한국심리학회지 : 인지 및 생물
The Korean Journal of Cognitive and Biological Psychology
2020, Vol. 32, No. 1, 85-99

Original Article
http://dx.doi.org/10.22172/cogbio.2020.32.1.006

Nasal Sounds are Lighter and More Yellowish than Glottal Sounds: Cross-modal Associations between Consonant Sounds and Colors*

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

[1]Department of Psychology, Korea University

[2]Department of English Language and Literature, Korea University

[3]Haskins Laboratories, USA

People share implicit cross-modal mappings for certain visual and auditory features such as pitch and human speech. The current study explored the role of phonetic features in intrinsic associations between consonant sounds and colors. For this purpose, we presented synthetic consonant sounds generated by parametrically manipulating oral and non-oral constriction gestures of speech organs, using an articulatory synthesizer. Participants were asked to choose a color after hearing each sound. Color-matching results showed that nasal sounds characterized by a velic gesture were associated with lighter and more yellowish colors than other sounds. The perceptual space of the consonant sounds from dissimilarity judgment ratings indicated that participants could capture the consonantal nature of the stimuli. These results imply the non-arbitrary association between phonetic features of consonants and colors.

Keywords: cross-modal association, color, consonant, articulatory synthesis

1 차원고접수 19.06.01; 수정본접수: 19.12.09; 최종게재결정 19.12.10

People share implicit cross-modal mappings between certain visual and auditory features. For example, it has been established in multiple studies that high-pitched sounds tend to be associated with smaller shapes (Bien, ten Oever, Goebel, & Sack, 2012; Gallace & Spence, 2006), more angular shapes (Karwoski & Odbert, 1938), and lighter shades (Marks, 1974), compared to low-pitched sounds. These mappings not only affect people's implicit preferences, but also promote multisensory integration in congruent cross-modal combinations (Parise & Spence, 2009).

Implicit cross-modal correspondences also exist for audio-visual features of human speech. A well-known example is the correspondence between size and vowel, where low vowels such as [o] or [a] tend to be matched with larger sizes while high vowels such as [e] or [i] tend to be matched with small sizes (Sapir, 1929). Speech sounds are also known to be associated with visual shape, often called the 'maluma-takete' (Köhler, 1929) or 'bouba-kiki' (Ramachandran & Hubbard, 2001) effect, i.e., the strong tendency to pair a round shape with 'maluma' or 'bouba', while pairing an angular shape with

* This study was supported by the National Research Foundation of Korea Grant funded by the Korean Government (No. NRF-2016R1A2B4011267) awarded to C-YK and by the Research Grant of Korea University (No. K1817661) awarded to HN

† Corresponding Authors: Chai-Youn Kim, Korea University, (02841), Anam-ro, Seongbuk-gu, Seoul, Korea
E-mail: chaikim@korea.ac.kr and Hosung Nam, Korea University, (02841), Anam-ro, Seongbuk-gu, Seoul, Korea
E-mail: hnam@korea.ac.kr

'takete' or 'kiki'. These non-arbitrary linkages between acoustic features of speech sounds and visual features have been demonstrated across speakers of different languages (e.g., Bremner et al., 2013, for the 'bouba-kiki' effect, and Shinohara & Kawahara, 2010, for vowel-size correspondence). Notably, such cross-modality of speech has been shown for pre-literate children (Maurer, Pathman, & Mondloch, 2006, for the 'bouba-kiki' effect) and even infants (Peña, Mehler, & Nespor, 2011, for vowel-size correspondence), suggesting a pre-linguistic basis (but also see Cuskley, Simner, & Kirby, 2017).

Cross-modal correspondences between phonetic and visual features are often dubbed 'phonetic symbolism' (Sapir, 1929), or more generally called 'sound symbolism', meaning a non-arbitrary linkage between sound and meaning (Cuskely & Kirby, 2013; Hinton, Nichols, & Ohala, 2006). Sound symbolism has attracted a wide range of interest from researchers in psychology and linguistics since it may have implications for the understanding of language in terms of acquisition (Asano et al., 2015; Imai, Kita, Nagumo, & Okada, 2008) and processing (Sučević, Savić, Popović, Styles, & Ković, 2015); it may even provide insights into the evolution of language (Cuskely & Kirby, 2013; Imai & Kita, 2014).

Beginning with previous findings on the cross-modality of human speech, the present study intends to explore the cross-modal associations between phonetic features inherent in sounds and colors. Our previous study has reported non-arbitrary associations between vowel sounds and colors (Kim, Nam, & Kim, 2018). In this study, we investigated cross-modal associations between consonant sounds and colors. For auditory stimuli, we employed consonant-like speech sounds that were synthetically generated using an articulatory synthesizer. Since the stimuli are synthetic (i.e., not natural), and not confined to a specific language, they are less sensitive to the influence of linguistic properties than human speech sounds. Thus, by using synthetic speech sounds, we tried to minimize the impact of linguistic factors in non-arbitrary mappings between speech sounds and colors.

For auditory stimuli, a set of synthesized vowel-consonant-vowel (VCV) sounds were generated using the Haskins Laboratories articulatory synthesizer (Nam, Goldstein, Saltzman, & Byrd, 2004; Rubin et al., 1996). Phonetic features of the auditory stimuli used in the present study were specified based on Articulatory Phonology (Browman & Goldstein, 1986). This account describes human speech in terms of coordinated articulatory movements of distinctive vocal tract organs (e.g., lips, tongue, velum, etc.) within a vocal tract space over time. Characteristic patterns of articulatory movements during speech production are defined as 'articulatory gestures'. In Articulatory Phonology, temporal organizations of articulatory gestures produced by distinct speech organs (e.g., lips, tongue, and glottis), including overlaps between them, constitute speech utterances. Individual articulatory gestures and the relations among them have been mathematically formalized within a computational speech model (the Task-Dynamic model: Saltzman & Kelso, 1983; Saltzman & Munhall, 1989), which has made it possible to generate synthetic speech sounds within the gestural framework. In the current study, two articulatory factors were considered for defining stimulus conditions: an 'oral constriction gesture' and a 'non-oral constriction gesture', which can be temporally overlapped during an utterance (see Methods for details). The flanking vowels were set to be equal

Table 1. The set of twelve VCV stimuli used in the experiment

		Non-oral constriction gesture			
		N/A	Velum	Glottis	Glottis (VOT>0)
Oral constriction gesture	Lips	[ebe]	[eme]	[epe]	[ep^{h}e]
	Tongue tip	[ede]	[ene]	[ete]	[et^{h}e]
	Tongue body	[ege]	[eŋe]	[eke]	[ek^{h}e]

across all VCV sequences (a vowel that sounds close to [e]), since the aim is to investigate sound-color associations related to consonants, and not the vowel quality of the stimuli. The resulting twelve VCV sounds generated by combinations of the two factors were used for the experiment (Kwak & Kim, 2018, also shown in Table 1).

While cross-modal correspondences between consonants and visual features such as shape or size have well been documented in the previous literature (e.g., the 'bouba-kiki' effect), such linkages between consonants and colors have seldom been addressed. Previous studies have tested synesthetes (who experience colors when hearing spoken words) in comparison to non-synesthetes, used as controls (Baron-cohen, Harrison, Goldstein, & Wyke, 1993; Beeli, Esslen, & Jäncke, 2008; Nunn et al., 2002; Paulesu et al., 1995). Moreover, those works employed spoken words, rather than isolated consonant sounds, as auditory stimuli, which suggests that their results might be confounded by linguistic contexts (e.g., Dixon, Smilek, Duffy, Zanna, & Merikle, 2006). Whilst other studies have examined sound-color associations in the general population using speech sounds without linguistic context, they had limitations in that they only

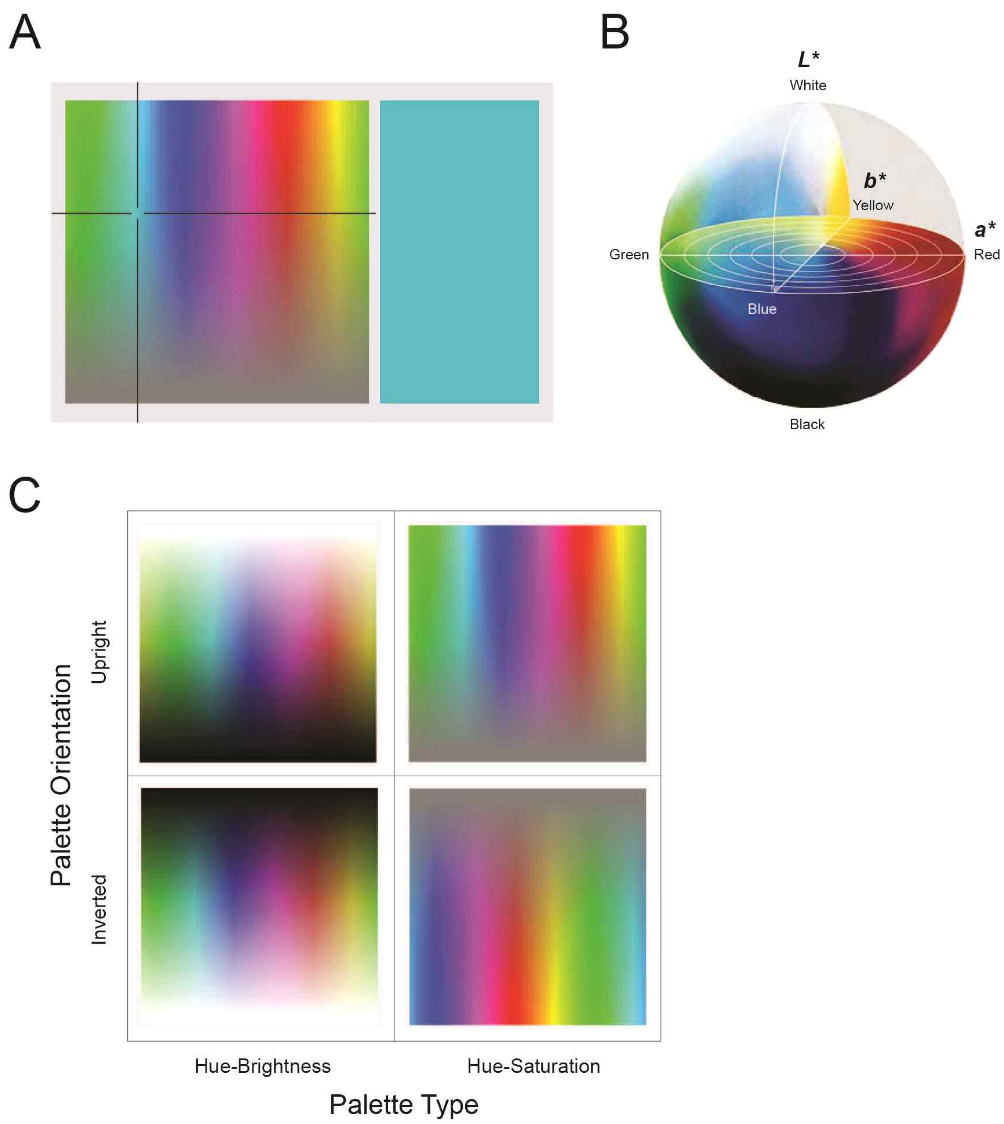


Figure 1. (A) An example of a color palette for the color-matching test. In each trial, participants heard an auditory stimulus and then selected a color best associated with it. (B) The CIELAB color space. The three dimensions L*, a*, and b* describe luminance, green-red axis, and blue-yellow axis, respectively. The RGB values of the matched colors were converted into Lab color coordinates for luminance and chromaticity analyses. (C) Four types of color palette used for the color-matching task.

used vowel sounds as stimuli. Specifically, they did not include consonant sounds (Kim et al., 2018; Moos, Smith, Miller, & Simmons, 2014), or used speech stimuli in a CVC context without manipulating the consonant parts (Wrembel, 2009). To our knowledge, the current work makes the first attempt to explore cross-modal associations between consonants and colors using synthetic speech sounds generated according to articulations of constricting speech organs.

The experiment comprised two sessions. In the first session, participants performed a color-matching task where in each trial they heard an auditory stimulus and had to select a color that best matches the sound from a color palette. The color palette comprised a continuous scale of colors, which allowed a great degree of freedom in choosing a color (Figure 1A). Accordingly, participants' color choices were not expected to be confined within a small number of representative colors. The chromaticity and luminance of the chosen colors were analyzed, which enabled us to examine whether the association between color and phonetic sound is systematically modulated by articulation (Figure 1B).

After completing the color-matching procedure, participants were tested with a dissimilarity judgment task where they heard two serially presented auditory stimuli and had to judge the degree of dissimilarity between them in each trial. Perceptual dissimilarity (or similarity) ratings can be used to reconstruct the perceptual space of participants' internal representations for given stimuli, which, in turn, can be compared to the physical space of the stimuli (Shepard, 1987; for example, Cooke, Jäkel, Wallraven, & Bülthoff, 2007; Lee Masson, Bulthé, Op de Beeck, & Wallraven, 2016). This task session was added to the experiment since the consonantal identities of the VCV sounds were not found to be readily discriminable from one another for most participants in a pilot experiment. Therefore, with the dissimilarity rating results, we sought to confirm whether participants represent the synthetic VCV sounds based on the articulatory factors by which they were generated, even when they are not explicitly informed about the consonantal characteristics of stimuli.

Methods

Participants

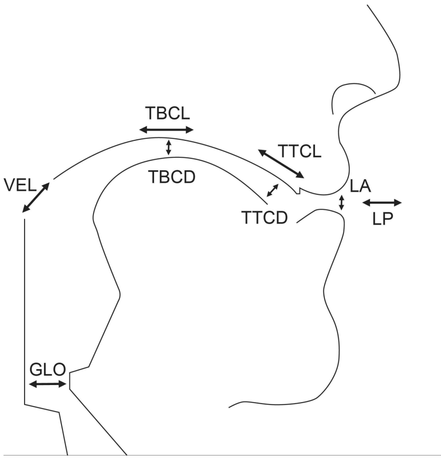
Forty participants (14 males, 19-28 years of age) took part in the experiment. The number of participants was determined based on previous studies (Kim et al., 2018; Moos et al., 2014) and on a pilot study which used stimuli and procedures analogous to the current study. All participants reported no form of synesthesia, which was verified 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 participants used Korean as their native language and provided written informed consent using forms approved by the Korea University Institutional Review Board (1040548-KU-IRB-15-67-A-2).

Apparatus

All auditory stimuli were delivered through SRH440 headphones. The color-matching and dissimilarity judgment procedures were run using Matlab (version 8.3, Mathworks, MA) on a 19-inch, color-calibrated CRT monitor (1024 × 768 resolution, 60 Hz frame rate). The experiment was conducted in a quiet, dark room.

Stimuli

Synthetic speech-like VCV sounds were used as auditory stimuli. The TAsk-Dynamic Application model (TADA; Nam, et al., 2004) was employed to create and manipulate consonants (Figure 2). TADA is a speech production model for mathematical and computational implementation of Articulatory Phonology (Browman & Goldstein, 1986). In this model, a speech utterance is expressed by a constellation of constriction gestures (i.e., opening-closing or closing-opening movements) produced by five constricting organs: lips, tongue tip, tongue body, velum, and glottis (Browman & Goldstein, 1986; 1988; 1990; Saltzman & Munhall, 1989). Lips, tongue tip, and tongue body are major oral constrictors while velum and glottis are non-oral constrictors that determine nasality



Organs	Tract variables
Lips	Lip aperture (LA) Lip protrusion (LP)
Tongue Tip	Tongue tip constriction location (TTCL) Tongue tip constriction degree (TTCD)
Tongue Body	Tongue body constriction location (TBCL) Tongue body constriction degree (TBCD)
Velum	Velic aperture (VEL)
Glottis	Glottal aperture (GLO)

Figure 2. Vocal tract representation with tract variables for generating consonant sounds (left) and a table listing five speech organs and their associated tract variables (right). Degrees of freedom for individual tract variables are indicated by arrows. To generate VCV sounds, articulatory gestures of oral constrictors (lips, tongue tip, and tongue body) and non-oral constrictors (velum, glottis) were manipulated. The vowel gesture and the associated tract variables were controlled across stimuli.

and voicing of a sound, respectively. For a given speech organ, a constriction gesture is defined by its vocal tract variables such as location or degree of the constriction (Figure 2). In TADA, a gestural action is implemented in the form of a gestural score, by which the presence or absence and the activation interval of constriction gestures are specified during an utterance.

A set of twelve VCV sequences was synthesized with manipulation of both oral and non-oral constriction gestures using TADA (see Table 1). The vowel gesture for the initial and final flanking vowels was fixed at its rest position across all the VCV sequences. For the consonant part, oral constriction gestures (lips, tongue tip, and tongue body) were manipulated with related tract variables (e.g., the constriction location and degree) controlled, which generated consonants that sound like [b], [d], and [g], respectively. Each type of oral gesture was independently coupled with non-oral constriction gestures to produce nasal consonants ([m], [n], and [ŋ] with a velic opening-closing gesture), and voiceless unaspirated consonants ([p], [t], and [k] with a glottal opening-closing gesture). Voiceless aspirated consonants ([p^h], [t^h], [k^h]) were also generated by manipulating inter-gestural timing between a glottal gesture and oral gestures, where the onset of a glottal gesture is temporally delayed with respect to that of an oral gesture, making the subsequent consonant aspirated.

The articulatory movement trajectories of the twelve VCV sequences were computed by the TADA procedure. The corresponding area functions, represented as the model articulator variables in the Haskins Laboratories Configurable Articulatory SYnthesis model (CASY: Rubin et al., 1996), were estimated from the computed vocal tract dynamics over time, and were then used to generate acoustic outputs through articulatory synthesis. The duration and pitch of the VCV sounds were set to 600 ms and 125 Hz, respectively.

Procedures

Participants engaged in a modified version of the standardized synesthesia battery (Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007; see Figure 1A) for the color-matching test. For each trial, participants were instructed to listen to an auditory stimulus and to pick a color associated with it by clicking on the color palette on the monitor screen using a mouse, with no time constraint. Participants were explicitly informed that the first part of an auditory stimulus (i.e., the first vowel of a VCV sound) was identical for every trial while the second part (i.e., the CV of a VCV sound) would sound different from trial to trial. They were thus encouraged to focus on the second part when listening to the sound and to select a color associated with that.

Two types of color palette with different color scales were used: one for hue-saturation (H-S) and the other for hue-brightness (H-B). Additionally, each type of color palette was presented in one of two orientations: brighter and more saturated colors were displayed either at the top (upright), or at the bottom (inverted), with respect to the vertical axis of the palette (Figure 1C). These methodological considerations were introduced to factor out any potential influences of the characteristics of the matching palette upon the color-matching results. The brightness or saturation of the color to be selected was adjusted using two keyboard buttons to increase ('→') or to decrease ('←') the brightness or saturation. Each auditory stimulus sequence was presented two times for each of the four different color palettes. Accordingly, the test consisted of 96 trials in total, which were presented in a randomized order.

Participants completed the dissimilarity judgment task in the next session. In each trial, a pair of either identical or different auditory stimuli were presented sequentially with an inter-stimulus interval of 1.2 s. Participants then rated the degree of dissimilarity between the sound pair on a seven-point scale with higher scores indicating greater dissimilarity; they were instructed to assign the value of 1 for pairs that sounded identical. During the session, each of the twelve differing pairs of auditory stimuli was presented twice, and each of the twelve pairs of identical stimuli was presented once. Therefore, the session included of 144 trials, presented in a random order.

Data analyses

For the color-matching data, the RGB values of the selected colors were converted into CIELAB color coordinates (Figure 1B) to examine luminance and chromaticity of the colors separately. The converted values were then averaged into a mean value for each stimulus/palette condition. The luminance and chromaticity of the matched colors were analyzed using their L^*, a^*, and b^* values. Since we found neither significant main effects nor interactions with regard to palette orientation (all $p > .05$), we collapsed the data across the palette orientation factor. However, we included the palette type factor in our main analyses since there were significant main effects for this, as well as interaction effects. Therefore, three dependent variables were entered into a three-way within-subjects repeated measures MANOVA [Palette Type (2: hue-saturation or hue-brightness type) X Oral Constriction Gesture (3: lip, tongue tip, or tongue body gesture) X Non-Oral Constriction Gesture (4: velic, glottal, delayed glottal, or no gesture)]. The Greenhouse–Geisser (GG) correction was applied to the degrees of freedom when necessary to correct for violations of sphericity. Bonferroni corrections were performed for all post-hoc multiple comparisons. All statistical analyses for the color-matching data were conducted using IBM SPSS Statistics (Version 23) software.

For the dissimilarity rating data, a dissimilarity matrix was made with the mean ratings of pairs for each participant. Since individual ratings across participants were highly correlated (mean r = 0.603), these matrices were averaged into a group dissimilarity matrix to perform multi-dimensional scaling (MDS) analysis (Ashby, Maddox, & Lee, 1994). Using the mdscale function in MATLAB, MDS was applied to the group matrix to obtain a two-dimensional perceptual space for VCV stimuli.

Results

Dissimilarity Rating Result

The MDS map of the group-averaged matrix was first analyzed to examine whether the participants had perceptual representations of the auditory stimuli according to the two gestural factors (i.e., oral and non-oral constriction gestures). Therefore, we applied the MDS for two-dimensional solutions (stress value = 0.054). The resulting two-dimensional perceptual space is shown in Figure 3. Consonants sharing the same oral or non-oral constriction gesture tended to be grouped together along the two dimensions. In particular, there was a clear perceptual distinction between VCV sounds with and without a glottal gesture. These results show that participants could capture the articulatory nature of

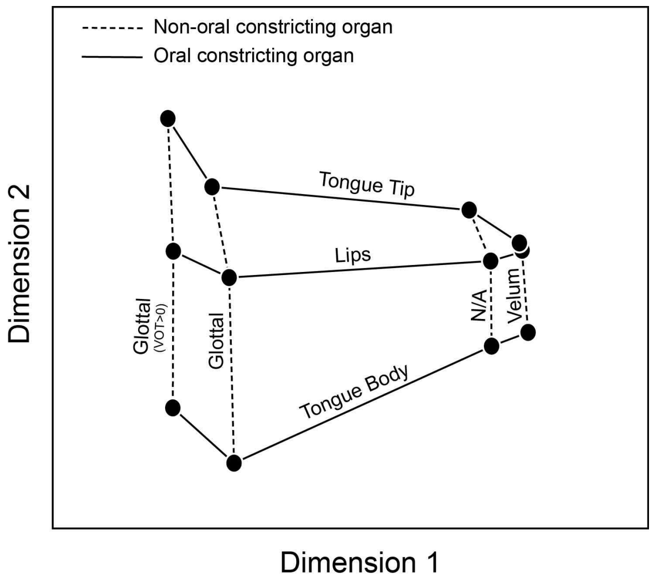


Figure 3. MDS results from the similarity judgment ratings for the VCV sounds. Solid and dashed lines connect consonants with the same oral and non-oral constriction gestures, respectively. Relevant speech organs are indicated next to the lines. The MDS for two-dimensional solutions yielded a perceptual space that, overall, conforms with the two constriction gesture factors.

synthesized sounds even though they were not explicitly informed of this.

Color-Matching Results

Group-averaged color values for each stimulus condition are displayed in the CIELAB color space (Figure 4). Consistent with the MDS results, there was a tendency to associate colors with VCV sounds based on the presence of a glottal gesture in their articulations. Furthermore, the matched color distribution showed non-arbitrary patterns within the non-oral constriction gesture factor. Statistical analysis confirmed this observation: a three-way repeated measures MANOVA of *L**, *a**, and *b** indicated significant main effects of non-oral constriction gesture [Wilks' l = .837, $F(3,117) = 2.36$, $p < .05$] and palette type [Wilks' l = .446, $F(1,39) = 15.29$, $p < .001$]. Also, we found a significant interaction between non-oral gesture and palette type [Wilks' l = .831, $F(3,117) = 2.46$, $p < .05$]. None of the other main and interaction effects was significant (all $p > .130$).

Since significant effects were obtained with the MANOVA, we next examined the color-matching results for individual color dimensions. In terms of luminance, a three-way repeated measures ANOVA of *L** revealed a significant main effect of non-oral constriction gesture

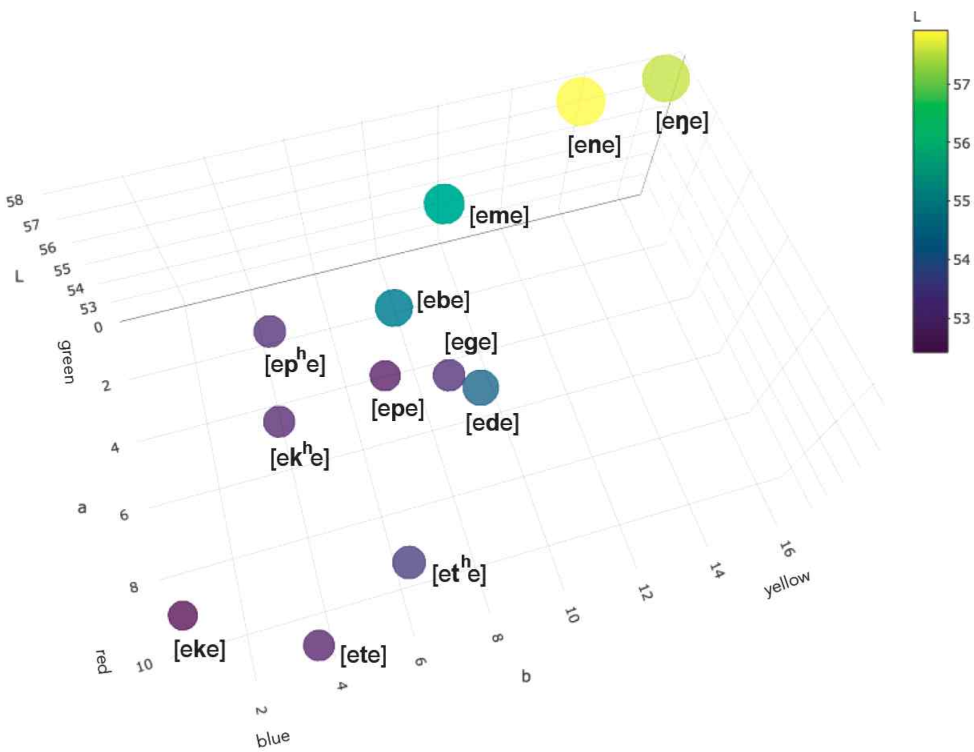


Figure 4. A visual representation of the group-averaged colors for each stimulus condition in the CIELAB color space (see Figure 1B). The circles are labelled with representative consonants of the stimuli. The size and color of each circle indicates the *L** value. Stimuli matched with lighter colors are shown with lighter, more yellowish, and bigger circles.

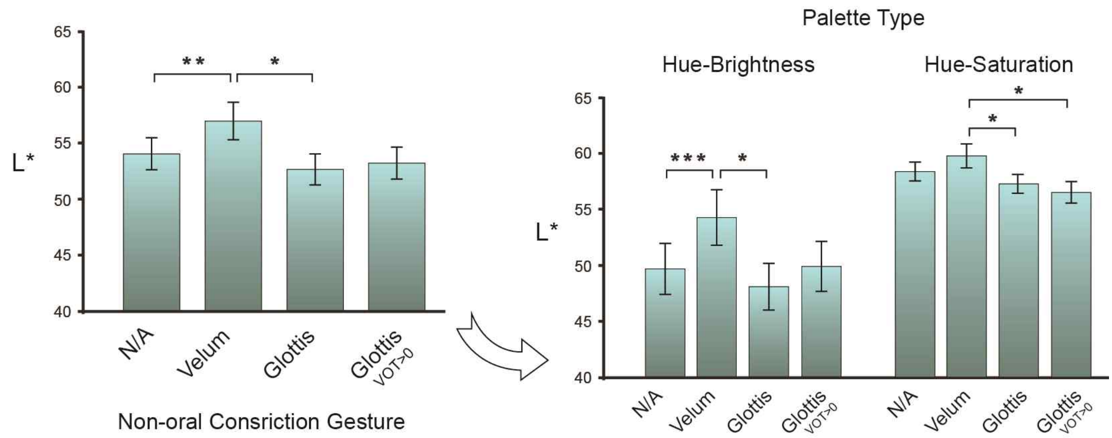


Figure 5. Luminance ($L*$) results. The group mean $L*$ values averaged for each non-oral constriction gesture involved in VCV sequences (left), and the L* values sub-divided based on the type of color palette (right). The error bars denote 1 SEM. *p 〈 .05, **p 〈 .01, ***p 〈 .001, Bonferroni corrections were applied for pair-wise comparisons.

[F(1.68,65.58) = 6.12, p 〈 .01; GG corrected]. Participants showed a tendency to select lighter colors when hearing VCV sounds with a velic gesture (i.e., nasal consonants that sound like [eme], [ene], and [eŋe]) than when hearing other VCV sounds (Figure 5). Post-hoc paired comparisons showed significant differences of lightness between the velum and N/A conditions (p 〈 .01), and between velum and glottis conditions (p 〈 .05). Also, the main effect of palette type was highly significant [F(1,39) = 26.61, p 〈 .001]; participants tended to choose lighter colors from a H-S palette than from a H-B palette. This effect was not surprising, considering the fact that the H-B palette provides much darker colors (closer to black), relative to the H-S palette (see Figure 1C). Of more relevance to the main purposes of the current study, a significant interaction effect was found between non-oral gesture and palette type [F(3,117) = 2.81, p 〈 .05], indicating that the effect of non-oral gesture differed depending on which type of palette was presented (Figure 5). Specifically, participants matched lighter colors to VCV sounds with a velic gesture (e.g., a sound like [eme]), compared to those without a non-oral gesture (e.g., a sound like [ebe]), and this was more evident, albeit with the same trends, when presented with the H-B palette than with the H-S palette. Post-hoc paired comparisons showed significant differences for the velum - N/A (p 〈 .001) and velum - glottis (p 〈 .05) pairs within the H-B palette condition, and for the velum - glottis (p 〈 .05) and velum - glottis (VOT〉0) (p 〈 .05) pairs within the H-S palette condition. All the other main and interaction effects for $L*$ were not significant (all p 〉 .264).

Next, the chromaticity of the matched colors along the blue-yellow color axis was analyzed using the $b*$ values. A three-way repeated measures ANOVA of $b*$ revealed a main effect of non-oral gesture [F(1.86,72.64) = 3.95, p 〈 .05; GG corrected]. Results showed that more yellowish colors tended to be associated with VCV sounds with a velic gesture, compared to the other sounds, albeit with no significant differences in paired comparisons (Figure 6). There was also a main effect of palette type [F(1,39) = 6.02, p 〈 .05], with more yellowish colors associated with selections on the H-S palette than on the H-B palette. More importantly, a two-way interaction between non-oral gesture and palette type was significant [F(3,117) = 3.50, p 〈 .05]. Unlike the same interaction term shown for $L*$, which indicated stronger non-oral gestural effects with the H-B palette, this effect was more evident with the H-S palette than with the H-B palette (Figure 6). In other words, nasal sounds were strongly associated with yellowish colors when the H-S palette was presented, and less so when the H-B palette was presented. This suggests that the H-S palette was more effective for detecting the non-arbitrary yellowness-color mappings shared within participants. Post-hoc paired comparisons showed

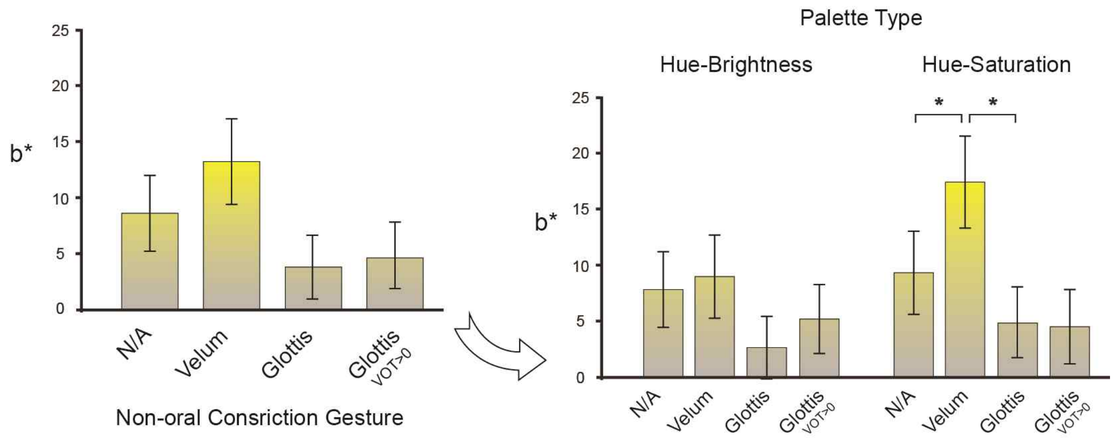


Figure 6. Chromaticity results along the blue-yellow continuum ($b*$). The group mean $b*$ values averaged for each non-oral constriction gesture involved in VCV sequences (left). The $b*$ values sub-divided based on the type of color palette (right). The error bars denote 1 SEM. *p 〈 .05, Bonferroni corrections were applied for pair-wise comparisons.

significant differences for the velum - N/A ($p < .05$) and velum - glottis ($p < .05$) pairs within the H-S palette condition. None of the other main effects or interactions for $b*$ were significant (all $p > .139$).

An analogous ANOVA performed on $a*$ values indicated no significant main effects or interactions for the chromaticity of the matched colors along the green-red color axis (all $p > .084$).

Discussion

In the current study, we explored cross-modal associations between sounds and colors based on the phonetic principles by which the sounds were generated using articulatory synthesis. Our experiment investigated the influence of consonant characteristics of synthetic speech sounds on the associated colors. The influence of consonants on colors was examined by exploiting VCV stimuli defined by articulatory gestures of oral and/or non-oral constrictors during consonant production. It was found that the nasality of sounds, characterized by a velic gesture, affected luminance and chromaticity of the matched colors. Specifically, nasal sounds-i.e., VCV sounds generated with a velic gesture-tended to be associated with lighter and/or more yellowish colors than did the non-nasal sounds[1]. The dissimilarity rating results support this articulatory interpretation for the consonant-color associations by showing that the perceptual space of VCV sounds, as reconstructed from MDS, conformed to two factors of oral and non-oral articulatory gestures. The current results imply cross-modal mechanisms between articulatory features and colors.

The articulatory approach taken by this study has several advantages in examining the cross-modality of speech sounds. Firstly, since the synthetic stimuli are both defined and generated in articulatory terms, they are not likely to depend on a specific language, and thus are not strongly affected by speaker or language characteristics. In this sense, a gestural approach such as phonetic principles may have the potential to be used for comparisons between multiple studies involving participants from different linguistic and cultural backgrounds. Secondly, in articulatory synthesis, it is possible to control articulatory components in a specific and systematic manner, enabling the generation of auditory stimuli. When producing a consonant sound, a constriction gesture produced by a speech organ can be affected by either degree or location of the constriction, and further by dynamic parameters

1) These results are consistent with the results of a pilot experiment conducted in our lab (Kim et al., VSS 2016), where CV sounds with a velic gesture (e.g., [me], [ne], [ŋe]) were associated with lighter and more yellowish colors than those with a glottal gesture (e.g., [pe], [te], [ke]), or without a non-oral gesture (e.g., [be], [de], [ge]). Such consistency across these two experiments further supports the finding of non-arbitrary associations between consonant sounds and colors.

such as gestural target, stiffness, and dampening (Saltzman & Munhall, 1989). However, the Haskins articulatory synthesizer can produce consonant acoustics specified solely by the intended articulatory variables, with the other variables controlled. Gestural identity (e.g., a lip gesture or a glottal gesture) and inter-gestural timing (e.g., onset time of a glottal gesture with respect to that of an oral gesture) were used to manipulate consonant conditions, and other task variables such as the vowel gesture and the constriction degree or location, were controlled at fixed values. These methodological advantages allow for a strict examination of cross-modal associations modulated by the targeted articulatory factors.

The current findings, which imply non-arbitrary association between nasality of sounds and lightness, are not easily reconciled with results from previous studies. There have only been a handful of studies addressing the consonant-lightness correspondence. For example, a classic work by Newman (1933) reported an experiment in which participants were asked to choose the 'darker' word out of a word pair including two non-words with different consonants. Results showed a tendency for words with 'm' to be judged as darker than those with 'b', while words with 'd' were judged as darker than those with 'n'. In other words, in articulatory terms, the influence of velic gesture (for 'm' and 'n', but not for 'b' and 'd') on the matched lightness, was different in accordance with the oral gestures involved (a lip gesture for 'b' and 'm' vs. a tongue tip gesture for 'd' and 'n', see Table 1). Another study demonstrated that voiced (e.g., [da]) and voiceless (e.g., [ta]) CV sounds were discriminated faster when accompanied by a black and a white square disks, respectively (Hirata, Ukita, & Kita, 2011), indicating that consonant-lightness correspondence is moderated by the presence of voice (i.e., the presence of a glottal gesture). In the present experiment, however, lighter colors were associated with VCV sounds with a velic gesture than they were with other sounds, and there were no significant differences in the matched luminance between voiced (without a non-oral gesture) and voiceless (with a glottal gesture) sounds. These results are not consistent with previous studies that reported the interaction effect between oral and non-oral gestures (Newman, 1933), and the effect of voice (Hirata et al., 2011). However, it might be premature to draw any conclusion from these differences, since there are a number of methodological discrepancies between the studies in terms of both presentation of consonant stimuli (synthetic vs. spoken or written) and experimental task (color-matching vs. two-alternative choice or discrimination), which could have contributed to the inconsistency of the results. Further investigation, including the comparison between different task settings and sets of stimuli, is needed to provide a more unified picture for consonant-lightness correspondences.

One might argue that the cross-modal effect of nasality on yellowness is confounded with the same effect on lightness because yellowish colors are intrinsically lighter than bluish colors. It should be noted, however, that the influence of nasality on the luminance and yellowness of matched colors differed according to the type of color palette: there was a more pronounced influence on L^* with the H-B palette while a more pronounced influence on b^* with the H-S palette (see Figures 4 and 5). This interaction presumably resulted from the fact that the H-B palette covers a wider range of lightness, which might be apt for capturing luminance-related effects, whereas a H-S palette displays continuous saturation values, which provides a wider range of chromaticity than luminance. Regardless of the reason behind this, the clear difference in the statistical results between L^* and b^* values suggests that the cross-modal effects of nasality on the luminance and chromaticity of colors might be dissociable, and may somewhat independently affect the associated color selections.

The nasality-yellowness mappings revealed here might be attributed to the phonological similarity between nasal sounds and the word 'yellow' in Korean, which is pronounced as [norang]. Indeed, it has been suggested in the grapheme-color synesthesia literature that the initial letter of a color name (e.g., 'y') often induces the corresponding color (e.g., 'yellow') (Rich, Bradshaw, &

Mattingley, 2005). This tendency of pairing between the first letter of a color name and the color was present even in individuals without synesthesia (Simner et al., 2005). A study which tested Korean grapheme-color synesthetes also showed that all participants tended to experience yellowish colors when viewing both the alphabet "N" or the *Hangul* (the Korean alphabet) "나" (pronounced as [na]; Shin & Kim, 2014). However, it should be noted that the stimuli used in the current study are neither natural nor confined to a specific language, and that no linguistic context was given to participants throughout the experiment. This suggests that participants in the current study might not as susceptible to linguistic influences as were those in previous studies. Nonetheless, future studies involving participants who speak languages in which nasal sounds are not semantically associated with yellow may be required to examine the semantic linkage of nasal sounds to yellowish colors.

It is noteworthy from our previous work that non-synesthetic participants tended to pick lighter and more yellowish colors when presented both with front and high vowels in nearly the same color-matching procedure (Kim et al., 2018). Combined with the current results, it seems that front or high vowels and nasal consonants are associated with colors in a similar way in terms of lightness and blueness-yellowness of the matched colors. These analogous effects, shown in vowel- and consonant-color associations, might be explained to some extent by the well-known relationship between the nasality of consonants and the height of vowels (Abramson, Nye, Henderson, & Marshall, 1981). Previous phonological studies have shown that nasalization of vowels alters the spectrum of the vowels, either lowering or raising perceived vowel height alongside interactions among several factors involving vowel height, vowel frontness, and vowel context (Beddor, 1993; Beddor, Krakow, & Goldstein, 1986). Furthermore, a vowel followed by a nasal consonant can be assumed to give rise to certain amount of expectation about the nasality of the vowel. The expectation of nasality is also assumed to be factored out from vowel quality, in turn affecting the perceived vowel height with the vowel perceived as lower or higher to an extent that compensates for the nasalization (Beddor et al., 1986; Ohala, 1986). Therefore, the alteration of perceived vowel height by a preceding nasal consonant could contribute to the effect whereby colors associated with nasal consonants are lighter and more yellowish, as shown for high vowels in Kim et al. (2018).

As an alternative hypothesis, certain aspects of the current findings can be explained by the symbolic correspondences between colors and emotions (D'Andrade & Egan, 1974; Valdez & Mehrabian, 1994). For example, lighter and more saturated colors are associated with positive connotations (e.g., 'happy') while darker and less saturated colors are associated with negative connotations (e.g., 'worried') (D'Andrade & Egan, 1974). In a similar vein, a recent study proposed that cross-modal correspondences between color and music are mediated by emotional aspects of music (Palmer, Schloss, Xu, &, Prado-León, 2013). In this study, faster music in major keys was preferentially matched to more saturated, lighter, and more yellowish colors than was slower music and minor keys; these music-color associations were mediated by correlations between emotional associations with music and with colors. Namely, fast tempo/major keys in music and saturated/light/yellow colors are correlated with positive emotions. According to this proposal, the cross-modal mappings between nasality and lightness/yellowness of color in the current study could be mediated by positive emotions associated with both sounds and colors. Such an interpretation is not far-fetched: there is evidence that artificial words containing nasal consonants (e.g., *m*, *n*, and *ng* as in 'sing') are more often selected as corresponding to euphonious English words (e.g., 'sweet'), than are words with other voiced or voiceless consonants (e.g., *t*, *d*, *ch* as in 'chap') (Johnson, Suzuki, & Olds, 1964; also see Auracher, Albers, Zhai, Gareeva, & Stavniychuk, 2011). Another study suggested that nasal sounds are preferentially included in English female nicknames than are other sounds (De Klerk & Bosch, 1997). In addition, marketing researchers have shown that the letter 'm' frequently occurs as the first letter of brand names (Schloss, 1981), and is frequently included within top brand names (Pogacar, Plant, Rosulek, & Kouril, 2015). Future investigations, including into emotional rating of consonant sounds and applications to consonant-color associations, may clarify the emotional mediation hypothesis.

To sum up, the present study reveals cross-modal associations between articulatory features of consonant sounds and colors by using articulatory synthesis. Our results showed that the nasality of speech sounds is associated with luminance and yellowness. We have discussed three hypotheses about the possible mechanisms of the nasality-yellowness mapping: semantic linkage, vowel height interaction, and emotional mediation. The specific mechanisms of the associations remain to be investigated. However, given the statistical difference in the gesture-palette interaction effects between results for L^* and b^*, nasality-luminance mapping might be independent from nasality-yellowness mapping, which implies that the semantic account may not provide a complete explanation for the cross-modal associations shown in this study. In addition to the semantic linkage of nasal sounds to yellowish colors, more intrinsic linkages between the acoustic and visual properties seem to be involved, which might constitute a basis for sound-symbolic phenomena.

References

Abramson, A. S., Nye, P. W., Henderson, J. B., & Marshall, C. W. (1981). Vowel height and the perception of consonantal nasality. *The Journal of the Acoustical Society of America, 70*, 329-339.

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex, 63*, 196-205.

Auracher, J., Albers, S., Zhai, Y., Gareeva, G., & Stavniychuk, T. (2010). P is for happiness, N is for sadness: Universals in sound iconicity to detect emotions in poetry. *Discourse Processes, 48*, 1-25.

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science, 5*, 144-151.

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

Beddor, P. S. (1993). The perception of nasal vowels. *Nasals, Nasalization, and the Velum, 5*, 171-196.

Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook, 3,* 197-217

Beeli, G., Esslen, M., & Jäncke, L. (2008). Time course of neural activity correlated with colored-hearing synesthesia. *Cerebral Cortex, 18*, 379-385.

Bien, N., ten Oever, S., Goebel, R. & 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.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition, 126*, 165-172.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219-252.

Browman, C. P., & Goldstein, L. M. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*, 140-155.

Browman, C. P., & Goldstein, L. (1990). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics, 18*(3), 411-424.

Cooke, T., Jäkel, F., Wallraven, C., & Bülthoff, H. H. (2007). Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia. 45*, 484-495.

Cuskley, C., & Kirby, S. (2013). Synaesthesia, cross-modality, and language evolution. In J. Simner & E. M. Hubbard (Eds), *Oxford handbook of synaesthesia* (pp. 869-899). Oxford: Oxford University Press.

Cuskley, C., Simner, J. & Kirby, S. (2017). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research, 81*, 119-130.

D'Andrade, R., & Egan, M. (1974). The colors of emotion. *American ethnologist, 1*, 49-63.

De Klerk, V., & Bosch, B. (1997). The sound patterns of English nicknames. *Language Sciences, 19*, 289-301.

Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., & 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., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods, 159*, 139-145.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*, 1191-1203.

Hinton, L., Nichols, J., & Ohala, J. J. (2006). *Sound symbolism*. Cambridge: Cambridge University Press.

Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and visual lightness using Garner's speeded classification task. *Perceptual and Motor Skills, 113*(3), 929-940.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B, 369.* 20130298.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54-65.

Johnson, R. C., Suzuki, N. S., & Olds, W. K. (1964). Phonetic symbolism in an artificial language. *The Journal of Abnormal and Social Psychology, 69*(2), 233.

Karwoski, T. F., & Odbert, H. S. (1938). Color-music. *Psychological Monographs, 50*(2, Whole No. 222).

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.

Kim, H.-W., Nam, H., & Kim, C.-Y. (2016). Audiovisual association between consonants and colors in non-synesthetes. *Journal of Vision*, *16*(12), 469-469.

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

Kwak, Y., & Kim, C.-Y. (2018). Cross-modal correspondence between acoustic feature and shape. *The Korean Journal of Cognitive and Biological Psychology, 30*, 269-275.

Lee Masson, H., Bulthé, J., Op de Beeck, H. P., & Wallraven, C. (2016). Visual and haptic shape processing in the human brain: Unisensory processing, multisensory convergence, and top-down influences. *Cerebral Cortex, 26*, 3402-3412.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, *87*, 173-188.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9*, 316-322.

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*(2), 132-142.

Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America, 115*, 2430-2430.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology, 45*, 53-75.

Nunn, J. A., Gregory, L. J., Brammer, M., Williams, S. C. R., Parslow, D. M., Morgan, M. J., . . . Gray, J. A. (2002). Functional magnetic resonance imaging of synesthesia: Activation of V4/V8 by spoken words. *Nature Neuroscience, 5*, 371-375.

Ohala, J. J. (1986). Phonological evidence for top-down processing in speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 386-397). Hillsdale, NJ: Lawrence Erlbaum.

Paulesu, E., Harrison, J., Baron-Cohen, S., Watson, J. D., Goldstein, L., Heather, J., . . . Frith, C. D. (1995). The physiology of coloured hearing A PET activation study of colour-word synaesthesia. *Brain, 118*, 661-676.

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music-color associations are mediated by emotion. *Proceedings of the National Academy of Sciences, 110*, 8836-8841.

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in nozn-synesthetes. *PLoS One, 4*(5), e5664.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science, 22*, 1419-1421.

Pogacar, R., Plant, E., Rosulek, L. F., & Kouril, M. (2015). Sounds good: Phonetic sound patterns in top brand names. *Marketing Letters, 26*, 549-563.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Rich, A. N., Bradshaw, J. L., & 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., & 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*. Autrans.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333-382.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225-239.

Schloss, I. (1981). Chickens and pickles. *Journal of Advertising Research*, 21, 47-49.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317-1323.

Shin, E. H., & 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.

Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. *Berkeley Linguistic Society, 36*(1), 396-410.

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

Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., & Ković, V. (2015). Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape-sound-meaning congruence in words, and shape-sound congruence in pseudowords. *Brain and Language, 145*, 11-22.

Wrembel, M. (2009). On hearing colours-Cross-modal associations in vowel perception in a non-synaesthetic population. *Poznań Studies in Contemporary Linguistics, 45*, 595-612.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General, 123*, 394-409.

비음이 성문음보다 더 밝고 노랗다: 자음과 색 간의 교차감각 연합 연구

김현웅[1], 남호성[2,3], 김채연[1]

[1]고려대학교 심리학과
[2]고려대학교 영어영문학과
[3]해스킨스 연구소

사람들은 음고와 발화 등의 청각 특징과 시각 특징에 대한 암묵적인 교차감각적 연합을 공유한다. 본 연구는 자음 소리와 색 간의 내재적 연합에 대한 음성학적 특징의 역할을 탐색하였다. 이를 위해 조음 기관의 구강 및 비구강 조음 동작을 모수적으로 조작하여 조음 합성기를 통해 생성한 합성 자음 소리를 제시하였다. 참가자들은 각 소리를 듣고 연상되는 색을 고르도록 지시받았다. 색-연합 과제 결과, 연막 조음 동작으로 정의되는 비음이 다른 소리보다 더 밝고 노랑 계열의 색과 연합되는 것으로 나타났다. 또한, 비유사성 판단 점수로부터 얻은 자음 소리의 지각 공간을 통해서 참가자들이 소리를 자극의 자음 특성에 따라 표상하고 있음을 확인하였다. 따라서 본 연구는 자음의 음성학적 특징과 색 간에 비무선적 연합이 존재함을 시사한다.

주제어: 교차감각 연합, 색, 자음, 조음 합성