한국심리학회지 : 인지 및 생물
The Korean Journal of Cognitive and Biological Psychology
2018, Vol. 30, No. 3, 269-275

Brief Report
http://dx.doi.org/10.22172/cogbio.2018.30.3.005

Cross-modal Correspondence Between Acoustic Feature and Shape*

Yuna Kwak[1], Chai-Youn Kim[1†]

[1]Department of Psychology, Korea University

Our brain tends to associate stimulus features across the senses in a non-random manner. For example, people show consistency in labelling a rounded shape ‘maluma’/‘bouba’ and a spiky shape ‘takete’/‘kiki’. Previous studies have attributed this phenomenon to the correspondence between sound and shape, but without controlling for other potential factors (i.e., linguistic/orthographical factors). The present study examines the role of acoustic aspect per se by manipulating articulatory gestures to generate synthetic speech sounds not confined to a specific language. Participants were asked to choose either a rounded or spiky shape to indicate the shape that better matched each synthetic speech sound. The results demonstrate that shape choice was systematically mapped on to the dimensions manipulated to generate the sounds. These results indicate that acoustic features indeed drive the association between sound and visual shape.

Keywords: cross-modal correspondence, sound, articulatory synthesis, shape

1 차원고접수 18.05.11; 수정원고접수 18.07.16; 최종게재결정 18.07.17

The human brain shows remarkable consistency in systematically associating a certain feature in one sensory modality with a feature in another modality. Dubbed “cross-modal correspondence”, this tendency is documented between a variety of stimulus features (e.g., auditory pitch and visual elevation, auditory pitch and visual brightness, tastes/odors and visual shape; see Spence, 2011 for review). Such intersensory association fascinated researchers nearly a century ago, with Köhler (1929, 1947) demonstrating for the first time that participants share overwhelming preference for labelling a rounded shape ‘maluma’ and a spiky shape ‘takete’ (renamed as the ‘bouba-kiki’ effect by Ramachandran & Hubbard, 2001).

Recent research on the correspondence between non-word and shape tends to focus on the sound of non-words, which is considered to be the driving force behind the non-random association. Studies have examined the relative contribution of consonant and vowel sounds by unpacking which classes of sounds (e.g., in terms of vowel frontness, consonant voicing, consonant place of articulation) are associated with which visual shape (Ahlner & Zlatev, 2010; D’Onofrio, 2014; Fort, Martin, & Peperkamp, 2015; Nielsen & Rendall, 2011, 2013).

What previous studies have not provided, however, is the compelling evidence for the role of acoustic aspect per se. Utilizing sound stimuli of which participants are easily aware as belonging to a specific language leaves much room for the possible influence of factors other

* This study was supported by a faculty research grant from the College of Liberal Arts at Korea University in 2015.

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

than sound, such as linguistic factors (e.g., sound eliciting activation of lexical items that are semantically associated with shape in a specific language). In addition, Cuskley, Simner, & Kirby (2017) suggests that orthography (e.g., the curvature of letters constituting non-words) exerts an influence on the systematic association between non-word and shape. Furthermore, although many studies have replicated the 'maluma-takete'/'bouba-kiki' effect across ages (Maurer, Pathman, & Mondloch, 2006; Ozturk, Krehm, & Vouloumanos, 2013) and cultures (Bremner et al., 2013; Davis, 1961; Tarte, 1974), some studies bring forward mixed results (Fort, Weiß, Martin, & Peperkamp, 2013; Styles & Gawne, 2017), hinting a potential role of factors other than acoustics.

Here we report that sound is the dominant feature mediating the non-random relationship between pseudo-words and visual shape. An articulatory synthesizer was employed to generate vowel-consonant-vowel (VCV) sequences not confined to any existing language, thereby minimizing the influence of aforementioned factors. Participants made a forced choice between a rounded and a spiky shape to indicate the one associated with the VCV sounds, across which only the consonant varied. Dissimilarity rating on all possible pairs of VCV sounds was also conducted, allowing us to reconstruct participants' perceptual space of the stimuli and examine its fit to the physical parameter space. The results of the present study indicate that it is indeed the acoustic aspect that elicits the systematic mapping between non-word and shape.

Methods

Participants

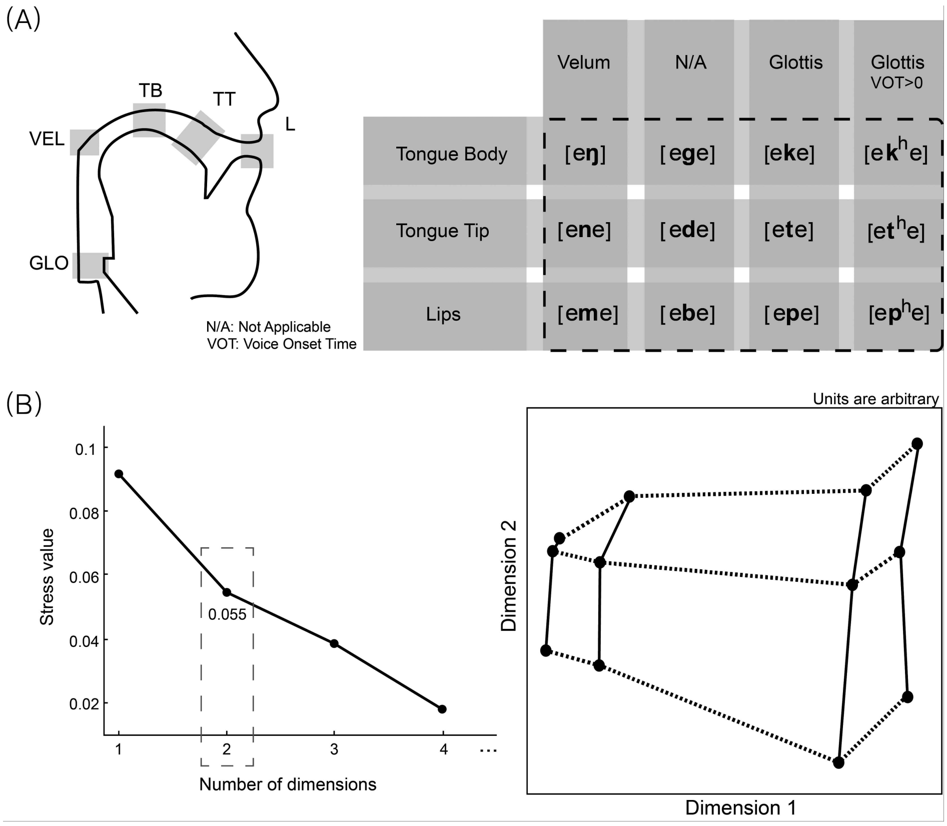
Forty individuals (15 males and 25 females, 19-28 years of age), all of whom gave informed consent approved by Korea University Institutional Review Board, participated in the study. All had normal or corrected-to-normal vision and hearing. Their native language was Korean.

Stimuli

Auditory Stimuli. For the auditory stimuli, Task-Dynamic Application model (TADA; Nam, Goldstein, Saltzman, & Byrd, 2004) was employed to generate synthetic speech-like VCV sounds. VCV sequences were generated instead of CV sequences because according to the subjective reports of participants in our pilot experiments, participants who listened to VCV sequences better perceived the consonants and thus better discriminated the stimuli compared to those who listened to CV sequences.

In TADA, an utterance is represented as a constellation of gestures of five constricting organs: lips, tongue tip, tongue body, velum, and glottis (Nam, Goldstein, Giulivi, Levitt, & Whalen, 2013). Using this model, a set of twelve VCV sequences was synthesized by manipulating the constriction gestures needed to generate consonants (Figure 1A; visit http://vcn.korea.ac.kr/consonants%20file.html to listen to the stimuli). The vowel gesture was kept identical at its rest position to control for the influence of vowel context (Donaldson & Kreft, 2006; Fort et al., 2015). Manipulating the oral constriction gestures (lips, tongue tip, and tongue body) generated consonants that sound similar to [b], [d], and [g], respectively. Then, each type of the oral gestures was coupled with each of the non-oral gestures (velum and glottis) which generated nasal consonants ([m], [n], [ŋ]) and voiceless unaspirated consonants ([p], [t], [k]), respectively. Voiceless aspirated consonants ([p^h], [t^h], [k^h]) were also generated by inserting a temporal delay between the onset of the oral gesture and the glottal gesture. To examine the effect of two factors, oral and non-oral gestures, sounds that did not fit into the 3-by-4 factorial design were excluded. Pitch and duration of the VCV sounds were set to 125 Hz and 600 msec, respectively.

Visual Stimuli. A rounded-spiky shape pair in Figure 2A was presented to participants. They subtended 12.9° × 12.9° on a black background.



	Velum	N/A	Glottis	Glottis VOT>0
Tongue Body	[eŋ]	[ege]	[eke]	[ek^{h}e]
Tongue Tip	[ene]	[ede]	[ete]	[et^{h}e]
Lips	[eme]	[ebe]	[epe]	[ep^{h}e]

Figure 1. (A) Twelve VCV sequences. For the consonants, the five constricting organs (lips [L], tongue tip [TT], tongue body [TB], velum [VEL], glottis [GLO]) were manipulated. The vowel gesture for the flanking vowels was fixed at its rest position. VOT (Voice Onset Time) 〉 0 indicates that the onset of the glottal gesture is temporally delayed with respect to that of the oral gesture. (B) Multidimensional scaling results (MDS) of VCV sounds. Left panel shows stress value as a function of number of dimensions. Right panel shows the two-dimensional space reconstructed from applying MDS to group-averaged dissimilarity matrix. The solid and dashed line connects VCV sounds with the same non-oral and oral constriction gestures respectively.

Apparatus

All visual stimuli were presented on a 19-inch CRT monitor (1024 × 768 resolution, 60 Hz refresh rate; viewing distance 52cm). Participants listened to the auditory stimuli through SRH440 headphones. Experiments were conducted in a quiet, dark room using MATLAB version 9.1 (The Mathworks, Inc., MA) and Psychophysics Toolbox version 3 (Brainard, 1997; Pelli, 1997).

Procedures

Participants participated in the shape matching task after performing the dissimilarity rating task, and this order was maintained across participants to minimize prior exposure to the VCV sounds for the shape matching task.

Dissimilarity Rating. On each trial, a pair of VCV sounds was presented sequentially, and participants had to rate the pair-wise dissimilarity on a scale from 1(same) to 7(different). Participants made the dissimilarity judgement for all possible pairs of 12 VCV sounds, with same sound pairs being presented once and different sound pairs being repeated twice. The total number of trials added up to 144 trials.

Shape Matching. Participants made a two-alternative forced choice judgement on the visual shape that was better associated with the VCV sound presented on each trial. The rounded shape and the spiky shape appeared on each side of the screen, and the position of each shape was randomly selected every trial. The onset of the VCV sound was simultaneous with the onset of the visual display. The experiment consisted of 20 repetitions for each sound, leading to a total of 240 trials.

Results

Dissimilarity Rating

The mean values of the pair-wise dissimilarity ratings were entered into a 12-by-12 matrix for each participant. A group-level dissimilarity matrix was created by averaging the individual matrices, since high correlation (mean r = 0.714) was observed across participants (Ashby, Maddox, & Lee, 1994; Gaiβert, Wallraven, & Bülthoff, 2010). We applied Multidimensional Scaling (MDS) analysis to reconstruct the participants' perceptual space of VCV stimuli using the MATLAB (version 9.2) built-in function mdscale (Lee Masson, Bülthoff, Op De Beeck, & Wallraven, 2016). Considering that two major factors - i.e., oral and non-oral constriction gestures - were manipulated, the perceptual space was reconstructed using two dimensions. The stress value for applying the two-dimensional solution was sufficiently low (stress = 0.055; Cox & Cox, 2001), which indicated a good fit with the data (Figure 1B, left panel).

As illustrated in Figure 1B (right panel), VCV sounds generated by manipulating the same oral or non-oral gesture were located close together along the two dimensions in the reconstructed perceptual space. Therefore, we concluded that the physical parameter space of the synthesized sounds was well recovered in perceptual space, confirming the effectiveness of the physical parameter manipulation.

Shape Matching

The mean percentage of trials in which the spiky shape was chosen for each VCV sound is shown in Figure 2B. The data were analyzed with IBM SPSS Statistics version 23.0. A two-way repeated measures ANOVA revealed main effects of both oral and non-oral constriction gestures on the associated shape, F (2, 78) = 10.403, p 〈 .001, F (1.344, 52.430) = 10.407, p 〈 .001, Greenhouse-Geisser corrected. The interaction effect between the two factors was not significant, F (4.472, 174.406) = 1.871, p = .087, Greenhouse-Geisser corrected.

In an attempt to determine whether each VCV sound was associated with one of the shapes at a statistically significant level, we compared each sound against chance using one-sample t-tests (see Figure 2B for p-values, FDR corrected). Replicating the results of previous studies reporting the "maluma-takete"/"bouba-kiki" effect, consonants sounding similar to [m] and [b] were associated with rounded shape, whereas those generated

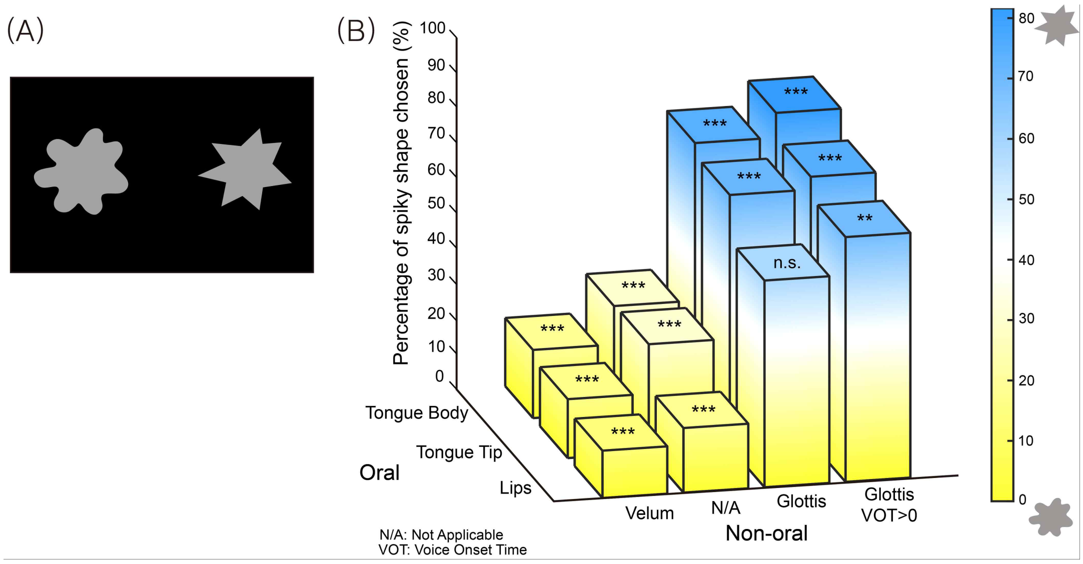


Figure 2. (A) The rounded-spiky shape pair presented to participants. (B) Shape matching results. Y-axis is the percentage of spiky shape choice for 12 VCV sounds. More yellowish color denotes higher percentage of rounded shape choice, whereas more bluish color denotes higher percentage of spiky shape choice. Asterisks indicate the p-values of one-sample t-test, comparing shape matching results to chance level (***p 〈 .001, **p 〈 .01, n.s. = not significant, FDR corrected).

by manipulating the articulatory gestures needed to produce [tʰ] and [kʰ] were associated with spiky shape, significantly above the chance level.

Discussion

Our results indicate that participants show consistency in matching either a rounded or spiky shape to the two factors determining the acoustic properties of consonants - oral and non-oral constriction gestures. With respect to the oral gestures, sounds including lip gestures (e.g., [eme], [ebe]) tended to be associated with rounded shape. In terms of the non-oral gestures, those with glottal gestures (e.g., [eke], [ekʰe]) were associated with spiky shape.

Since the current work focuses on the acoustic feature of consonants, it does not provide direct evidence on how vowel context (i.e., which vowels are presented with consonants; Donaldson & Kreft, 2006) influences sound-shape correspondence, as explored in Fort et al. (2015). However, the current work is meaningful in that we parametrically manipulated articulatory gestures to examine the assumption upon which earlier studies are based - i.e., that such results are due to the association between shape and acoustic feature per se. We observed systematic mapping of shape to sound despite the fact that participants were not fully aware of the dimensions manipulated to generate the speech sounds: some did not recognize the stimuli as speech sounds while others reported to have listened to three to eight types of sounds when asked after the experiment. Thus, it is likely that participants relied heavily on the acoustic features to perform the tasks, minimizing the influence of linguistic factors. Therefore, our results consolidate the role of acoustics in sound-shape correspondence.

On a final note, the current work has implications for sound symbolism, which refers to the non-random mapping between sound and meaning (Hinton, Nichols, & Ohala, 1994; Taitz et al., 2018). Our results suggest that these observations in high cognitive function (i.e., language) can be traced down to the association between distinct sensory dimensions that exists in the human brain.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 298-348.

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.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.

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.

Cox, T. F., & Cox, M. A. (2001). *Multidimensional scaling* (2nd ed.). London, UK: Chapman & Hall.

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

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*, 259-268.

Donaldson, G. S., & Kreft, H. A. (2006). Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users. *Ear and Hearing*, *27*, 658-677.

D’Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the Bouba-Kiki paradigm. *Language and Speech*, *57*, 367-393.

Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are more important than vowels in the Bouba-kiki Effect. *Language and Speech*, *58*, 247-266.

Fort, M., Weiß, A., Martin, A., Peperkamp, S. (2013). Looking for the bouba-kiki effect in pre-lexical infants. in: *Proceedings of the 12th international conference on auditory-visual speech processing*, 71-76.

Gaißert, N., Wallraven, C., & Bülthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *Journal of Vision*, *10*, 1-20.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). *Sound Symbolism*. Cambridge, UK: Cambridge University Press.

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

Köhler, W. (1947). *Gestalt psychology* (2nded.). NewYork: Liveright.

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.

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

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

Nam, H., Goldstein, L. M., 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.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, *65*, 115-124.

Nielsen, A., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, *67*, 153-163.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*, 173-186.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.

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

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971-995.

Styles, S. J., & Gawne, L. (2017). When does Maluma/Takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception*, *8*, 1-17.

Taitz, A., Assaneo, M. F., Elisei, N., Trípodi, M., Cohen, L., Sitt, J. D., & Trevisan, M. A. (2018). The audiovisual structure of onomatopoeias: An intrusion of real-world physics in lexical creation. *PLoS ONE*, *13*, e0193466.

Tarte, R. (1974). Phonetic symbolism in adult native speakers of Czech. *Language and Speech*, *17*, 87-94.

소리의 음성적 특성과 형태 간의 교차양태 관련성

곽유나[1], 김채연[1†]

[1]고려대학교 심리학과

교차양태 관련성이란, 서로 다른 감각 양태에 존재하는 자극 속성들 간의 비무선적인 연합을 의미한다. 예를 들어, 둥근 모양과 뾰족한 모양의 시각 자극을 제시하고 어울리는 이름을 고르도록 한 과제에서 대부분의 참가자들은 둥근 모양을 'maluma'/'bouba', 뾰족한 모양을 'takete'/'kiki'라고 명명하는 경향을 보인다. 이는 시청각 교차 관련성의 대표적인 사례로 소개되지만, 대부분의 선행 연구들은 이러한 연합에 영향을 미칠 수 있는 소리 외의 철자법 등 언어적인 요인을 통제하지 못했다는 한계를 지닌다. 따라서 본 연구에서는 소리의 음성적 특징과 형태 간의 연합에 대해 알아보고자, 조음 기관을 체계적으로 조작하여 특정 언어에 국한되지 않은 합성 소리를 참가자들에게 제시하고 둥근/뾰족한 모양 중 각 소리와 어울리는 모양을 고르게 하였다. 그 결과, 참가자들의 모양 선택은 합성 자음 소리를 생성하기 위해 조작된 차원에 의해 조절되었다. 따라서 본 연구 결과는 소리의 음성적 특징이 명칭과 형태 간의 연합의 중요한 요인임을 제시한다.

주제어: 교차양태 관련성, 소리, 조음 합성, 형태