Perception of Sound Symbolism in Mimetic Stimuli: The Voicing Contrast in Japanese and English*

Kotoko Nakata
The University of Kansas

Abstract
The current study investigated sound symbolism in Japanese mimetic stimuli. We examined whether the voicing contrast in consonants (/d, g, z/ vs. /t, k, s/) affects perception both in Japanese native speakers and in English native speakers. Stimuli were evaluated on 4 different dimensions: size (big-small), shape (round-spiky), and two evaluative dimensions (good-bad, graceful-clumsy). The voicing on the stimuli was manipulated, creating a continuum from voiced to voiceless endpoints in order to examine the categorical nature of the perception. The current study found that both Japanese and English speakers tended to associate voiced sounds with largeness, badness, and clumsiness and voiceless sounds with smallness, goodness, and gracefulness. However, for the shape dimension, English speakers only tended to associate voiced stop consonants with roundness and voiceless stops with spikiness. The present results show systematic sound symbolic relationships in terms of voicing for Japanese and English.

Keywords
Sound Symbolism, Japanese, Mimetics, Consonantal voicing

* I am deeply thankful to my committee chair, Dr. Joan Sereno and committee members, Dr. Allard Jongman and Dr. Utako Minai whose encouragement, guidance and support from the initial to the final level enabled me to complete the study.

1. Introduction

Sound symbolism is a concept in which the sound of a word and the meaning of the word are systematically related. Although it is generally acknowledged that the relation between sound and meaning is arbitrary (de Saussure, 1915), a number of studies have been conducted to investigate this relationship in many languages since the 1920s. Some aspects of sound symbolism have been found to be language-specific and some to be shared across different languages.

The current paper first examines whether the voicing contrast affects two types of sound symbolism, size symbolism (big-small dimensions) and shape symbolism (round-spiky dimension) by conducting a perception experiment with native speakers of Japanese and native speakers of English who have no knowledge of Japanese. Evaluative sound symbolism (good-bad dimension and graceful-clumsy dimension) was also examined in similar manner.

2. Sound Symbolism

Sapir (1929) found that given two non-words "mil" and "mal" and two tables of different size, English speakers tend to associate the vowel /i/ in "mil" with a small table and the vowel /a/ in "mal" with a large table. This size symbolism has been found in other languages such as Japanese (Hamano, 1998), Chinese, and Thai (Huang et al. 1969). Such size symbolism has been found not only in vowels but also in consonants. Hamano (1998) found size symbolism in voicing of initial obstruents in Japanese mimetic words, with voiceless obstruents /p, t, s, k/ indicating smallness/lightness and voiced obstruents /b, d, z, g/ indicating largeness/heaviness.

Shinohara and Kawahara (2016) tested size symbolism by investigating how the images of size are affected by phonetic factors (the height and the backness of vowels, and voicing in obstruents) in four languages: Chinese, English, Japanese, and Korean. They found size symbolism held across languages and the three phonetic factors do affect sound symbolism: (i) low vowels are associated with larger images than high vowels, (ii) back vowels are associated with larger images than front vowels, and (iii) voiced obstruents are associated with larger images than voiceless images. They provided articulatory/acoustic explanations to explain size symbolism for each of the three phonetic factors.

Koher (1929) found shape symbolism when subjects were presented with two non-words, "baluma" and "takete", and two types of shapes (a curvy round shape and a spiky angular shape), one tends to associate "baluma" with the curvy round shape and "takete" with the spiky angular shape. Kohler (1929) and Westbury (2005) found that English speakers tend to associate continuant consonants with roundness and stop consonants with spikiness. Aveyard (2012) found that non-plosive consonants tend to represent roundness and plosive consonants tend to represent spikiness. However, the stimuli used in these experiments may have had other differences other than the stop/continuant contrast that have been found to affect sound symbolism.

3. Sound Symbolism in Japanese mimetics

Unlike Indo-European languages, which have few mimetic words, Japanese has the second largest number of mimetic words with over 5000 mimetic words (BMFT Publisher Japan 2012). Mimetic words are commonly used in daily life providing speakers with rich means of expression that reveal subtle sensitivity. Japanese mimetic words consistently are one of word classes that are found to be sound symbolic because of the specific semantic information and iconic images they carry and their frequent use by Japanese speakers. According to studies on Japanese mimetic words such as Hamano (1998) and Ivanova (2006), particular phonemes are systematically used to express certain meanings. There is some agreement in their findings, but not all are congruent. Some sound sybolism such as voicing contrast in consonants (Hamano, 1998) are found to be a sound symbolic in other languages such as English and Chinese (Shinohara & Kawahara, 2016).

Iwasaki, Vinson, and Vigliocco (2007) examined sound symbolism in Japanese sound-imitating words (giongo) and mimetic words (gitaigo). They investigated whether speakers of other languages are able to access the meaning of the two types of Japanese onomatopoeic words from their sounds. They asked native English speakers who had no prior knowledge of Japanese to listen to real Japanese mimetic words and rate each word's meaning on semantic differential scales. They found that English and Japanese speakers' responses to onomatopoeic words correlated for a number of words and different semantic dimensions.

The correlation is more significant for audio manner mimetic words such as 'laughing' words than for visual manner mimetic words such as 'walking' words.

They claimed that sound symbolism such as big-small, masculine-feminine, noisy-quiet, wet-dry, and soft-hard symbolism are shared across languages but some symbolism such as beautiful-ugly, graceful-clumsy, and good-bad are demonstrated only by Japanese speakers, which suggested that some sound symbolism is language-specific.

4. Current Study

The current study investigated sound symbolism in Japanese using Japanese mimetic stimuli. In this study, voicing of consonants was examined and vowel influence was controlled. We examined whether the voicing contrast in consonants (/d, g, z/ vs. /t, k, s/) affects the perception in both Japanese native speakers and English native speakers (who had no knowledge of Japanese). Two additional manipulations were also included. First, stimuli were evaluated on 4 different dimensions including both size and shape dimensions as well as evaluative dimensions, in order to examine the generality of the sound symbolism. Second, voicing was manipulated, creating a continuum from voiced to voiceless endpoints, in order to examine the categorical nature of the perception.

The stimuli were 3 pairs of novel Japanese mimetic non-words that contrast in the voicing of the first and third consonants (deze deze & tese tese, gede gede & kete kete, zege zege &

seke seke). Unlike Iwasaki et al. (2007) in which they used real mimetic words as stimuli, we used mimetic non-words in order to avoid associations in the real words. In order to put focus on the voicing contrast in the consonants, we controlled vowels contrast by using only one vowel /e/ in the stimuli. Previous research showed that the height and the frontness of vowels influenced how the sound is perceived (Shinohara & Kawahara, 2016). We used the vowel /e/ since it is the only vowel among Japanese vowels (/i, a, e, o, u/) that is neither extremely high or low nor front or back.

For each of the 3 pairs of mimetic non-words, a voicing continuum of 4 members was created. Each pair had a voiced member (deze deze), an ambiguous voiced member, an ambiguous voiceless member, and a voiceless member (tese tese). These voicing continua were created in order to investigate the nature of the transition from voiced sounds to voiceless sounds.

As we introduced earlier, the stop vs. continuant contrast is found to affect the perception of shapes. The current study further examined whether the manner of articulation affects the perception of shape among Japanese speakers and English speakers.

We expected both Japanese and English speakers to associate voiced sounds with roundness and voiceless sounds with spikiness. In addition to the voicing contrast, we also examined whether having stops or fricatives would affect perceived shape of an image. Since previous studies found perceptional distinction between stops and continuants on the round-spiky dimension in English speakers, we expected English participants to associate stops with spikiness. There were no previous studies that examined shape symbolism in fricatives, therefore the current study explored whether fricatives would play any role in perceived shape of an image.

The other two dimensions, good-bad and graceful-clumsy, were chosen to test the finding by Iwasaki et al. (2007) that sound symbolism in evaluative dimensions tend to be language-specific. Both good-bad and graceful-clumsy were found to be language-specific in Iwasaki et al. (2007); Japanese speakers tended to associate voiced sounds with badness and clumsiness whereas English speakers tended to do so with voiceless sounds. The current study examined whether the voicing affects evaluative perception, and whether Japanese and English speakers share the perceptional tendency towards goodness and gracefulness. We expected Japanese speakers to associate voiced sounds with badness and clumsiness, and voiceless sounds with goodness and gracefulness, while we expected English speakers to associate voiced sounds with goodness and gracefulness, and voiceless sounds with badness and clumsiness.

The current study investigated three research questions. First, does the voicing contrast in consonants affect the perception of Japanese and English speakers? If so, on what semantic (big-small, round-spiky) or evaluative (good-bad, graceful-clumsy) dimensions does the voicing influence each language group? Second, does stop (/d, g, t, k/) and fricative (/z, s/) affect perceived shape of an image? Third, does manipulation of voicing continuum from

voiced to voiceless affect perception on different dimensions? By conducting the current perception experiment, we investigated how speakers of different languages (Japanese, English) access meaning from their sounds.

4.1 Participants

Participants were 12 native speakers of Japanese and 16 native speakers of English who had no prior knowledge of Japanese.

The Japanese speakers (N=12, 7 females, 5 males, average age: 33) were recruited either at University of Kansas or in Fukuoka, Japan. 3 of the participants lived in the U.S. for 2 years, 5 of them lived in the U.S.A. for 6 months, and 4 of them never lived outside Japan.

All the native speakers of English (N=16, 12 females, 4 males, average age: 24) were students recruited at University of Kansas. 3 of them were English monolinguals, 12 of them were either beginner or intermediate learners of another language (Spanish, French, and Italian). None had any knowledge of the Japanese language.

4.2 Stimuli

The stimuli pronounced by a female Japanese speaker consisted of 3 voiced-voiceless pairs that are novel Japanese mimetic non-words in C1VC2V-C3VC4V form (deze deze & tese tese; gede gede & kete kete; zege zege & seke seke). None of those words exist in a Japanese dictionary of onomatopoeias (Atouda & Hoshino, 2009).

Among the three voiced-voiceless pairs, the “gede-kete” pair consisted of stops (/g, d, k, t/) and the vowel /e/, whereas the other two pairs, “deze-tese” and “zege-seke” consisted of stops, fricatives (/z, s/) and the vowel /e/.

For each of the 3 voiced-voiceless stimulus pairs, an ambiguous continuum was created for each pair by adjusting the stop and fricative duration of the first and the third consonants (e.g. /d/ in deze deze, /t/ in tese tese, /g/ in gede gede, /k/ in kete kete, /z/ in zege zege, /s/ in seke seke) usgin Praat (Boersma and Weenink, 2007). One of the two ambiguous members is perceptually closer to the voiced stimulus (ambiguous voiced) and the other is closer to the voiceless stimulus (ambiguous voiceless). Therefore, a 4-member continuum for each of the 3 voiced-voiceless pairs was tested.

All the stimulus words had the same accentuation pattern of LHHH (low-high-high-high) which is the typical pattern for mimetic adjectives in CVCV-CVCV form.

4.3 Procedure

The participants first answered a questionnaire on their language background. Then the participants were instructed to sit in front of a computer screen, wear headphones, and were provided with 4 answer sheets (one for each dimension).

In the instructions, Japanese speakers were told that they would listen to some Japanese non-words and English speakers were told that they would hear some words of an unknown language. Both participants were directed to listen to a word by clicking a speaker icon shown on the screen and they were allowed to listen to the stimuli as many times as they wanted to.

The participants listened to and rated each word along 4 dimensions. For each dimension, a four-point scale was created and each point was indicated with the degree as in “big”, “relatively big”, “relatively small”, and “small” for the big-small dimension. The participants circled one of the four points on a scale. 12 (3 pairs x 4 continuum members) stimuli were presented randomly in each of the 4 dimensions. Each stimulus word was embedded in a Power Point slide. In order to gain reliability, there were two successive repeated sessions per participant (12 stimuli x 4 dimensions x 2 sessions). The order of the stimuli in both sessions was randomized for each dimension and for each participant. The participants were allowed to take a break. The whole procedure took about 30 minutes for the Japanese and 45 minutes for the English.

The rating results for all 3 voiced-voiceless stimulus pairs on four-point scales were converted to 1 to 4 to express polarity with positive values corresponding to the label of the scale (e.g. 1: big, 2: relatively big, 3: relatively small, 4: small). The converted numbers were averaged across participants within each subject group. In each subject group, there were averaged voiced results, ambiguous voiced results, ambiguous voiceless results, and voiceless results for each of the 4 dimensions.

5. Results

A repeated measures 2 x 4 ANOVA (language x voicing) was conducted for each dimension (big-small, round-spiky, good-bad, graceful-clumsy). Language (Japanese and English) was a between subjects factor and voicing (voiced, ambiguous voiced, ambiguous voiceless, voiceless) was a within subjects factor.

5.1 Big-Small dimension (size symbolism)

The overall mean rating was 2.43 for the 12 Japanese participants and 2.40 for the 16 English participants. There was no significant difference across language background for the big-small dimension; $F(1,26)=0.17$, $p=.684$. The ratings for the Japanese and the English participants are shown in figure 1 for big-small categorization.

There was a significant main effect of voicing; $F(3,78)=57.79$, $p<.001$. Planned comparison tests revealed that there were significant voicing effects between all voiced and voiceless comparisons.

The significant voicing x language interaction showed a strong trend; $F(3,78)=2.30$, $p=.084$. The English participants showed a greater effect of voicing ($p=.095$) than the Japanese participants.

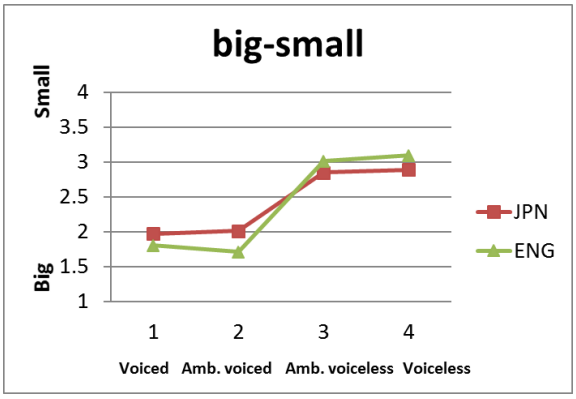

Figure 1: Mean rating for Japanese and English speakers for big-small dimension

Overall, for the big-small dimension, both Japanese speakers and English speakers showed a significant voicing effect, with the English speakers show a slight stronger effect. These results suggest that both Japanese speakers and English speakers associated both the voiced and the ambiguous voiced stimuli with bigness, and they associated the voiceless and ambiguous voiceless stimuli with smallness.

5.2 Round-Spiky dimension (shape symbolism)

The overall mean rating was 2.87 for the 12 Japanese participants and 2.45 for the 16 English participants. There was a significant difference across language background for round-spiky dimension; $F(1,26)=24.11$, $p<.001$, with the Japanese participants having overall more spiky responses than the English participants. The ratings for the Japanese and the English participants are shown in figure 2 for round-spiky categorization.

Overall, there was no significant main effect of voicing; $F(3,78)=0.024$, $p=.995$. There was no significant difference in voicing was observed between the voiced and voiceless stimuli.

Only the English participants showed a significant main effect of voicing in the only-stop stimulus “gede-kete” pair. The English participants associated voiced stops (/g, d/) with roundness and voiceless stops (/k, t/) with spikiness. There was a significant difference observed between the two ambiguous stimuli ($p=.076$) as well as between the two endpoints ($p=.001$). No such voicing effect was observed either in the two mixed (stop & fricative) pairs “deze-tese” or “zege-seke”.

There was a trend for voicing x language interaction; $F(3,78)=2.36$, $p=.078$, with Japanese showing more spiky responses especially for voiced stimuli.

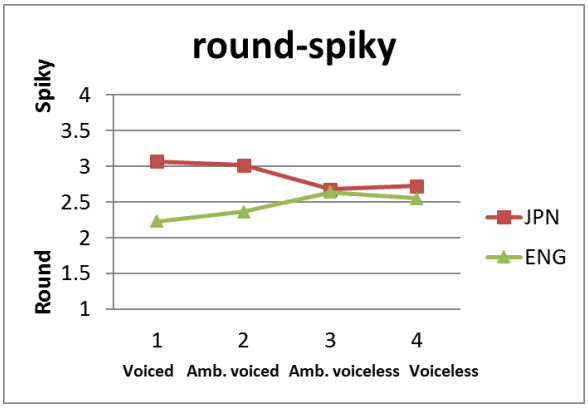

Figure 2: Mean rating for Japanese and English speakers for round-spiky dimension

To sum up, for the round-spiky dimension, the Japanese speakers did not show a significant voicing effect in any of the three stimulus pairs, whereas the English speakers showed a significant voicing effect only in the only-stop stimulus pair "gede-kete". The English participants associated voiced stops with roundness and voiceless stops with spikiness. No such tendencies were observed when there were both stops and fricatives in a stimulus. While the Japanese participants showed a slightly different pattern across the voicing continua, none of the differences was significant.

5.3 Good-Bad dimension

The overall mean rating was 2.76 for the 12 Japanese participants and 2.53 for the 16 English participants. There was a significant difference across language background for good-bad dimension; $F(1,26)=9.36$, $p<.01$, with the Japanese participants having more bad responses than the English participants. The ratings for the Japanese and the English participants are shown in figure 3 for good-bad categorization.

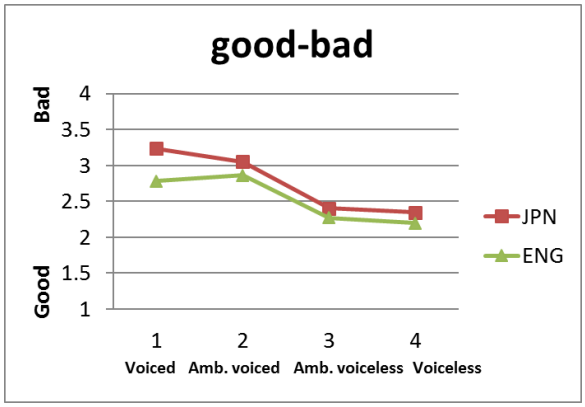

Figure 3: Mean rating for Japanese and English speakers for good-bad dimension

There was a significant main effect of voicing; F(3,78)=16.23, p<.001. Planned comparison tests revealed that significant voicing effects were observed between the voiced and voiceless stimuli while there was no significant voicing difference between the voiced and the ambiguous voiced (p=.609) and between the ambiguous voiceless and voiceless (p=.305) stimuli.

There was no overall significant voicing x language interaction; F (3, 78)= 0.60, p=0.620.

Overall, for the good-bad dimensions, both Japanese speakers and English speakers showed a significant voicing effect. The results suggest that both Japanese speakers and English speakers associated the voiced and ambiguous voiced stimuli with badness, and they associated the voiceless and ambiguous voiceless stimuli with goodness.

5.4 Graceful-Clumsy dimension

The overall mean rating was 2.91 for the 12 Japanese participants and 2.52 for the 16 English participants. There was a significant difference across language background for graceful-clumsy dimension; F(1,26)=13.398, p<.05, with the Japanese participants overall using more clumsy responses than the English participants. The ratings for the Japanese and the English participants are shown in figure 4 for graceful-clumsy categorization.

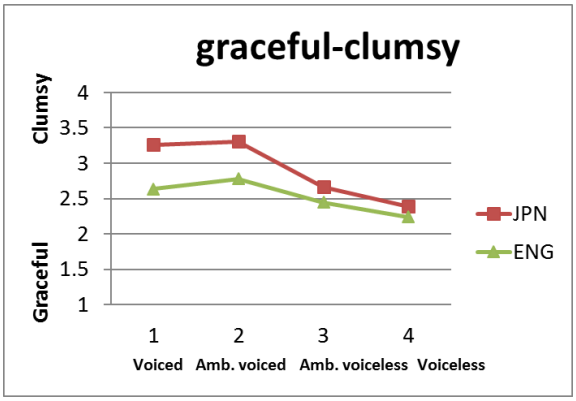

Figure 4: Mean rating for Japanese and English speakers for graceful-clumsy dimension

There was a significant main effect of voicing; F(3,78)=10.78, p<.001. Planned comparison tests revealed that significant voicing effects were observed between the voiced and voiceless stimuli (p<.005). There was no significant voicing difference between the voiced and the ambiguous voiced stimuli (p=.281). However, there was a difference between the ambiguous voiceless and voiceless stimuli (p=.018).

There was no significant voicing x language interaction; F(3,78)=1.25, p=.297.

Overall, for the graceful-clumsy dimension, both Japanese speakers and English speakers showed a significant voicing effect. While the Japanese overall used more clumsy responses than the English participants, the results suggest that both Japanese speakers and English

speakers associated voiced and the ambiguous voiced stimuli with clumsiness, and they associated voiceless stimuli with goodness.

6. Conclusions and Discussion

The current study showed the existence of sound symbolism in a voicing contrast shared by both Japanese and English speakers. Voiced consonants /d, g, z/ tend to be associated with largeness, badness, and clumsiness whereas voiceless consonants /t, k, s/ tend to be associated with smallness, goodness, and gracefulness. Unlike Iwasaki et al. (2007), sound symbolism in the voicing contrast on evaluative dimensions such as good-bad and graceful-clumsy are found to be shared by Japanese and English participants; both listener groups tended to link voiced sounds with badness and clumsiness, and voiceless sounds with goodness and gracefulness.

For the big-small dimension, both Japanese speakers and English speakers showed a significant voicing effect. Both language groups associated the voiced (deze deze, gede gede, zege zege) and ambiguous voiced stimuli with bigness, and they associated the voiceless (tese tese, kete kete, seke seke) and ambiguous voiceless stimuli with smallness. There was a clear perceptional distinction between the ambiguous voiced and ambiguous voiceless stimuli for both language groups. Additionally, English speakers showed a slightly greater effect of voicing, observed clearly at the ambiguous boundary, than Japanese speakers.

For the good-bad and graceful-clumsy dimensions, both Japanese and English speakers associated the voiced stimuli with badness and clumsiness and the voiceless stimuli with goodness and gracefulness. Interestingly, the distinction between ‘bad’ and ‘good’ and between ‘clumsy’ and ‘graceful’ on the voiced versus voiceless stimuli was less extreme than that for the big-small distinction. The difference between the voiced and voiceless stimuli on good-bad dimension was 0.89 for Japanese and 0.58 for English speakers, and the difference on graceful-clumsy dimension was 0.87 for Japanese and 0.40 for English speakers. On the other hand, the difference on the big-small dimension was 0.92 for Japanese speakers and 1.29 for English speakers. Moreover, there was a significant language effect observed for good-bad and graceful-clumsy dimensions. This was due to English participants who provided fewer ‘bad’ (more ‘good’) and fewer ‘clumsy’ (more ‘graceful’) ratings than the Japanese participants.

With regard to the sound symbolism on round-spiky shape dimension, no significant effect was found in the voicing contrast for the Japanese participants, but there was an effect of voicing in the only-stop stimulus pair “gede-kete” for the English participants. Kohler (1929) claimed that continuant strings (“baluma”) tend to represent roundness and stop strings (“takete”) tend to represent spikiness. However, there is also a voicing contrast in their consonants - “baluma” has three voiced consonants and “takete” has three voiceless consonants. The current data show that the English listeners associated voiced stops with roundness and voiceless stops with spikiness. Interestingly, such voicing effects were not observed when the stimuli had both stops and fricatives. In addition, it needs to be noted that there were differences

across listener group, with a trend for Japanese listeners associating voiced sounds with spikiness and English speakers associating voiced sounds with roundness. A possible explanation for this pattern is that the Japanese participants might have associated the voiced non-word stimuli ("zege zege") with an existing mimetic word "giza giza" ('sharp' and 'jaggy') which contains two voiced consonants /g/ and /z/.

In the current study, the participants showed strong perceptional distinctions between big-small, round-spiky, good-bad, graceful-clumsy when hearing either voiced and voiceless sounds as well as when hearing the ambiguous voiced and ambiguous voiceless sounds.

Since Ferdinand de Saussure's publication, studies have examined the relationship between the 'signifier' (sound) and the 'signified' (meaning), finding a relationship between the sound and the meaning of a word. The present study adds data to address issues of sound symbolism in Japanese and English listeners, finding effects of voicing in size categorization (voiced-big, voiceless-small), good-bad categorization (voiced-bad, voiceless-good), and graceful-clumsy categorization (voiced-clumsy, voiceless-graceful) in both Japanese and English.

References

Atouda, T. and Hoshino, K. (2009) *Usage guide to Japanese onomatopoeias.* Sotakusha.

Aveyard M. E. (2012) Some consonants sound curvy: Effects of sound symbolism on object recognition. *Mem Cognition* 40, 83–92.

BMFT Publisher Japan. (2012) *Words expressing taste.*

Hamano, S. (1998) *The sound symbolic system of Japanese.* Stanford, CA: CSLI Publications; Tokyo: Kuroshio.

Iwasaki, N., Vinson, D.P., and Vigliocco, G. (2007) What do English speakers know about gera-gera and yota-yota: A cross linguistic investigation of mimetic words for laughing and walking. *World's Japanese Language Education* 17, 53-78.

Ivanova J. (2006) Sound-symbolic approach to Japanese mimetic words. *Toronto Working Papers in Linguistics* 26, 103-114.

Kohler, W. (1929) *Gestalt Psychology.* New York, USA: Liveright.

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

de Saussure F. (1983) *Course in general linguistics.* R. Harris, trans. Open Court, La Salle, IL (Original work published 1916).

Shinohara, K. and Kawahara, S. (2016) A cross-linguistic study of sound symbolism: The image of size. *Proceedings of the 36th Annual Meeting of Berkeley Linguistics Society.* 396-410.

Westbury C. (2005) Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language* 93. 10–19.