Accepted with minor revisions, Language and Cognition
Revised version.

Iconic prosody enhances the depictive power of ideophones

Journal:	*Language and Cognition*
Manuscript ID	LCO-2025-0064.R1
Manuscript Type:	Article
Keywords:	evolution of language, iconicity, ideophones, prosody, Japanese
Abstract:	Prosody not only signals the speaker's cognitive states but can also imitate various concepts. However, previous studies on the latter, iconic function of prosody have mostly analyzed novel words and nonlinguistic vocalizations. To fill this gap in the literature, the current study has examined the iconic potential of the prosodic features of existing Japanese imitative words known as ideophones. In Experiment 1, female Japanese speakers pronounced 20 sentences containing ideophones in infant-directed speech. They used higher f0 to express faster and more pleasant movements. Similar iconic associations were observed in Experiment 2, in which Japanese speakers chose the best-matching pitch–intensity–duration combination for each of the ideophones. In Experiment 3, Japanese speakers chose the best-matching voice quality—creaky voice, falsetto, harsh voice, or whisper—for the ideophones. Falsetto was preferred for a light object's fast motion, harsh voice for violent motion, and whisper for quiet motion. Based on these results, we entertain the possibility that the iconic prosody of ideophones provides a missing link in the evolutionary theory of language that began with iconic vocalizations. Ideophones with varying degrees of iconic prosody can be considered to be located between nonlinguistic vocalizations and arbitrary words in this evolutionary path.

1. Introduction

The same word can be pronounced in different ways according to the intended message. We would shout *Yes!* when our favorite table tennis player wins a long rally, but we might whisper *Yes…* when we respond to our friend's question while watching a movie. In Peirce's (1932) semiotics, these prosodic features of speech sound are *indexical* of our emotional state or attitude: there is a causal relationship between our inner state and the phonetic details of our pronunciation.

On the other hand, we might whisper the adjective *quiet* when we exaggerate the quietness of the car we want to sell, as in *This car is* {W *quiet* W}.[1] This whisper is *iconic* of the car's quietness: the voice quality resembles the car's sonic attribute. While most studies on the functions of suprasegmental features investigates its indexical use (Anderson, Klofstad, Mayew, & Venkatachalam, 2014; Esling, Moisik, Benner, & Crevier-Buchman, 2019; Gussenhoven, 2016; Hancil & Hirst, 2013; Hübscher, Borràs-Comes, & Prieto, 2017; Laver, 1994), its iconic use is also worth special attention in the current cognitive science inquiries, since both indexicality and iconicity are viewed as relevant to the origin of human language (Everett, 2017; Imai & Kita, 2014; Perniss, Thompson, & Vigliocco, 2010; Vigliocco, Perniss, & Vinson, 2014).

To this end, the current study explores the iconic prosody, as it relates to the use of Japanese ideophones, which are imitative words that themselves iconically represent various sensory and emotional information, such as *piyopiyo* 'cheeping', *sutasuta* 'walking briskly', *sarari* 'dry and smooth', *zukin* 'one's head throbbing', and *ukiuki* 'buoyant'. Ideophones are often accompanied by prominent prosody, such as distinctly high or low pitch and marked voice quality (Childs, 1994; Dingemanse et al., 2016). For example, according to Dingemanse and Akita's (2017) study using a multimodal corpus of Japanese, about 30% of adverbial ideophones marked by a quotative particle were "phonationally foregrounded" i.e., pronounced with marked voice quality. For instance, in the following quote, the female speaker pronounces the ideophone *hyoi* 'unexpected' in a falsetto voice, which appears to emphasize the unexpectedness or suddenness of her son's answering her phone call.

> *Soshite, nan-byak-kai-ka nan-zen-kai-ka wakar-anai-n-des-u-kedo, keitai shi-tara,* **{F *hyoi* F}***-to koo yuugata, daibu kuraku nat-te-kara, guuzen tsunagat-ta-n-des-u.*
> 'And, I don't know how many phone calls I made, but when I made a phone call on my cellphone, *unexpectedly* in the evening, when it was very dark, [my son] answered it by chance.'
>
> https://www2.nhk.or.jp/archives/movies/?id=D0026020063_00000

Expanding on these previous observations, in particular the one by Dingemanse and Akita

[1] For voice quality symbols, see Ball, Howard, and Miller (2018).

(2017), the current study reports on three experiments, which together explore what the prosodic features of ideophones can represent iconically in Japanese.

The organization of this paper is as follows. Section 2 summarizes previous experimental studies on the iconic properties of speech sound. Section 3 reports on an experiment in which native speakers of Japanese pronounced ideophones expressively, and Sections 4 and 5 report on perception experiments in which native Japanese speakers chose pitch–intensity–duration combinations and voice qualities that intuitively suited individual ideophones. Section 6 discusses the implications that the current findings may offer for the theory of language evolution. Section 7 is the conclusion.

2. Previous studies

There are several studies on the production and perception of the iconic functions of speech prosody, but they are so far limited both in number and scope. Shintel, Nusbaum, and Okrent (2006) for example asked English speakers to describe the direction of motion of an animated dot and found that the participants tend to use higher fundamental frequency (f_0) to describe upward motion than to describe downward motion, and faster speech rate to describe faster motion than to describe slower motion.

Nygaard, Herold, and Namy (2009) asked three female speakers of English to pronounce sentences with novel adjectives in infant-directed speech: *Can you get the {blicket/seebow/daxen/foppick/tillen/riffel} one?* The participants pronounced the novel words as having positive ('happy', 'hot', 'big', 'yummy', 'tall', 'strong'), negative ('sad', 'cold', 'small', 'yucky', 'short', 'weak'), or neutral (i.e., unspecified) meanings. The acoustic analysis of the recorded materials revealed that, for instance, the novel words were pronounced with higher f_0, higher intensity and shorter duration, when the speakers intended to express 'happiness', and with higher intensity and longer duration, when they intended to express 'largeness' (see also Ferrara, Lu, & Goldin-Meadow, 2025; Herold, 2006; Herold, Nygaard, Chicos, & Namy, 2011; Kunihira, 1971; Michaelini & Nygaard, 2025 for related experiments).

Similar experiments have also been conducted on iconic vocalizations (Ćwiek et al., 2021; Ćwiek & Fuchs, 2019; Perlman, 2026; Perlman & Cain, 2014; Perlman, Dale, & Lupyan, 2015; Perlman & Lupyan, 2018). In this series of experiments, English speakers were instructed to use nonlinguistic vocalizations to express various meanings, such as 'tiger', 'water', 'small', 'many', and 'that'. For example, 'water' was iconically represented by mimicking the sound of pouring water into a glass, and 'tiger' by mimicking its roar. In forced-choice tasks, speakers of both English and several other languages showed accuracy that was greater than chance at selecting the intended meanings of the obtained vocalizations.

A few studies on iconic prosody can also be found in the recent literature on sound

symbolism. According to Akita (2021) and Motoki, Pathak, and Spence (2022), Japanese speakers associate novel words pronounced with creaky voice with largeness, spikiness, and bitterness, those pronounced in a falsetto with roundedness, brightness, and sweetness, and those pronounced in a whisper with smallness, roundedness, and darkness (see also Lacey et al., 2020; Villegas, Akita, & Kawahara, 2023).

While these studies are starting to unveil an important role of iconic prosody in natural languages, what is yet to be addressed is its role in *real words*, such as *hyoi* 'unexpected' discussed in Section 1. It remains an open question how prosodic features interact with the lexical meanings of conventional words other than the directional words (i.e., *up* and *down*) examined in Shintel et al. (2006) (cf. Stolarski, 2019). We would like to emphasize here that iconic prosody has not so far been tested with ideophones, iconic words that are characterized by their marked prosody. The current study fills this gap in the literature by experimentally examining the production and perception of Japanese ideophones.

3. Experiment 1: Production

To investigate whether—and if so, how—iconic prosody can contribute to ideophonic utterances, we first built upon Nygaard et al.'s (2009) nonword-based study, asking Japanese speakers to produce sentences with ideophones in infant-directed speech (henceforth IDS). We focused on IDS, as it generally employs exaggerated prosody, such as heightened pitch, a wide pitch range, and lengthened vowels (Fernald et al., 1989; Garnica, 1977: Igarashi, Nishikawa, Tanaka, & Mazuka, 2013; Mazuka, Igarashi, Martin, & Utsugi, 2015).

3.1. Method

3.1.1. Participants

Thirty female monolingual speakers of Japanese (age: 23–63, M = 38.00, SD = 9.43) were recruited on CrowdWorks, a Japanese crowdsourcing platform. Twenty-four of them had childcare experience, but this factor did not significantly improve the fit of a regression model reported below; hence this factor was not considered in the subsequent analysis. They were paid 300 JPY for their participation.

3.1.2. Stimuli

We prepared a total of 20 simple sentences containing an ideophone, as listed in Table 1. The current experiment focused on those ideophones that represent manners of motion, such as walking, running, and floating. They constitute a major semantic domain in the Japanese ideophone inventory, and their meanings have been extensively described in the literature (Akita, 2020a; Ibarretxe-Antuñano, 2019; Saji et al., 2019; Toratani, 2012). We selected motion ideophones from Akita (2020a) that have a non-reduplicated form and end with a so-called "sokuon" /Q/ (phonetically realized as the first half of a geminate when followed by a

consonant, as in the current experiment). We used this particular morphological shape because it is known that expressive, emphatic prosody appears most frequently with ideophones of this type (Akita, 2020b). The 20 sentences were presented in a random order on Google Forms.

Table 1. Stimulus sentences for Experiment 1, with abbreviated semantic labels for cross-referencing in parentheses.

*Kanban-ga **bataQ**-to taore-ta-ne.* 看板がバタッと倒れたね。
'A signboard slammed down.' (signboard)
*Kooen-o **buraQ**-to sanpo shi-ta-ne.* 公園をブラッと散歩したね。
'We took a walk in the park.' (walk)
*Hako-ga **dokaQ**-to ochi-ta-ne.* 箱がドカッと落ちたね。
'A box thudded down.' (box)
*Papa-ga **goroQ**-to ne-korogat-ta-ne.* パパがゴロッと寝転がったね。
'Dad lay down.' (dad)
*Kooen-o **guruQ**-to mawat-ta-ne.* 公園をグルッと回ったね。
'We walked around the park.' (walk around)
*Sakura-ga **hiraQ**-to ochi-te ki-ta-ne.* 桜がヒラッと落ちてきたね。
'A cherry blossom petal fluttered down.' (petal)
*Namida-ga **horoQ**-to kobore-ta-ne.* 涙がホロッとこぼれたね。
'A tear dropped.' (tear)
*Donguri-ga **koroQ**-to ochi-ta-ne.* どんぐりがコロッと落ちたね。
'An acorn dropped.' (acorn)
*Kazamidori-ga **kuruQ**-to mawat-ta-ne.* 風見鶏がクルッと回ったね。
'A weathercock spun quickly once.' (weathercock)
*Yuge-ga **mowaQ**-to de-ta-ne.* 湯気がモワッと出たね。
'A cloud of steam came out.' (steam)
*Ame-ga **paraQ**-to fut-te ki-ta-ne.* 雨がパラッと降ってきたね。
'It started to rain lightly.' (rain)
*Gohan-ga **poroQ**-to kobore-ta-ne.* ご飯がポロッとこぼれたね。
'Rice dropped.' (rice)
*Mizu-ga **potaQ**-to ochi-ta-ne.* 水がポタッと落ちたね。
'Water dropped quietly.' (water)
*Matsubokkuri-ga **potoQ**-to ochi-ta-ne.* 松ぼっくりがポトッと落ちたね。
'A pine cone dropped.' (pine cone)
*Awa-ga **pukaQ**-to ukan-da-ne.* 泡がプカッと浮かんだね。
'A bubble floated up.' (bubble)
*Kaeru-ga **pyokoQ**-to hane-ta-ne.* カエルがピョコッと跳ねたね。
'A frog hopped once.' (frog)
*Himo-ga **suruQ**-to nuke-ta-ne.* 紐がスルッと抜けたね。
'A string went off quietly.' (string)

*Ase-ga **taraQ**-to tare-ta-ne.* 汗がタラッと垂れたね。
'A sweat dropped.' (sweat)
*Pengin-san-ga **tsuruQ**-to subet-ta-ne.* ペンギンさんがツルッと滑ったね。
'Mr. Penguin slipped.' (penguin)
*Ashi-ga dobu-ni **zuboQ**-to hamat-ta-ne.* 足がドブにズボッと嵌ったね。
'Your foot fell into the ditch.' (ditch)

In order to quantitatively explore the possible correlations between the semantic features of these ideophones and the use of particular prosodic patterns, a different group of 20 monolingual speakers of Japanese (female: 13, male: 7; age: 22–55, M = 40.05, SD = 8.80) rated each of the 20 ideophones on six 7-point semantic scales, adapted from Ibarretxe-Antuñano (2019) and Saji et al. (2019): size (from 0 'small' to 6 'large'), speed (from 0 'slow' to 6 'fast'), weight (from 0 'light' to 6 'heavy'), intensiveness (from 0 'moderate' to 6 'intensive'), pleasantness (from 0 'unpleasant' to 6 'pleasant'), and noise (from 0 'quiet' to 6 'noisy'). Although no additional instructions were given as to how to interpret these scales, the ratings were no more variable for highly subjective scales (e.g., pleasantness: mean SD = 1.07) than less subjective scales (e.g., speed: mean SD = 1.28). Using Google Forms, the ideophones were visually presented with the example sentences in Table 1. The order of the ideophones was randomized. The mean semantic ratings for the ideophones are shown in Table 2.

Table 2. Mean semantic ratings, with standard deviation in parentheses, for all the ideophones that were examined.

	Size	Speed	Weight	Intensiveness	Pleasantness	Noise
bataQ 'a two-dimensional object slamming down' (signboard)	4.85 (1.09)	4.20 (1.24)	5.25 (0.79)	5.10 (0.72)	1.45 (1.10)	5.15 (0.67)
buraQ 'taking a walk' (walk)	2.95 (1.36)	1.55 (1.10)	2.35 (1.42)	0.95 (1.00)	4.40 (1.19)	1.20 (1.01)
dokaQ 'thudding down' (box)	5.40 (0.60)	4.05 (1.47)	5.55 (0.83)	5.15 (0.81)	1.15 (0.81)	5.25 (0.55)
goroQ 'lying down' (dad)	4.65 (0.93)	1.60 (0.60)	5.20 (0.83)	2.80 (1.24)	2.80 (1.01)	2.90 (1.62)
guruQ 'going around, drawing a large circle' (walk around)	4.45 (1.28)	2.85 (1.27)	3.45 (1.15)	2.80 (1.11)	3.60 (0.94)	2.50 (1.28)
hiraQ 'a light thin object fluttering down' (petal)	0.85 (0.81)	2.10 (1.52)	0.15 (0.37)	0.40 (0.50)	5.20 (0.89)	0.35 (0.49)
horoQ 'a light teardrop dropping' (tear)	0.75 (0.79)	2.80 (1.54)	0.75 (0.79)	1.00 (0.86)	3.70 (0.92)	0.45 (0.69)
koroQ 'a light object rolling once' (acorn)	0.55 (0.69)	3.75 (1.33)	0.45 (0.69)	1.10 (0.79)	4.35 (0.88)	0.90 (0.79)
kuruQ 'a light object spinning quickly once' (weathercock)	2.20 (1.06)	5.00 (0.80)	1.35 (1.04)	2.75 (1.45)	3.95 (0.83)	1.80 (0.83)
mowaQ 'steam/smoke coming out' (steam)	3.45 (1.47)	2.05 (1.36)	2.30 (1.89)	1.65 (1.18)	2.85 (1.50)	1.05 (1.10)
paraQ 'small light drops falling' (rain)	1.20 (0.89)	4.10 (1.12)	0.85 (0.67)	1.85 (1.18)	3.15 (1.23)	1.55 (1.05)
poroQ 'a small light object dropping' (rice)	0.70 (0.57)	3.60 (1.64)	0.65 (0.93)	1.25 (1.02)	2.95 (1.15)	0.85 (0.67)
potaQ 'liquid dropping quietly' (water)	0.75 (1.37)	3.65 (1.60)	0.50 (0.61)	1.00 (0.80)	3.85 (1.18)	1.05 (0.83)

potoQ 'a small light object dropping' (pine cone)	1.35 (0.75)	3.55 (1.36)	1.50 (1.10)	1.60 (0.94)	3.80 (1.24)	1.30 (0.73)
pukaQ 'a light object floating' (bubble)	1.60 (0.94)	2.40 (1.19)	0.55 (0.83)	1.15 (0.99)	4.05 (1.15)	1.25 (0.64)
pyokoQ 'a little frog hopping once' (frog)	0.75 (0.72)	4.50 (1.05)	0.70 (0.57)	2.15 (1.31)	4.00 (0.97)	1.65 (1.09)
suruQ 'a light object going off quietly' (string)	1.35 (0.99)	4.60 (1.23)	0.75 (0.72)	1.65 (1.57)	3.80 (1.24)	0.70 (0.80)
taraQ 'liquid dropping' (sweat)	1.10 (0.97)	2.80 (1.70)	1.15 (0.99)	1.70 (1.08)	1.70 (1.13)	0.65 (0.81)
tsuruQ 'slipping' (penguin)	2.45 (0.83)	5.05 (0.89)	1.55 (0.95)	3.30 (1.22)	3.25 (0.97)	1.85 (0.93)
zuboQ 'one's foot falling into a ditch' (ditch)	4.55 (0.95)	3.55 (1.54)	4.90 (0.85)	4.65 (1.04)	0.95 (1.05)	3.80 (1.24)

Some of these scales were found to be strongly correlated with each other. For example, the Spearman correlation between size and weight was 0.78. Analyzing all these scales in a single regression analysis was not desirable, which would have resulted in a collinearity problem. Hence, a principal component analysis was run, which revealed that the first three components account for 89.52% of the variability in the data. As shown in Table 3, these dimensions are primarily characterized by speed and pleasantness, and the other four scales primarily contribute to PCs 4 to 6. Therefore, we only used the speed and pleasantness in the following statistical analyses.

Table 3. Principal components' loadings. Boldface > |0.50|.

	PC1	PC2	PC3	PC4	PC5	PC6
Size	0.45	0.23	0.30	**0.53**	–0.07	**0.60**
Speed	0.05	**–0.91**	0.12	0.37	0.11	–0.07
Weight	0.47	0.23	0.15	0.31	0.01	**–0.78**
Intensiveness	0.46	–0.24	0.08	–0.46	**–0.71**	0.03
Pleasantness	–0.37	0.05	**0.91**	–0.13	–0.08	–0.09
Noise	0.47	–0.08	0.19	**–0.51**	**0.68**	0.11

3.1.3. Procedure

The 30 participants, recruited for the main experiment, were instructed to complete the experiment alone in a quiet room, read the sentences aloud so that even a 1-year-old infant could understand what they meant, and record their pronunciation on their smartphone or other devices. They were allowed to pronounce each sentence as many times as they liked. In this experiment, as well as in Experiments 2 and 3 reported below, the participants read the consent form before they began their task.

3.1.4. Predictions

The previous literature on sound symbolism allows us to make some specific predictions about how the prosodic features of ideophones may be used to express the speed and pleasantness of motion. Notably, the Frequency Code Hypothesis (Ohala, 1984)—one of the most influential hypotheses in the literature on sound symbolism—states that higher frequency sounds signal a

smaller vocalizer than lower frequency sounds; for example, a small mouse makes a higher-frequency voice than a large elephant. From this hypothesis, Ohala and his colleagues proposed to derive several sound–meaning associations, as quoted below:

> "… high tones, vowels with high second formants (notably /i/), and high-frequency consonants are associated with high-frequency sounds, small size, sharpness, and rapid movement; low tones, vowels with low second formants (notably /u/), and low-frequency consonants are associated with low-frequency sounds, large size, softness, and heavy, slow movements."
>
> (Hinton, Nichols, & Ohala, 1994, p. 10)

If Japanese ideophones behave according to the Frequency Code, it is predicted, for instance, that fast motion is expressed by high f_0.

In addition, according to Nygaard et al.'s (2009) nonce word experiment, high f_0 might also be used to express pleasant motion of some kind. It may also be reasonable to expect faster motion to be expressed by shorter duration, as in Shintel et al. (2006) (see also Knoeferle, Li, Maggioni, & Spence, 2017; Perlman & Cain, 2014).

3.1.5. Analysis

We analyzed the last repetition of each ideophone unless it was mispronounced. Ideophones whose form was changed from the intended target form, as in *buraburaburaaQ* or *buran* for *buraQ* 'taking a walk', were excluded from the data. A total of 557 recordings entered the following acoustic and statistical analysis (30 participants x 20 sentences – 43 exclusions; exclusion percentage = 7%). Although the participants were from different areas of Japan, we did not exclude any recordings, as the obtained pronunciation of the ideophones did not exhibit conspicuous dialectal variations.

Using Praat version 6.3.18 (Boersma & Weenink, 2023), we obtained the mean f_0 of the second vowel (V2) of each ideophone (e.g., /a/ of *buraQ*) and the mean intensity and duration of the entire ideophone (i.e., from the initial consonant to the beginning of the quotative particle *-to*) and standardized them within each participant. We focused on the f_0 of V2 specifically, as it is the locus of the pitch accent.

Using the brms package (Bürkner, 2017) with R version 4.4.0 (R Core Team, 2024), Bayesian mixed effects models were fit, with mean f_0, mean intensity, and duration as the dependent variables and the two mean ratings (speed and pleasantness), initial voicing of the ideophones, and V2 as fixed effects, as well as a random intercept for participant and a random slope for participant associated with each of the three fixed factors. The consonantal and vocalic factors were included in the current models, as they may influence f_0 in non-trivial ways. In particular, in the sound-symbolic system of Japanese ideophones, voiceless obstruents

in the word-initial position represent small, light, fine objects, as in *korokoro* 'a light object rolling' vs. *gorogoro* 'a heavy object rolling' (Hamano, 1998). The Frequency Code Hypothesis would predict that ideophones with voiceless obstruents are pronounced with higher f_0. There are also acoustic bases for the inclusion of these segmental factors. Voiceless obstruents are known to raise the f_0 of adjacent vowels, and voiced obstruents lower it (Kingston & Diehl, 1994). High vowels, such as [i] and [u], tend to have higher pitch than low vowels, such as [ɑ] (Ohala, 1973).

We used default weakly informative priors for the intercept and group-level standard deviations. Four MCMC chains were run with 2,000 iterations each, but the first 1,000 iterations were discarded as warm-up (burn-in) iterations. Convergence was assessed via R-hat statistics (all R-hat values = 1.00) and visual inspection of trace plots. To improve sampling efficiency and avoid divergent transitions, we set the target acceptance rate as adapt_delta = 0.97 and the maximum tree depth as max_treedepth = 15. All the analytical details can be checked as the R Markdown file at the project's OSF repository at https://osf.io/xrd96/?view_only=669137be82c14b27832361972155a6c2.

3.2. Results

3.2.1. Prosodic tendencies of individual ideophones

Overall, the prosodic properties of the pronounced ideophones were consistent with some previously reported sound–meaning associations in Japanese and beyond. Figure 1 presents the mean f_0 of the V2 of the 20 ideophones. Ideophones with an initial voiceless obstruent (/p, t, k, s, h/) tended to be pronounced with a higher f_0 than those with voiced obstruents onset (/b, d, g, z/)—note that these are not due to phonetic f_0 perturbation effects (Kingston & Diehl, 1994), as f_0 was measured at V2.

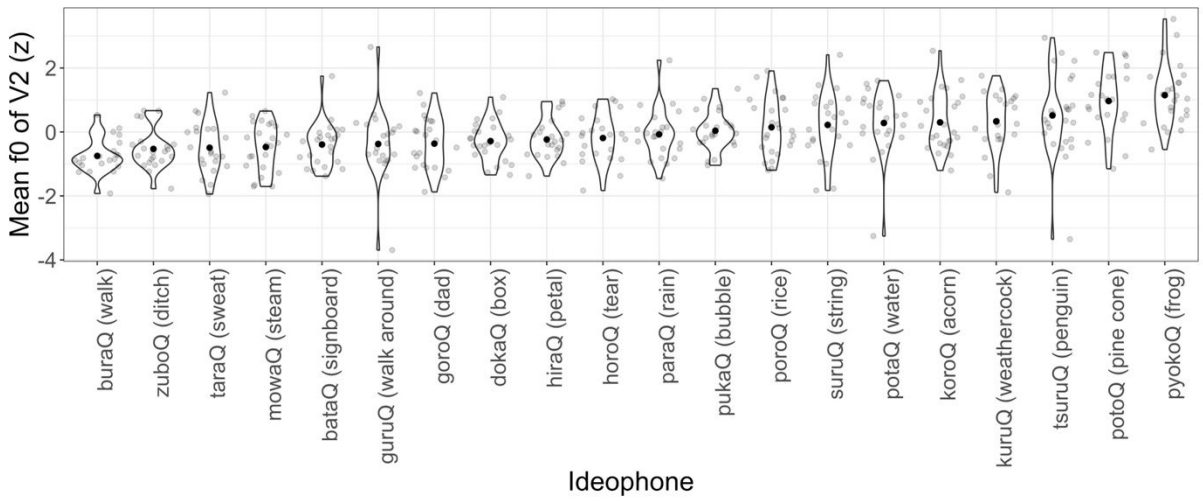


Figure 1. Mean standardized f_0 of the V2 of ideophones, from the lowest to the highest.

In Japanese ideophones, smallness and lightness are often represented by word-initial voiceless obstruents (Hamano, 1998). The current results suggest that the same size and weight information is also expressible by f_0: the higher the fundamental frequency, the smaller and lighter the referent. For example, *kuruQ* 'a light object spinning quickly once' was pronounced with a higher f_0 (M = 0.33) than *guruQ* 'going around, drawing a large circle' (M = –0.37). Similarly, *potoQ* represents the falling motion of a small light object and was generally produced with high f_0 (M = 0.97), whereas *bataQ* represents that of a heavy two-dimensional object and was produced with low f_0 (M = –0.39). These results appear to accord well with the Frequency Code Hypothesis.

Figure 2 shows the mean intensity of the 20 ideophones. It appears that ideophones that depict heavy objects' forceful movements tended to be pronounced strongly. For example, *goroQ* represents a person's lazy rolling movement on the floor and tended to be produced with high intensity (M = 0.73), whereas *pukaQ* represents a light object's floating motion and was generally produced with low intensity (M = –1.04).

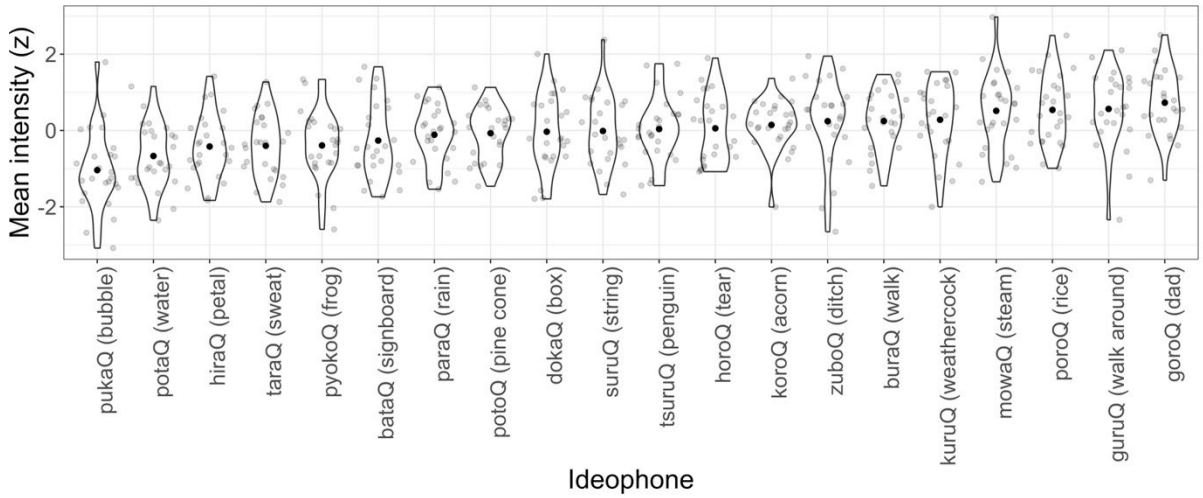


Figure 2. Mean standardized intensity of ideophones, from the lowest to the highest.

Figure 3 shows the mean duration of the 20 ideophones. /Q/-ending ideophones generally tend to represent quick movements. However, the relatively long duration of *mowaQ* 'steam/smoke coming out' (M = 1.23) reflects the slow movement it depicts.

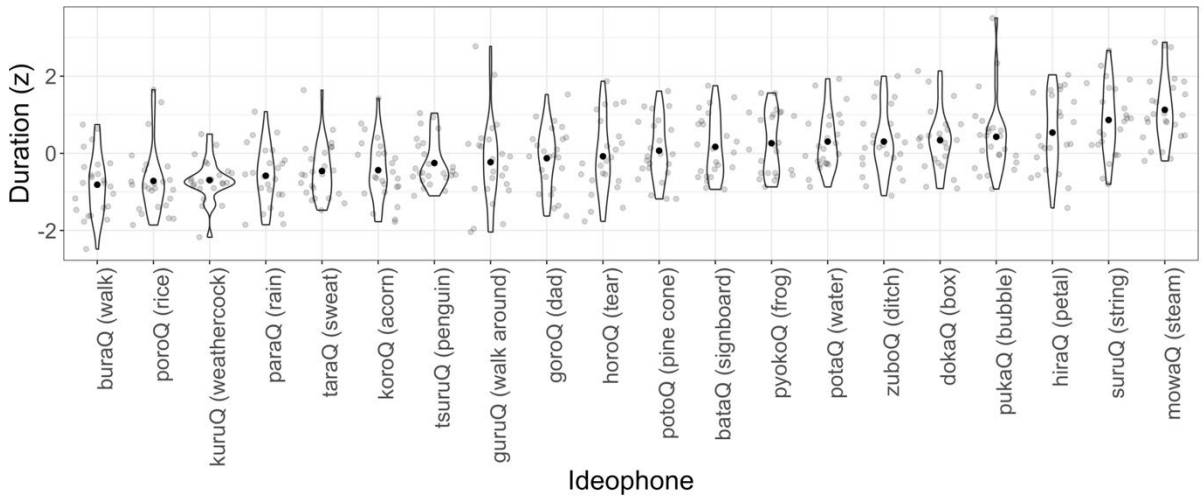


Figure 3. Mean standardized duration of ideophones, from the lowest to the highest.

3.2.2. Ideophone semantics and prosody

The subjective ratings obtained for these ideophones (see Table 1) allow us to quantitatively assess how ideophones' meanings and their prosodic features are correlated with each other. Figure 4 shows positive correlations between the two selected semantic dimensions of motion ideophones (i.e., speed and pleasantness) and the mean f_0 of V2. Ideophones that represent faster and pleasant motion, such as *pyokoQ* 'a little frog hopping once', tend to be pronounced with higher f_0 than those that represent slower and unpleasant motion, such as *goroQ* 'a heavy object rolling down, lying down'.

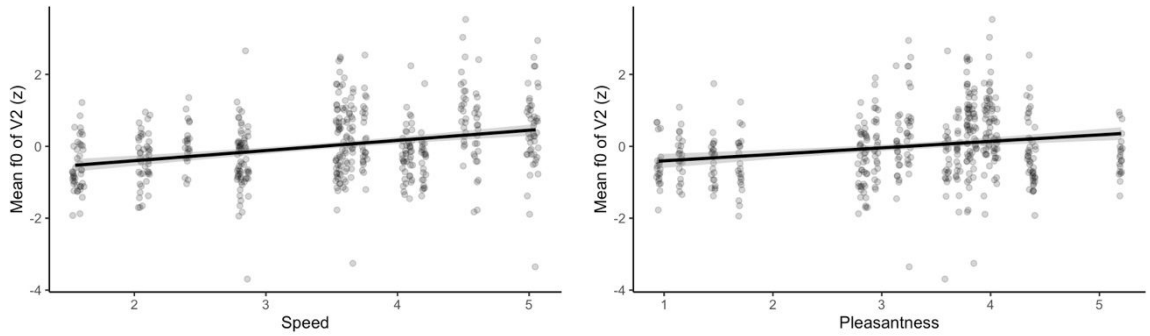


Figure 4. The speed and pleasantness of ideophones and the mean standardized f_0 of their V2.

Table 4 shows the results of the Bayesian regression model. The last column shows the probability that the posterior samples are either positive or negative, depending on the skew of the posterior distribution. These probabilities represent the certainty of the effects being credible. The Bayesian mixed regression model revealed positive, very credible effects of speed and pleasantness on f_0 at V2. Moreover, ideophones with V2 /o/ (e.g., *pyokoQ*, *potoQ*) tended to be pronounced with higher f_0 than those with V2 /a/. It might be that the small and

"inconspicuous" image associated with the vowel /o/ in Japanese ideophones (Hamano, 1998) were produced with higher f_0 via the Frequency Code.[2]

Table 4. The results of the Bayesian mixed regression model for the mean f_0 of the V2 of ideophones.

	Estimate	SE	95% CrI	$p_{direction}$
Intercept	–1.94	0.27	[–2.47, –1.41]	100%
Speed	0.33	0.06	[0.21, 0.45]	100%
Pleasantness	0.20	0.05	[0.10, 0.31]	100%
Voiceless	0.14	0.13	[–0.12, 0.39]	85.22%
/o/	0.31	0.10	[0.11, 0.51]	99.92%
/u/	–0.15	0.16	[–0.47, 0.16]	82.93%

Figure 5 shows the weak inverse relationship between the two semantic scales and mean intensity.

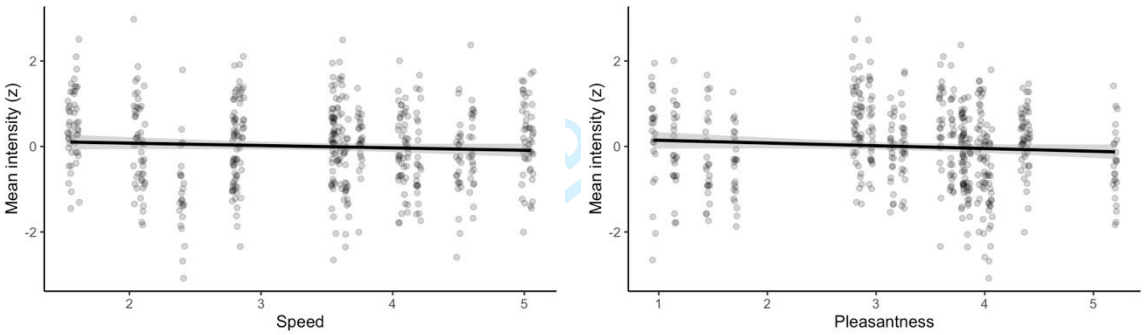


Figure 5. The speed and pleasantness of ideophones and their mean standardized intensity.

As shown in Table 5, a Bayesian regression model revealed that ideophones with an initial voiceless obstruent (e.g., *koroQ*) tended to be pronounced with lower intensity than those with an initial voiced obstruent (e.g., *goroQ*). Moreover, ideophones with V2 /o/ and those with V2 /u/ tended to be pronounced with higher intensity than those with V2 /a/. The effects of the two semantic scales, speed and pleasantness, were modest at best, however.

Table 5. The results of the Bayesian mixed regression model for the intensity of ideophones.

	Estimate	SE	95% CrI	$p_{direction}$
Intercept	0.34	0.27	[–0.21, 0.86]	89.40%
Speed	–0.09	0.06	[–0.21, 0.03]	92.73%
Pleasantness	–0.02	0.05	[–0.12, 0.09]	61.88%
Voiceless	–0.46	0.14	[–0.73, –0.19]	99.95%

[2] One may wonder whether this difference is due to their inherent f_0 difference, where mid vowels tend to be produced with higher pitch than low vowels. While this interpretation is also possible, we also note that [u] did not credibly differ from [a], despite the fact that [u] should show higher f_0 than [a] (Ohala, 1973).

/o/	0.53	0.10	[0.33, 0.72]	100%
/u/	0.67	0.17	[0.32, 1.02]	100%

Figure 6 shows the relationships between the two semantic scales and mean duration.

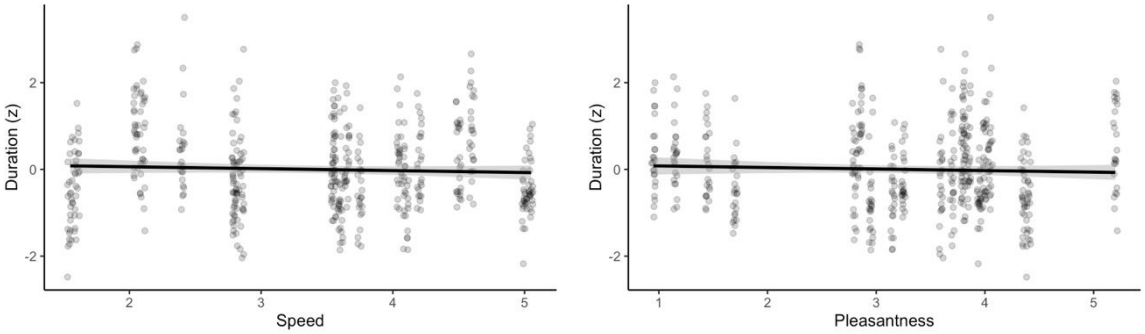


Figure 6. The speed and pleasantness of ideophones and their mean standardized duration.

As shown in Table 6, a Bayesian regression model revealed that ideophones with V2 /o/ tended to be pronounced with shorter duration that those with V2 /a/. No noticeable effects were found for the two semantic scales in the regression model.

Table 6. The results of the Bayesian mixed regression model for the duration of ideophones.

	Estimate	SE	95% CrI	$p_{direction}$
Intercept	0.07	0.29	[–0.50, 0.63]	58.80%
Speed	0.03	0.06	[–0.10, 0.15]	67.62%
Pleasantness	0.02	0.06	[–0.09, 0.14]	64.83%
Voiceless	–0.19	0.15	[–0.47, 0.10]	89.85%
/o/	–0.21	0.11	[–0.42, –0.01]	97.75%
/u/	–0.22	0.16	[–0.53, 0.09]	91.53%

3.3. Discussion

The production experiment demonstrated the iconic use of expressive prosody in Japanese ideophones. It showed that the association between pleasantness and f_0 that Nygaard et al. (2009) found with native speakers of English also holds with Japanese speakers producing ideophones. We also found that f_0 was associated with speed.

On the other hand, the effects of the semantic scales were modest—if not entirely absent—in terms of the intensity and duration of ideophones, which were more strongly affected by consonant and vowel types. It may be the case that the modest effects of intensity can partly be attributed to the uncontrolled recording environment; the distance between the participants' mouths and the recording device may not have been fixed across the tokens. A perception experiment using controlled audio stimuli, such as Experiment 2, will address this potential confound.

In addition to these findings, it was observed that some participants used marked voice quality to express nuanced semantic differences between motion ideophones. For example, one participant pronounced *dokaQ* 'a heavy object thudding' with a harsh, pressed voice to emphasize the violent sound and motion. Falsetto was used for *potoQ* 'a small light object dropping' and *pyokoQ* 'a little frog hopping once', both of which represent a light motion of a small entity. One participant used a voiceless pronunciation for the quick, violent motion expressed by *bataQ* 'a two-dimensional object slamming down'.

Ideophones generally have highly specific, holistic, multisensorial meanings, as suggested by their multiword translations used in this paper (Nuckolls, 2019). Therefore, it might be that the relationship between prosody and ideophones' semantic properties can be captured more intuitively in terms of specific voice quality categories, such as falsetto and creaky voice, at least more so than general prosodic features, such as f_0 and intensity. The relationship between ideophones and specific voice qualities will be further explored in Experiment 3, after testing the sound-symbolic significance of pitch, intensity, and duration in perception in Experiment 2.

4. Experiment 2: Perceived meaning of f_0, intensity, and duration

Experiment 1 demonstrated that Japanese speakers can utilize iconic prosody to highlight the meaning of ideophones. Considering that production experiments generally have a high degree of freedom, the iconic effects of each prosodic feature may manifest themselves in a clearer fashion in perception experiments. To this end, Experiments 2 and 3 examined whether Japanese speakers use iconic prosody in understanding ideophones.

4.1. Method

4.1.1. Participants

A total of 50 monolinguals speakers of Japanese (female: 26, male: 24; age: 22–62, $M = 42.04$, $SD = 9.19$) were recruited on CrowdWorks. None of them participated in Experiment 1. They were paid 300 JPY for their participation.

4.1.2. Stimuli

The same set of 20 simple ideophone sentences as Experiment 1 was used, but this time without IDS-like expressions, such as *papa* 'dad', *Pengin-san* 'Mr. Penguin', and the sentence-final particle *-ne*, which conveys "a soft tone of voice". The first author, who is a male native speaker of Japanese, pronounced all ideophones in eight (2 x 2 x 2) distinct prosodic patterns: high ($M = 145.10$ Hz, $SD = 5.16$) vs. low f_0 ($M = 118.86$ Hz, $SD = 2.99$), high ($M = 66.99$ dB, $SD = 1.83$) vs. low intensity ($M = 63.04$ dB, $SD = 1.51$), long ($M = 0.57$ s, $SD = 0.03$) vs. short duration ($M = 0.42$ s, $SD = 0.02$). Long pronunciation involved an extra-long V2, as in *pukaaaQ*, a strategy that is commonly found in the emphatic usage of ideophones. The same recording

of the non-ideophonic part of each sentence, which was pronounced in normal tones, was used for all eight pronunciations of each ideophone. All the sound files are available at the project's OSF repository mentioned above.

The sentences were presented in a random order, one sentence per page, on Google Forms, but the order of the eight audio files, labeled "A" to "H", was fixed due to the technical limitations of the platform: A: high f_0–high intensity–long duration, B: high–high–short, C: high–low–long, D: high–low–short, E: low–high–long, F: low–high–short, G: low–low–long, H: low–low–short.

4.1.3. Procedure

The participants were instructed to wear headphones or earphones and perform the task alone in a quiet environment. They were also instructed to listen to each recording as many times as they liked and choose the most suitable (or, if all sounded unnatural, most acceptable) pronunciation, from A to H, for the sentence.

4.1.4. Analysis

Separate Bayesian logistic regression models were fit for f_0 (high (A, B, C, D) vs. low (E, F, G, H)), intensity (high (A, B, E, F) vs. low (C, D, G, H)), and duration (long (A, C, E, G) vs. short (B, D, F, H)). The two mean semantic ratings (speed and pleasantness) for the ideophones, initial voicing, and V2 were included as fixed effects, a random intercept of participant as well as random slopes for the two semantic ratings, initial voicing, and V2. We used default weakly informative priors for the intercept and group-level standard deviations. In the f_0 model, four MCMC chains were run with 5,000 iterations each, of which the first 4,000 were discarded as warm-ups. In the intensity and duration models, four MCMC chains were run with 3,000 iterations each, of which 2,000 iterations were warm-up iterations. Convergence was assessed via R-hat statistics (R-hat values = 1.00) and visual inspection of trace plots. The other details are identical to those in Experiment 1. See the R Markdown file for further details.

4.2. Results

In general, we observe those patterns that accord well with the overall results of Experiment 1. Figure 7 shows the proportions of high- and low-frequency choices for each ideophone. High f_0 was preferred for ideophones with initial voiceless obstruents, which generally represent fast and pleasant motion, such as *pyokoQ* 'a little frog hopping once', and low f_0 was preferred for those with initial voiced obstruents, which represent slow and unpleasant motion, such as *goroQ* 'a heavy object rolling once, lying down'.

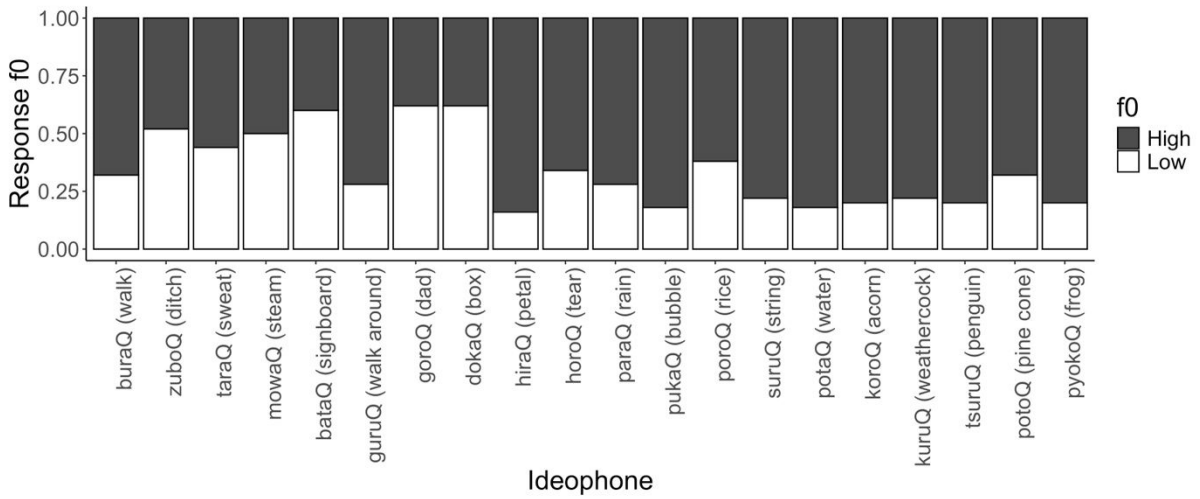


Figure 7. Proportions of high and low f_0 sounds (A, B, C, D vs. E, F, G, H) preferred for the 20 ideophones, ordered in the same way as the corresponding figure in Experiment 1.

A Bayesian regression model, summarized in Table 7, revealed credible effects of pleasantness and voicing. High f_0 was preferred for those with initial voiceless obstruents and those that represent pleasant motion.

Table 7. The results of the Bayesian mixed regression model for the preferred f_0 level.

	Estimate	SE	95% CrI	$p_{direction}$
Intercept	–1.95	0.56	[–3.05, –0.82]	99.92%
Speed	0.18	0.13	[–0.06, 0.42]	92.22%
Pleasantness	0.62	0.13	[0.38, 0.88]	100%
Voiceless	0.67	0.32	[0.06, 1.30]	98.50%
/o/	–0.31	0.20	[–0.71, 0.09]	93.80%
/u/	0.26	0.33	[–0.37, 0.97]	78.90%

Figure 8 shows the proportions of high- and low-intensity sounds chosen for each ideophone, which showed results that are more straightforward than the corresponding results of Experiment 1. Two ideophones for violent movements (*dokaQ* 'a hard object thudding down', *bataQ* 'a two-dimensional object slamming down') exhibited a strong preference for pronunciation with high intensity. In contrast, ideophones for light objects' quiet motion, such as *hiraQ* 'a light thin object fluttering down', *horoQ* 'a light teardrop falling', *potaQ* 'liquid dropping quietly', and *potoQ* 'a small light object dropping', preferred pronunciation with low intensity.

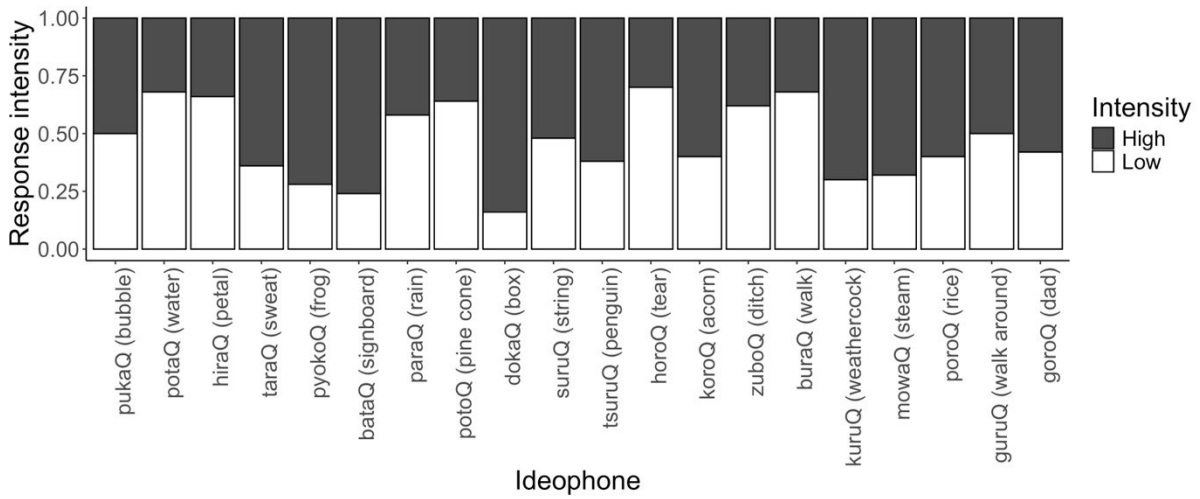


Figure 8. Proportions of high and low intensity sounds (A, B, E, F vs. C, D, G, H) preferred for the 20 ideophones, ordered in the same way as the corresponding figure in Experiment 1.

As shown in Table 8, a Bayesian regression model revealed credible effects of speed and pleasantness. Tokens with high intensity were preferred for ideophones for fast and unpleasant motion.[3]

Table 8. The results of the Bayesian mixed regression model for the preferred intensity level.

	Estimate	SE	95% CrI	$p_{direction}$
Intercept	–0.08	0.45	[–0.93, 0.80]	57.17%
Speed	0.32	0.10	[0.12, 0.52]	99.85%
Pleasantness	–0.17	0.09	[–0.35, 0.01]	97.00%
Voiceless	–0.37	0.22	[–0.79, 0.06]	95.12%
/o/	–0.16	0.17	[–0.49, 0.17]	83.25%
/u/	–0.03	0.25	[–0.51, 0.47]	55.25%

Figure 9 shows the preference for long vs. short rendition for each ideophone. As was the case with Experiment 1, our /Q/-ending ideophones generally represent quick motion and prefer short pronunciation, but long pronunciation was also chosen frequently for ideophones that represent relatively slow movements, such as *mowaQ* 'steam/smoke coming out', *buraQ* 'taking a walk', and *goroQ* 'a heavy object rolling once, lying down'.

[3] We do not have a straightforward explanation as to why intensity was associated with slower motion in Experiment 1 but with faster motion in the current experiment. One possible reason may lie in the ambivalent relationship between energy and speed: while generating high speed typically requires more energy (i.e., $F = ma$), slower movement may also imply motion against greater resistance.

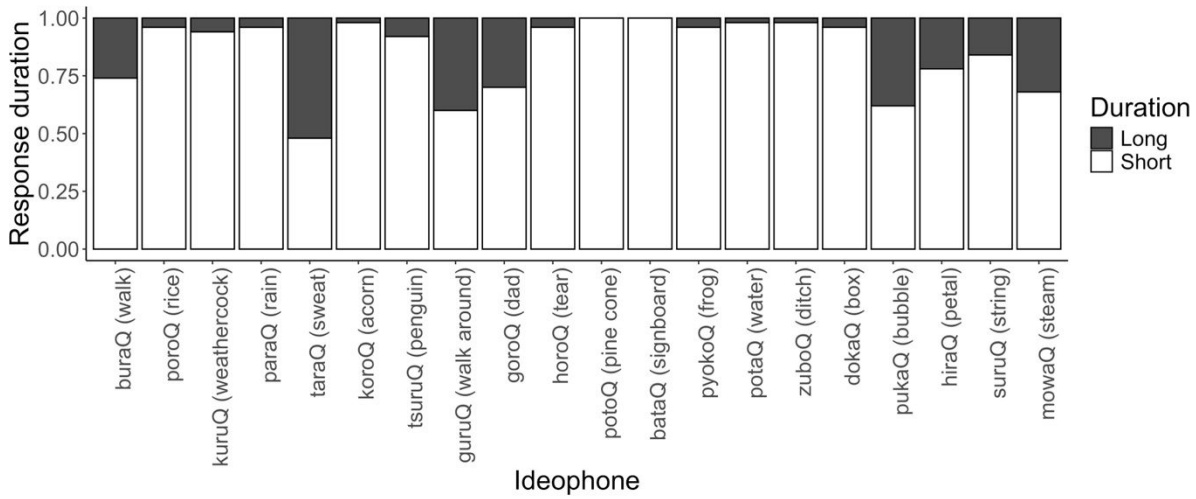


Figure 9. Proportions of long and short sounds (A, C, E, G vs. B, D, F, H) preferred for the 20 ideophones, ordered in the same way as the corresponding figure in Experiment 1.

As shown in Table 9, a Bayesian regression model revealed credible effects of speed, pleasantness, voicing, V2 /o/, and /u/. Long duration was preferred for those with initial voiceless obstruents, V2 /u/, and slow and unpleasant motion, whereas short duration was preferred for V2 /o/ and fast and pleasant motion.

Table 9. The results of the Bayesian mixed regression model for the preferred duration.

	Estimate	SE	95% CrI	$p_{direction}$
Intercept	7.20	1.14	[5.04, 9.46]	100%
Speed	–2.71	0.35	[–3.44, –2.07]	100%
Pleasantness	–1.10	0.18	[–1.46, –0.76]	100%
Voiceless	2.32	0.43	[1.51, 3.19]	100%
/o/	–1.95	0.63	[–3.39, –0.93]	100%
/u/	3.77	0.58	[2.65, 4.91]	100%

4.3. Discussion

The current perception experiment provided additional, and arguably even clearer, evidence for the sound-symbolic relevance of prosody in Japanese ideophones. Higher f_0 was, like Experiment 1, associated with more pleasant motion. Higher intensity was associated with faster and less pleasant motion, which may involve greater energy. Longer duration was associated with slower motion and, contrary to the results of Experiment 1, greater pleasantness, extending the scope of Shintel et al.'s (2006) findings. This association between slow motion and pleasantness—rather than unpleasantness—may be due to specific motion contexts where someone or something moves in a relaxed, leisurely manner, illustrated by *buraQ* 'taking a walk', *goroQ* 'lying down', and *pukaQ* 'a light object floating'.

5. Experiment 3: Perceived meaning of voice quality

In the final experiment, we went a step further into the nature of iconic prosody and investigated how different voice quality categories may interact with the meanings of Japanese ideophones. People do not actively use marked voice quality in non-interactive experimental settings such as Experiment 1. Therefore, in order to explore the question of whether certain voice qualities are favored for particular meanings, we recorded ideophones pronounced with marked voice quality and asked Japanese speakers to evaluate them in a perception experiment.

5.1. Method

5.1.1. Participants

Fifty-one monolingual speakers of Japanese who did not participate in either Experiment 1 or 2 (female: 21, male: 29, prefer not to answer: 1; age: 22–66, M = 42.82, SD = 10.24) were recruited on CrowdWorks. They were paid 300 JPY for their participation.

5.1.2. Stimuli

The same set of 20 simple ideophone sentences as Experiment 2 was used. The first author, who is a male native speaker of Japanese, pronounced all ideophones with four types of marked phonation: creaky voice (a low-pitched voice with audible pulses, as in the end of an old man's utterance), harsh voice (a rough, pressed voice produced with tensed vocal folds, similar to the sound that one often makes when lifting a heavy object), falsetto, and whisper. The rest of the sentences was pronounced in a modal voice. These sounds are available at the OSF repository. The order of the sentences was randomized on Google Forms, but the order of the four voice qualities for each sentence was fixed: creaky voice, harsh voice, falsetto, and whisper.

5.1.3. Procedure

The participants were instructed to wear headphones or earphones and complete the task alone in a quiet environment. They were also instructed to listen to each recording as many times as they liked and choose the most suitable (or, if all sounded unnatural, most acceptable) pronunciation for the sentence: the first, second, third, or fourth.

5.1.4. Analysis

A Bayesian multinomial logistic regression model was fit, with the four voice quality categories as a dependent variable, the two mean semantic ratings (speed and pleasantness) for the ideophones, initial voicing, and V2 as fixed effects; the random effect structure was identical to that of Experiments 1 and 2. The model consisted of four chains with 2,000 iterations, with a warm-up period of 1,000 iterations. We set the target acceptance rate as adapt_delta = 0.97 and the maximum tree depth as max_treedepth = 15. See the R Markdown file at the study's OSF repository for details.

5.2. Results

As shown in Figure 10, different ideophones exhibited different preferences for the four voice qualities, and the iconic basis of these sound–meaning associations appears to be fairly straightforward to interpret. Harsh voice was preferred for ideophones that represent a violent sound and motion, such as *zuboQ* 'one's foot falling into a ditch', *dokaQ* 'a hard object thudding down', and *bataQ* 'a two-dimensional object slamming down'. Falsetto was preferred for ideophones that represent a light object's fast motion, such as *tsuruQ* 'slipping', *koroQ* 'a light object rolling once', *kuruQ* 'a light object spinning quickly once', *potoQ* 'a small light object dropping', and *pyokoQ* 'a little frog hopping once'. This result is again consistent with the Frequency Code Hypothesis (Ohala, 1984).[4] Whisper was preferred for ideophones that represent quiet motion, such as *mowaQ* 'steam/smoke coming out', *paraQ* 'small light drops falling', *hiraQ* 'a light thin object fluttering down', *horoQ* 'a light teardrop falling', *suruQ* 'a light object going off quietly', and *potaQ* 'liquid dropping quietly'.

No straightforward semantic generalization appears to hold for ideophones that preferred creaky voice: *buraQ* 'taking a walk', *goroQ* 'a heavy object rolling once', *guruQ* 'going around', *taraQ* 'liquid dropping', *pukaQ* 'a light object floating', and *poroQ* 'a small light object dropping'. This is probably because creaky voice (i.e., a very low-pitched voice) generally sounded most natural and least marked for the male speaker.

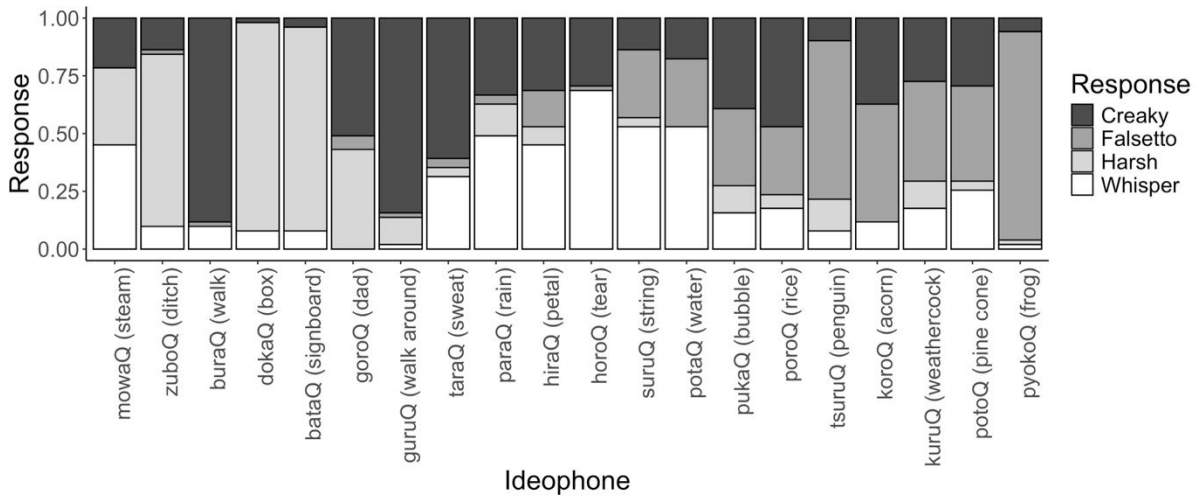


Figure 10. Proportions of the four voice qualities preferred for the 20 ideophones.

As shown in Table 10, a Bayesian regression model revealed that, consistent with the results of Experiments 1 and 2, falsetto (i.e., a distinctly high-pitched sound) was preferred to creaky voice for faster and more pleasant motion. This result may be related to the actual

[4] Acoustic measurement of the stimulus ideophones revealed that, replicating Experiment 1, high f_0 was preferred for ideophones for quick motion, such as *kuruQ* 'a light object spinning quickly once' and *pyokoQ* 'a little frog hopping once'.

example of an ideophonic utterance cited in Section 1, in which *hyoi* pronounced in a falsetto appeared to express suddenness and unexpected happiness. On the other hand, harsh voice was preferred to creaky voice for ideophones depicting faster but less pleasant motion. Whisper was preferred to creaky voice for ideophones for faster motion.

Table 10. The results of the Bayesian regression model for the preferred voice qualities, with creaky voice as a baseline. The factors that are of particular interest are highlighted in bold.

		Estimate	SE	95% CrI	$p_{direction}$
Intercept	Falsetto	–9.69	0.98	[–11.65, –7.84]	100%
	Harsh	–0.90	0.79	[–2.50, 0.63]	87.42%
	Whisper	–2.23	0.59	[–3.43, –1.05]	100%
Falsetto	**Speed**	**1.54**	**0.20**	**[1.15, 1.95]**	**100%**
	Pleasantness	**0.85**	**0.18**	**[0.50, 1.22]**	**100%**
	Voiceless	0.84	0.59	[–0.27, 2.04]	93.00%
	/o/	0.56	0.28	[0.01, 1.10]	97.60%
	/u/	–1.08	0.42	[–1.92, –0.25]	99.75%
Harsh	**Speed**	**1.08**	**0.19**	**[0.70, 1.45]**	**100%**
	Pleasantness	**–0.59**	**0.16**	**[–0.91, –0.27]**	**99.98%**
	Voiceless	–2.60	0.45	[–3.52, –1.77]	100%
	/o/	–0.25	0.30	[–0.83, 0.35]	79.75%
	/u/	–1.47	0.48	[–2.52, –0.59]	99.95%
Whisper	**Speed**	**0.52**	**0.15**	**[0.22, 0.82]**	**100%**
	Pleasantness	0.04	0.11	[–0.19, 0.25]	61.48%
	Voiceless	0.76	0.30	[0.18, 1.36]	99.52%
	/o/	–0.70	0.22	[–1.14, –0.27]	99.88%
	/u/	–1.32	0.36	[–2.02, –0.62]	99.98%

5.3. Discussion

The current results show that the complex semantics of motion ideophones can be iconically associated with specific voice quality categories. Crucially, raw acoustic features, such as f_0 and intensity, do not fully account for these associations. Specifically, while in Experiments 1 and 2, fast speed was generally associated with pleasantness, harsh voice was associated with fast but unpleasant motion. Likewise, while low intensity evoked a slow image in Experiment 2, whisper, which is a low intensity sound, was associated with fast motion. These results indicate that Japanese speakers can refer to the acoustic details of ideophones—demonstrably both in terms of raw acoustic features such as f_0 as well as voice qualities—in their iconic interpretation of these words.

As an anonymous reviewer pointed out, it is worth mentioning that semantic scales and voice quality categories do not always have one-to-one correspondence. For example, in Experiment 1, one participant pronounced *bataQ* 'a two-dimensional object slamming down' without regular vibrations of the vocal folds (i.e., in a whisper voice). However, in the current experiment, this ideophone exhibited a strong preference for harsh voice, arguably due to the violent movement it represents. This oscillation indicates that iconic prosody can not only

emphasize the semantic components inherent in individual ideophones but may also be able to add a new dimension to them.

6. General discussion

The three experiments demonstrated that the prosody of real words can be used and understood iconically. These findings confirm the importance of iconic prosody in human communication, which has been primarily investigated using novel words and nonlinguistic vocalizations in the previous studies. Iconic prosody is attracting broad attention in cognitive science, as it is one of the major candidates for the original form of human language (Arbib, Liebal, & Pika, 2008; Ć wiek et al., 2021; Haiman, 2018; Perlman, 2026). The iconic prosody of ideophones examined in the current study may be able to provide a missing link in this evolutionary theory.

Ideophones are imitative lexemes that are iconic and conventionalized at the same time. Unlike iconic vocalizations (Ćwiek et al., 2021), which are primarily prosodic and do not consist of consonants and vowels, ideophones are segmentally specified in the lexicon of an individual language. One can speculate that ideophones inherited iconic prosody from nonlinguistic vocalizations and gradually lost it to form a conventionalized system of non-imitative *symbols* (i.e., signs whose form–meaning relationship is arbitrary; Peirce, 1932).

This hypothesis gains additional support from the fact that the iconic prosody of ideophones can be adjusted in a graded manner. For example, *pyokoQ-to* 'with a light hop' can be pronounced in plain prosody [pʲokót:o], expressive prosody [pʲǒkőt:o] (extra-short V1, extra-high-pitched V2), and even more expressive prosody [{F pʲǒkőt:o F}] (extra-short V1, falsetto) (for a related observation, see Rhode, 1994). Thus, as depicted in Figure 11, iconic prosody allows us to draw a fine-grained evolutionary path from nonlinguistic vocalizations to non-imitative, symbolic words via ideophones with varying degrees of expressiveness. Analyzing the iconic function of ideophone prosody might reveal how this evolution may have taken place—for example, what type of meaning is lexicalized first.

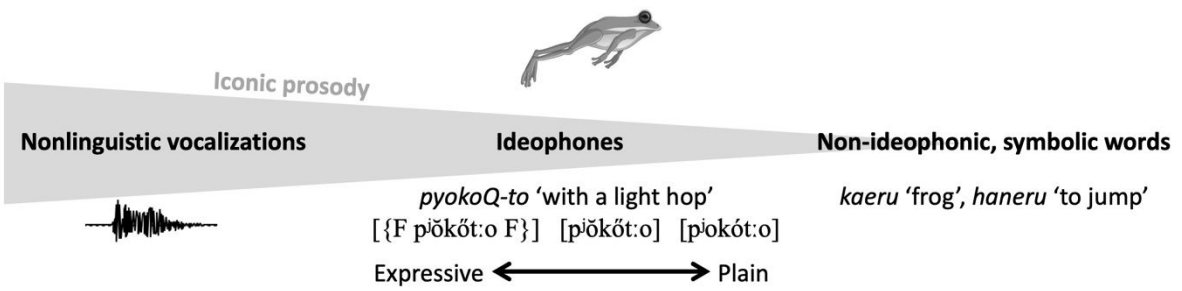


Figure 11. Possible evolutionary path from nonlinguistic vocalizations to non-ideophonic, symbolic words via ideophones.

7. Conclusion

This paper has examined the relevance of prosody in both the production and perception of ideophonic expressions in Japanese. We have demonstrated that Japanese speakers utilize iconic prosody for enhancing the depictive power of each ideophone and that they also have a clear preference regarding which acoustic properties should accompany what kinds of meanings. It is also worth noting that iconic prosody can be observed at different levels of abstraction. The association between high f_0 and speed was confirmed in all three experiments, but the results of Experiment 3 also indicated that specific semantic aspects of ideophones can be iconically linked with a specific type of voice quality. The iconic prosody of ideophones can be considered a remnant of nonlinguistic vocalizations as an early form of spoken language. This remnant may connect iconic vocalizations and arbitrary words, which are otherwise far apart in the relevant evolutionary theory.

This study is the first step toward a comprehensive examination of iconic prosody in real words and opens up research opportunities in various respects. For example, future research should examine how iconic prosody works in semantic domains other than motion, including more abstract ones such as pain and emotion (McLean, 2021). Another future direction would be a finer-grained psychoacoustic analysis of iconic prosody using continuous measures of voice quality instead of discrete categories (Lacey et al., 2020; Villegas et al., 2023). Furthermore, a cross-linguistic comparison of prosody–meaning associations may enrich our evolutionary considerations. The hypothesis outlined in Section 6 would predict that iconic prosody should manifest themselves across languages, which should be empirically tested in future studies. Finally, a qualitative analysis of actual ideophone uses in specific discourse, such as the ideophone in a falsetto cited in the beginning of this paper, may help us to better understand the meaning of iconic prosody.

Acknowledgments

We are grateful to Laura Speed and the anonymous reviewer for their insightful comments on an earlier version of this paper.

Competing interests

The authors have no competing interests to declare.

Data availability

All stimuli, experimental instructions, data, and code are available at https://osf.io/xrd96/?view_only=669137be82c14b27832361972155a6c2.

References

Akita, K. (2020a). Modality-specificity of iconicity: The case of motion ideophones in Japanese. In P. Perniss, O. Fischer, & C. Ljungberg (Eds.), *Operationalizing iconicity* (pp. 3–20). Amsterdam: John Benjamins.

Akita, K. (2020b). System integration of Japanese ideophones. Ms., Nagoya University. https://drive.google.com/file/d/1-fYXYjbRXFhkr0C7gnKq3kZ9RvzwcAnV/view?usp=sharing

Akita, K. (2021). Phonation types matter in sound symbolism. *Cognitive Science*, *45*(5), e12982.

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.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076.

Ball, M. J., Howard, S. J., & Miller, K. (2018). Revisions to the extIPA chart. *Journal of the International Phonetic Association*, *48*(2), 155–164.

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer. Version 6.3.17, retrieved 8 October 2023.

Bürkner, P-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

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.

Ćwiek, A. et al. (2021). Novel vocalizations are understood across cultures. *Scientific Reports*, *11*, 10108.

Ć wiek, A., & Fuchs, S. (2019). Iconic prosody is rooted in sensori-motor properties: Fundamental frequency and the vertical space. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 1572–1578. Merced, CA: University of California, Merced.

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.

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

Everett, D. L. (2017). *How language began: The story of humanity's greatest invention*. New York: Liveright.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501.

Ferrara, C., Lu, J. C., & Goldin-Meadow, S. (2025). Playing with language in the manual modality: Which motions do signers gradiently modify? *Cognitive Science*, *49*(4), e70051.

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 63–88). Cambridge, UK: Cambridge University Press.

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, *8*(2), 425–434.

Haiman, J. (2018). *Ideophones and the evolution of language*. Cambridge: Cambridge University Press.

Hamano, S. (1998). *The sound-symbolic system of Japanese*. Tokyo: Kurosio.

Hancil, S., & Hirst, D. (Eds.). 2013. *Prosody and iconicity*. Amsterdam: John Benjamins.

Herold, D. S. (2006). *Acoustic correlates to word meaning in infant directed speech*. Ph.D. dissertation, Emory University.

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology*, *108*, 229–241.

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.

Ibarretxe-Antuñano, I. (2019). Towards a semantic typological classification of motion ideophones: The motion semantic grid. In K. Akita & P. Pardeshi (Eds.), *Ideophones, mimetics, and expressives* (pp. 137–166). Amsterdam: John Benjamins.

Igarashi, Y., Nishikawa, K., Tanaka, K., & Mazuka, R. (2013). Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *The Journal of the Acoustical Society of America*, *134*, 1283–1294.

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.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*, 5562.

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal*

Learning and Verbal Behavior, *10*, 427–429.

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*(9), e12883.

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

Mazuka, R., Igarashi, Y., Martin, A., & Utsugi, A. (2015). Infant-directed speech as a window into the dynamic nature of phonology. *Laboratory Phonology*, *6*(3-4), 281–303.

Michelini, L., & Nygaard, L. C. (2025). Size-sound iconicity in English-like pseudowords influences referent labeling and prosody. *Cognitive Science*, *49*(2), e70042.

McLean, B. (2021). Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology*, *25*(3), 507–549.

Motoki, K., Pathak, A., & Spence, C. (2022). Tasting prosody: Crossmodal correspondences between voice quality and basic tastes. *Food Quality and Preference*, *100*, 104621.

Nuckolls, J. B. (2019). The sensori-semantic clustering of ideophonic meaning in Pastaza Quichua. In K. Akita & P. Pardeshi (Eds.), *Ideophones, mimetics and expressives* (pp. 167–198). 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*(1), 127–146.

Ohala, J. J. (1973). Explanations for the intrinsic pitch of vowels. In *Monthly Internal Memorandum* (pp. 1–14). Berkeley, CA: University of California.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.

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

Perlman, M. (2026). Iconic prosody is deeply connected to iconic gesture, and it may occur just as frequently. In O. Fischer, K. Akita, & P. Perniss (Eds.). *The Oxford handbook of iconicity in language*. Oxford: Oxford University Press.

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350.

Perlman, M., Dale, R., & 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.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227.

R Core Team. (2024). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rhodes, R. (1994). Aural images. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 276–292). Cambridge, UK: Cambridge University Press.

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.

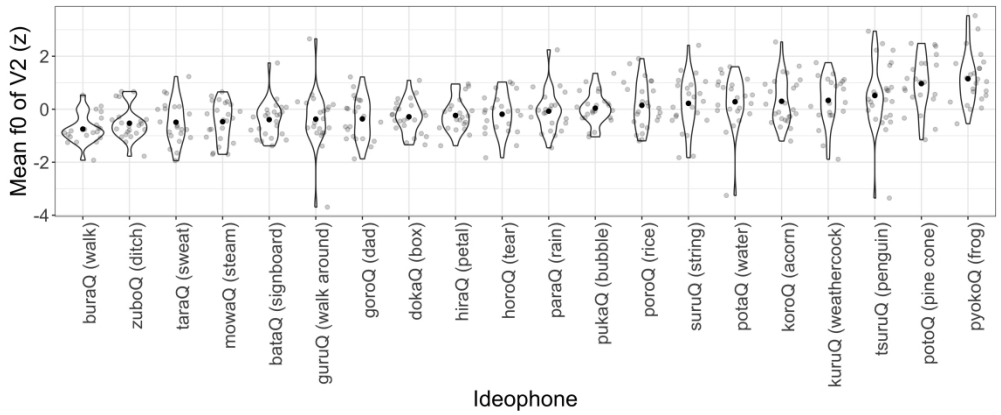
Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*, 167–177.

Stolarski, Ł. (2019). Correlations between positive or negative utterances and basic acoustic features of voice: A preliminary analysis. *Research in Language*, *20*(2), 153–178.

Toratani, K. (2012). The role of sound-symbolic forms in Motion event descriptions: The case of Japanese. *Review of Cognitive Linguistics*, *10*, 90–132.

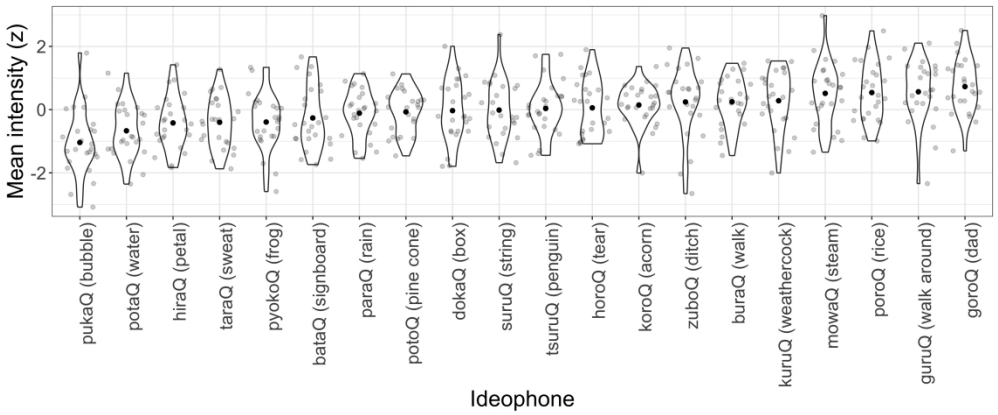
Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B*, *369*, 20130292.

Villegas, J., Akita, K. & Kawahara, S. (2023). Psychoacoustic features explain subjective size and shape ratings of pseudo-words. *Proceedings of the 10th Convention of the European Acoustics Association: Forum Acusticum 2023*.



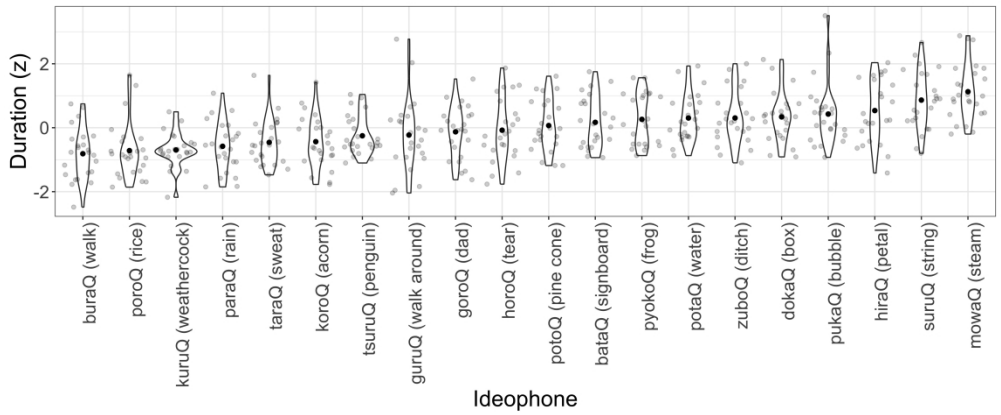
Mean standardized f0 of the V2 of ideophones, from the lowest to the highest.

1481x617mm (72 x 72 DPI)



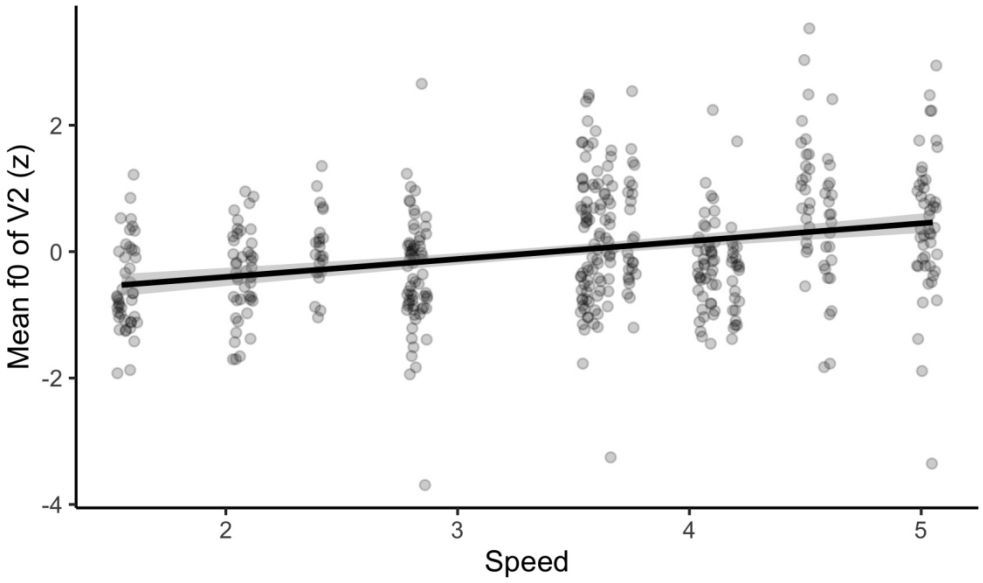
Mean standardized intensity of ideophones, from the lowest to the highest.

1481x617mm (72 x 72 DPI)



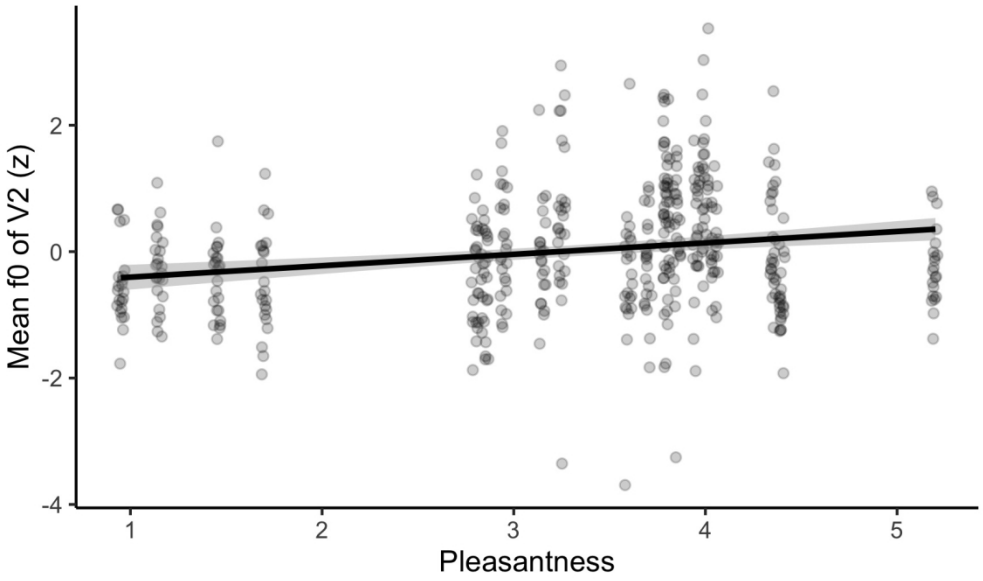
Mean standardized duration of ideophones, from the lowest to the highest.

1481x617mm (72 x 72 DPI)



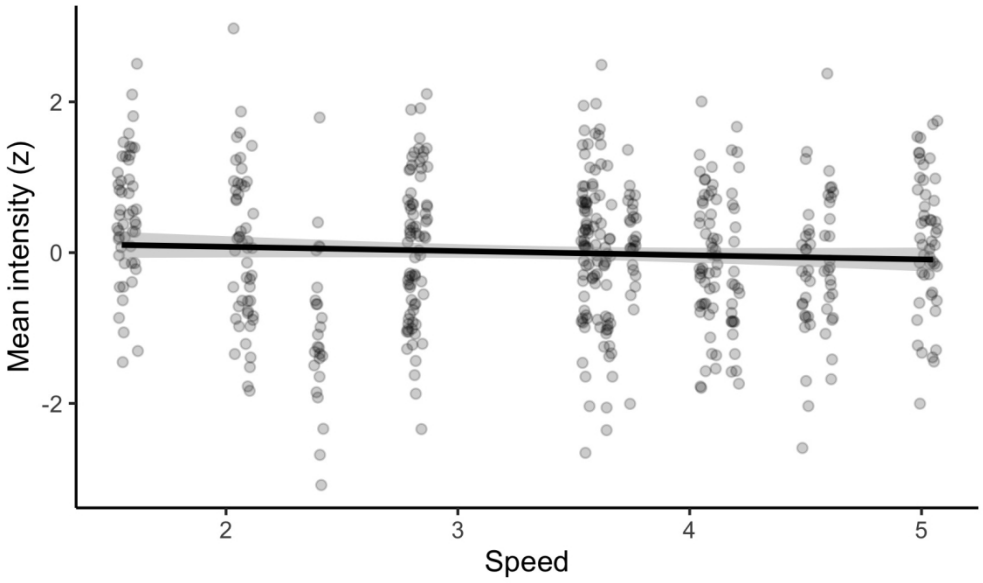
The speed and pleasantness of ideophones and the mean standardized f0 of their V2.

617x370mm (72 x 72 DPI)



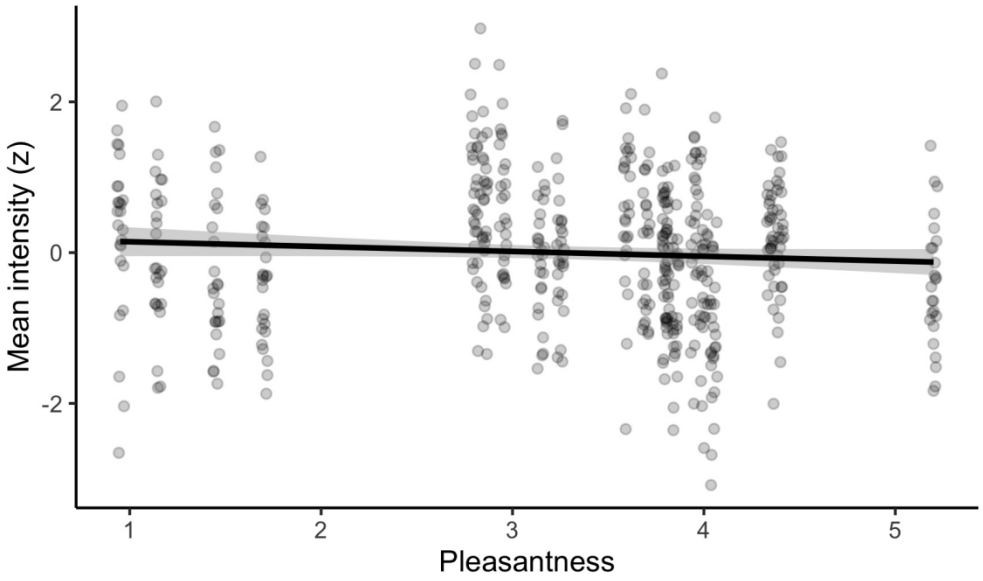
The speed and pleasantness of ideophones and the mean standardized f0 of their V2.

617x370mm (72 x 72 DPI)



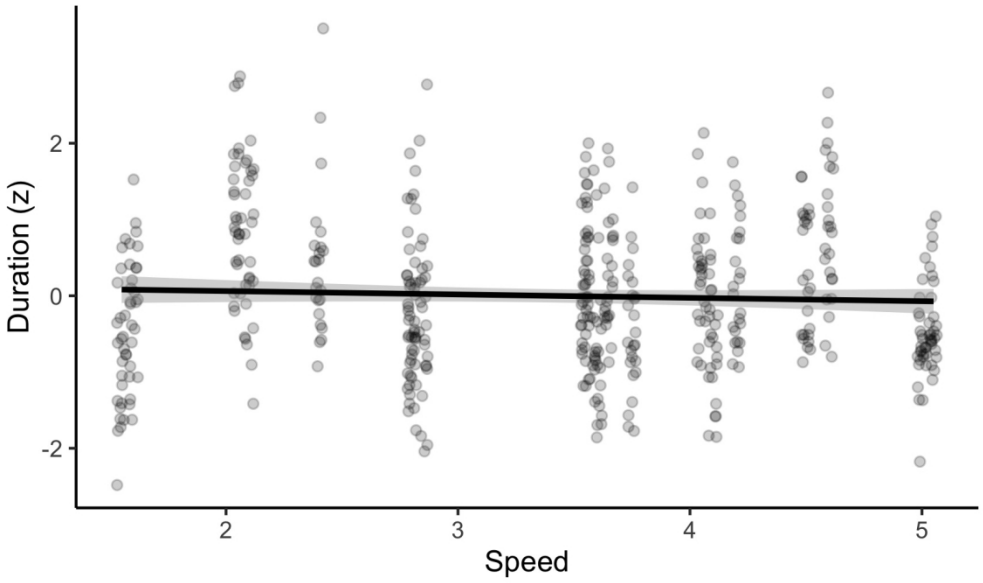
The speed and pleasantness of ideophones and their mean standardized intensity.

617x370mm (72 x 72 DPI)



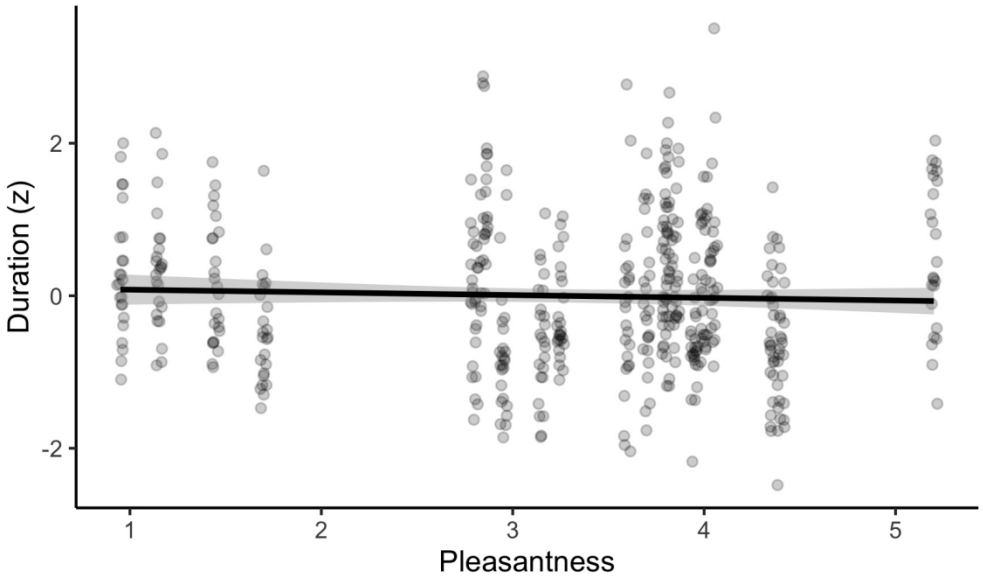
The speed and pleasantness of ideophones and their mean standardized intensity.

617x370mm (72 x 72 DPI)



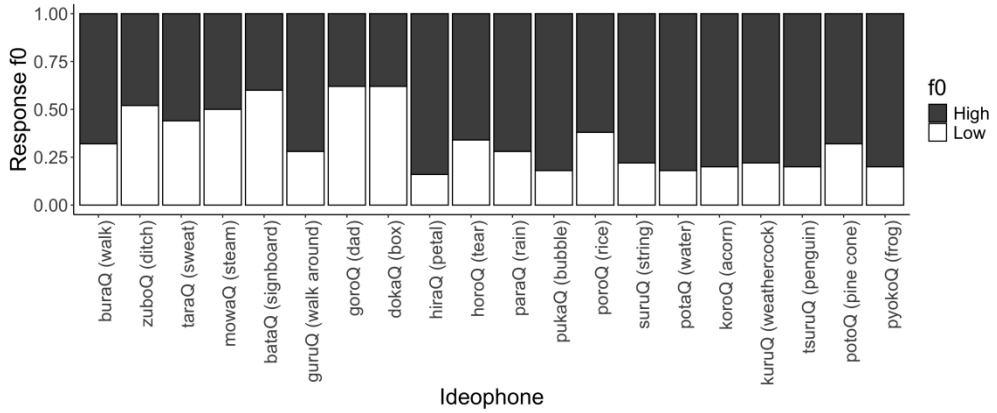
The speed and pleasantness of ideophones and their mean standardized duration.

617x370mm (72 x 72 DPI)



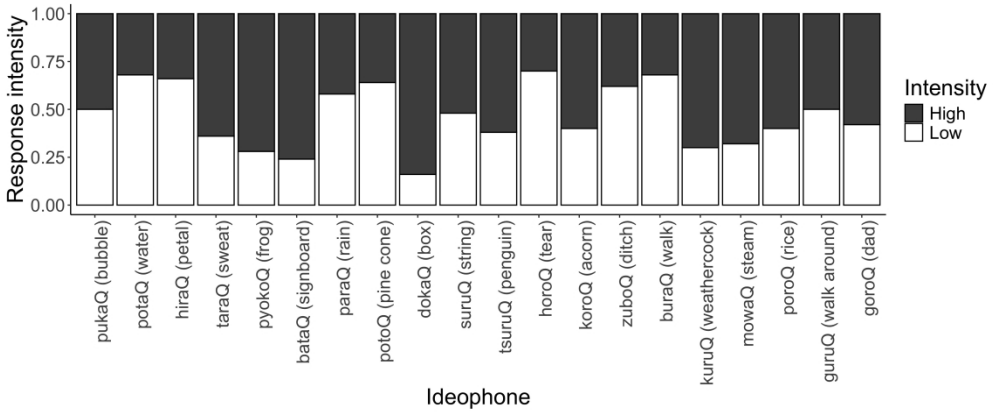
The speed and pleasantness of ideophones and their mean standardized duration.

617x370mm (72 x 72 DPI)



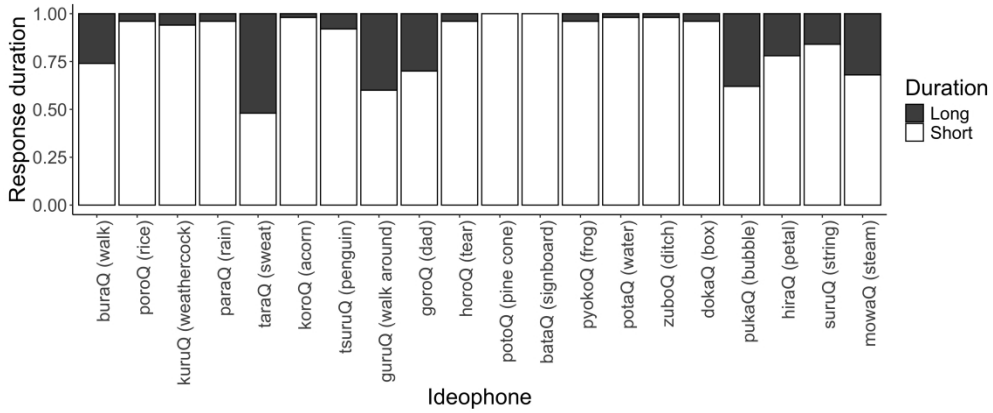
Proportions of high and low f0 sounds (A, B, C, D vs. E, F, G, H) preferred for the 20 ideophones, ordered in the same way as the corresponding figure in Experiment 1.

1481x617mm (72 x 72 DPI)



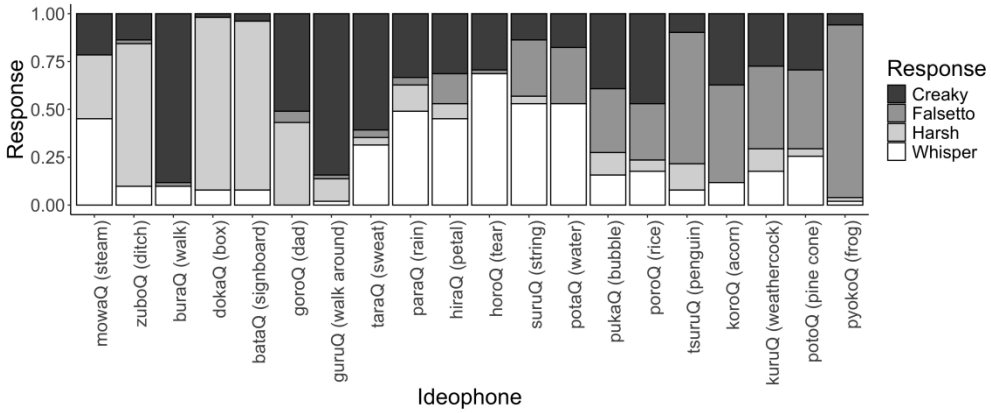
Proportions of high and low intensity sounds (A, B, E, F vs. C, D, G, H) preferred for the 20 ideophones, ordered in the same way as the corresponding figure in Experiment 1.

1481x617mm (72 x 72 DPI)



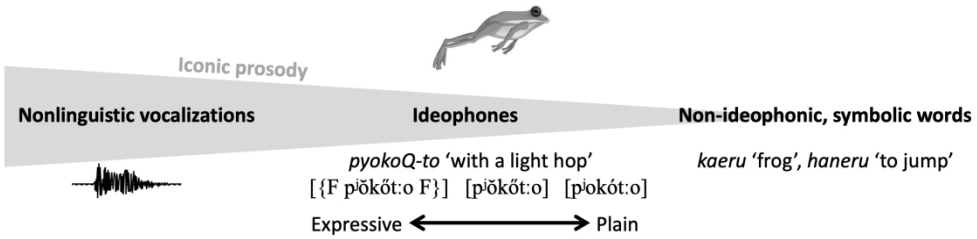
Proportions of long and short sounds (A, C, E, G vs. B, D, F, H) preferred for the 20 ideophones, ordered in the same way as the corresponding figure in Experiment 1.

1481x617mm (72 x 72 DPI)



Proportions of the four voice qualities preferred for the 20 ideophones.

1481x617mm (72 x 72 DPI)



Possible evolutionary path from nonlinguistic vocalizations to non-ideophonic, symbolic words via ideophones.

438x112mm (144 x 144 DPI)