Multisensory Research

Cross-Modal Correspondence between Speech Sound and Visual Shape Influencing Perceptual Representation of Shape: the Role of Articulation and Pitch

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

[1] Department of Psychology, Korea University, Seoul 02841, Korea

[2] Department of English Language and Literature, Korea University, Seoul 02841, Korea

[3] Haskins Laboratories, New Haven, CT 06511, USA

Received 14 June 2018; accepted 21 October 2019

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

Supplementary material

Supplementary Text

Multidimensional scaling analysis of the dissimilarity ratings of pitch-varying vowel sounds.

1. Methods

1.1. Participants

Thirty-six individuals (23 males, 13 females, 20–29 years of age) from Korea University participated in the experiment after giving informed consent approved by the Korea University Institutional Review Board (KU-IRB-17-85-A-1). All participants had normal hearing. Their native language was Korean.

1.2. Stimuli

See main text, Section 2.1.2 — *Auditory Stimuli*.

1.3. Apparatus

The apparatus is identical to that used in Experiment 1.

1.4. Procedures

Participants were asked to rate the pairwise dissimilarity between all possible pairs of 25 vowel sounds on a seven-point scale. On each trial, two vowel sounds were presented sequentially, and participants responded on a scale from 1 (same) to 7 (different). The same sound pairs were presented once, and different sound pairs were repeated twice, yielding a total of 625 trials. The order of the trials was randomized for each participant.

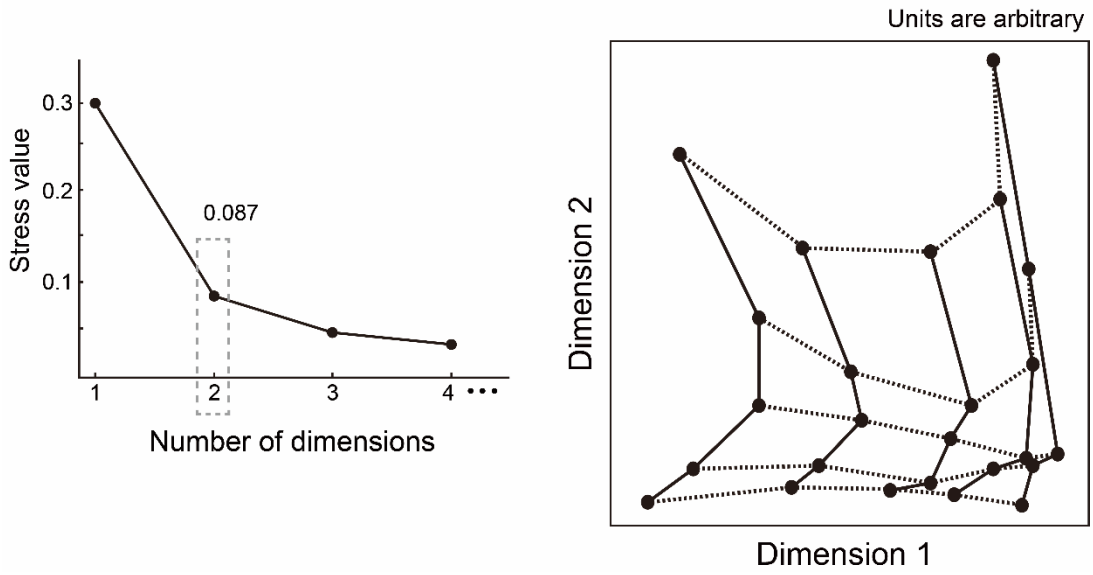

Figure S1. Multidimensional scaling results (MDS) of vowel sounds. The right panel shows the two-dimensional space reconstructed by applying MDS to the group-averaged dissimilarity matrix. The solid and dashed lines connect vowel sounds with the same frontness and height of the tongue body's center, respectively. The left panel demonstrates stress values as a function of number of dimensions. The dashed box indicates the 'statistical elbow', the point beyond which the stress value does not show sharp decrements.

2. Results

For the dissimilarity rating data, mean ratings of sound pairs were used to create a 25 × 25 dissimilarity matrix for each participant. Since individual ratings across participants showed high correlation (mean $r = 0.733$), matrices were averaged across participants to obtain a group dissimilarity matrix (Ashby *et al.*, 1994; Gaißert *et al.*, 2010). Multidimensional scaling (MDS) was applied to the group matrix to reconstruct participants' perceptual space of vowel stimuli using the MATLAB (version 9.2; MathWorks, Natick, MA, USA) built-in function mdscale (Lee Masson *et al.,* 2016). We used non-metric MDS, which is more appropriate for human similarity/dissimilarity data than classical metric MDS (Cooke *et al.*, 2007; Gaißert *et al.,* 2010). The function mdscale also gives as output the stress value for each number of dimensions (Supplementary Fig. S1, left panel), which is crucial for determining the number of dimensions necessary to explain the data.

Since the physical parameter space was defined according to the two factors frontness and height, we applied MDS for two-dimensional solutions (stress value = 0.087). A good fit is characterized by a sufficiently low stress value and a clear 'statistical elbow', the point beyond which the stress value does not show sharp decrements (Cox and Cox, 2001). Based on these criteria, we concluded that the data can be adequately explained with two dimensions, corresponding to frontness and height.

Supplementary Fig. S1 (right panel) shows the two-dimensional perceptual space reconstructed to determine how well the physical parameter space of synthesized vowels is recovered in participants' perceptual space. Vowels sharing the same frontness of the tongue body's position tended to be grouped together with respect to one dimension, and vowels sharing the same height of the tongue body's position were grouped together with respect to the other dimension. That the topology of the physical parameter space was well preserved in participants' perceptual space led to the conclusion that participants could capture the

articulatory nature of the synthesized sounds although such information was not provided. This confirmed the effectiveness of the physical parameter manipulation.

References

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

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

Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling*, 2nd ed. Chapman and Hall, London, UK.

Gaißert, N., Wallraven, C. and Bülthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans, *J. Vision* **10**, 2. doi. 10.1167/10.11.2

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