BRILL

Multisensory Research 31 (2018) 419–437

brill.com/msr

[i] is Lighter and More Greenish Than [o]: Intrinsic Association Between Vowel Sounds and Colors

Hyun-Woong Kim[1], **Hosung Nam**[2,3,*] **and Chai-Youn Kim**[1,*]

[1] Department of Psychology, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, South Korea

[2] Department of English Language and Literature, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, South Korea

[3] Haskins Laboratories, 300 George St., New Haven, CT 06511, USA

Received 8 October 2016; accepted 11 May 2017

Abstract

It has recently been reported in the synesthesia literature that graphemes sharing the same phonetic feature tend to induce similar synesthetic colors. In the present study, we investigated whether phonetic properties are associated with colors in a specific manner among the general population, even when other visual and linguistic features of graphemes are removed. To test this hypothesis, we presented vowel sounds synthesized by systematically manipulating the position of the tongue body's center. Participants were asked to choose a color after hearing each sound. Results from the main experiment showed that lightness and chromaticity of matched colors exhibited systematic variations along the two axes of the position of the tongue body's center. Some non-random associations between vowel sounds and colors remained effective with pitch and intensity of the sounds equalized in the control experiment, which suggests that other acoustic factors such as inherent pitch of vowels cannot solely account for the current results. Taken together, these results imply that the association between phonetic features and colors is not random, and this synesthesia-like association is shared by people in the general population.

Keywords

Vowel, color, cross-modal correspondence, synesthesia, articulatory synthesizer, CASY

1. Introduction

Synesthesia is a condition under which stimulation in one sensory or cognitive pathway evokes additional perceptual or cognitive experience in another,

---

[*] To whom correspondence should be addressed. E-mail: chaikim@korea.ac.kr; hnam@korea.ac.kr

DOI:10.1163/22134808-00002581

irrelevant pathway. One of the most common types of synesthesia includes color experiences induced by individual letters or digits, known as grapheme–color synesthesia (Rich *et al.*, 2005; Simner *et al.*, 2006). Though individual grapheme–color associations are idiosyncratic, it has been suggested that synesthetic associations are not just random nor simple. For example, several studies emphasized the importance of visual features by showing that graphemes of similar shape tend to induce similar synesthetic colors (Brang *et al.*, 2011; Watson *et al.*, 2012). It is noteworthy that when confronted with an ambiguously shaped character (e.g., a shape interpreted either as ‘5’ or as ‘S’), the induced synesthetic color might be different depending on its semantic context (e.g., in an array of digits or letters, Dixon *et al.*, 2006) or how the character is recognized (Kim *et al.*, 2013). Another study also suggested the importance of meaning by showing the tendency for the initial letter of a color word to be associated with the color represented by the word. For instance, ‘y’ is often associated with yellow (Rich *et al.*, 2005). It has also been suggested that more frequently used graphemes in written and spoken language tend to trigger lighter and more saturated synesthetic colors (Beeli *et al.*, 2007; Kim and Kim, 2014; Smilek *et al.*, 2007). Together, these previous studies imply that linguistic properties inherent in graphemes influence the synesthetic association between graphemes and colors (Simner, 2007).

Sound is another candidate feature as a determining factor of synesthetic color. It has been reported that graphemes associated with similar sounds tend to induce similar synesthetic colors (Asano and Yokosawa, 2011, 2012; Shin and Kim, 2014). A recent study built on those previous works and found the influence of acoustic-phonetic features on grapheme–color associations (Kang *et al.*, 2017). In the study, multilingual Korean grapheme–color synesthetes, who reported not to have any explicit, phonetic knowledge of English and Korean, participated in a color-matching procedure. Results showed that graphemes sharing a phonetic feature tended to induce similar synesthetic colors across languages. For one synesthete who participated in the study, the *Hangul* (the Korean alphabet) consonant ‘ㄴ’ (pronounced as [n]) and Latin-alphabet ‘D’ (pronounced as [d]), which are both alveolar in place of articulation, were associated with yellowish colors. It should be noted that the common phonetic feature (i.e., place of articulation) of the two graphemes was not readily noticeable.

The influence of phonetic properties of graphemes on synesthetic colors in grapheme–color synesthesia may bear implications on cross-modal associations between acoustic and visual features shared by individuals in the general population. Indeed, it has been well established that people share implicit cross-modal mappings between certain visual and auditory features. For example, sounds of higher pitch tend to pair with lighter colors (Marks, 1974) and more angular (Karwoski and Odbert, 1938; Parise and Spence, 2009) and smaller shapes (Bien *et al.*, 2012; Gallace and Spence, 2006; Parise and Spence, 2009) than do sounds of lower pitch. Such cross-modal correspondences seem to exist for audio-visual features of human speech as well. An early finding showed a tendency for pseudo-words with low vowels such as [o] or [a] (e.g., mal) to be associated with large size whereas pseudo-words with high vowels such as [e] or [i] (e.g., mil) are associated with small size (Sapir, 1929). The most well-known example of this line of studies is 'bouba–kiki', the strong tendency where people prefer naming a round shape with 'bouba' while naming an angular shape with 'kiki' (Köhler, 1929; Ramachandran and Hubbard, 2001). These non-arbitrary linkages between acoustic features of speech sound and visual shapes have been found among pre-literate children (Maurer *et al.*, 2006, for 'bouba–kiki' effect) and infants (Peña *et al.*, 2011, for vowel-size correspondence), indicating the pre-linguistic nature of the cross-modality of speech (also see Cuskley *et al.*, 2017).

The current study intends to explore cross-modal associations between phonetic and visual features as an extension of our previous studies on grapheme–color synesthesia (Kang *et al.*, 2017; Shin and Kim, 2014). Based on the previous findings of audio-visual, cross-modal correspondences of human speech among the general population, we investigated the implicit association between auditory and visual features shared by people without synesthesia. We excluded the intervention of visual and linguistic features from phonetic ones since our previous study suggesting the role of phonetic features of graphemes on determining synesthetic colors was limited by visual and linguistic factors intermingling with phonetic factors (Kang *et al.*, 2017). For example, an individual grapheme may or may not correspond to a single phoneme based on the orthographic depth of a given language.

To carry out the study, we presented a series of auditory stimuli to participants without synesthesia and asked them to match the sounds with colors on a color palette that immediately came to mind. For auditory stimuli, vowel-like speech sounds were synthetically generated using an articulatory synthesizer (online Supplementary Movie1). The 'vowel' aspect of our stimuli links the current work to our previous studies on grapheme–color synesthesia. More importantly, the 'synthetic' aspect of our stimuli makes them less susceptible to the influence of linguistic factors. The vowel acoustics of our stimuli were produced by manipulating only the position of the tongue body, which allowed our stimuli to be dependent only on a speech articulator. Position of the tongue body was defined along two-dimensional axes — i.e., 'height' and 'frontness' — constrained in a physiologically plausible space, which modulated vowel quality (Fig. 1). The resulting 42 stimuli are far more than the number of vowels that exist in a particular language and therefore enabled us to prevent our results from being language-specific.

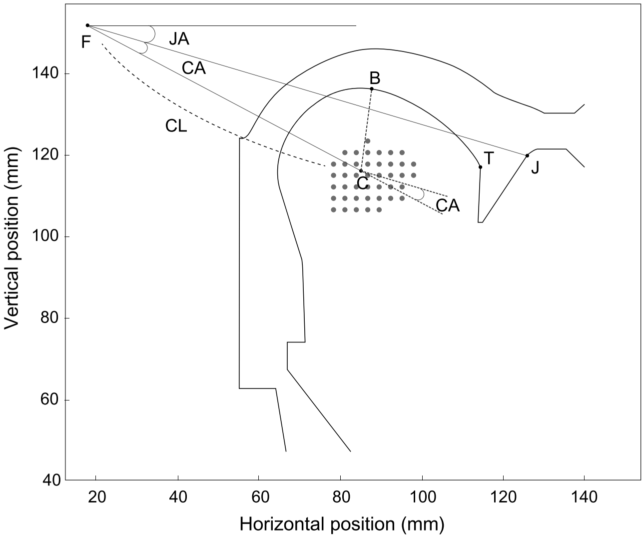

Figure 1. Vocal tract representation of CASY with its articulator variables for generating vowel sounds. F: mandibular condyle, C: tongue body center, B: tongue blade, T: tongue tip, J: jaw. The jaw position is given by the angle (JA) from a horizontal line at the joint with the constant distance from F. The tongue body is represented as an imaginary circle with a fixed radius. The position of the tongue body is given by the position of the circle's center, which is determined by the angle (CA) from the line F–J and the length (CL) of the line F–C. The tongue blade and tip are attached to the tongue body's circle. The tongue body's articulator variables (CA and CL) were modulated parametrically to manipulate vowel sounds. All the other articulatory variables were fixed, including the jaw position that affects mouth openness. The positions of the tongue body's center used for generating the 42 auditory stimuli are superimposed on the outline of the vocal tract.

We are not the first to investigate intrinsic cross-modal associations between vowels and colors. A couple of early studies on synesthesia of the phoneme–color type reported the non-arbitrary relationship between vowels and colors (Jacobson, 1962; Marks, 1975). However, the results were mainly based on case reports lacking quantitative measures and statistical tests. More recently, the influence of phonetic properties on vowel–color mappings has been examined in the general population through statistical analyses introducing phonetic terms such as vowel categories (Wrembel, 2009) or vowel formants (Moos *et al.*, 2014). However, the vowel stimuli used in these studies were limited within the range of ordinary languages, leaving phonetic and other linguistic factors intermingled. For example, Moos *et al.* (2014) employed eight primary cardinal vowels and several of their variants. Cardinal vowels are reference vowels used for describing sounds that exist in human languages and the sounds produced when the tongue is in the extreme corner of the vowel (or tongue body) space. In contrast, our stimuli, generated by manipulating the position of the tongue body equi-distantly, could cover the whole vowel space. Therefore, we expected to analyze vowel–color mappings more systematically and exhaustively, and thus gain a better understanding of the intrinsic association between purely auditory features and colors.

We conducted two experiments. In the main experiment, the stimuli were synthesized based on Articulatory Phonology (Browman and Goldstein, 1986). In those synthesized vowels, acoustic properties such as pitch and intensity also vary with the tongue body's position in a systematic manner (Fig. 2). The variation in pitch and intensity is not an artificial by-product of synthesis, but a result reflecting the actual articulatory process (Fairbanks *et al.*, 1950; Whalen and Levitt, 1995). Nonetheless, we also conducted a control experiment where pitch and intensity of vowel stimuli were equal. By doing so, we attempted to touch upon novel, intrinsic audio-visual correspondences going beyond the previously shown pitch–color associations (e.g., Ward *et al.*,

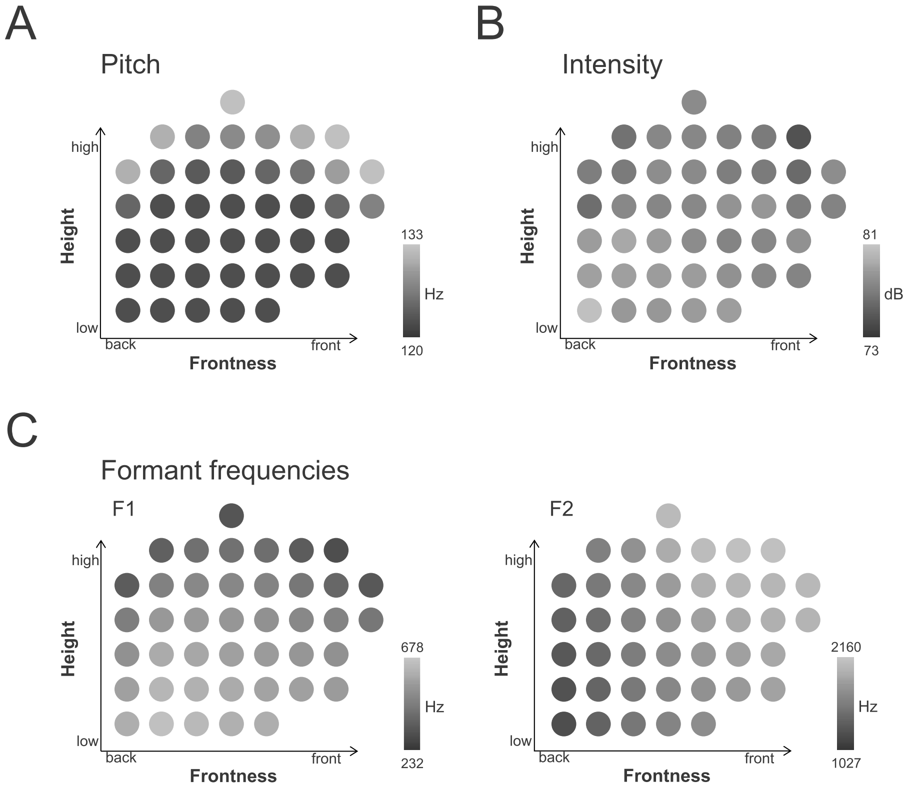

Figure 2. Acoustic variations of the auditory stimuli in the main experiment generated by articulatory synthesis. Acoustic properties of the vowel stimuli were displayed as brightness of each circle. (A) pitch, (B) intensity, (C) the first (left) and the second (right) formants. Stimuli with higher intensities and frequencies are represented with lighter shades.

2006). We analyzed the chromaticity and luminance of the colors matched from the two experiments and examined whether the color association is systemically modulated by the tongue body position factor. The results from the two experiments are reported below.

2. Material and Methods

2.1. Participants

Twenty-four individuals (7 males, 19–29 years of age) participated in the main experiment, and an additional 24 individuals (8 males, 19–26 years of age) participated in the control experiment. All participants reported no form of synesthesia, which was reconfirmed after the experiments when they reported that they did not experience synesthetic colors when hearing the auditory stimuli during the experiments. They had normal or corrected-to-normal visual acuity and normal color vision. All of them used Korean as their native language. They consented to participating in the study, which was approved by the Korea University Institutional Review Board (1040548-KU-IRB-15-67-A-2).

2.2. Stimuli

For auditory stimuli, synthetic vowel-like sounds were used in the main and the control experiments. We employed the Haskins Laboratories Configurable Articulatory SYnthesis model (CASY; Rubin *et al.*, 1996) to parametrically manipulate vowel sounds. In CASY, speech synthesis is implemented in articulatory terms, based on Mermelstein's articulatory model (Mermelstein, 1973). In this model, speech events are represented as temporal variations of the vocal tract shape by the model articulator variables, which are either joints or segments. These variables specify the positions of articulators (e.g., lips, jaw, tongue body, or tongue tip) and thus determine the geometric representation of the vocal tract in the mid-sagittal plane (Fig. 1).

In this study, we generated vowel stimuli by manipulating the tongue body articulator variables (CL, CA) with all the other articulator variables fixed (see Fig. 1; Nam *et al.*, 2013). A set of 42 points for the tongue body's center positions was created at physiologically plausible spaces for vowels. All points were at an equal distance (2.8 mm) from one another. The spatial property of the tongue body in the vocal tract was defined as height and frontness, which are equivalent to the vertical and horizontal space of the tongue body's center position, respectively (Fig. 1). The area functions computed from each vowel's vocal tract shape were used to calculate formant frequencies (F1, F2) (see Fig. 2C). Formant values were then used as inputs to HLsyn™ (Sensimetrics Inc., Malden, MA), a high-level, quasi-articulatory speech synthesizer (Hanson and Stevens, 2002), to generate the acoustic output of the vowels. The input fundamental frequency (F0) and duration of vowels were set to 120 Hz and 500 ms, respectively. Vowel sounds generated by HLsyn™ were used as auditory stimuli in the main experiment.

In articulatory synthesis, not only is the input parameter F0 adjusted to account for changes in the intrinsic pitch of vowels, but amplitude is also affected by variation in oral pressure (Hanson and Stevens, 2002), resulting in systematic differences in pitch and intensity of acoustic outputs. As shown in Fig. 2, output pitch (120–133 Hz) and intensity (73–81 dB) of the vowel stimuli generated by HLSyn™ were modulated by height and frontness. To control for such systematic variations in pitch and intensity, another set of vowel sounds was generated from the same formant values calculated by CASY using Praat software (Boersma and Weenink, 2013). Pitch (110 Hz) and intensity (77 dB) of acoustic waveforms generated by Praat were set identically for all vowel stimuli. The stimulus set generated by Praat was used in the control experiment.

2.3. Apparatus

All stimuli were auditorily presented through SRH440 headphones. The visual display for the color-matching procedure was presented on a 19-inch, color-calibrated CRT monitor (1024 × 768 resolution, 60 Hz frame rate). Experiments were conducted using Matlab (version 8.3, Mathworks, MA) in a quiet, dark room.

2.4. Procedures

Participants engaged in a modified Matlab version of the standardized synesthesia battery (Eagleman *et al.*, 2007) for the color-matching test both in the main and the control experiments. For each trial, participants were instructed to listen to an auditory stimulus and to select a color best matched to it with no time constraint. The character of the stimuli (i.e., vowel quality) was not indicated to prevent potential influence of the awareness of linguistic properties. After hearing the stimulus, participants selected a color by clicking on the color palette on the monitor screen using a mouse. The color palette displayed a continuous scale for hue and saturation of colors. The brightness of color was adjusted using two keyboard buttons to increase (‘→’) or to decrease (‘←’) the brightness. Each of the 42 auditory stimuli was repeated three times, resulting in a total of 126 trials. The trials were presented in a randomized order.

2.5. Data Analysis

The RGB values of matched colors for each auditory stimulus were converted to CIE Lab color coordinates. The data from the three trials for each auditory stimulus were averaged into a mean value. Luminance of the matched colors was analyzed using L^* values. The a^* (green–red axis) and b^* (blue–yellow axis) values were used for chromaticity analysis. These three dependent variables were analyzed separately. The effects of the tongue body's position in association with vowel sounds and colors were statistically examined through a repeated measures ANOVA with the two within-subject factors — i.e., height and frontness. Since the two factors were not fully crossed due to the physiological constraint in tongue body positions (Fig. 1), the most back/front (78.4/98.0 mm) and low/high (106.4/123.2 mm) vowel stimuli were trimmed off. Accordingly, data for a total of 30 positions (5 levels for height and 6 levels for frontness) were entered into the statistical analyses, which was still a far greater number of vowels than cardinal vowels. Greenhouse–Geisser correction was used to adjust degrees of freedom when the sphericity assumption was violated.

3. Results

3.1. Luminance (L^*)

Mean L^* values for each level of height and frontness of the tongue body's positions are displayed in Fig. 3.

For the data from the main experiment, a two-way repeated measures ANOVA revealed a statistically significant main effect of height on L^* [$F(1.84, 42.33) = 24.90$, $p < 0.001$: the green curve in the top panel of Fig. 3A]. Luminance of the matched colors was higher for high vowels (e.g., stimuli that sound close to [u] or [i]) than for low vowels (e.g., stimuli that sound close to [o] or [e]). The main effect of frontness was also significant [$F(2.83, 65.15) = 6.49$, $p < 0.001$: the green curve on the bottom panel of Fig. 3A], with colors of greater luminance matched to front vowels (e.g., stimuli that sound close to [e] or [i]) rather than back vowels (e.g., stimuli that sound close to [u] or [o]). We also found a statistically significant two-way interaction between height and frontness [$F(8.38, 155.93) = 4.01$, $p < 0.001$]; the frontness' effect on L^* was more markedly pronounced with high vowels than with low vowels (see the top panel of Fig. 3B). The results showed that luminance of the matched colors is systematically modulated by articulatory gestures of the tongue (i.e., tongue body position) from which vowels are produced. In other words, both height and frontness of the tongue implicated in vowel acoustics affected luminance of the associated colors.

In the control experiment, where pitch and intensity of vowel sounds were equalized, the luminance effect was also significant for frontness [$F(2.44, 56.08) = 4.18$, $p < 0.05$: the blue curve in the bottom panel of Fig. 3A], but not for height nor interaction between them (both $p > 0.372$: the blue curve in the top panel of Fig. 3A and also see the bottom panel of Fig. 3B). Given that there is a high correlation between tongue body's height

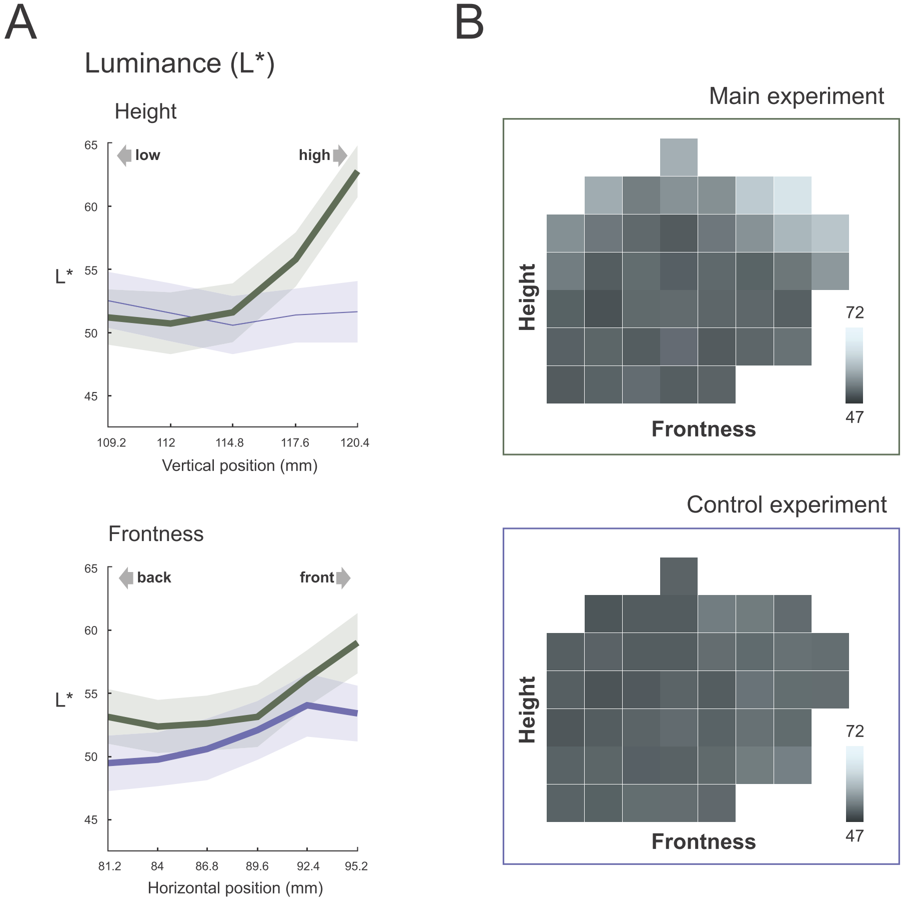

Figure 3. Luminance (L^*) results. (A) The relationship between the tongue body's position (height and frontness) and L^* values of the matched colors. The dark green and blue lines indicate results from the main and the control experiment, respectively. The bold lines indicate statistical significance ($p < 0.05$, F-test). The shades indicate ±1 standard error of the mean (SEM). (B) The group mean L^* values for each of the 42 auditory stimuli based on the tongue body's position are shown as the level of lightness. Stimuli matched with lighter colors (larger L^*) were shown in lighter shades.

and pitch, conflicting results between the two experiments appear to reflect the influence of pitch on luminance of the matched colors. In contrast, the colors associated with front vowels were lighter than with back vowels regardless of the intrinsic acoustic variation of vowels in pitch and intensity.

3.2. Chromaticity (a^* and b^*)

Mean a^* and b^* values for each level of height and frontness of the tongue body's position are shown in Figs 4 and 5, respectively.

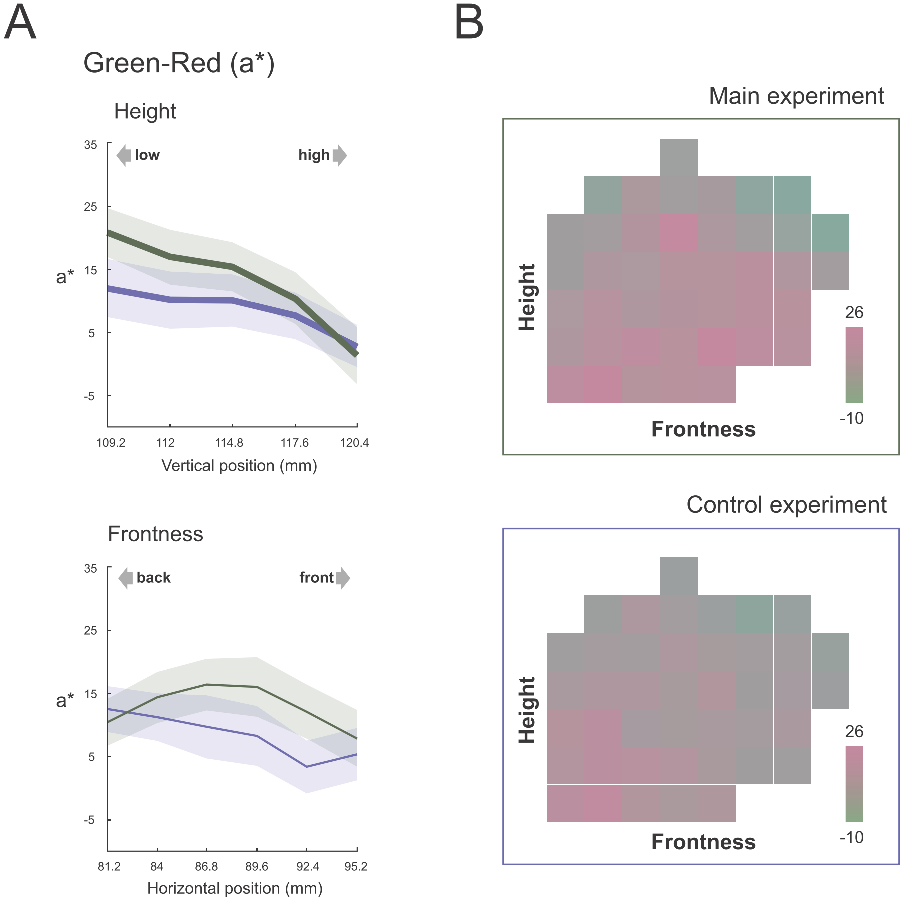

Figure 4. Chromaticity results along the green–red continuum (a^*). (A) The relationship between the tongue body's position (height and frontness) and a^* values of the matched colors. Dark green and blue lines indicate results from the main and the control experiment, respectively. The bold lines indicate statistical significance ($p < 0.05$, F-test). The shades denote ±1 SEM. (B) The group mean a^* values for each of the 42 auditory stimuli based on the tongue body's position are represented as colors with reference to the green–red color axis. More reddish colors denote larger a^* values whereas more greenish colors denote smaller a^* values.

Concerning a^*, the chromaticity index on the red–green axis, a two-way repeated measures ANOVA of the data from the main experiment revealed a statistically significant main effect of height [$F(2.54, 58.32) = 9.99$, $p < 0.001$: the green curve in the top panel of Fig. 4A]. More reddish colors on the green–red color axis were associated with lower vowels. In the control experiment, where pitch and intensity of vowel sounds were equalized, the chromaticity effect indexed by a^* was also significant for height in the same direction [$F(2.06, 47.38) = 3.37$, $p < 0.05$: the blue curve in the top panel of Fig. 4A].

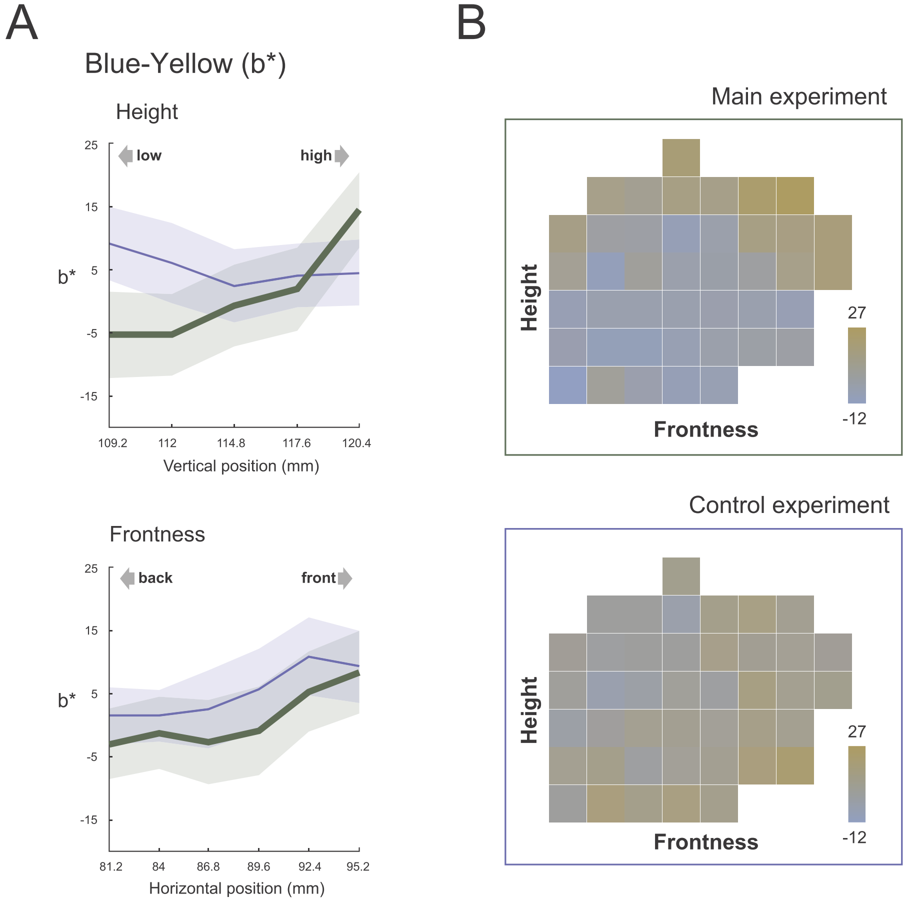

Figure 5. Chromaticity results along the blue–yellow continuum (b^*). (A) The relationship between the tongue body's position (height and frontness) and b^* values of the matched colors. The dark green and blue lines indicate results from the main and the control experiment, respectively. The bold lines indicate statistical significance ($p < 0.05$, F-test). The shades denote ±1 SEM. (B) The group mean b^* values for each of the 42 auditory stimuli based on the tongue body's position are represented as colors with reference to the blue–yellow color axis. More yellowish colors denote larger b^* values whereas more bluish colors denote smaller b^* values.

Taken together, the results indicate the green–red color modulation in height (i.e., reddish colors associated with low vowels) was not solely determined by inherent pitch variations along the vowel height.

Neither the main effect of frontness nor the two-way interaction effect between height and frontness on the green–red color axis was significant both in the main and in the control experiments (all $p > 0.123$: bottom panel of Fig. 4A and also see Fig. 4B).

Turning now to b^*, the chromaticity index on the blue–yellow axis, a two-way repeated measures ANOVA of the data from the main experiment revealed a statistically significant main effect of height [$F(2.30, 52.89) = 7.48$, $p < 0.001$: the green curve on the top panel of Fig. 5A] and frontness [$F(3.57, 82.16) = 3.48$, $p < 0.05$: the green curve in the bottom panel of Fig. 5A]. That is, more yellowish colors were associated with high and front vowels more than with low and back vowels on the blue–yellow color axis. The two-way interaction effect between height and frontness was not statistically significant ($p = 0.613$, see Fig. 5B).

In the control experiment, however, neither the main effect of height nor the main effect of frontness on b^* were significant statistically (all $p > 0.108$: the blue curves in the panels of Fig. 5A). This suggests that the blue–yellow color modulation by the tongue body's position reflects the influence of variations in other acoustic features (presumably pitch), not just vowel quality modulated by the articulatory organ.

4. Discussion

The present study demonstrates non-random associations between vowel sounds and colors in a non-synesthetic population. The associated colors showed systematic modulations in accordance with the spatial mapping — height and frontness — of the tongue body's center position in the vocal tract, from which acoustics of synthetic vowels are generated. Specifically, lighter colors tended to be associated with high and front vowels more than with low and back vowels. In addition, vowel–color association was observed on the two opponent-color dimensions (green–red and blue–yellow axes). Participants tended to select more greenish and yellowish colors when high vowels were presented. They also showed a tendency to associate yellowish colors with front vowels. Even with vowel stimuli of which intrinsic pitch and intensity were controlled, some of the color-matching results showed the same tendency: (1) the more front, the lighter, and (2) the higher, the more greenish. This indicates our results cannot be solely attributed to the acoustic variation in pitch or intensity. Taken together, the results imply cross-modal mechanisms between articulatory features and colors in people without synesthesia.

Our results are in line with the results from previous studies reporting non-arbitrary mappings between vowels and colors. For example, studies on phoneme–color synesthesia showed that front vowels tended to induce lighter colors than back vowels (Jacobson, 1962; Marks, 1975). In another study where only non-synesthetes were tested, participants were more likely to select green and yellow for high and front vowels (e.g., [i]) (Wrembel, 2009). A more recent study examined vowel–color associations in both grapheme–color synesthetes and non-synesthetes and found a tendency to associate low vowels and more reddish colors, as well as front vowels and lighter colors for both groups (Moos *et al.*, 2014).

Although the current study is in line with these previous studies, it is more innovative in several aspects. First, we used a denser set (thus larger number) of auditory stimuli and a color palette in which a color was chosen on a continuous rather than a discrete scale. These methodological considerations were taken to avoid forced or guided categorical judgment of a given auditory stimulus and colors. Second, we systematically employed phonetic principles for both the generation of vowel sounds and analyses of the matched colors. A couple of previous works have tried to explain the vowel–color associations in terms of the first two formant frequencies (F1 and F2), the most important acoustic properties to distinguish vowels (Marks, 1975; Moos *et al.*, 2014). For example, the stimuli in Moos *et al.*'s (2014) study were the cardinal vowels produced by a trained phonetician. Height and frontness in cardinal vowels are highly related to F1 and F2, respectively. However, reference vowels, even when produced by a phonetician, can be highly affected by a speaker's idiosyncrasy such as physiological differences (e.g., size of oral cavity and head, length of vocal tract) and native language. In addition, such human speech cannot be perfectly controlled acoustically or articulatorily. For example, [i] and [u] are both high vowels but no phonetician could produce the vowels at the same height (and thus the same F1). Consequently, any further analysis based on acoustics and articulation might be imperfect and/or indirect. In contrast, our current articulatory approach can minimize the potential problems of the previous studies. Our vowel sounds were synthesized by systematically manipulating articulatory parameters, and we analyzed the associated colors with the same articulatory factors (i.e., height and frontness of the tongue body's position). Since our stimuli are synthetic, they do not include speaker variation and are not influenced by a specific language.

Some of the articulatory effects obtained in the main experiment were no longer significant when acoustic variation of stimuli in pitch and intensity was equalized in the control experiment. It is tempting to claim that these effects are purely from pitch and intensity and should thus be excluded for the spatial effect of articulation. Indeed, intensity of sound is known to affect the brightness of colors (Lewkowicz and Turkewitz, 1980; Marks *et al.*, 1986) though its influence is inconsistent in direction (Marks, 1974) and far less salient (Marks, 1989) than the influence of pitch. More importantly, it has been shown in multiple studies that sounds of higher pitch in general tend to be associated with lighter colors than sounds of lower pitch in both non-synesthetes and synesthetes (Hubbard, 1996; Melara, 1989; Ward *et al.*, 2006). Also, a finding from a classic study buttresses those reports by showing that children tended to pair high-frequency tones with yellowish colors and low-frequency tones with bluish colors (Simpson *et al.*, 1956). The pitch information accompanied by a vowel at a particular tongue body height presumably mediates the subsequent effects of height on lightness and yellowness of the matched colors in the main experiment.

It should be noted, however, that the variation in pitch and intensity is not just an artifact that is independent from vowel quality, but rather a direct acoustic consequence of vowel articulation (Fairbanks *et al.*, 1950; Whalen and Levitt, 1995). In particular, the intrinsic pitch of vowels is prevalently shown across languages so that high vowels have higher pitches than low vowels (Whalen and Levitt, 1995). It has also been reported that articulation (e.g., activity to produce vowels and consonants) and phonation (e.g., activity to control voicing and pitch, etc.) inevitably interact with each other at the muscle level (Honda, 1983). Unlike most acoustic synthesizers, the Haskins articulatory synthesizer is more physiologically plausible, for example, by reflecting the height–pitch coupling. The stimuli in the control experiment might lack such height–pitch coupling and result in making the effect observed in the main one disappear.

In the current study, we found a significant effect of the associated color with articulatory features by testing individuals without synesthesia. This finding bears significance in understanding the implicit association between multisensory information shared by individuals in the general population, and thus extends previous findings in the synesthesia literature. The correspondence between vowel sounds and colors suggests the involvement of multi-modal cortical regions. Candidate regions include the occipito-temporo-parietal junction (OTPJ, Ramachandran and Hubbard, 2001), posterior parietal cortex (Noppeney *et al.*, 2007; Revill *et al.*, 2014), and the posterior portion of the superior temporal sulcus (Beauchamp *et al.*, 2004; Noesselt *et al.*, 2007). It seems less plausible to posit direct connections between uni-modal regions such as the speech-specific auditory region and color-processing visual region as in the case of aberrant connections between inducer- and concurrent-processing regions in synesthetic brains. Some theories of multisensory correspondences imply the brain network subserving emotion as a link between relevant unimodal regions (Palmer *et al.*, 2013).

Previous studies have reported common cross-modal mechanisms across synesthetes and non-synesthetes (e.g., Fernay *et al.*, 2012; Simner and Ludwig, 2012; Simner *et al.*, 2005; Ward *et al.*, 2006), implying that synesthesia might be a stronger or exaggerated manifestation of implicit cross-modal association present in non-synesthetes (Brang *et al.*, 2012; Cohen Kadosh and Henik, 2007; Martino and Marks, 2001; see also Deroy and Spence, 2013). In particular, Moos *et al.* (2014) reported that synesthetes and non-synesthetes showed significant correlations between vowel sounds and colors in the same direction. Importantly, some of these associations were more consistent and stronger for the synesthete group. This suggests that certain aspects of phonetic effects revealed in the present study might be based on a common cross-modal mechanism that underlies synesthesia (Bankieris and Simner, 2015). Comparisons based on synesthesia were not possible in the current work due to the unavailability of synesthetes who experience colors upon hearing specific sounds such as our auditory stimuli. It will be informative to compare synesthetic and non-synesthetic sound–color associations in future studies.

On a final note, our findings have implications for sound symbolism, a non-arbitrary linkage between sound and meaning (Cuskley and Kirby, 2013; Hinton *et al.*, 2006). Vowels have been linked to other visual properties such as shape (Nielsen and Rendall, 2013) and size (Thompson and Estes, 2011). For example, people preferred matching syllables with back vowels to a rounded shape while they tended to match syllables with front vowels to a jagged shape (Nielsen and Rendall, 2013; Spector and Maurer, 2013). In another study, people tended to pair a large visual object with the name ‘wodolo’ whereas they tended to pair a small visual object with the name ‘kitete’ (Thompson and Estes, 2011). A phonetic feature of the vowels as well as that of the consonants in those names indicates the frontness of articulatory gestures; [o] is articulated at the back compared to [i] or [e], which are both articulated at the front. Researchers have also noted the importance of articulatory tongue position producing vowels in vowel–size correspondence (Newman, 1933; Sapir, 1929; see also Ohala, 1994), which is in line with our findings.

Given a sound symbol (e.g. [i]), all the cross-modal associations between the symbol and the associated color (yellow), size (small), pitch (high), and shape (unrounded) are related to its articulation. Producing [i] in articulation involves the small size and a less rounded cross section in the oral cavity, and results in a higher pitch (Newman, 1933; Sapir, 1929). It seems that direct control of articulatory parameters and the articulation-based analysis can provide a unified principle — articulation — to seemingly unrelated associations, suggesting complex, multi-directional, cross-modal correspondences ingrained in the minds of people in the general population.

Acknowledgements

We thank Sujin Kim, Yuna Kwak, and Minsun Park for their helpful comments on the manuscript. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (No. NRF-2016R1A2B4011267) awarded to C-YK and by National Institutes of Health-National Institute on Deafness and Other Communication Disorders (NIH-NIDCD) Grant No. DC-002717 to Haskins Laboratories.

References

Asano, M. and Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes, *Consc. Cogn.* **20**, 1816–1823.

Asano, M. and Yokosawa, K. (2012). Synesthetic colors for Japanese late acquired graphemes, *Consc. Cogn.* **21**, 983–993.

Bankieris, K. and Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition* **136**, 186–195.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H. and Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex, *Nat. Neurosci.* **7**, 1190–1192.

Beeli, G., Esslen, M. and Jäncke, L. (2007). Frequency correlates in grapheme–color synaesthesia, *Psychol. Sci.* **18**, 788–792.

Bien, N., Ten Oever, S., Goebel, R. and Sack, A. T. (2012). The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study, *NeuroImage* **59**, 663–672.

Boersma, P. and Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Version 5.3.51. http://www.praat.org/. Retrieved 2 June 2013.

Brang, D., Rouw, R., Ramachandran, V. S. and Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme–color synesthesia, *Neuropsychologia* **49**, 1355–1358.

Brang, D., Williams, L. E. and Ramachandran, V. S. (2012). Grapheme–color synesthetes show enhanced crossmodal processing between auditory and visual modalities, *Cortex* **48**, 630–637.

Browman, C. P. and Goldstein, L. M. (1986). Towards an articulatory phonology, *Phonol. Yearb.* **3**, 219–252.

Cohen Kadosh, R. and Henik, A. (2007). Can synaesthesia research inform cognitive science? *Trends Cogn Sci.* **11**, 177–184.

Cuskley, C. and Kirby, S. (2013). Synaesthesia, cross-modality, and language evolution, in: *Oxford Handbook of Synaesthesia*, J. Simner and E. M. Hubbard (Eds), pp. 869–907. Oxford University Press, Oxford, UK.

Cuskley, C., Simner, J. and Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect, *Psychol. Res.* **81**, 119–130.

Deroy, O. and Spence, C. (2013). Why we are not all synesthetes (not even weakly so), *Psychonom. Bull. Rev.* **20**, 643–664.

Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P. and Merikle, P. M. (2006). The role of meaning in grapheme–colour synaesthesia, *Cortex* **42**, 243–252.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D. and Sarma, A. K. (2007). A standardized test battery for the study of synesthesia, *J. Neurosci. Meth.* **159**, 139–145.

Fairbanks, G., House, A. S. and Stevens, E. L. (1950). An experimental study of vowel intensities, *J. Acoust. Soc. Am.* **22**, 457–459.

Fernay, L., Reby, D. and Ward, J. (2012). Visualized voices: a case study of audio-visual synesthesia, *Neurocase* **18**, 50–56.

Gallace, A. and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size, *Percept. Psychophys.* **68**, 1191–1203.

Hanson, H. M. and Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn, *J. Acoust. Soc. Am.* **112**, 1158–1182.

Hinton, L., Nichols, J. and Ohala, J. J. (2006). *Sound Symbolism*. Cambridge University Press, Cambridge, UK.

Honda, K. (1983). Relationship between pitch control and vowel articulation, in: *Haskins Laboratories Status Report on Speech Research, SR, 73*, pp. 269–282. Haskins Laboratories, New Haven, CT, USA.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval, *Am. J. Psychol.* **109**, 219–238.

Jacobson, R. (1962). *Selected Writings I: Phonological Studies*. Mouton, The Hague, The Netherlands.

Kang, M.-J., Kim, Y., Shin, J.-Y. and Kim, C.-Y. (2017). Graphemes sharing phonetic features tend to induce similar synesthetic colors, *Front. Psychol.* **8**, 337. DOI:10.3389/fpsyg.2017.00337.

Karwoski, T. F. and Odbert, H. S. (1938). Color–music, *Psychol. Monogr.* **50**, 1–60.

Kim, S., Blake, R. and Kim, C.-Y. (2013). Is 'Σ' purple or green? Bistable grapheme–color synesthesia induced by ambiguous characters, *Consc. Cogn.* **22**, 955–964.

Kim, Y. and Kim, C.-Y. (2014). Correlation between grapheme frequency and synesthetic colors in color-graphemic synesthesia, *Korean J. Cogn. Biol. Psychol.* **26**, 133–149.

Köhler, W. (1929). *Gestalt Psychology*. Liveright, New York, NY, USA.

Lewkowicz, D. J. and Turkewitz, G. (1980). Cross-modal equivalence in early infancy: auditory–visual intensity matching, *Dev. Psychol.* **16**, 597–607.

Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness, *Am. J. Psychol.* **87**, 173–188.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions, *Psychol. Bull.* **82**, 303–331.

Marks, L. E. (1989). On cross-modal similarity: the perceptual structure of pitch, loudness, and brightness, *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 586–602.

Marks, L. E., Szczesiul, R. and Ohlott, P. (1986). On the cross-modal perception of intensity, *J. Exp. Psychol. Hum. Percept. Perform.* **12**, 517–534.

Martino, G. and Marks, L. E. (2001). Synesthesia: strong and weak, *Curr. Dir. Psychol. Sci.* **10**, 61–65.

Maurer, D., Pathman, T. and Mondloch, C. J. (2006). The shape of boubas: sound–shape correspondences in toddlers and adults, *Dev. Sci.* **9**, 316–322.

Melara, R. D. (1989). Dimensional interaction between color and pitch, *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 69–79.

Mermelstein, P. (1973). Articulatory model for the study of speech production, *J. Acoust. Soc. Am.* **53**, 1070–1082.

Moos, A., Smith, R., Miller, S. R. and Simmons, D. R. (2014). Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder, *i-Perception* **5**, 132–142.

Nam, H., Goldstein, L. M., Giulivi, S., Levitt, A. G. and Whalen, D. H. (2013). Computational simulation of CV combination preferences in babbling, *J. Phon.* **41**, 63–77.

Newman, S. S. (1933). Further experiments in phonetic symbolism, *Am. J. Psychol.* **45**, 53–75.

Nielsen, A. K. and Rendall, D. (2013). Parsing the role of consonants *versus* vowels in the classic Takete–Maluma phenomenon, *Can. J. Exp. Psychol.* **67**, 153–163.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J. and Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices, *J. Neurosci.* **27**, 11431–11441.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J. and Friston, K. J. (2007). The effect of prior visual information on recognition of speech and sounds, *Cereb. Cortex* **18**, 598–609.

Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch, in: *Sound Symbolism*, L. Hinton, J. Nichols and J. J. Ohala (Eds), pp. 325–347. Cambridge University Press, Cambridge, UK.

Palmer, S. E., Schloss, K. B., Xu, Z. and Prado-León, L. R. (2013). Music–color associations are mediated by emotion, *Proc. Natl Acad. Sci.* **110**, 8836–8841.

Parise, C. V. and Spence, C. (2009). ‘When birds of a feather flock together’: Synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS ONE* **4**, e5664. DOI:10.1371/journal.pone.0005664.

Peña, M., Mehler, J. and Nespor, M. (2011). The role of audiovisual processing in early conceptual development, *Psychol. Sci.* **22**, 1419–1421.

Ramachandran, V. S. and Hubbard, E. M. (2001). Synaesthesia — a window into perception, thought and language, *J. Consc. Stud.* **8**, 3–34.

Revill, K. P., Namy, L. L., DeFife, L. C. and Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: evidence from fMRI and DTI, *Brain Lang.* **128**, 18–24.

Rich, A. N., Bradshaw, J. L. and Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: implications for the role of early experience in lexical–colour associations, *Cognition* **98**, 53–84.

Rubin, P. E., Saltzman, E., Goldstein, L. M., McGowan, R. S., Tiede, M. K. and Browman, C. P. (1996). CASY and extensions to the task-dynamic model, in: *Proceedings of the 1st ESCA ETRW on Speech Production Modeling and 4th Speech Production Seminar*, pp. 125–128. Autrans, Grenoble, France.

Sapir, E. (1929). A study in phonetic symbolism, *J. Exp. Psychol.* **12**, 225–239.

Shin, E. H. and Kim, C. Y. (2014). Both ‘나’ and ‘な’ are yellow: cross-linguistic investigation in search of the determinants of synesthetic color, *Neuropsychologia* **65**, 25–36.

Simner, J. (2007). Beyond perception: synaesthesia as a psycholinguistic phenomenon, *Trends Cogn. Sci.* **11**, 23–29.

Simner, J. and Ludwig, V. U. (2012). The color of touch: a case of tactile–visual synaesthesia, *Neurocase* **18**, 167–180.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L. and Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations, *Cogn. Neuropsychol.* **22**, 1069–1085.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C. and Ward, J. (2006). Synaesthesia: the prevalence of atypical cross-modal experiences, *Perception* **35**, 1024–1033.

Simpson, R. H., Quinn, M. and Ausubel, D. P. (1956). Synesthesia in children: association of colors with pure tone frequencies, *J. Gen. Psychol.* **89**, 95–103.

Smilek, D., Carriere, J. S., Dixon, M. J. and Merikle, P. M. (2007). Grapheme frequency and color luminance in grapheme–color synaesthesia, *Psychol. Sci.* **18**, 793–795.

Spector, F. and Maurer, D. (2013). Early sound symbolism for vowel sounds, *i-Perception* **4**, 239–241.

Thompson, P. D. and Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function, *Q. J. Exp. Psychol. (Hove)* **64**, 2392–2404.

Ward, J., Huckstep, B. and 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. and Enns, J. T. (2012). Second-order mappings in grapheme–color synesthesia, *Psychonom. Bull. Rev.* **19**, 211–217.

Whalen, D. H. and Levitt, A. G. (1995). The universality of intrinsic F_0 of vowels, *J. Phon.* **23**(3), 349–366.

Wrembel, M. (2009). On hearing colours — cross-modal associations in vowel perception in a non-synaesthetic population, *Poznań Stud. Contemp. Linguist.* **45**, 595–612.