Cross-modal Effect of an Aurally Presented Phoneme on the Judgment of a Visual Object's Size

Sachi ITAGAKI* and Kohta I. KOBAYASI*

(Received October 19, 2016)

Sound symbolism is the idea that sound itself makes an impression and is the basis for learning language. Stimuli in most of the previous studies have been presented visually using letters, and subjects were required to read the stimulus either silently or aloud, and then presented directly to the sound. As these research frameworks involved many unquantifiable and confounding factors, the neural basis of sound symbolism has not been fully investigated. The purpose of this experiment was to determine whether sound symbolism can be observed even when a sound stimulus is presented aurally and to establish an experimental framework applicable to cognitive neuroscience (e.g., brain imaging research). We examined sound symbolism as subject judged the magnitude of a visual stimulus. Phonemes were presented aurally simultaneously with a target visual stimulus, and subjects indicated whether the size of the visual stimulus was smaller or larger than a standard. Result showed that reaction time during the congruent trial was shorter than that during the incongruent trial, and the difference was systematically related with the difficulty of judging the size of the visual stimulus. Our data confirm that sound symbolism occurs even when a sound stimulus is presented aurally, and suggest that sound symbolism affects the judgment of visual size. Future research using our behavioral framework, will reveal the brain regions involved in sound symbolism.

Key words : sound symbolism, fMRI, Bouba/Kiki effet

1. Introduction

In natural language, no particular regularity is assumed to exist in the in correspondence between an object and the acoustic property of the word that describes that object [1]. However, the association between meanings and sounds (i.e., phonemes), a phenomenon termed sound symbolism or phonetic symbolism has been confirmed by previous studies [2,3]. The idea is that a sound itself makes a particular impression, which then serves as the psychological basis for the word-meaning association.

The most famous example is the so-called Bouba/Kiki effect [4]. To demonstrate this effect, a participant names spiky and round shapes using only the sounds "Bouba" and "Kiki". In previous research, the round shapes were named "Bouba" by approximately 95% of the subjects and the spiky shapes were named "Kiki". Thus, for most people, "Bouba" and "round" comprised an associated pair, and "Kiki" and "spiky"

* Graduate School of Life and Medical Sciences, Doshisha University, Kyo-tanabe, Kyoto, 610-0321
Telephone; +81-774-65-6439, E-mail; dmp1006@mail4.doshisha.ac.jp

another. This effect is observed regardless of subject age or native language. Similarly, most sound symbolism studies have used a relatively naturalistic approach [5-17]. In one study, subjects were asked to describe their impressions of an artificial word, after they see the word in an attempt to illustrate the effect of each phoneme [5]. These studies have successfully revealed associations between particular acoustic features and the impressions made by those features, but the neural basis of sound symbolism remains largely unknown. One of the major obstacles has been the experimental framework that most of these studies employed. Because they used a naturalistic approach that involved unquantifiable and confounding factors, previous frameworks were not suitable for identifying brain regions specifically related to sound symbolism. In this study, we measured the effect of sound symbolism quantitatively and tried to establish an experimental framework applicable to cognitive neuroscience. We focused on judgments of the size of a visual stimulus, and examined the association between the visual stimulus and the phoneme, the smallest unit of a word. Our results will shed light on the neural basis supporting the connection between sound and meaning.

2. Materials and Methods

2.1 *Subjects*

Seven (four females and three males; age, 21–24 years) participated in a behavioral experiment after providing informed written consent. All subjects were right-handed and native Japanese speakers. None of the participants had any knowledge of sound symbolism or the experiment.

2.2 *Experimental environment*

The experiment was conducted in a soundproof room. Sound stimuli were presented to subjects via headphones (ATH-A900; Audio-Technica, Inc., Tokyo, Japan) through a D-A converter (OCTA-CAPTURE; Roland Co., Osaka, Japan) from a PC, and visual stimuli were presented by a LCD display (VL-150VA; Fujitsu, Tokyo, Japan). The subject was seated 50 cm from the display, and used a standard keyboard to respond to the task with their left hand. The PC and D-A converter were located outside the soundproof room to prevent noise being heard by the subjects. We used stimulus presentation software (Presentation; Neurobehavioral Systems, Inc., Albany, CA, USA).

2.3 *Stimuli*

2.3.1 *Visual stimulus*

This experiment examined sound symbolism using judgments of the size of a visual stimulus. As the visual stimulus, different-sized gray circles that looked like doughnuts were presented on a LCD display with a black background. The standard stimulus had an outer circle of 300 pixels and an inner circle of 280 pixels. The target size was either smaller or larger than the standard by ±5%, ±10%, or ±20% in diameter. The overall visual stimulus had one size of standard and six sizes of targets. Each stimulus was lit twice for 200 ms, with a 120 ms interstimulus interval (ISI; total, 520 ms) between stimuli. The screen size was 1024 × 768 pixels. A red cross (34 pixels) was always presented as a fixation point at the center of the screen.

2.3.2 *Sound stimulus*

The sound stimuli were “bobo” and “pipi”. A publicly available sound dataset (FW03; NTT Communication Science Laboratories, Kanagawa, Japan) was used to create the sounds. All sounds were recorded at a sampling frequency of 48 kHz and quantization of 16 bits. The single syllable utterances “bo” and “pi” were spoken by a male speaker, and these were duplicated to produce the sound stimuli “bobo” and “pipi”. The sound duration was 520 ms, and stimulus amplitude was 80 dB SPL. The sound stimulus was presented synchronously with the visual stimulus.

2.4 *Experimental framework*

Subjects were required to determine the difference in visual size between the standard and the target stimulus. Each trial began with a 3,000 ms rest period.

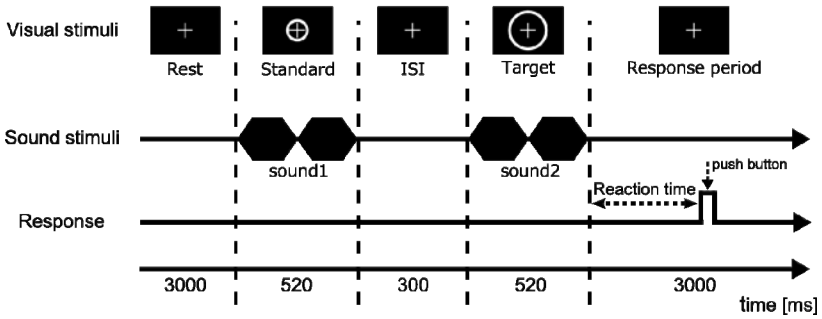

Fig. 1. Flow chart for the stimulus presentation and behavioral response.
Each trial started with a rest period (3,000 ms). Then, the standard stimulus (520 ms) was presented, followed by the target stimulus (520 ms), with a 300-ms interstimulus interval (ISI). Subjects were instructed to respond to the task during the response period (3,000 ms) with their left hand using a keyboard. A red cross was always presented as a fixation point.

Then, the 520 ms standard stimulus was presented, followed by the 300 ms ISI. Next, the 520 ms target stimulus was presented, followed by a response period of 3,000 ms (Figs. 1 and 2). The screen was black during the rest and response periods. After the subject responded to the task by pressing a button, the next trial began automatically. If the sound (sound 2) presented with the target was identical to the sound (sound 1) presented with the standard, the subject was instructed to press the middle button with the middle finger regardless of the magnitude of any difference between the visual stimuli (control task). If the sound stimulus presented with the target was different from that presented with the standard, the subject was instructed to respond differently depending on whether the target circle was smaller or larger than standard, using the index finger and ring finger to press the left or right button, respectively (comparison task). The button (right or left) that represented particular answer (larger or smaller) was changed between subjects. There were six visual (one standard × six targets) and four sound (two Sound 1 × two Sound 2) possible combination for each stimulus. Thus, there were 24 stimulus combinations overall. The entire stimulus set was randomized to create one block (24 trials), and each session consisted of three blocks. In the behavioral experiment, each subject completed three sessions (216 trials) with a 5 min break between sessions.

2.5 *Analysis*

We analyzed the reaction time (RT) on the comparison tasks. If a subject made a mistake in size judgment, the RTs for those trials were excluded from

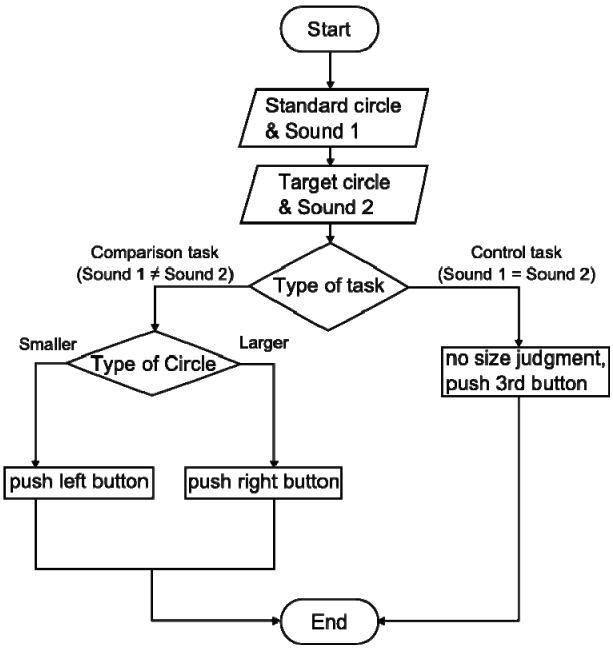

Fig. 2. Flowchart illustrating the trial process.
Subjects were instructed to push the middle button to respond to the control task, and to push the left or right button to respond to the comparison task. The left and right key assignments were changed between subjects.

the analysis. Previous studies have indicated that "p" and "i" produce smaller impressions, whereas "b" and "o" produce larger impressions [7]. Therefore, the stimuli "bobo" and "pipi" should produce larger and smaller impressions, respectively. We defined the congruent and incongruent conditions as follows and analyzed the data accordingly. The congruent condition occurred when the target stimulus was consistent with the impression made by the sound (i.e., the larger target was presented with "bobo" or the smaller with "pipi"). The incongruent condition occurred when the target stimulus was inconsistent with the impression of the sound (i.e., the

larger target was presented with “pipi” or the smaller one with “bobo”).

3. Results

All subjects performed well on the comparison task, the mean correct response rate was 96.8 % (732 / 756). A larger difference in target size yielded a higher correct answer rate (Table 1). The differences in RT between congruent and incongruent conditions for the various target size combinations are depicted in Fig. 3. The RT Z-score was calculated for each subject (Table 2). RT became shorter as the difference in target size increased from ±5 to ±20%. The mean RT under the incongruent condition was longer than that under the congruent condition for all target size conditions; the difference for the ±5% and ±20% target size difference was not statistically significant, the difference for the ±10% target size difference was statistically significant ($t = -2.27$, $n = 14$, $p < 0.05$).

Table 1. Mean correct answer rate for each subject.

Subject	Correct answer rate [%]					
	±5%		±10%		±20%	
	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Sub1	83.3	88.9	100.0	100.0	100.0	100.0
Sub2	100.0	94.4	100.0	100.0	100.0	100.0
Sub3	100.0	83.3	88.9	94.4	94.4	100.0
Sub4	88.9	88.9	100.0	100.0	100.0	100.0
Sub5	94.4	77.8	100.0	100.0	100.0	100.0
Sub6	100.0	94.4	100.0	100.0	100.0	100.0
Sub7	94.4	100.0	100.0	100.0	100.0	100.0

Table 2. Mean reaction time for each subject.

Subject	Reaction time [ms]					
	±5%		±10%		±20%	
	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Sub1	1579.9	1489.1	1384.3	1348.1	1250.2	1207.4
Sub2	859.0	918.4	808.3	840.3	769.0	723.1
Sub3	1282.9	1388.0	1422.6	1420.1	1353.4	1477.7
Sub4	1447.1	1296.1	1171.8	1227.0	1105.7	1206.6
Sub5	993.9	1063.1	894.0	966.6	785.1	930.0
Sub6	1132.9	1135.6	1131.0	1067.0	996.8	1041.2
Sub7	1388.8	1327.3	966.1	1155.9	930.1	884.0

4. Discussion

Subjects took longer to push the key as the difference in size between the standard and target

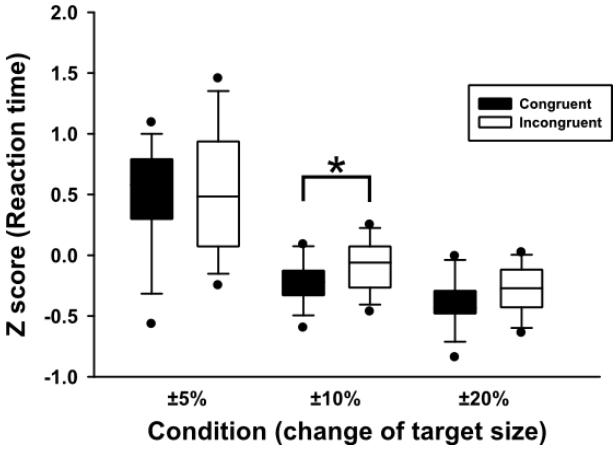

Fig. 3. Difference in reaction time between the congruent and incongruent conditions under different target sizes. The reaction time Z-scores under each condition are depicted in the box plots (outliers: 5% and 95%; bar: 10% and 90%; box: 25%, 50%, and 75%; white or black lines: median). Black boxes represent the congruent condition, and white boxes represent the incongruent condition. The difference at a ±10% target difference was significant ($t = -2.265$, $n = 14$, $p < 0.05$).

decreased, indicating that our framework was suitable for quantifying the difficulty of the cognitive task. Several studies have shown that the phonemes “b”, “d”, “g” and “o” make large impressions, whereas “p”, “t”, “k” and “i” make smaller ones [7]. We defined the congruent and incongruent conditions following this idea. The efficacy of sound symbolism differed by visual target size. Only the ±10% condition yielded a significant difference. Considering that the difference in target size affected the RT more than the congruent-incongruent difference did (Fig. 3), it is not surprising that the effect of sound symbolism was not prominent under some of the target conditions. A discussion of exactly how the difference came about is beyond the scope this paper. However, we propose that when the task was too easy (i.e., ±20% target change), sound symbolism did not produce a measurable conflict with the task, whereas RT inevitably lengthened when the task was too difficult (i.e., ±5% target change), and the prolongation may have masked the effect of sound symbolism. In conclusion, sound symbolism affected the judgment of visual size even when the stimulus was presented aurally. As our experimental framework was

composed of a relatively simple stimulus and task, this scheme is suitable for quantifying sound symbolism and applicable to cognitive neuroscience research.

References

1) F. de Saussure, *Course in general linguistics*, (McGraw-Hill Book Co., New York, 1966).

2) W. Köhler, *Gestalt psychology 2nd ed*., (Liveright, New York, 1947).

3) J. Auracher, S. Albers, Y. Zhai, G. Gareeva, and T. Stavniychuk, "P Is for Happiness, N Is for Sadness: Universals in Sound Iconicity to Detect Emotions in Poetry", *Discourse Processes*, **48**, 1-25 (2011).

4) V. S. Ramachandran, and E. M. Hubbard, "Synesthesia – A Window into Perception, Thought and Language", J. *Consciousness Stud*., **8**, 3-34 (2001).

5) S. Hirata, J. Ukita, and S. Kita, "Compatibility between Pronunciation of Voiced / Voiceless Consonants and Brightness of Visual Stimuli", *Cognitive Studies*, **18[**3], 470-476 (2011).

6) A. Gallace, and C. Spence, "Multiseneory Synesthetic Interactions in the Speeded Classification of Visual Size", *Percept. and Psychophys.*, **68**[7], 1191-1203 (2006).

7) S. S. Newman, "Further Experiments in Sound Symbolism", *Am. J. Psychol.*, **45**, 53-75 (1933).

8) J. Kanero, M. Imai, J. Okuda, H. Okada, and T. Matsuda, "How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words", *Plos ONE*, **9**, e97905 (2014).

9) E. Sapir, "A Study in Phonetic Symbolism", *J. Exp. Psychol.*, **12**, 225-239 (1929).

10) L. C. Nygaard, A. E. Cook, and L. L. Namy, "Sound to Meaning Correspondences Facilitate Word Learning", *Cognition*, **112**, 181-186 (2009).

11) V. U. Ludwig, I. Adachi, and T. Matsuzawa, "Visuoauditory Mappings between High Luminance and High Pitch are Shared by Chimpanzees (Pan troglodytes) and Humans", *Proc. Natl. Acad. Sci.*, **108**, 20661-20665 (2011).

12) C. V. Parise, and C. Spence, "Audiovisual Crossmodal Correspondences and Sound Symbolism: A Study Using the Implicit Association Test", *Exp. Brain Res.*, **220**, 319-333 (2012).

13) C. L. E. Paffen, M. J. V. der Smagt, and T. C. W. Nijboer, "Cross-modal, Bidirectional Priming in Grapheme-Color Synesthesia", *Conscious. Cogn.*, **33**, 325-333 (2015).

14) T. Oyama, and J. Haga, "Common Factors between Figural and Phonetic Symbolism", *Psychologia*, **6**, 131-144 (1963).

15) U. Noppeney, O. Josephs, Julia Hocking, C. J. Price, and K. J. Friston, "The Effect of Prior Visual Information on Recognition of Speech and Sounds", *Cereb. Cortex*, **18**, 598-609 (2008).

16) D. Maurer, T. Pathman, and C. J. Mondloch, "The Shape of Boubas: Sound-Shape Correspondences in Toddlers and Adults", *Developmental Sci.*, **9**[3], 316-322 (2006).

17) K. Shinohara, N. Yamauchi, S. Kawahara, and H. Tanaka, "Takete and Maluma in Action: A Cross-Modal Relationship between Gestures and Sounds", *Plos ONE*, **11**, e0163525 (2016).