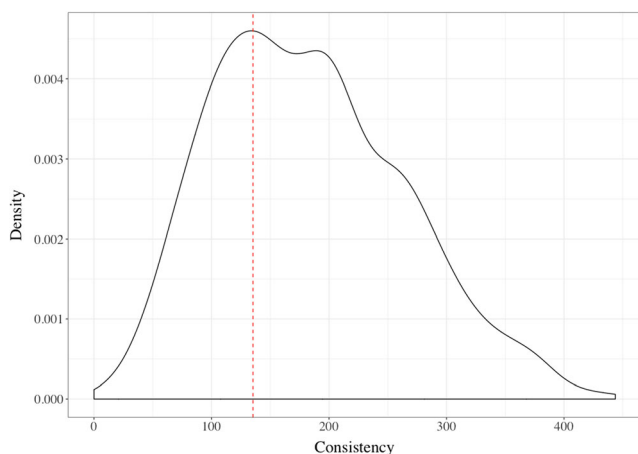


Fig. 2 Density plot of all consistency scores for vowel–color associations. The red dashed line indicates the cutoff for synesthesia from Rothen et al. (2013), with all participants to the left of the line being classified as vowel–color synesthetes

Table 1 shows the synesthetic status of the 1,130 participants based on the *c* cutoff value in each task. The rate of synesthesia in our sample, for both grapheme–color and vowel–color synesthesia, is significantly higher than would be expected in a random sample (e.g., Simner & Carmichael, 2015; Simner et al., 2006). This is likely because the task was specifically advertised for people interested in cross-sensory associations and synesthesia, resulting in higher participation rates of synesthetes than in a truly random sample (Dingemans & van Leeuwen 2015).

Although the overall prevalence of synesthesia in our sample was likely skewed, we did not find a significantly higher proportion of female synesthetes, though we did have more female participants overall (Fig. 3). This echoes recent studies showing that the early reports of much higher rates of synesthesia in females were likely due to sampling methods that relied on targeted recruitment of self-reported synesthetes (Simner & Carmichael, 2015), as contrasted with the present study, which invited participants interested in cross-modality associations generally. The use of consistency scores to automatically identify synesthetes without self-report may also

Table 1 Synesthetes and nonsynesthetes in the sample, as defined by the consistency score cutoffs reported in Rothen et al. (2013)

		Vowel–Color		Total
		Nonsynesthete	Synesthete	
Grapheme–Color	Synesthete	38	33	71
	Nonsynesthete	114	49	163
	Unknown	613	283	896
	Total	765	365	

have resulted in false positives, an issue we will address in part with our structure measure.

To examine the role that categorical perception might play in respondents' color choices relative to acoustic measures, we used the F1 and F2 values of the vowel stimuli provided by Moos et al. (2014) as our acoustic predictor. For categorical phoneme perception, each of the stimuli was categorized as a Dutch vowel phoneme. Using F1–F2 values from Dutch vowels taken from Adank, van Hout, and Smits (2004), each stimulus was categorized as an instance of its nearest Euclidean neighbor in F1–F2 space, on the basis of the overall mean F1–F2 values for each monophthong Dutch vowel.

Figure 4 shows each Dutch vowel phoneme plotted in F1–F2 space. We used a Voronoi tessellation (dashed lines), such that every point within each cell is closer to the central phoneme of the cell than to any other phoneme (measured in Euclidean distance). Vowel stimuli were categorized according to the cell they fell into, detailed in Table 6 in Appendix 1, which also details their F1–F2 values and grapheme category (according to the general Dutch orthography conventions from Nunn, 2006, p. 15).

Results

Model analyses

We used linear models to analyze participants' responses in each dimension of the CIELuv space, adopting a mixed-effect approach to account for repeated measures, with random effects for participants and trial. Where individual fixed effects or interactions did not significantly improve the model fit, they were dropped, using the `step()` function from the `lmerTest` package (Kuznetsova, Brockho, & Christensen, 2018). All of the models reported below have significantly improved fit over a null model or simpler alternatives. The models were analyzed in R using the `lme4` package (Bates, Maechler, Bolker, & Walker, 2013); *F* and *p* values were estimated using Straitherwaite approximations and the `anova()` function in the `lmerTest` package (Kuznetsova et al., 2018).

We examined the continuous predictors of F1 and F2 vowel formants, as well as the categorical predictors of vowel phoneme category and grapheme category, summarized in Table 2. We compared models with different predictors using the `sem.model.fits` function in the `piecewiseSEM` package (Lefcheck, 2016), which compares Akaike information criterion (AIC) values across models and estimates the marginal (fixed effect) and conditional (fixed + random effect) R^2 values for each model. Although phonemic, graphemic, and acoustic factors all play roles, we found that vowel phoneme category is the best overall predictor of color choice, and

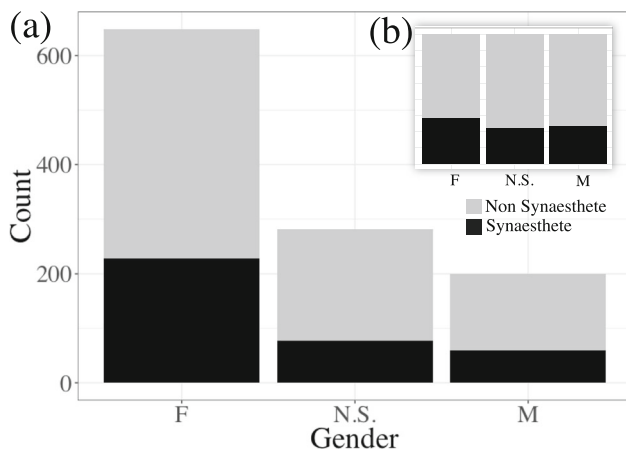


Fig. 3 Rates of synesthesia, according to consistency scores, among female (F), male (M), and other (N.S., not selected) participants in our sample. The inset shows the proportions for each group, showing that despite the greater number of females overall, the rates of synesthesia among the three groups were comparable

that grapheme category accounts for more variation than acoustic factors. Details of the grapheme models are included in Appendix 2, but they will not be discussed at length here, since vowel category is a better predictor of responses. In the following analyses, we focus first on acoustic factors for a comparison with earlier work, followed by further analyses of vowel phoneme category.

Acoustic factors Figure 5 shows the results in color space in the same way as in Moos et al. (2014). As in Moos et al.'s study, we found that synesthetes chose a generally wider range of colors, but the general shape of the mapping between vowels and colors was shared between synesthetes and nonsynesthetes. Figure 6 shows more exact color values chosen in each dimension for each item, plotted in F1–F2 space.

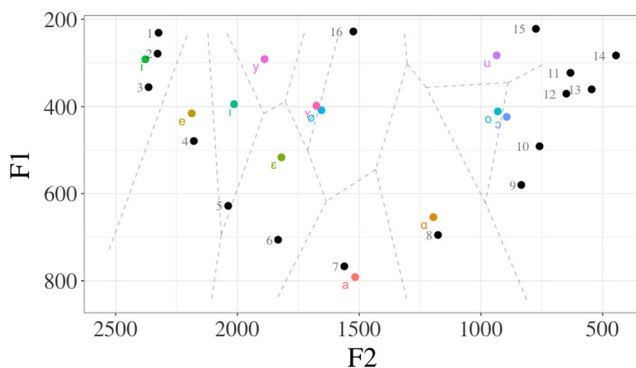


Fig. 4 Canonical Dutch monophthong vowels plotted in F1–F2 space, taken from Adank et al., 2004. The dashed lines indicate a Voronoi tessellation: Each cell contains all points that are closest to the canonical vowel within that cell. Thus, the vowel stimuli (numbered and indicated by black dots) were classified according to their nearest phoneme neighbor in Euclidean space

In the formant models described below, F1, F2, and synthetic status were tested as fixed-effect predictors of L , u , and v . The models tested interactions with each acoustic predictor and synesthetic status (but not between acoustic predictors) as fixed effects, after Moos et al. (2014). Below we describe the results for L , u , and v , respectively.

For acoustic factors, more variance was accounted for in lightness choices ($\sim 15\%$, marginal $R^2 = .155$) than in the u and v dimensions, where the fixed effects accounted for less than 10% of variation (u , red–green, marginal $R^2 = .017$; v , blue–yellow, marginal $R^2 = .08$).

Participants generally chose lighter colors for more front vowels and darker colors for more back vowels (Fig. 6, top left), reflected by a strong significant main effect of F2 in the model (Table 3). The model also showed that participants chose slightly lighter colors for lower vowels. Synesthetes were likely to choose slightly darker colors overall, but an interaction between F2 and synesthetic status showed that synesthetes tended to choose slightly lighter colors for front vowels than nonsynesthetes. On the other hand, the interaction between F1 and synesthetic status showed that synesthetes chose slightly darker colors for lower vowels (higher values of F1) than nonsynesthetes. Although the synesthete/nonsynesthete contrasts are strongly significant, they are difficult to detect in Fig. 6, since the color swatches reflect means of choices from a large, continuous color palette instead from a predefined set of 16 colors, as in Moos et al. (2014). For the specific estimates and p values of the fixed effects, see Table 3. For examples of the individual response patterns, see Fig. 10 below.

In terms of the u (green–red) scale (Fig. 6, top right), acoustic factors accounted for the least amount of variation in the u (green–red) scale, with less than 2% of the variation in responses being predicted by variation in F1, F2, and synesthetic status. Participants generally chose redder colors for lower vowels (higher F1) and more back vowels (higher F2). There was no main effect of synesthetic status in this dimension; however, there were significant interactions with F1 and F2. Synesthetes chose colors that were greener for low and back vowels than did nonsynesthetes. The detailed estimates, along with F and p values, are outlined in Table 4.

On the blue–yellow (v) dimension (Fig. 6, bottom), the interaction between F1 and synesthetic status was dropped, since it did not improve model fit. Participants preferred yellower colors for high (low F1) and front (high F2) vowels. Synesthetes preferred bluer colors overall; however, they chose significantly yellower colors for front vowels than did nonsynesthetic participants, reflected in the significant interaction between F2 and synesthetic status. The detailed estimates, along with F and p values, are outlined in Table 5.

Table 2 Marginal and conditional R^2 values for each predictor in each dimension of color space

		Marginal R^2	Conditional R^2	dAIC	Fixed effects
L	Acoustic	.155	.336	456	F1 + F2 + vSyn + F2 × vSyn + F1 × vSyn
	Vowel	.161	.342	–	vowel × vSyn
	Grapheme	.154	.340	463	grapheme × graphemeSyn
u	Acoustic	.017	.131	379	F1 + F2 + vSyn + F2 × vSyn + F1 × vSyn
	Vowel	.024	.138	–	vowel × vSyn
	Grapheme	.023	.136	62.12	grapheme × graphemeSyn
v	Acoustic	.080	.219	264	F1 + F2 + vSyn + F2 × vSyn
	Vowel	.085	.223	–	vowel × vSyn
	Grapheme	.081	.220	190	grapheme × graphemeSyn

dAIC stands for delta AIC (Akaike information criterion), which is an estimate of fit. Lower values of AIC indicate better model fit; where this number is omitted, it indicates the lowest AIC value. For all models, trial and participant were included as random effects. In the fixed effects described in the last column, *vSyn* stands for vowel synesthete, and *graphemeSyn* stands for grapheme synesthete, whereas *vowel* and *grapheme* indicate the respective categorical predictors. In all dimensions, vowel category predicts the most variation in responses.

Phonemes and graphemes In the categorical predictor models described in this section, phoneme (and grapheme) category and synesthetic status were fixed-effect predictors of L , u , and v . As with acoustic predictors, these models also tested interactions between phoneme (and grapheme) category and synesthetic status as fixed effects. For grapheme category, grapheme synesthete status was used as the synesthetic predictor instead of vowel synesthete status. Since vowel category was a better predictor of responses, we will not detail the grapheme results here, but they are provided in Appendix 2. The detailed results of the vowel models are provided below.

For vowel category, we observed significant main effects of this variable in all three color dimensions (L , $F = 1,493$, $p < .001$; u , $F = 166$, $p < .001$; v , $F = 660$, $p < .001$), as well as a significant main effect of synesthetic

status in the L and v dimensions (L , $F = 10.73$, $p = .001$; u , $F = 0.75$, $p = .39$; v , $F = 660$, $p < .001$). These results echo those found for acoustic factors, with synesthetes choosing lighter (estimate = 1.34, $SE = 0.49$) and yellower (estimate = 6.92, $SE = 2.71$) colors than nonsynesthetes.

To assess specific effects for vowels, we calculated contrasts between all vowel categories in each dimension, with Bonferroni adjustments for all reported p values, using the `lsmeans` package (Lenth, 2016). The differences between vowel categories were mostly highly significant for all color dimensions and are fully listed in Appendix 3, with estimated means, confidence intervals, t ratios, and p values. Since the contrasts were mostly significant, Fig. 7 shows which vowels were *not* different from one another, using black lines for uncorrected nonsignificant contrasts

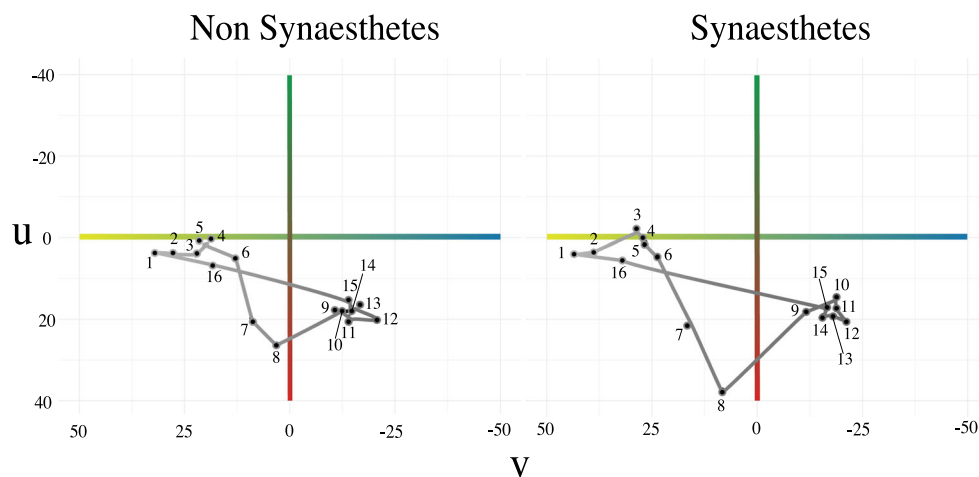


Fig. 5 Vowels plotted in color space for nonsynesthetes (left) and synesthetes (right), after Moos et al. (2014). Note that the colors represented here are slightly idealized, for comparison with the plots in Moos et al. (2014), and the lightness choices indicated are an approximate

gradient interpolation between the lightest and darkest values chosen (i.e., the midrange values are not exact). For more precise color and lightness choices, see Fig. 6

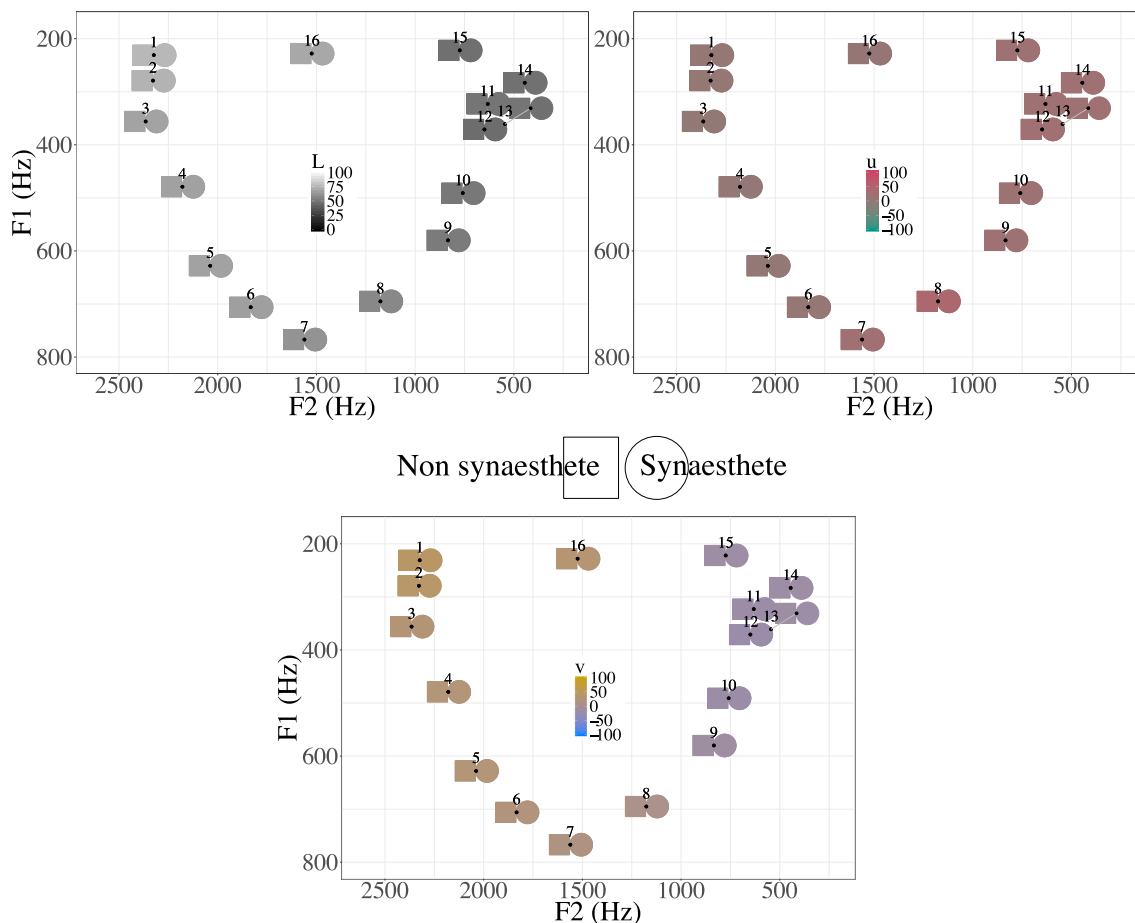


Fig. 6 Mean *L* (top left), *u* (top right), and *v* (bottom) color values chosen for vowel items, as plotted in F1–F2 space for synesthetes (squares) and nonsynesthetes (circles). Axis labels indicate vowel quality (vowel height on the F1 axis and front/back vowels on the F2 axis). For each color scale,

the values shown reflect the mean for the item in the relevant dimension, with the other dimensions held at middle values (the middle values for *L*, *u*, and *v*, are 50, 0, and 0, respectively)

($p > .05$), a dotted line for a contrast in which $p < .05$, and a dashed line for a contrast in which $p < .01$. In other words, the more solid a line between two vowels is, the more similarly participants responded to them. For all unmarked contrasts, $p < .001$. Vowels are plotted with their canonical Dutch phoneme values from Adank et al. (2004). Figure 7 shows the results in each dimension.

In all dimensions, the vowels /e, i, ε, ø/ tended not to be significantly different, forming a “mid- front” grouping for which participants generally chose similar colors that were lighter, greener, and yellower than those of low back vowels. In both the *L* and *v* spaces, the vowel /i/ stood apart from this group, being even lighter and yellower. The vowels /u/ and /ɔ/ were also informally grouped in the *u* and *v* dimensions. In the

Table 3 Model: $L \sim F1 + F2 + vSyn + F2 \times vSyn + F1 \times vSyn + (1 | Participant) + (1 | Trial)$

	Est.	SE	F	p
(Intercept)	43.62	0.454	–	–
F1	0.001	0.0004	5.55	.018
F2	0.010	0.0001	11,318	< .001
vowelSyn	– 1.80	0.728	6.11	.013
F1 × vowelSyn	– 0.003	0.001	248	< .001
F2 × vowelSyn	0.004	0.0002	24.56	< .001

Table 4 Model: $u \sim F1 + F2 + vSyn + F2 \times vSyn + F1 \times vSyn + (1 | Participant) + (1 | Trial)$

	Est.	SE	F	p
(Intercept)	20.16	1.09	–	–
F1	0.012	0.001	126	< .001
F2	– 0.009	0.0004	864	< .001
vowelSyn	– 0.2295	1.93	0.01	.905
F1 × vowelSyn	0.006	0.003	5.37	.020
F2 × vowelSyn	0.001	0.0007	4.43	.035

Table 5 Model: $v \sim F1 + F2 + \text{vowelSyn} + F2 \times \text{vowelSyn} + (1 | \text{Participant}) + (1 | \text{Trial})$

	Est.	SE	F	p
(Intercept)	-26.72	1.26	-	-
F1	0.005	0.001	12.17	< .001
F2	0.024	0.0004	860.09	< .001
vowelSyn	-6.25	1.99	2.201	.002
F2 \times vowelSyn	0.007	0.0007	3.35	< .001

u dimension, these also converged with /a/, although /a/ was set apart as the reddest of all the vowels.

Contrasts between synesthetes and nonsynesthetes were not calculated for the *u* dimension, due to the lack of a main effect of synesthetic status. For the other dimensions, significant contrasts between synesthetes and nonsynesthetes are marked in Fig. 7 as follows: *** $p < .001$, ** $p < .01$, * $p < .05$.

The vowels in the “mid-front” grouping described above were generally lighter for synesthetes than for nonsynesthetes, and this was also true for the highest front vowel /i/. In the *v* dimension, synesthetes’ choices were generally yellower than nonsynesthetes’ for the same “mid-front” grouping, and also yellower for the high front vowel /i/. They were also slightly yellower for the low vowels /a/ and /ɑ/.

Summary In a large sample of Dutch speakers, we found evidence of shared vowel–color associations. As in earlier work, our data showed that the acoustic factors F1 and F2 were predictive of color choices: Higher values of F1 (i.e., lower vowels) are darker, redder, and bluer; higher values of F2 (i.e., more front vowels) are lighter, greener, and yellower. These results echo those found for English speakers by Moos et al. (2014), and for Korean–English bilinguals in Kim et al. (2018). Although the general shape of associations is shared across synesthetes and nonsynesthetes (Fig. 5), synesthetes

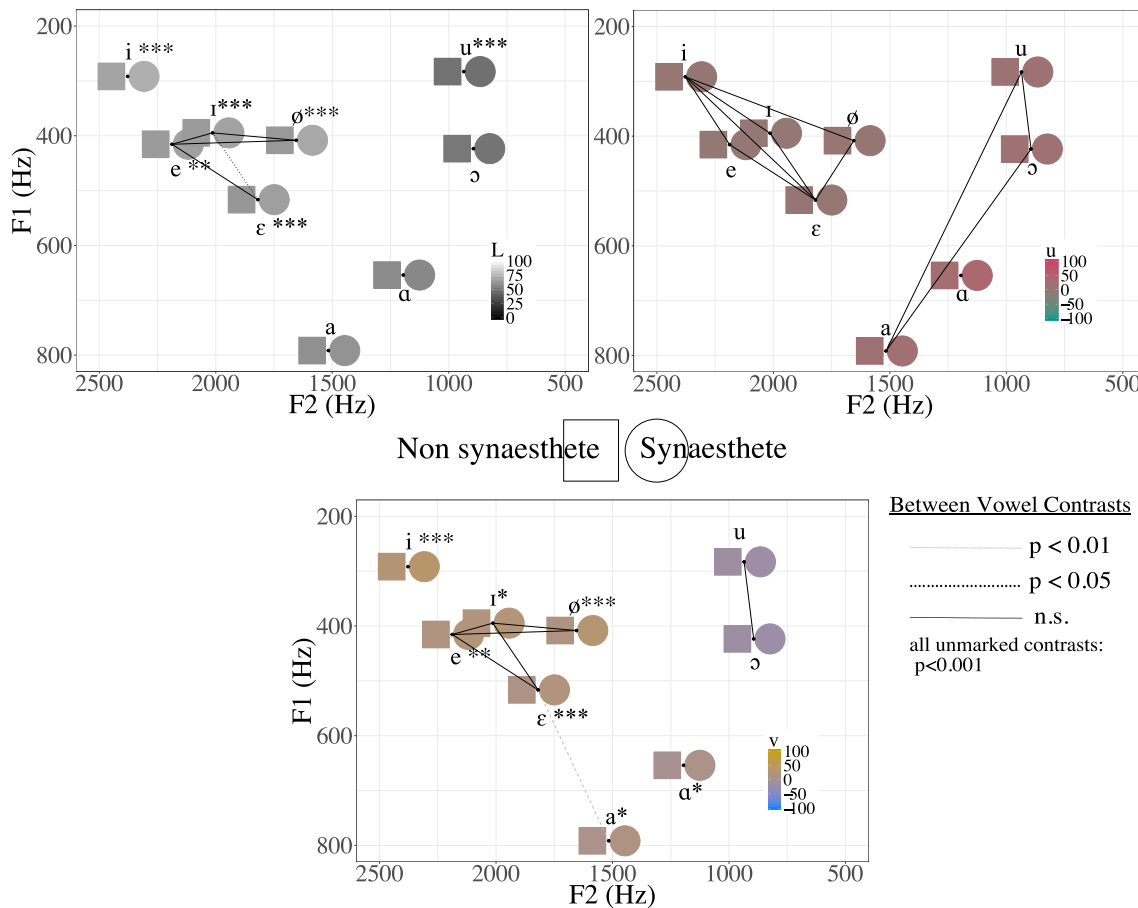


Fig. 7 Mean color choices for and contrasts between vowel categories plotted in F1–F2 space, labeled with canonical Dutch phoneme values from Adank et al. (2004). The axis labels indicate vowel quality (vowel height on the F1 axis, and front/back vowels on the F2 axis). For each color scale, the values shown reflect the mean for the item in the relevant dimension with the other dimensions held at middle values (the middle values for *L*, *u*, and *v*, are 50, 0, and 0, respectively). Solid lines between

vowels indicate nonsignificant contrasts, as per the legend. Thus, clusters of vowels connected by lines did not elicit significantly different color choices from participants, and can be interpreted as informal groups. Asterisks next to a phoneme indicate that synesthetes and nonsynesthetes chose significantly different colors for that phoneme (see the text for details)

show more extreme color and lightness choices, selecting lighter and yellower colors for high values of F2 especially. The differences between synesthetes and nonsynesthetes in our results are not as marked as those reported by Moos et al.; we address several potential reasons for this in the General Discussion.

However, we also found that an approximation of phoneme category is a better predictor of color choice than the acoustic measures, indicating that categorical perception can shape the structure of cross-modal associations. As would be expected, given the acoustic results, front vowels are lighter, greener, and yellower, whereas low and back vowels are darker, redder, and bluer. There were generally no significant differences between the “mid-front” group of /e/, /ɪ/, /ø/, and /ɛ/. However, particularly for this group, synesthetes chose slightly lighter and yellower colors than did nonsynesthetes. They also chose lighter and yellower colors for the high front vowel /i/, and darker colors for the low back vowel /u/.

Although the values from the phoneme model are very close to those from the grapheme model, comparisons of the two predictors showed that the phoneme model is significantly superior in every color dimension. It may be that grapheme category is only a good predictor of color choices insofar as it is a fairly good, though imperfect, predictor of vowel category. Although a rough mid-front vowel grouping emerged in the phoneme category analyses, this does not correspond to a larger grapheme grouping, since the four relevant phonemes map onto three different graphemes (/e/ and /ɛ/ to *e*, /ɪ/ to *i*, and /ø/ to *u*) in Dutch orthography.

Mapping structure

So far, our analyses were largely concerned with the kind of questions asked in classic cross-modal association studies, linking specific colors or color dimensions to acoustic and phonemic features. We also looked at contrasts between synesthetes and nonsynesthetes, based largely on consistency across trials. Although the temporal consistency of mappings is rightly considered a benchmark of genuine synesthesia, and some earlier studies have considered how the mappings of synesthetes relate to those of nonsynesthetes, less consideration has been given to the internal structure of synesthetic and cross-modal mappings. The traditional approach tells us something about whether synesthetes choose colors for sounds that are different from or similar to the kinds chosen by nonsynesthetes, but it is less adept at detecting overall structure in cross-modal mappings or telling us whether the mappings of synesthetes are more internally structured than those of nonsynesthetes. Are there structural regularities in how we link one sensory domain to another? Does the shape of the vowel space map onto

the color space more reliably for synesthetes than for nonsynesthetes?

We operationalized structure in this context by comparing paired distances across spaces, using a method borrowed from ecology (Mantel, 1967). This method has been used extensively in iterated artificial-language learning studies to detect structured mappings between form and meaning spaces (e.g., Kirby, Cornish, & Smith, 2008). The Mantel test in this context measures whether distances in form correlate with distances in meaning. To the extent that they do, we can say there is structure in the mappings between two spaces.

In the context of the present data, for example, a mapping would be structured when pairs of vowels that are similar in F1–F2 acoustic space map onto pairs of colors that are similar in three-dimensional color space, and when pairs of vowels that are dissimilar in F1–F2 acoustic space map onto pairs of colors that are dissimilar. Thus, structure implies a degree of isomorphism across (multidimensional) sensory spaces—in this case, acoustic and color spaces.

Whereas our earlier consistency measures had used CIELuv space to align with prior work (Rothen et al., 2013), for this measure we used the related CIELab space. In CIELab space, the *L* dimension is identical, whereas *a* corresponds to a green–red continuum (similar to *u*) and *b* corresponds to a blue–yellow continuum (similar to *v*). The benefit of CIELab space for the structure measure is that it allows us to use more perceptually realistic distances, specifically ΔE_{2000} (Sharma, Wengcheng, & Dalal, 2005). The ΔE_{2000} distance takes into account that Euclidean distances have nonuniform perceptual effects, particularly at the edges of the color space. For example, as lightness increases to the white point, the perceptual differences between chroma shrink and eventually disappear, even though their plain Euclidean distance in the CIELab or CIELuv spaces would be identical to those of two perceptually distinct colors elsewhere in the space. Therefore, our structure measure relies on ΔE_{2000} distances in CIELab space. For vowel distances, we used Euclidean distance in F1–F2 space using the canonical phoneme values shown in Fig. 7.⁴ Since Euclidean distance can be high-dimensional, this allowed us to use all three dimensions of a participants’ color response to an item at once.

⁴ ΔE_{2000} distances may also be more accurate for detecting synesthesia using consistency scores. However, we were unable to use this measure to detect synesthetes, because there is no benchmark for this in the literature. Rothen et al. (2013) provided an evaluation of CIELab-based consistency scores that used straight Euclidean distance in CIELab space (also known as ΔE_{1976}), but they did not test ΔE_{2000} . They reported superior performance in detecting synesthetes using CIELuv distances, so these were preferred for our consistency scores and were used to classify synesthetes in the sample. However, for detecting structural isomorphism, we used ΔE_{2000} because this is likely to better capture real perceptual distances.

Where there is structure, the pairwise distances within each space will be correlated with one another. Once we had pairwise distances for every mapping in each space from our participants' data, we permuted the vowel–color mapping between the two spaces and then recalculated the pairwise distances in order to get a distribution of potential correlations between the two spaces. Using this distribution, we obtained a z score indicating where a given participant's mappings were on the actual distribution, and a p value that indicated the likelihood that any random mapping of the vowel–color space would be more structured than the actual one.⁵ In other words, $p < .05$ in this case means that fewer than 5% of mappings generated in the simulation were more structured than the real one.

Pairwise distances were calculated between all sounds and all colors chosen by a particular participant, and a veridical correlation was calculated between these distance matrices. To account for multiple responses to the same stimuli, the responses were first shuffled within a particular item, and then across items. These matrices were then shuffled into 10,000 random permutations, each with its own r , allowing us to calculate a z score as described above. Python code for performing the Mantel structure analyses is available online at <http://github.com/mdingemane/colouredvowels>. Figure 8 shows a density plot of z scores of synesthetes and nonsynesthetes in the vowel–color association task.

Three findings stand out. First, the mappings of synesthetes tend to be more structured than those of nonsynesthetes ($t = -7.09$, $df = 660$, $p < .001$). Second, the majority of participants' mappings are more structured than would be expected by chance: All participants to the right of the vertical dotted line had correlations between the vowel and color spaces greater than 95% of random permutations generated by the Mantel test. Third, there is correlation between the structure score and CIELuv consistency scores: Participants with more consistent associations across trials (i.e., lower consistency scores) tended to have more structured mappings across the vowel and color spaces ($r = -.313$, $t = -11.07$, $p < .001$; Fig. 9).

This measure provides a new way to quantify the structure of cross-modal mappings and is a valuable quantitative complement to traditional unimodal consistency scores. Figure 10 shows individual participants that fall

in specific parts of the consistency–structure space. The participant in panel a, who was classified as a synesthete according to consistency, appears to have achieved this by having high consistency across items (i.e., choosing the same color regardless of stimulus or trial), rather than by having structured, temporally consistent associations. This indicates that participants with high consistency and low structure are less likely to be genuine synesthetes, perhaps explaining the slight peak of unstructured synesthetes in Fig. 8. The participant in panel b has both low structure and low consistency, having chosen idiosyncratic colors on each trial and across the space, and sometimes mapping distant vowels (e.g., low-central and high-back/high-front vowels) with similar colors.

A nonsynesthete participant with middling consistency and significant but not especially high structure is shown in panel c. This participant shows structured mappings for some parts of the space—for instance, showing similar yellow/green choices for the cluster of high-front vowels, and brownish choices for the cluster of high back vowels, resulting in significant structure. However, this participant was inconsistent across trials for the same item, distinguishing the participant from highly structured synesthetes like the participant in panel e, showing categorical, structured associations that are highly consistent across trials. Finally, panel d shows a participant with high structure but low consistency: This participant made structured mappings across the space, but seems to have done so differently on each trial, as indicated by the inversions of green/blue in mid-front vowels and red/blue in back vowels across trials.

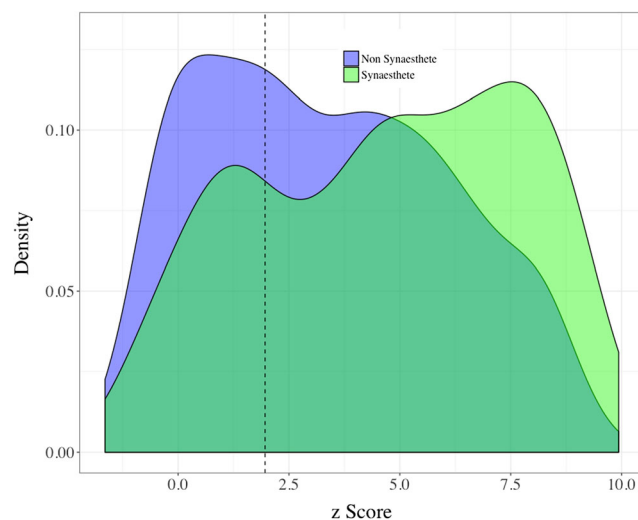


Fig. 8 Distributions of structure scores among synesthetes and nonsynesthetes. Higher z -score values indicate more structured mappings. Values to the right of the dashed line indicate mappings that are significantly structured

⁵ Participants' z scores and p values are provided as part of their structure score in the data repository at <https://github.com/mdingemane/colouredvowels>, along with Python code to calculate pairwise distances and run the Mantel simulations. However, note that if recomputed, these values would differ very slightly for each participant, since they are based on a simulated distribution created when the scores were calculated. Recomputing these values would also involve creating new distributions, which would slightly alter *exactly* where participants fall on those distributions. However, the overall picture described in this section would hold.

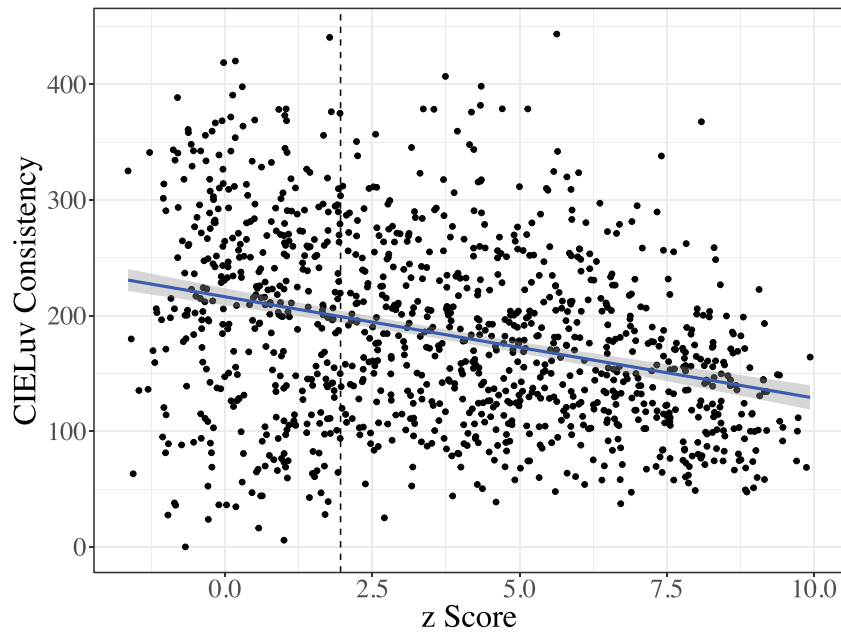


Fig. 9 Relationship between structure score and consistency score, showing that participants with more consistency (i.e., lower CIELuv color distance) across trials tend to have more structured mappings (i.e., higher z scores)

Discussion

We examined vowel–color associations in a large number of participants in an online study. Four key findings emerge.

First, acoustic factors (F1 and F2) predict some of the variance in color responses, replicating a result by Moos et al. (2014), but in a different and larger population. Participants chose lighter colors for more front vowels (i.e., vowels with higher F2 values), redder colors for lower vowels (i.e., vowels

with higher F1 values), and greener colors for high and front vowels (i.e., lower F1 and higher F2 values, respectively). Synesthetes showed the same patterns but were slightly more extreme, choosing even lighter and yellower colors for high vowels in particular, an effect also found by Moos et al. Overall, the locations of associations in the color space were not identical to those found by Moos et al., but the general differences in shape were comparable: Synesthetes’ choices were generally more extreme than those of nonsynesthetes.

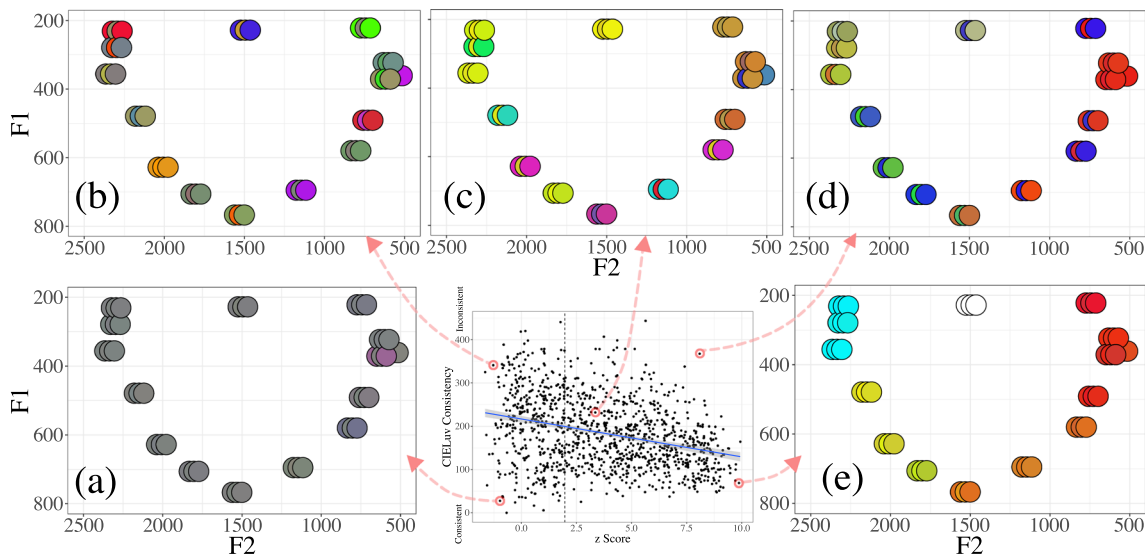


Fig. 10 Mappings of individual participants showing, clockwise from bottom left, (a) a participant with very low structure yet high consistency across trials and items, probably indicating a false positive synesthete, (b) a typical nonsynesthete with inconsistent and unstructured mappings, (c)

a middling participant with significant structure but inconsistent choices across trials, and (d) a highly structured but inconsistent participant, and (e) a typical vowel–color synesthete, with highly structured, consistent and categorical mappings

Our findings resemble those of previous studies of vowel–color mappings in smaller samples of (non)synesthetes, with lighter colors (yellow, green) being associated with more front vowels (higher F2; e.g., /i/, /e/), darker colors (e.g., red, brown, blue) being associated with back vowels (/o/, /u/), and redder colors being associated with low front vowels (e.g., /a/) (Marks, 1975; Wrembel, 2009; see Guillamon, 2014, for a cross-linguistic overview, particularly their Table 5, p. 44). The association of front vowels with high F2 to lighter colors could be related to more widespread cross-modal phenomena such as pitch–lightness associations (e.g., Ward et al., 2006), although this potential relationship requires further study, since vowel quality and pitch are independent (see Ohala, 1994, for an account that integrates these). Size–sound symbolism may also play a role: The space in the oral cavity is smaller for higher and more front vowels, which may in turn be associated with lightness and brightness (Cuskey & Kirby, 2013).

Second, color associations are predicted better by phonemes than by acoustic features, showing that cross-modal associations between vowel sounds and colors are modulated to an important degree by categorical perception (Goldstone & Hendrickson, 2010). Prior work on sound–color associations has mostly focused on how gradual changes in chromaticity are associated with low-level acoustic factors (Moos et al., 2014; Ward et al., 2006). Here we have shown that shifts in color associations correspond to category boundaries in participants' vowel systems. The importance of categorical structure in vowel–color associations has implications for the underlying nature of synesthetic associations, pointing to the important role of learning (Mroczko-Wasowicz & Nikolić, 2014; Simner, 2007).

For example, recent studies have shown that learning can play an influential role in synesthesia: Specifically, the presence of colored fridge magnets in childhood was formative for grapheme–color synesthesia (Witthoft & Winawer, 2013; Witthoft, Winawer, & Eagleman, 2015). However, the learning implied in the categorical effects that we observed is perhaps qualitatively different: Rather than being based on highly specific early perceptual experiences, our effects arise from a categorical warping of the vowel space that is a crucial part of spoken language acquisition.

Although our results indicate that acquired categories play a key role in vowel–color associations, there remains an important role for lower-level acoustic perceptual cues. For example, in the *u* dimension, we observed synesthetic effects for acoustic factors but not for phoneme category. This may point to the relevance of lower-level acoustic properties for some synesthetes specifically. It is possible that synesthetes react to acoustic factors in a way that nonsynesthetes do not, making acoustic factors a stronger predictor of color choices for synesthetes than for nonsynesthetes. Further targeted studies will be required in order to address this question.

Third, we have introduced a novel measure of structure in cross-modal associations. Most participants showed a significant degree of structure, implying that vowel–color associations rely at least in part on establishing structural isomorphism across perceptual domains. We found that this measure correlates with consistency across trials: Participants with more temporally consistent associations tend to have more structured mappings, with synesthetes being at the most consistent and structured extreme. Although synesthetic associations are idiosyncratic to some degree, the prevalence of structure shows that these associations share similarities with cross-modal associations in the general population (Simner et al., 2005). This implies that private, involuntary synesthetic associations, as well as overt, elicited cross-modal associations, may be underpinned to a significant degree by common principles of isomorphic mapping across sensory modalities. Our structure measure makes it possible to probe cross-modal and synesthetic associations in a way that is at least partly independent of consistency. This has the potential to provide a valuable complementary measure of genuineness in synesthesia (Simner, 2012). The structure measure can be applied in any domain in which the perceptual features of stimuli and responses (or of inducers and concurrents) are quantifiable in terms of some distance measure.

Fourth, we have shown that an online task can be used to learn about cross-modal associations and to infer synesthesia for a subset of participants along the way, providing a scalable method for identifying synesthetes and for studying the relation between cross-modal associations and synesthetic mappings from a population perspective. Although online tasks have become a fixture of synesthesia studies at least since Eagleman et al. (2007), they have mostly been used to test preselected sets of participants. Here we have shown that a widely advertised survey can succeed in capturing a broad and diverse sample of the population, including synesthetes. To foster more of this work, the code underlying our web application is openly available.

Limitations and future work

Although acoustic and vowel phoneme predictors accounted for significant amounts of variation in color choices, as shown by the analyses above, the conditional R^2 values in Table 2 indicate that individual variation accounted for at least as much variation in responses (the marginal-conditional R^2 quantifies, roughly, the variance accounted for by both fixed and random effects). This indicates that, although there were obvious trends and although our structure measures point to a shared capacity for structure mappings, the specific identity of mappings may still vary considerably across participants. However, our structure results show that the capacity to structure mappings is strong and is shared across much of the population.

Further targeted studies should be done to tease apart the relative contributions of vowel category and grapheme category, especially the extent to which these may interact with how stimuli are presented. The data analyzed here were responses to auditory stimuli. Although there is evidence that auditory stimuli automatically activate graphemic representations to some extent (Ziegler & Ferrand, 1998), auditory presentation privileges acoustic features over graphemes (Cuskley et al., 2017), and graphemic representations may be less likely to be activated when auditory stimuli are presented devoid of a word-like context (e.g., within a nonword). Synesthetes tend to associate similar colors to similarly shaped graphemes (Brang, Rouw, Ramachandran, & Coulson, 2011; Eagleman, 2010; Jürgens, Mausfeld, & Nikolic, 2010; Watson, Akins, & Enns, 2012), implying an important role for visual grapheme shape in mediating the associations. However, other factors, such as ordinality in the alphabet and frequency in language, also play a role (e.g., van Leeuwen, Dingemans, Todil, Agameya, & Majid, 2016; Watson et al., 2012).

Although most of the broad patterns found in Moos et al. (2014) were replicated in our data, the differences between the associations of synesthetes and nonsynesthetes was less pronounced than in their study (Fig. 5, and cf. Fig. 2 in Moos et al., 2014, p. 136). Two important differences between the studies are sheer sample size and online presentation, both of which added more variability to the responses in our study. Another possibility is that the confined set of 16 color swatches used by Moos et al. constrained responses in such a way that the divisions between synesthetes and nonsynesthetes became clearer. It may also be that Dutch speakers differ from English speakers in the finer details of how they map vowels onto color space. However, despite the differences, our results show quantitative differences between synesthetes and nonsynesthetes, as well as an important degree of shared mappings.

Although we expect the general role of categorical perception to be replicated across languages (mediated by known acoustic factors), we do not expect the precise groupings of vowel sounds to be replicated, because this is a function of how acoustic space is carved into a language-specific phoneme inventory. This opens up the possibility of a degree of linguistic relativity in cross-modal associations and synesthetic experience. Just as lexicalization patterns in the domain of color can shape low-level processes of color perception (Roberson, Pak, & Hanley, 2008), so phonemic structure may shape cross-modal associations. This is one place where linguistic diversity in phonetics, phonology, and orthography can be used to learn more about the mechanisms underlying vowel–color associations and to tease apart the roles of acoustic, phonemic, and graphemic features

in cross-modal associations (Root et al., 2018; van Leeuwen et al., 2016).

Finally, further work will be needed to combine consistency scores with structure scores to create reliable behavioral indicators of genuine synesthesia, especially for use with large-scale online methodologies. Earlier reports of consistency scores had used them primarily to confirm self-identified synesthetes, rather than to detect synesthesia in a random sample. Figure 10a shows that our consistency cutoff to identify synesthetes (taken from Rothen et al., 2013) likely resulted in some false positives. A few potential issues may have contributed to this problem. First, Rothen et al. measured consistency in a grapheme–color task, which had 36 items. It may be that with fewer items (16 vowels), the consistency threshold was easier to pass. More importantly, consistency across items also needs to be taken into account: If a participant is consistent across trials and items, this is a likely flag for a false positive synesthete. Our structure measure captured this well, since cross-item distances form the core of the measure. Further work with confirmed synesthetes and nonsynesthetes will be needed in order to fully understand how best to combine structure and consistency scores to reliably detect synesthetes.

Conclusions

In the first half of the 20th century, a common term for synesthesia was *audition colorée*, or “colored hearing”, after one of the most commonly reported forms of synesthesia: a connection between vowels and colors. Colored hearing was famously described by Nabokov (1989) and studied by such noted linguists as Roman Jakobson (Reichard, Jakobson, & Werth, 1949). This early focus on sound–color associations probably was one of the reasons for a fruitful period of experimental work on cross-modal correspondences between sound and color (e.g., Marks, 1978). In contrast, most modern work on synesthesia has tested graphemes in alphabetic writing systems (written representations of speech sounds), and indeed, grapheme–color synesthesia is likely by far the most studied variant of synesthesia to date.

Here we have brought these traditions together in pursuit of fundamental questions into the nature of synesthesia and cross-modal correspondences. We studied associations between vowel sounds and colors in over 1,100 people, including hundreds of synesthetes. We replicated earlier findings on the relation between acoustic cues and color choices, but additionally showed an important role for categorical perception over and above such cues. Our findings underline the roles of learned categories and structural isomorphisms in the cross-modal associations made by synesthetes and nonsynesthetes alike. The measure of structural isomorphism we have introduced can help create more nuanced diagnostic tools for synesthesia. As work on synesthesia and cross-modal associations grows to accommodate larger samples and more varied

measures, it will provide fundamental insights into how mappings across sensory modalities are made and maintained.

Acknowledgments Open access funding provided by Max Planck Society.

Author note C.C. performed the analyses and led in writing the manuscript, M.D. and T.v.L. designed the study and collected the data, M.D. developed the analysis pipeline, and S.K. devised the structure analyses.

All authors contributed to the write-up. We thank Anja Moos for sharing the auditory stimuli with us. Financial support for this work came from NWO Veni grants to M.D. (016.154.087) and T.v.L. (451.14.025), and from a Groot Nationaal Onderzoek grant from NWO, with further support from NTR/De Kennis van Nu, Quest, the Max Planck Institute for Psycholinguistics, the Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, and the British Academy (C.C., Postdoctoral Fellowship Grant No. pf150065).

Appendix 1: Audio file categorization

Table 6 F1 and F2 values of audio stimuli, as well as canonical Dutch phoneme categories and graphemes. F1 and F2 values of assigned Dutch phonemes are also listed

Stimulus number	Original Moos et al. (2014)		Canonical Dutch by phoneme Category from Adank et al. (2007)		Phoneme category	Grapheme category
	F1	F2	F1	F2		
1	231	2,324	291.75	2,378	i	i
2	279	2,328	291.75	2,378	i	i
3	356	2,365	291.75	2,378	i	i
4	479	2,179	415.50	2,187.75	e	e
5	628	2,038	394.75	2,013.75	I	i
6	706	1,833	516.5	1,819.25	ε	e
7	767	1,561	791.75	1,516.5	a	a
8	695	1,176	654	1,195	ɑ	a
9	580	834	423.5	894	ɔ	o
10	491	759	423.5	894	ɔ	o
11	323	632	423.5	894	ɔ	o
12	371	649	423.5	894	ɔ	o
13	361	545	423.5	894	ɔ	o
14	283	445	423.5	894	ɔ	o
15	222	774	283	935	u	œ
16	228	1,524	408.25	1,654	ø	u

Appendix 2: Grapheme predictor models

Lightness (*L*)

Formula: $L \sim \text{graphcat} \times \text{graphSyn} + (1 \mid \text{Participant}) + (1 \mid \text{Trial})$

Random effects:

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	72.8006 8.5323	
Trial	(Intercept)	0.1121	0.3348
Residual		266.0102 6.3098	
Number of obs: 52,320	groups: profileid	1,130; Trial	3

Fixed effects:

	Estimate	SE	<i>t</i> Value
(Intercept)	58.43613	1.30452	44.795
graphcat[e]	6.58091	1.13204	5.813
graphcat[i]	10.37372	0.97758	10.612
graphcat[o]	-9.50099	0.92226	-10.302
graphcat[u]	8.18896	1.39438	5.873
graphcat[œ]	-11.44930	1.38716	-8.254
graphSyn[No]	-0.08566	1.54476	-0.055
graphSyn[Unknown]	0.74399	1.34078	0.555
graphcat[e] × graphSyn[No]	-2.16767	1.35470	-1.600
graphcat[i] × graphSyn[No]	-2.62095	1.1698	-2.240
graphcat[o] × graphSyn[No]	3.0334	1.10324	2.750
graphcat[u] × graphSyn[No]	-2.29062	1.66649	-1.375
graphcat[œ] × graphSyn[No]	2.05050	1.65938	1.236
graphcat[e] × graphSyn[Unknown]	-1.52903	1.17683	-1.299
graphcat[i] × graphSyn[Unknown]	-1.49775	1.01636	-1.474
graphcat[o] × graphSyn[Unknown]	0.19959	0.95891	0.208
graphcat[u] × graphSyn[Unknown]	-2.22422	1.44914	-1.535
graphcat[œ] × graphSyn[Unknown]	0.02415	1.44240	0.017

Green-red (*u*)

Formula: $u \sim \text{graphcat} \times \text{graphSyn} + (1 \mid \text{Participant}) + (1 \mid \text{Trial})$

Random effects:

Groups	Name	Variance	Std. Dev.
profileid	(Intercept)	3.435e+02	18.53454
Trial	(Intercept)	2.321e-03	0.04817
Residual		2.619e+03	51.17159
Number of obs: 52,320	groups:	profileid,1130	Trial,3

Fixed effects:

	Estimate	SE	<i>t</i> Value
(Intercept)	31.1034	3.3360	9.324
graphcat[e]	-30.4597	3.5516	-8.576
graphcat[i]	-27.0697	3.0670	-8.826
graphcat[o]	-7.0261	2.8935	-2.428
graphcat[u]	-21.2296	4.3746	-4.853
graphcat[œ]	1.1076	4.3520	0.254
graphSyn[No]	-9.7412	3.9929	-2.440
graphSyn[Unknown]	-5.9603	3.4673	-1.719
graphcat[e] × graphSyn[No]	10.8361	4.2502	2.550
graphcat[i] × graphSyn[No]	8.0058	3.6703	2.181
graphcat[o] × graphSyn[No]	6.2345	3.4612	1.801
graphcat[u] × graphSyn[No]	7.9684	5.2283	1.524
graphcat[œ] × graphSyn[No]	-2.9329	5.2061	-0.563
graphcat[e] × graphSyn[Unknown]	9.1329	3.6921	2.474
graphcat[i] × graphSyn[Unknown]	5.2102	3.1886	1.634
graphcat[o] × graphSyn[Unknown]	-0.8595	3.0084	-0.286
graphcat[u] × graphSyn[Unknown]	1.9343	4.5463	0.425
graphcat[œ] × graphSyn[Unknown]	-11.9767	4.5253	-2.647

Blue-yellow (v)

Formula: $v \sim \text{graphcat} \times \text{graphSyn} + (1 \mid \text{Participant}) + (1 \mid \text{Trial})$

Random effects:

Groups	Name	Variance	Std. Dev.
profileid	(Intercept)	608.9975	24.6779
Trial	(Intercept)	0.2484	0.4984
Residual		3408.6133	58.3833
Number of obs: 52,320	groups:	profileid,1130	Trial,3

Fixed effects:

	Estimate	SE	t Value
(Intercept)	6.0206	4.1046	1.467
graphcat[e]	14.2310	4.0522	3.512
graphcat[i]	23.1225	3.4993	6.608
graphcat[o]	-16.5514	3.3013	-5.014
graphcat[u]	18.0172	4.9912	3.610
graphcat[œ]	-10.1347	4.9654	-2.041
graphSyn[No]	1.2632	4.9018	0.258
graphSyn[Unknown]	2.3055	4.2557	0.542
graphcat[e] × graphSyn[No]	-5.9137	4.8493	-1.220
graphcat[i] × graphSyn[No]	-2.9596	4.1876	-0.707
graphcat[o] × graphSyn[No]	-4.6218	3.9491	-1.170
graphcat[u] × graphSyn[No]	-0.8866	5.9653	-0.149
graphcat[œ] × graphSyn[No]	-13.5783	5.9399	-2.286
graphcat[e] × graphSyn[Unknown]	-2.7891	4.2125	-0.662
graphcat[i] × graphSyn[Unknown]	-2.8695	3.6381	-0.789
graphcat[o] × graphSyn[Unknown]	-7.1203	3.4325	-2.074
graphcat[u] × graphSyn[Unknown]	-4.1091	5.1872	-0.792
graphcat[œ] × graphSyn[Unknown]	-12.6501	5.1631	-2.450

Appendix 3: Pairwise contrasts for vowel category model

Vowel Contrasts

All results are averaged over the levels of: SynVowel. Confidence level used: .95; Conf-level adjustment: Bonferroni method for 36 estimates; p value adjustment: Bonferroni method for 36 tests.

Lightness (L)

Contrast	Estimate	SE	df	Lower CL	Upper CL	t ratio	p value
a-ɑ	2.9732174	.4325377	51203.17	1.5903401	4.35609472	6.874	<.0001
a-e	-4.7163037	.4336548	51213.20	-6.1027525	-3.32985492	-10.876	<.0001
a-ɛ	-3.5210589	.4345513	51197.25	-4.9103740	-2.13174374	-8.103	<.0001
a-i	-9.0659630	.3544838	51244.92	-10.1992923	-7.93263368	-25.575	<.0001
a-ɪ	-4.9762427	.4328268	51206.59	-6.3600442	-3.59244123	-11.497	<.0001
a-ø	-5.6924734	.4343398	51258.29	-7.0811121	-4.30383477	-13.106	<.0001
a-ɔ	10.7007156	.3321399	51236.03	9.6388227	11.76260852	32.217	<.0001
a-u	12.6825523	.4360642	51226.74	11.2884002	14.07670437	29.084	<.0001
ɑ-e	-7.6895211	.4313307	51202.01	-9.0685394	-6.31050282	-17.827	<.0001
ɑ-ɛ	-6.4942762	.4322595	51186.53	-7.8762640	-5.11228850	-15.024	<.0001
ɑ-i	-12.0391804	.3515084	51214.46	-13.1629968	-10.91536393	-34.250	<.0001
ɑ-ɪ	-7.9494601	.4305068	51195.72	-9.3258442	-6.57307602	-18.465	<.0001
ɑ-ø	-8.6656908	.4319932	51261.86	-10.0468274	-7.28455429	-20.060	<.0001
ɑ-ɔ	7.7274982	.3288896	51193.57	6.6759968	8.77899959	23.496	<.0001
ɑ-u	9.7093349	.4336279	51204.60	8.3229720	11.09569777	22.391	<.0001
e-ɛ	1.1952448	.4332962	51192.02	-0.1900573	2.58054700	2.758	.2091
e-i	-4.3496593	.3528033	51217.67	-5.4776158	-3.22170276	-12.329	<.0001
e-ɪ	-0.2599390	.4314470	51194.10	-1.6393293	1.11945126	-0.602	1.0000
e-ø	-0.9761697	.4331532	51278.00	-2.3610150	.40867548	-2.254	.8720
e-ɔ	15.4170193	.3305336	51231.99	14.3602618	16.47377680	46.643	<.0001
e-u	17.3988560	.4348562	51224.95	16.0085662	18.78914578	40.011	<.0001
ɛ-i	-5.5449041	.3541070	51218.02	-6.6770287	-4.41277953	-15.659	<.0001
ɛ-ɪ	-1.4551839	.4324424	51184.49	-2.8377565	-0.07261128	-3.365	.0276
ɛ-ø	-2.1714146	.4341534	51268.77	-3.5594575	-0.78337167	-5.001	<.0001
ɛ-ɔ	14.2217745	.3317993	51210.15	13.1609706	15.28257836	42.863	<.0001
ɛ-u	16.2036111	.4358015	51212.11	14.8102992	17.59692304	37.181	<.0001
ɪ-ɪ	4.0897202	.3517211	51204.59	2.9652237	5.21421677	11.628	<.0001
ɪ-ø	3.3734895	.3533586	51270.04	2.2437578	4.50322125	9.547	<.0001
ɪ-ɔ	19.7666786	.2161304	51281.31	19.0756826	20.45767456	91.457	<.0001
ɪ-u	21.7485153	.3557603	51240.26	20.6111049	22.88592569	61.132	<.0001
ɪ-ø	-0.7162307	.4322979	51267.44	-2.0983412	.66587975	-1.657	1.0000
ɪ-ɔ	15.6769583	.3294043	51215.65	14.6238114	16.73010532	47.592	<.0001
ɪ-u	17.6587950	.4340359	51219.43	16.2711279	19.04646215	40.685	<.0001
ø-ɔ	16.3931890	.3314333	51336.56	15.3335551	17.45282298	49.461	<.0001
ø-u	18.3750257	.4353557	51278.07	16.9831390	19.76691247	42.207	<.0001
ɔ-u	1.9818367	.3333068	51211.68	.9162131	3.04746021	5.946	<.0001

Green-Red (*u*)

Contrast	Estimate	SE	df	Lower CL	Upper CL	<i>t</i> ratio	<i>p</i> value
a-a	-11.4947594	1.3615416	51219.48	-15.8477785	-7.14174031	-8.442	<.0001
a-e	18.8711194	1.3650100	51230.83	14.5070114	23.23522741	13.825	<.0001
a-ε	15.3977264	1.3679115	51211.33	11.0243420	19.77111086	11.256	<.0001
a-i	16.9820216	1.1156590	51278.84	13.4151199	20.54892334	15.222	<.0001
a-r	18.4977927	1.3624309	51225.62	14.1419304	22.85365500	13.577	<.0001
a-ø	14.0750251	1.3669368	51287.62	9.7047567	18.44529339	10.297	<.0001
a-ο	2.1079468	1.0453726	51267.12	-1.2342403	5.45013384	2.016	1.0000
a-u	3.6939508	1.3725145	51253.33	-0.6941502	8.08205179	2.691	.2563
α-e	30.3658788	1.3577469	51218.49	26.0249920	34.70676574	22.365	<.0001
α-ε	26.8924858	1.3607458	51200.18	22.5420109	31.24296080	19.763	<.0001
α-i	28.4767810	1.1064221	51239.41	24.9394110	32.01415106	25.738	<.0001
α-r	29.9925521	1.3551781	51214.00	25.6598780	34.32522627	22.132	<.0001
α-ø	25.5697845	1.3595479	51284.74	21.2231394	29.91642955	18.808	<.0001
α-ο	13.6027062	1.0353117	51209.58	10.2926850	16.91272736	13.139	<.0001
α-u	15.1887102	1.3649611	51223.99	10.8247584	19.55266203	11.128	<.0001
ε-ε	-3.4733930	1.3639889	51203.42	-7.8342366	.88745053	-2.546	.3918
ε-i	-1.8890978	1.1104876	51241.77	-5.4394657	1.66127011	-1.701	1.0000
ε-r	-0.3733267	1.3581588	51205.30	-4.7155307	3.96887721	-0.275	1.0000
ε-ø	-4.7960944	1.3631160	51304.42	-9.1541471	-0.43804158	-3.518	.0156
ε-ο	-16.7631727	1.0403331	51261.67	-20.0892479	-13.43709740	-16.113	<.0001
ε-u	-15.1771686	1.3687209	51251.20	-19.5531410	-10.80119624	-11.089	<.0001
ε-i	1.5842952	1.1145831	51246.41	-1.9791666	5.14775700	1.421	1.0000
ε-r	3.1000663	1.3613340	51196.42	-1.2522892	7.45242180	2.277	.8200
ε-ø	-1.3227013	1.3663048	51296.25	-5.6909491	3.04554642	-0.968	1.0000
ε-ο	-13.2897797	1.0443956	51238.40	-16.6288432	-9.95071608	-12.725	<.0001
ε-u	-11.7037756	1.3717577	51237.12	-16.0894570	-7.31809420	-8.532	<.0001
i-r	1.5157711	1.1071361	51224.59	-2.0238817	5.05542383	1.369	1.0000
i-ø	-2.9069965	1.1120477	51289.91	-6.4623525	.64835941	-2.614	.3222
i-ο	-14.8740748	.6801159	51336.63	-17.0484906	-12.69965913	-21.870	<.0001
i-u	-13.2880708	1.1196934	51274.87	-16.8678710	-9.70827063	-11.868	<.0001
r-ø	-4.4227676	1.3604741	51293.61	-8.7723738	-0.07316145	-3.251	.0414
r-ο	-16.3898459	1.0368359	51245.09	-19.7047403	-13.07495155	-15.808	<.0001
r-u	-14.8038419	1.3661603	51247.66	-19.1716276	-10.43605618	-10.836	<.0001
ø-ο	-11.9670783	1.0427712	51378.29	-15.3009486	-8.63320802	-11.476	<.0001
ø-u	-10.3810743	1.3700462	51304.73	-14.7612837	-6.00086480	-7.577	<.0001
ο-u	1.5860040	1.0491445	51233.86	-1.7682422	4.94025034	1.512	1.0000

Blue-Yellow (v)

Contrast	Estimate	SE	df	Lower CL	Upper CL	t ratio	p value
a-a	6.92191901	1.5515752	51214.23	1.96133868	11.88249934	4.461	.0003
a-e	-10.69387791	1.5555506	51225.20	-15.66716801	-5.72058780	-6.875	<.0001
a-ε	-5.78344466	1.5588183	51206.90	-10.76718205	-0.79970728	-3.710	.0075
a-i	-19.61860180	1.2714684	51266.29	-23.68364562	-15.55355797	-15.430	<.0001
a-r	-10.66658308	1.5525998	51219.10	-15.63043921	-5.70272695	-6.870	<.0001
a-ø	-12.54198175	1.5578594	51277.08	-17.52265360	-7.56130991	-8.051	<.0001
a-ο	27.98820877	1.1913479	51255.52	24.17932022	31.79709732	23.493	<.0001
a-u	27.18913797	1.5641450	51243.64	22.18837030	32.18990563	17.383	<.0001
α-e	-17.61579692	1.5472487	51213.16	-22.56254501	-12.66904882	-11.385	<.0001
α-ε	-12.70536367	1.5506299	51195.67	-17.66292167	-7.74780567	-8.194	<.0001
α-i	-26.54052081	1.2608780	51230.40	-30.57170584	-22.50933577	-21.049	<.0001
α-r	-17.58850208	1.5443111	51207.59	-22.52585833	-12.65114584	-11.389	<.0001
α-ø	-19.46390076	1.5494362	51277.62	-24.41764263	-14.51015889	-12.562	<.0001
α-ο	21.06628976	1.1797971	51204.16	17.29433063	24.83824889	17.856	<.0001
α-u	20.26721896	1.5554795	51217.19	15.29415590	25.24028202	13.030	<.0001
e-ε	4.91043325	1.5543335	51200.21	-0.05896569	9.87983218	3.159	.0570
e-i	-8.92472389	1.2655154	51233.26	-12.97073540	-4.87871238	-7.052	<.0001
e-r	.02729483	1.5476939	51202.35	-4.92087651	4.97546617	.018	1.0000
e-ø	-1.84810384	1.5535431	51296.11	-6.81497573	3.11876804	-1.190	1.0000
e-ο	38.68208668	1.1855966	51250.75	34.89158562	42.47258773	32.627	<.0001
e-u	37.88301588	1.5598175	51241.67	32.89608377	42.86994798	24.287	<.0001
ε-i	-13.83515714	1.2701886	51235.72	-17.89610931	-9.77420497	-10.892	<.0001
ε-r	-4.88313842	1.5512934	51192.73	-9.84281787	.07654104	-3.148	.0593
ε-ø	-6.75853709	1.5571590	51287.07	-11.73696953	-1.78010465	-4.340	.0005
ε-ο	33.77165343	1.1901893	51227.32	29.96646914	37.57683772	28.375	<.0001
ε-u	32.97258263	1.5632491	51227.72	27.97467920	37.97048606	21.092	<.0001
i-r	8.95201872	1.2616687	51217.62	4.91830583	12.98573161	7.095	<.0001
i-ø	7.07662005	1.2673776	51284.09	3.02465486	11.12858523	5.584	<.0001
i-ο	47.60681057	.7751548	51314.73	45.12854359	50.08507754	61.416	<.0001
i-u	46.80773977	1.2760588	51261.69	42.72801993	50.88745961	36.681	<.0001
r-ø	-1.87539868	1.5505089	51285.04	-6.83256986	3.08177251	-1.210	1.0000
r-ο	38.65479185	1.1815846	51233.76	34.87711762	42.43246607	32.714	<.0001
r-u	37.85572104	1.5568908	51236.99	32.87814607	42.83329602	24.315	<.0001
ø-ο	40.53019052	1.1885629	51363.82	36.73020581	44.33017523	34.100	<.0001
ø-u	39.73111972	1.5614419	51296.31	34.73899413	44.72324531	25.445	<.0001
ο-u	-0.79907080	1.1955964	51225.82	-4.62154231	3.02340071	-0.668	1.0000

Synesthete–Nonsynesthete contrasts

Synesthete–Nonsynesthete contrasts (*Estimate* indicates difference with Synesthetes as the intercept). Degrees-of-freedom method: Satterthwaite; Confidence level used: .95; Conf-level adjustment: Bonferroni method for two estimates; *p* value adjustment: Bonferroni method for two tests.

Lightness (L)

Phoneme	Estimate	SE	df	Lower CL	Upper CL	<i>t</i> Ratio	<i>p</i> Value
a	− 1.3425405	0.8207041	4890.22	− 2.9514891	0.2664081	− 1.636	.1019
ɑ	1.4107140	0.8158002	4778.76	− 0.1886209	3.0100488	1.729	.0838
e	− 2.5960336	0.8180872	4831.38	− 4.1998520	− 0.9922152	− 3.173	.0015
ε	3.4268784	0.8204578	4881.72	− 5.0353442	− 1.8184126	− 4.177	< .0001
i	− 4.3373060	0.6468907	1937.04	− 5.6055023	− 3.0691097	− 6.705	< .0001
I	− 2.7163752	0.8162240	4789.28	− 4.3165410	− 1.1162095	− 3.328	.0009
ø	− 5.9132909	0.8182063	4845.37	− 7.5173428	− 4.3092389	− 7.227	< .0001
ɔ	1.6846702	0.5983915	1419.51	0.5115540	2.8577863	2.815	.0049
u	0.3076004	0.8230901	4942.52	− 1.3060260	1.9212268	0.374	.7086

Green-Red (u)

There was no significant main effect of synesthetic status in the *u* dimension, so comparisons between synesthetes and nonsynesthetes were not calculated.

Blue-Yellow (v)

Phoneme	Estimate	SE	df	Lower CL	Upper CL	<i>t</i> Ratio	<i>p</i> Value
a	− 6.922706	2.709932	7418.02	− 12.2349421	− 1.61047039	− 2.555	.0107
ɑ	− 5.299904	2.690772	7226.73	− 10.5745819	− 0.02522639	− 1.970	.0489
e	− 7.304734	2.699739	7317.78	− 12.5969884	− 2.01247963	− 2.706	.0068
ε	− 11.170913	2.708874	7400.58	− 16.4810748	− 5.86075175	− 4.124	< .0001
i	− 9.354233	2.015447	2398.50	− 13.3050817	− 5.40338345	− 4.641	< .0001
I	− 6.094542	2.692466	7246.25	− 11.3725388	− 0.81654528	− 2.264	.0236
ø	− 13.694378	2.700581	7351.95	− 18.9882834	− 8.40047350	− 5.071	< .0001
ɔ	2.790157	1.812159	1571.02	− 0.7621881	6.34250210	1.540	.1238
u	2.302293	2.719161	7504.97	− 3.0280339	7.63261997	0.847	.3972

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America*, *116*, 1729–1738. <https://doi.org/10.1121/1.1779271>
- Asano, M., & Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes. *Consciousness and Cognition*, *20*, 1816–1823. <https://doi.org/10.1016/j.concog.2011.05.012>
- Asano, M., & Yokosawa, K. (2012). Synesthetic colors for Japanese late acquired graphemes. *Consciousness and Cognition*, *21*, 983–993. <https://doi.org/10.1016/j.concog.2012.02.005>
- Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: Prevalence and familiarity. *Perception*, *25*, 1073–1079.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and s4. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.3.21) [Computer program]. Retrieved from www.praat.org
- Brang, D., Rouw, R., Ramachandran, V., & Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme–color synesthesia. *Neuropsychologia*, *49*, 1355–1358. <https://doi.org/10.1016/j.neuropsychologia.2011.01.002>
- Chan, X. B. V., Goh, S. M. S., & Tan, N. C. (2014). Subjects with colour vision deficiency in the community: What do primary care physicians need to know? *Asia Pacific Family Medicine*, *13*, 10. <https://doi.org/10.1186/s12930-014-0010-3>
- Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*, e57410. <https://doi.org/10.1371/journal.pone.0057410>
- Cuskley, C., & Kirby, S. (2013). Synesthesia, cross-modality, and language evolution. In J. Simner & E. M. Hubbard (Eds.), *Oxford handbook of synesthesia* (pp. 869–900). Oxford, UK: Oxford University Press.
- Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, *81*, 119–130. <https://doi.org/10.1007/s00426-015-0709-2>
- Dingemans, M., & van Leeuwen, T. M. (2015). Taal als samenspel van de zintuigen. *Onze Taal*, *12*, 344–345.
- Dingemans, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*, 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*, 613–621. <https://doi.org/10.1177/0956797612457374>
- Eagleman, D. M. (2010, March). *What has large-scale analysis taught us*. Paper presented at the 2010 Meeting of the UK Synaesthesia Association, Brighton, UK.
- Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods*, *159*, 139–145. <https://doi.org/10.1016/j.jneumeth.2006.07.012>
- Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 69–78. <https://doi.org/10.1002/wcs.26>
- Guillamon, P. M. (2014). Vowel–colour symbolism in English and Arabic: A comparative study. *Miscelánea*, *47*, 31–52.
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language. *Philosophical Transactions of the Royal Society B*, *369*, 20130298. <https://doi.org/10.1098/rstb.2013.0298>
- Johnson, K. (2011). *Acoustic and auditory phonetics*. Chichester, UK: Wiley-Blackwell.
- Jürgens, U., Mausfeld, R., & Nikolic, D. (2010, March). *Grapheme–color synaesthesia: Similar shapes take similar colors*. Paper presented at the 2010 Meeting of the UK Synaesthesia Association, Brighton, UK.
- Kim, H.-W., Nam, H., & Kim, C.-Y. (2018). [i] is lighter and more greenish than [o]: Intrinsic association between vowel sounds and colors. *Multisensory Research*, *31*, 419–437. <https://doi.org/10.1163/22134808-00002581>
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.
- Kuznetsova, A., Brockho, P. B., & Christensen, R. H. B. (2018). lmerTest: Tests in linear mixed effects models. Retrieved from <https://CRAN.R-project.org/package=lmerTest>
- Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*, 573–579. <https://doi.org/10.1111/2041-210X.12512>
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*, 1–33.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.
- Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin*, *82*, 303–331. <https://doi.org/10.1037/0033-2909.82.3.303>
- Marks, L. E. (1978). *The unity of the senses: Interrelations among the modalities*. New York, NY: Academic Press.
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*, 61–65. <https://doi.org/10.1111/1467-8721.00116>
- Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *i-Perception*, *5*, 132–142.
- Mroczko-Wasowicz, A., & Nikolić, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Frontiers in Human Neuroscience*, *8*, 509. <https://doi.org/10.3389/fnhum.2014.00509>
- Nabokov, V. (1989). *Speak, memory: An autobiography*. New York, NY: Vintage.
- Nunn, A. (2006). *Dutch orthography: A systematic investigation of the spelling of Dutch words*. The Hague, The Netherlands: Holland Academic Graphics.
- Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (p. 325–346). Cambridge, UK: Cambridge University Press.
- Reichard, G. A., Jakobson, R., & Werth, E. (1949). Language and synesthesia. *Word*, *5*, 224–233.
- Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated:

- Evidence from Korean. *Cognition*, 107, 752–762. <https://doi.org/10.1016/j.cognition.2007.09.001>
- Root, N. B., Rouw, R., Asano, M., Kim, C.-Y., Melero, H., Yokosawa, K., & Ramachandran, V. S. (2018). Why is the synesthete's "a" red? Using a five-language dataset to disentangle the effects of shape, sound, semantics, and ordinality on inducer–concurrent relationships in grapheme–color synesthesia. *Cortex*, 99, 375–389. <https://doi.org/10.1016/j.cortex.2017.12.003>
- Rosner, B., & Pickering, J. (1994). *Vowel perception and production*. Oxford, UK: Oxford University Press.
- Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013). Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity. *Journal of Neuroscience Methods*, 215, 156–160. <https://doi.org/10.1016/j.jneumeth.2013.02.009>
- Rouw, R., Case, L., Gosavi, R., & Ramachandran, V. (2014). Color associations for days and letters across different languages. *Frontiers in Psychology*, 5, 369. <https://doi.org/10.3389/fpsyg.2014.00369>
- Sharma, G., Wencheng, W., & Dalal, E. (2005). The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, 30, 21–30.
- Simner, J. (2007). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Sciences*, 11, 23–29. <https://doi.org/10.1016/j.tics.2006.10.010>
- Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, 103, 1–15. <https://doi.org/10.1348/000712610X528305>
- Simner, J., & Carmichael, D. A. (2015). Is synaesthesia a dominantly female trait? *Cognitive Neuroscience*, 6, 68–76. <https://doi.org/10.1080/17588928.2015.1019441>
- Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, 39, 553–569.
- Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., ... Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, 35, 1024–1033. <https://doi.org/10.1068/p5469>
- Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. (2005). Non-random associations of graphemes to colours in synaesthetic and normal populations. *Cognitive Neuropsychology*, 22, 1069–1085.
- Smilek, D., Moffatt, B. A., Pasternak, J., White, B., Dixon, M. D., & Merikle, P. M. (2002). Synaesthesia: A case study of discordant monozygotic twins. *Neurocase*, 8, 338–342.
- Styles, S., & Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception*, 8, 2041669517724807. <https://doi.org/10.1177/2041669517724807>
- Taylor, G. (2017). Python colormath. Retrieved 13 March, 2017, from <http://python-colormath.readthedocs.io/en/latest/>
- van Leeuwen, T. M., & Dingemanse, M. (2016). Colour associations in synaesthetes and nonsynaesthetes: A large-scale study in Dutch. *Perception*, 45, 333–334.
- van Leeuwen, T. M., Dingemanse, M., Todil, B., Agameya, A., & Majid, A. (2016). Non-random associations of graphemes with colors in Arabic. *Multisensory Research*, 29, 223–252. <https://doi.org/10.1163/22134808-00002511>
- 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.
- Watson, M. R., Akins, K. A., & Enns, J. T. (2012). Second-order mappings in grapheme–color synesthesia. *Psychonomic Bulletin & Review*, 19, 211–217. <https://doi.org/10.3758/s13423-011-0208-4>
- Witthoft, N., & Winawer, J. (2013). Learning, memory, and synesthesia. *Psychological Science*, 24, 258–265. <https://doi.org/10.1177/0956797612452573>
- Witthoft, N., Winawer, J., & Eagleman, D. M. (2015). Prevalence of learned grapheme–color pairings in a large online sample of synesthetes. *PLoS ONE*, 10, e0118996. <https://doi.org/10.1371/journal.pone.0118996>
- Wrembel, M. (2009). On hearing colours: Cross-modal associations in vowel perception in a non-synaesthetic population. *Poznan Studies in Contemporary Linguistics*, 45, 581–598. <https://doi.org/10.2478/v10010-009-0028-0>
- Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, 5, 683–689. <https://doi.org/10.3758/BF03208845>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.