Phonation Types Matter in Sound Symbolism

Kimi Akita

Department of English Linguistics, Graduate School of Humanities, Nagoya University

Abstract

Sound symbolism is a non-arbitrary correspondence between sound and meaning. The majority of studies on sound symbolism have focused on consonants and vowels, and the sound-symbolic properties of suprasegmentals, particularly phonation types, have been largely neglected. This study examines the size and shape symbolism of four phonation types: modal and creaky voices, falsetto, and whisper. Japanese speakers heard 12 novel words (e.g., /íbi/, /ápa/) pronounced with the four types of phonation and rated the size and roundedness/pointedness each of the 48 stimuli seemed to represent on seven-point scales. The results showed that phonation types as well as consonantal and vocalic features influenced the ratings. Creaky voice was associated with larger and more pointed images than modal voice, which was in turn associated with larger and more pointed images than whisper. Falsetto was also associated with roundedness but not with smallness. These results shed new light on the acoustic approaches to sound symbolism and suggest the significance of phonation types and other suprasegmental features in the phenomenon.

Keywords: Sound symbolism; Iconicity; Bouba-kiki effect; Phonation; Voice quality

1. Introduction

In the last century, it was widely assumed that the relationship between the form and meaning of linguistic signs is essentially arbitrary (de Saussure, 1916). For example, what is called *dog* in English is called *Hund* in German and *inu* in Japanese, both phonologically dissimilar to *dog*. However, numerous linguists and psychologists now acknowledge the ubiquity of

non-arbitrariness in the lexicon and discuss its role in the typology, acquisition, and evolution of language (Imai & Kita, 2014; Nielsen & Dingemanse, 2020; Perlman & Lupyan, 2018).

Sound symbolism is "the direct linkage between sound and meaning" (Hinton, Nichols, & Ohala, 1994, p. 1) that typically involves "iconicity," perceived resemblance between the signifier and the signified (Emmorey, 2014; Peirce, 1932). An enormous number of experimental studies have tested the iconic correspondences between segmental sounds (e.g., consonants and vowels) and visual images (for a review, see Lockwood & Dingemanse, 2015). For example, front and back vowels (or high and low vowels) are known to evoke smaller and larger images, respectively, as exemplified by Sapir's (1929) non-word pair *mil* and *mal*. Studies since Köhler (1929, 1947) have shown that sonorant and obstruent consonants, illustrated by *maluma* and *takete*, are associated with rounded and pointed shapes, respectively (for a fine-grained examination of the *maluma/takete* effect (aka "the *bouba/kiki*" effect" since Ramachandran & Hubbard, 2001), see Nielsen & Rendall, 2011; Nobile, 2015). Some recent studies suggest that even infants might be sensitive to these sound-symbolic mappings (Asano et al., 2015; Peña, Mehler, & Nespor, 2011; but see Fort et al., 2018).

Sound symbolism is not unique to non-words. Ideophones, such as English onomatopoeia (e.g., *bowwow*, *kaboom*) and Japanese mimetics (e.g., *nikoniko* "smiling," *dokin* "one's heart thumping"), are considered sound-symbolic lexemes whose phonological forms imitate sensory meanings. Moreover, large-scale cross-linguistic studies have demonstrated that even some non-ideophonic basic words involve non-arbitrary sound–meaning correspondences. For example, Johansson, Anikin, Carling, and Holmer (2020) identified such correspondences between the nasal consonant /n/ and the concept "nose" (e.g., *hana* in Japanese, *núːsnu* in Nez Perce) and between the voiceless velar stop /k/ and the concept "hard" (e.g., *katai* in Japanese, *kəwa* in Estonian).

These sound-symbolic correspondences can be iconic in that they appear to have an articulatory or acoustic basis. For example, the size of the oral cavity is smaller in the articulation of /i/ than in the articulation of /a/, and this articulatory contrast may contribute to our perception of size sound symbolism. Another possible motivation for size sound symbolism is acoustic. According to Ohala's (1984) "frequency code hypothesis," high/low acoustic frequency is innately associated with a small/large vocalizer as illustrated by animals (e.g., chicks vs. elephants) and musical instruments (e.g., violas vs. contrabasses). As the F2 of /i/ is higher than that of /a/, the symbolic values of these vowels match the frequency code. Consonants' shape symbolism also has acoustic motivations. Sonorants (e.g., *maluma*) involve smoother air pressure changes than obstruents and, therefore, are readily associated with rounded shapes.

The current paper investigates a neglected aspect of speech sound in the context of sound symbolism: phonation types. Phonation is "the production of sound by the vibration of the vocal folds" (Keating & Esposito, 2006, p. 1) or, more precisely, "the use of the laryngeal system, with the help of an airstream provided by the respiratory system, to generate an audible source of acoustic energy that can then be modified by the articulatory actions of the rest of the vocal apparatus" (Laver, 1994, p. 184; see also Laver, 1980). Natural speech involves various types of phonation, from modal (normal) to creaky voice (aka vocal fry, a very low voice one often uses at the end of an utterance with falling intonation), breathy voice (aka murmured voice, a sigh-like voice that leaks more air than modal voice), harsh voice (aka pressed voice, a type of voice one uses when carrying something heavy), whisper (a voiceless sound

one uses when speaking quietly), and falsetto (a type of voice one uses when singing notes beyond the pitch range of one's modal voice). Creaky and breathy voices are phonemic (i.e., are used to distinguish different meanings) in some languages (Gordon & Ladefoged, 2001), but even languages without a phonemic creaky or breathy voice may use marked phonation to indicate attitude, personality, or emotion. For example, English speakers sometimes whisper the interjection *wow* ([{W waʊ W}], with a whispered phonation) to stress their surprise. The current study explores the sound-symbolic properties of phonation types and their general implications for the mechanisms underlying sound-symbolic perceptions.

This paper is organized as follows. Section 2 summarizes selected previous studies on the functions of marked phonation types. Sections 3 and 4 report on experiments on the sound symbolism of four phonation types in the domains of size and shape. Section 5 concludes this paper by briefly discussing the acoustic basis for both segmental and suprasegmental sound symbolism.

2. Previous studies

The majority of sound symbolism studies focus on segments, and the sound symbolism of suprasegmentals, including phonation types, has not been extensively investigated (for rare exceptions, see Dingemanse, Schuerman, Reinisch, Tufvesson, & Mitterer, 2016; Lacey et al., 2020; Nygaard, Herold, & Namy, 2009; Perlman, Dale, & Lupyan, 2015). However, the non-arbitrary uses of some phonation types have been repeatedly noted in the phonetic and sociolinguistic literature, not necessarily in relation to sound symbolism (see Esling, Moisik, Benner, & Crevier-Buchman, 2019; Laver, 1994).

Winter and Grawunder (2010) report on their role-play experiments in which female, but not male, Korean speakers used more breathy voice in formal settings than in informal settings (see also Brown, Winter, Idemaru, & Grawunder, 2014; Hübscher, Borràs-Comes, & Prieto, 2017). Breathy voice is also common in heroes, but not villains, in Japanese animation (Teshigawara, 2003; see also Starr, 2015). Yuasa (2010) shows that American undergraduates perceive female creaky voice as not only "hesitant, non-aggressive, and informal but also educated, urban-oriented, and upwardly mobile" (p. 315), while Anderson, Klofstad, Mayew, and Venkatachalam (2014) provide supplementary data that suggest that American adults, especially women, perceive female creaky voice as "less competent, less educated, less trustworthy, less attractive, and less hirable" (p. 5). All these phonational characteristics are "indexical," rather than iconic, of particular communicative intention (e.g., politeness) or personality traits (e.g., upward mobility) in the sense that they point to or signal these internal properties. Indexical sounds are often placed outside the scope of sound symbolism research (cf. Hinton et al., 1994; Sidhu & Pexman, 2018), and the same was true for these indexical uses of phonation.

Studies on ideophones provide related insights. Ideophones are often uttered with marked phonation (Childs, 1994, p. 184). For example, Dingemanse and Akita (2017, p. 510) show that in an interview about the Great East Japan Earthquake in 2011, a Japanese speaker pronounced the adverbial ideophone *paat-to* "with a rush" as [pḁ:t:o] (with a voiceless vowel) to vividly depict her dog running away from the tsunami into the house. We can interpret

that this special phonation represents the dog's fast, quiet movement and perhaps the tense situation. Dingemanse and Akita report that approximately 30% of adverbial ideophones in their interview data involved marked phonation ("phonational foregrounding" in their terms). Similar iconic uses of phonation are noted for quotations, in which the speaker can mimic the voice quality of the quoted speech (Clark & Gerrig, 1990, p. 775).

The current study experimentally investigates the iconic properties of three marked phonation types—creaky voice, falsetto, and whisper—in two well-studied semantic domains: size and shape. Despite the primarily exploratory nature of the experiments, the acoustic properties of the three phonation types allow us to predict some of their sound-symbolic values. Specifically, the frequency code predicts that creaky voice, which is a particularly low-frequency sound, is associated with large size, whereas falsetto and whisper, which are high-frequency sounds, are associated with small size. Moreover, the audible pulses creaky voice makes may be readily associated with pointed shapes, whereas the smooth pressure changes in falsetto may evoke rounded images.

3. Experiment 1: Size sound symbolism

We conducted an online experiment with native speakers of Japanese, a language in which none of the three types of marked phonation are phonemic (e.g., [inɯ] (modal) and [ḭn̰ɯ̰] (creaky) both mean "dog"). The participants listened to novel size words and rated the size they seemed to represent (for a similar experimental design, see Shinohara & Kawahara, 2016).

3.1. Participants

Forty-one Japanese monolinguals participated in the experiment (female = 23, male = 18; age = 19–73, $M = 38.27$, $SD = 14.04$). They were recruited through CrowdWorks, a crowd-sourcing service in Japan. They gave informed consent before starting the experiment. At the end of the experiment, they were asked whether they knew or heard about "sound symbolism" or "the *bouba-kiki* effect." One participant answered yes and was not included in the current dataset of 41 participants. All participants were paid 400 JPY for their participation.

3.2. Stimuli

Forty-eight audio stimuli were created. They all consisted of two identical vowels and a consonant in between (i.e., $/V_1CV_1/$), which constitute a common word structure in Japanese. The vowels were either /i/ (unrounded high front) or /a/ (unrounded low non-front; more precisely, the Japanese low central vowel /ä/), and the consonant was either a voiced (/b, d, g/) or voiceless stop (/p, t, k/). A male Japanese speaker pronounced these 12 phoneme sequences, listed in Table 1, in four distinct phonation types: modal and creaky voices, falsetto, and whisper. Breathy voice, another frequently mentioned phonation type, was also recorded but not included, as a Japanese speaker who checked the quality of the stimuli found it hard to distinguish from modal voice. All stimuli were pronounced with an initial accent: the first mora was higher-pitched than the second mora. The variations in word length and intensity were kept minimum. The whole stimulus set is available at https://osf.io/jghs2/.

Table 1
The stimulus words. Each of the 12 words was pronounced in four phonation types: modal and creaky voices, falsetto, and whisper. They all had an initial pitch accent

	Voiced Bilabial	Alveolar	Velar	Voiceless Bilabial	Alveolar	Velar
/i/	/íbi/	/ídi/	/ígi/	/ípi/	/íti/	/íki/
/a/	/ába/	/áda/	/ága/	/ápa/	/áta/	/áka/

According to previous studies (e.g., Shinohara & Kawahara, 2016), Japanese speakers associate non-front vowels (e.g., /a/) and voiced obstruents (e.g., /b/) with larger images than front vowels (e.g., /i/) and voiceless obstruents (e.g., /p/). It has also been suggested that obstruent voicing plays a particularly important role in the sound-symbolic system of Japanese (Akita & McLean, 2021; Kawahara, Noto, & Kumagai, 2018; Saji, Akita, Kantartzis, Kita, & Imai, 2019). Keeping segmental properties consistent and manipulating acoustic parameters, the present stimulus set allows us to see whether phonation types influence Japanese speakers' sound-symbolic intuition and, if so, how strong the influence may be in comparison with these previously documented cases of segmental sound symbolism.

3.3. Procedure

The experiment was conducted on a Google Form. The participants read the following instructions on the first page (for the original instructions in Japanese, see the supplementary material):

> In this experiment, you'll hear a language that is spoken somewhere in the world. This language has as many as 48 words that describe size. Please guess the size each word represents from their sound alone and rate it on a 7-point scale from 0 "extremely small" to 6 "extremely large." The same speaker pronounces all sounds. You may listen to the sounds as many times as you like, but you don't need to think too much. Please just trust your intuition.

The participants were asked to wear headphones or earphones and stay alone in a quiet room during the experiment. They were also asked to listen to a sentence meaning "Please adjust the volume on your computer" to adjust the volume on their computer and not to change it during the experiment.

The 48 stimuli were presented randomly, with one on each page. The Google Form required the participants to rate all stimuli.

3.4. Results

The sound-symbolic effects of both segmental and phonational features were obtained. Fig. 1 shows the distribution of the ratings by phonation type and vowel quality.
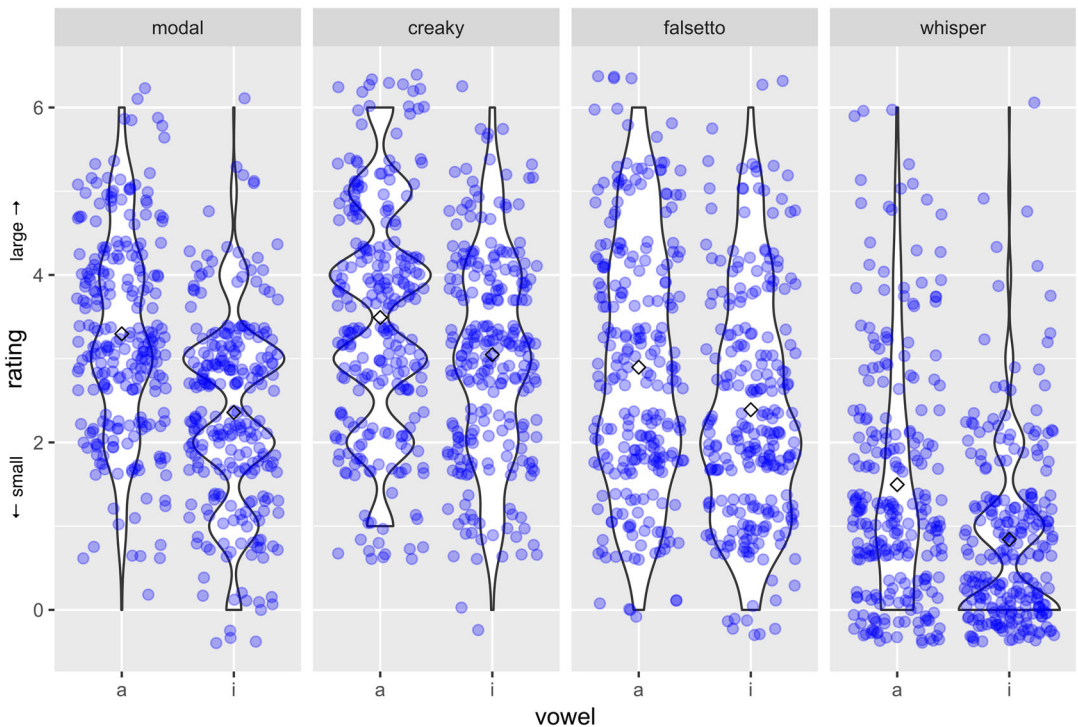
Fig. 1. Size ratings (from extremely small to large) by phonation type and vowel quality. The jitters represent the distribution of the ratings, and the width of violins corresponds to their relative density on the scale. The white squares are the means. Although /a/ tended to be associated with larger images than /i/, the ratings were more dependent on the phonation types. Creaky voice tended to be rated particularly large, and whisper particularly small.

Using the "ordinal" package (Christensen, 2019) on R version 4.0.4 (R Core Team, 2021), a cumulative link mixed model analysis was run, with size rating as the response variable and phonation type, vowel quality, obstruent voicing, and place of consonant articulation as fixed effects (no interactions considered). Random intercepts were included for stimulus and participant and a random slope for phonation type by participant. The model is given in Table 2.

Replicating the previous findings, /a/ was judged as larger than /i/ ($p < .001$). More crucially, the creaky voice was rated as larger than modal voice ($p < .01$), which was in turn rated as larger than whisper ($p < .001$). No significant difference was obtained between modal voice and falsetto ($p = .28$). The estimated coefficients suggest that whisper has a particularly strong sound-symbolic effect among these phonetic properties. Interestingly, the effect of obstruent voicing, which is known to play the central role in the Japanese sound-symbolic system, did not reach statistical significance ($p = .18$). This result is presumably attributed to the weakened voicing contrast in the present stimuli. Creaky voice made the voiceless stops partly voiced, and whisper made the voiced stops voiceless.

Table 2
The cumulative link mixed model for size ratings. The baseline is /ába/ (/a/, voiced bilabial) pronounced in modal voice. /i/ and whisper tended to be rated smaller, and creaky voice larger

		Estimate	SE	z	p
Vowel	/i/	−1.23	0.19	−6.56	< .001
Consonant	Voiceless	−0.25	0.19	−1.35	.18
	Alveolar	0.03	0.23	0.12	.90
	Velar	−0.24	0.23	−1.06	.29
Phonation type	Creaky	0.80	0.27	2.96	< .01
	Falsetto	−0.35	0.32	−1.08	.28
	Whisper	−3.79	0.40	−9.40	< .001

3.5. Discussion

The current results suggest that phonation types do matter in sound symbolism. Given that actual speech uses diverse phonational types, the previously attested sound–meaning correspondences might be better viewed as idealized results that are more or less different from our sound-symbolic intuition in actual communication. Despite the novelty of the current findings, the sound symbolism of each phonation type is largely consistent with the existing approaches to sound symbolism. The largeness of creaky voice and the smallness of whisper serve as new evidence for the frequency code hypothesis (see Section 1). Since creaky voice has a low fundamental frequency (F0; Gordon & Ladefoged, 2001), it obtains a large image via the frequency code. Whisper is voiceless, and voiceless sounds have higher acoustic frequency than their voiced counterparts. Thus, the frequency code assigns a small image to whisper. However, the frequency code fails to account for the intermediate size ratings for falsetto. Falsetto has a higher F0 than modal voice, and the frequency code would associate it with smallness.

Another possible account of part of the present results would refer to pressure variations. Whisper involves small pressure variations (i.e., a narrow amplitude range with a low maximum amplitude), which may be associated with a small image. In fact, Japanese describes high- and low-amplitude sounds as *ōkii oto* "big sound" and *chīsai oto* "small sound," respectively, and these conventional sound–meaning links may have helped the participants to get large and small images. This acoustic account is also compatible with the not-so-low size ratings for falsetto, which tends to have high amplitude. To further illuminate the sound-symbolic values of these phonation types, the next section examines another semantic domain.

4. Experiment 2: Shape sound symbolism

We conducted another online experiment using the same setting as Experiment 1. Japanese-speaking participants rated the roundedness/pointedness of the 48 novel shape words seemed to represent.
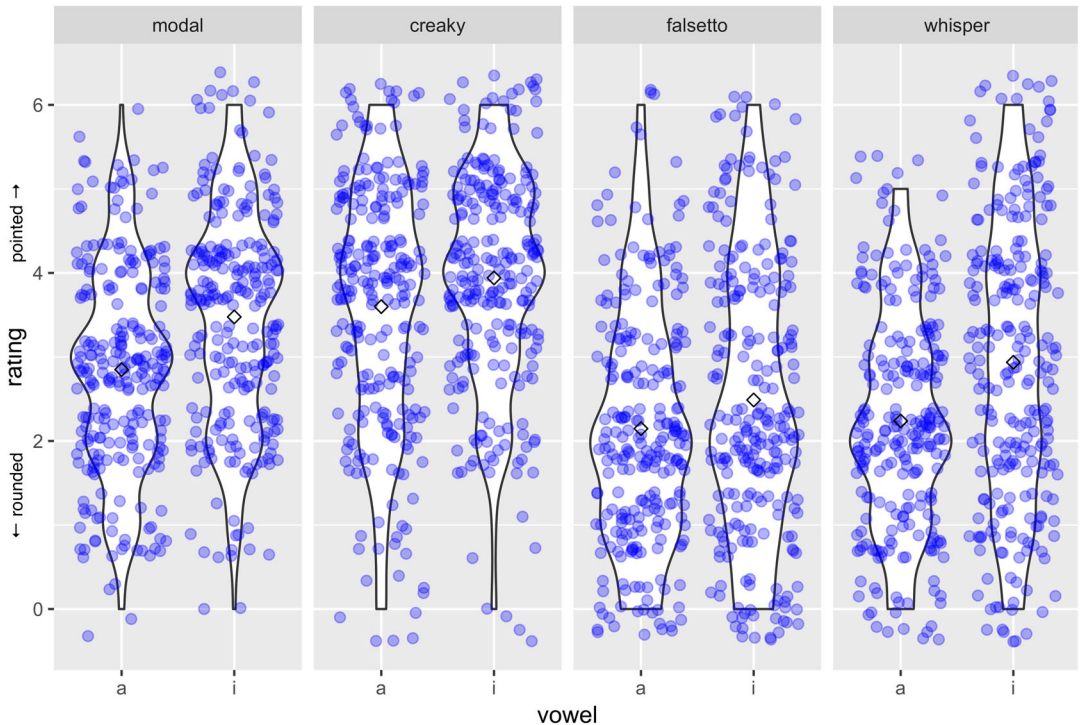
Fig. 2. Shape ratings (from extremely rounded to pointed) by phonation type and vowel quality. /a/ tended to be rated more rounded than /i/, while creaky voice tended to be judged as pointed; falsetto and whisper tended to be judged as rounded.

4.1. Method

Forty-two Japanese monolinguals (female = 26, male = 16; age = 21–65, M = 35.12, SD = 9.73) were recruited through CrowdWorks. None of them participated in Experiment 1 or reported that they knew sound symbolism or the bouba-kiki effect. They were paid 400 JPY for their participation.

The same stimulus set as Experiment 1 was used. The participants rated the 48 stimuli on a seven-point scale from 0 "extremely rounded" to 6 "extremely pointed" (for the actual instructions, see the supplementary material). Previous studies on shape sound symbolism allow us to expect /a/ and the bilabials /b, p/ to receive more "rounded" ratings than /i/ and the alveolars /d, t/ and velars /g, k/ (D'Onofrio, 2014; Nielsen & Rendall, 2011; Nobile, 2015).

4.2. Results

Again, we obtained the effects of both segmental and suprasegmental features as shown in Figs. 2 and 3.

A cumulative link mixed model analysis was run, with shape rating as the response variable, the phonational, consonantal, and vocalic features as fixed effects, and stimulus and
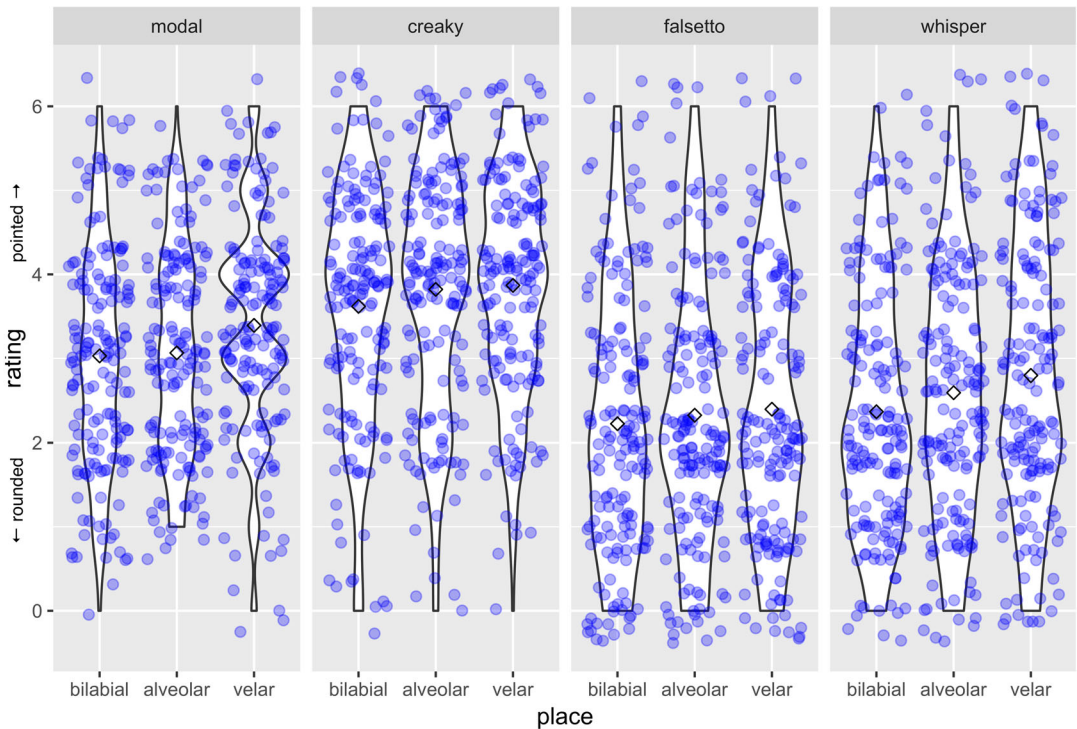
Fig. 3. Shape ratings (from extremely rounded to pointed) by phonation type and place of consonant articulation. The bilabial consonants tended to be rated more rounded than the velar consonants.

Table 3
The cumulative link mixed model for shape ratings. The baseline is /ába/ (/a/, voiced bilabial) pronounced in modal voice. /i/, velars, and creaky voice tended to be rated more pointed, and falsetto and whisper more rounded

		Estimate	SE	z	p
Vowel	/i/	0.69	0.12	5.98	< .001
Consonant	Voiceless	−0.19	0.11	−1.67	.09
	Alveolar	0.23	0.14	1.64	.10
	Velar	0.43	0.14	3.07	< .01
Phonation type	Creaky	0.93	0.21	4.35	< .001
	Falsetto	−1.33	0.32	−4.17	< .001
	Whisper	−0.90	0.28	−3.26	< .01

participant as random effects, with a random slope for phonation type by participant. The model is detailed in Table 3.

As expected, the high front vowel /i/ was associated with more pointed shapes than the low non-front vowel /a/ ($p < .001$). The rating contrast between the bilabials and the velars ($p < .01$) is also consistent with the previous findings. Creaky voice was associated with more pointed shapes than modal voice ($p < .001$), which was in turn rated as more pointed
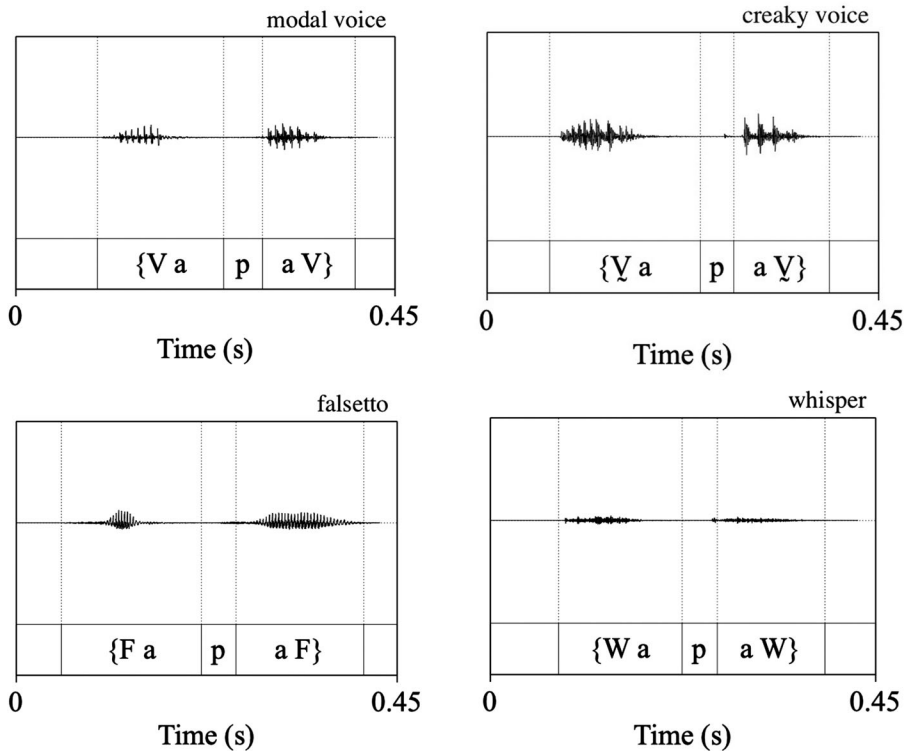
Fig. 4. Waveforms of /ápa/ in the four types of phonation. Creaky voice creates audible irregular pulses. Falsetto involves seemingly smooth pressure changes. Whisper creates limited pressure variations.

than falsetto ($p < .001$) and whisper ($p < .01$). The effect of obstruent voicing was again not significant ($p = .09$).

4.3. Discussion

The different shape ratings obtained for the four types of phonation are interpretable in terms of the pressure variations they create (Fig. 4). Creaky voice involves irregular vocal fold vibration that causes irregular pulses, a likely source of a pointed image. In contrast, having high F0, falsetto causes a seemingly gradual change in pressure, which is readily associated with a rounded image. The roundedness of whisper is presumably ascribable to the consistent absence of pressure variations. Taken together with Experiment 1, it appears to be safe to conclude that, as with segmental symbolism, phonational symbolism has an acoustic basis.

5. General discussion

In this paper, we have opened up a new area of sound symbolism research by reporting on our first explorations in iconic phonational symbolism. The two experiments revealed

that phonation types make a considerable contribution to Japanese speakers' sound-symbolic ratings of size and shape. We argued that phonational symbolism has relatively clear acoustic grounds and, therefore, reinforces the acoustic, rather than articulatory, accounts of sound symbolism. Generally, acoustic phonetics enables us to discuss segmental and suprasegmental features within the same framework and arrive at broad, fundamental generalizations that may be applied cross-linguistically. In this respect, it will be worthwhile to extend the current study to other suprasegmental features, such as pitch range, loudness, speech rate, and rhythm. Since, in our everyday conversation, we say the same word with various suprasegmental adjustments, scrutinizing the sound-symbolic relevance of suprasegmentals will mean approaching the reality of sound symbolism in human language.

To be more specific, the relatively limited significance of voicing, a central segmental feature in Japanese sound symbolism, observed for size and shape sound symbolism suggests that suprasegmental symbolism can override segmental symbolism. Moreover, the greater sound-symbolic significance of amplitude than F0 observed for the size symbolism of falsetto indicates that suprasegmentals may reveal unknown aspects of sound symbolism. Thus, the consideration of suprasegmentals may allow us to discover what we cannot easily discover in segmental symbolism.

Furthermore, it will be worth discussing whether the observed phonational symbolism is universal or language-specific. The likely acoustic basis of the sound–meaning mappings suggests their cross-linguistic availability. However, given the language- or culture-specificity of some indexical uses of marked phonation, such as female Americans' creaky voice (Section 2), the iconic values of phonation types might also differ from language to language.

Finally, it is expected that two more approaches will advance the study of phonational (and other types of suprasegmental) symbolism. One is a lexical-typological approach. The present study targeted speakers of Japanese, in which the four phonation types are not phonemically distinctive. A lexical survey in languages with phonemic breathy or creaky voice (e.g., Newar, Jalapa Mazatec) might reveal the lexical uses of phonational symbolism. For example, words for large concepts might be more likely than words for small concepts to contain creaky voice. The semantic properties of phonationally marked ideophones (see Section 2) will also be worth close attention. The other possible approach to phonational symbolism is an acoustic one. The current study used descriptive categories for phonation types. Employing a set of specific acoustic measures will make a more fine-grained observation possible (see Lacey et al., 2020; Perlman & Lupyan, 2018).

Open Research Badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/jghs2/.

Notes

1. There will be at least two other possible mechanisms for the sound symbolism of the observed phonation types: the indexicality of individuals typically using these phonation types and the iconicity of situations in which these phonation types would be used (Sidhu & Pexman, 2018).

2. Johansson et al. (2020) explored the sound–meaning correspondences in the words for 344 near-universal concepts in 245 languages that represent 245 language families. In their database, we can find phonemic creaky voice in only one large property word ([*waȷ̣at*] "deep" in Nez Perce (Sahaptian, North America)) and two small property words ([*taʔaɰ*] "empty" and [*hɪcaaɰɪc*] "light (of weight)" in Nez Perce).

References

Akita, K., & McLean, B. (2021). Onshōchō chikaku no nichi/ei taishō: Imi chizu ni yoru kentō [A contrastive study of Japanese and English speakers' sound-symbolic perception: A semantic map approach]. In K. Shinohara & R. Uno (Eds.), *Jikken ninchi gengogaku no shinka [Progress in experimental cognitive linguistics]*. (165–189). Tokyo: Hituzi Syobo.

Anderson, R. C., Klofstad, C. A., Mayew, W. J., & Venkatachalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. *PLoS ONE*, *9*(5), e97506.

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.

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, *66*, 45–60.

Childs, G. T. (1994). African ideophones. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 178–204). Cambridge, UK: Cambridge University Press.

Christensen, R. H. B. (2019). *ordinal: Regression models for ordinal data*. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstations. *Language*, *66*(4), 764–805.

de Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.

Dingemanse, M., & Akita, K. (2017). An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics*, *53*(3), 501–532.

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*, e117–e133.

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the *bouba-kiki* paradigm. *Language and Speech*, *57*(3), 367–393.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B*, *369*, 20130301.

Esling, J. H., Moisik, S. R., Benner, A., & Crevier-Buchman, L. (2019). *Voice quality: The laryngeal articulator model*. Cambridge, UK: Cambridge University Press.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, *21*(5), e12659.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). Introduction: Sound-symbolic processes. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 1–12). Cambridge, UK: Cambridge University Press.

Hübscher, I., Borràs-Comes, J., & Prieto, P. (2017). Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics*, *65*, 145–159.

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.

Johansson, N. E., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, *24*(2), 253–310.

Kawahara, S., Noto, A., & Kumagai, G. (2018). Sound symbolic patterns in Pokémon names. *Phonetica*, *75*, 219–244.

Keating, P. A., & Esposito, C. (2006). Linguistic voice quality. *Proceedings of the Eleventh Australasian International Conference on Speech Science and Technology*, Auckland, Australia.

Köhler, W. (1929/1947). *Gestalt psychology*. New York: Horace Liveright.

Lacey, S., Jamal, Y., List, S. M., McCormick, K., Sathian, K., & Nygaard, L. C. (2020). Stimulus parameters underlying sound-symbolic mapping of auditory pseudowords to visual shapes. *Cognitive Science*, *44*, e12883.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge, UK: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge, UK: Cambridge University Press.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*, 1246.

Nielsen, A. K. S., & Dingemanse, M. (2020). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52–72.

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*(2), 115–124.

Nobile, L. (2015). Phonemes as images: An experimental inquiry into shape-sound symbolism applied to the distinctive features of French. In M. K. Hiraga, W. J. Herlofsky, K. Shinohara, & K. Akita (Eds.), *Iconicity: East meets West* (pp. 71–91). Amsterdam: John Benjamins.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*, 127–146.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica*, *41*, 1–16.

Peirce, C. S. (1932). *Collected papers of Charles Sanders Peirce, Vol. 2*, Cambridge, MA: Harvard University Press.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*(11), 1419–1421.

Perlman, M., Dale, R. D., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*, 150152.

Perlman, M., & Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific Report*, *8*, 2634.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

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

Saji, N., Akita, K., Kantartzis, K., Kita, S., & Imai, M. (2019). Cross-linguistically shared and language-specific sound symbolism in novel words elicited by locomotion videos in Japanese and English. *PLoS ONE*, *14*(7), e0218707.

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

Shinohara, K., & Kawahara, S. (2016). A cross-linguistic study of sound symbolism: The image of size. *Berkeley Linguistics Society*, *36*, 396–410.

Sidhu, D., & Pexman, P. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*, 1619–1643.

Starr, R. L. (2015). Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society*, *44*(1), 1–34.

Teshigawara, M. (2003). Voices in Japanese animation: A phonetic study of vocal stereotypes of heroes and villains in Japanese culture *(Doctoral dissertation)*. University of Victoria, Canada.

Winter, B., & Grawunder, S. (2010). The polite voice in Korean: Searching for acoustic correlates of *contaymal* and *panmal*. In S. Y. Cheon (Ed.), *Japanese/Korean linguistics, Vol. 19* (pp. 419–431). Stanford, CA: CSLI Publications.

Yuasa, I. P. (2010). Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech*, *85*(3), 315–337.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information