MDPI

Article

Exploring the Positional Effects in Sound Symbolism: The Case of Hardness Judgments by English and Japanese Speakers

Kazuko Shinohara * and **Ryoko Uno**

Institute of Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan; ryokouno@cc.tuat.ac.jp
* Correspondence: k-shino@cc.tuat.ac.jp

Abstract: Sound symbolism refers to the systematic and iconic relationships between sounds and meanings. While the research on this topic is growing rapidly, one issue that is understudied in the literature is whether segments in psycholinguistically salient positions (e.g., word-initial position) exhibit stronger sound–symbolic effects. We report two experiments that, together, show Japanese speakers are more sensitive to the sound–symbolic values of segments in initial syllables, whereas English speakers, rather surprisingly, are more sensitive to the sound–symbolic values of segments in final syllables. This cross-linguistic difference is intriguing, and we believe it opens up opportunities for future experimentation.

Keywords: sound symbolism; hardness; positional effects; English; Japanese

1. Introduction

How (in)direct is the relationship between sounds and meanings? It is safe to say that this question is almost as old as how far back we can track the intellectual history of human beings: Plato deals extensively with this topic in the dialogue *Cratylus* (Harris and Taylor 1989). Therein, two thinkers present arguments from two different points of view—one in which the relationships between sounds and meanings are completely arbitrary and only socially determined, and the other in which the relationships between sounds and meanings can be systematic and intrinsic. Socrates sides with the latter view, suggesting that words are tools to represent meanings and that tools should generally be shaped in such a way that they can serve their purposes; therefore, it only makes sense that the sounds of the words represent what they mean in one way or another. They list several concrete cases of sound–meaning correspondences, such as [a] being large, [i] being small, and [o] being round, which, as it turns out, would be corroborated by modern psychological research with robust experimental methodology about 2500 years later.

The very debate presented in *Cratylus* has been repeatedly discussed in various guises throughout the history of human intellectual development. However, modern linguistic theories in the 20th century were more or less dominated by Saussure's (1916) influential dictum that the relationships between sounds and meanings are in principle arbitrary, perhaps with the exception of "marginal" vocabulary items such as onomatopoetic expressions. Saussure (1916) actually raises this notion of arbitrariness as the first principle of natural languages. To boil down the argument by Saussure, if the relationship between sounds and meanings were fixed, and then human languages should use the same sound sequences to denote the same meaning, the logical consequence would be that all languages should use the same word to represent the same object, which is obviously wrong from an empirical perspective. Since different languages use different sets of sounds (phonemes) and have different sets of phonotactic restrictions, this logic cannot be taken at face value (see Styles and Gawne 2017, for a related discussion); nevertheless, it is also true that different languages use widely different sets of names to denote the same object. Hockett (1959) is

another influential work, fostering this idea that sound–meaning relationships in human languages are arbitrary. They take this arbitrariness as one of the design features of human languages, which distinguishes them from other animal communication systems. It is safe to say that the arbitrariness of the relationships between sounds and meanings had been taken for granted by most if not all linguists in the 20th century.

While the principle of arbitrariness appears to be an important feature of natural languages, there are cases in which some sounds do systematically evoke some particular meanings, patterns that are generally known as "synaesthetic sound symbolism" or, more simply, "sound symbolism". To borrow the definition from the existing literature, Hinton et al. (1994, p. 4), for example, define synaesthetic sound symbolism as "the process whereby certain vowels, consonants, and suprasegmentals are chosen to consistently represent visual, tactile, or proprioceptive properties of objects, such as size or shape" (for an almost synonymously defined term "phonaesthesia", see Berlin (2006, p. 26)). In the rest of this paper, we use the term "sound symbolism" for the sake of simplicity. Perhaps one of the most well-known cases of sound symbolism is, as Socrates already pointed out in *Cratylus*, that [i] is known to represent smallness in various languages. The English diminutive suffix <-y> (as in <dog> and <doggy>) is a typical example. A cross-linguistic survey by Ultan (1978) shows that this connection between [i] and smallness seems to hold in the lexicon of various languages (see also Alderete and Kochetov (2017) as well as Blasi et al. (2016) for more recent and more extensive cross-linguistic studies). An experimental work by Sapir (1929), which is now generally considered to be a classic experimental study in this research domain, famously shows that the nonce word mil tends to be judged as smaller than the nonce word mal by native speakers of English and possibly by native speakers of Chinese as well. The emerging observation is that while languages constitute a system that can connect sounds and meanings in arbitrary fashions, there exist some systematic relationships between sounds and meanings. Such relationships are often referred to as sound symbolism.

The influence of Saussure was rather substantial in modern linguistic theories, and, hence, sound symbolism tended to be considered a marginal phenomenon at best. However, recent studies in psychology, linguistics, and cognitive science reveal increasingly more cases of sound–symbolic connections that seem to hold across many different languages. We are now witnessing a surprisingly fast-growing body of studies on this topic (see Nielsen and Dingemanse (2020) for some quantitative data on this research trend). We believe there are a number of reasons behind these growing interests. The first reason is the simple realization that the Saussurian dictum of arbitrariness was too strong, since as we carefully examine the lexicon of several languages, we seem to observe interesting patterns of systematic sound–meaning relationships (see Winter and Perlman (2021) for an excellent demonstration of this point), which can also be replicated with studies using nonce words. This means that the fundamental architecture of linguistic knowledge, as envisioned by Saussure and their followers, missed something important.

The second reason is that these sound–symbolic relationships make phonetic sense (Ohala 1994). For example, sounds with high-frequency energy (e.g., [i] having high second formant values and high tones having a high fundamental frequency) tend to be associated with smallness, which follows from the physical law that a small vibrator/resonator emits high-frequency sounds. Thus, studying sound symbolism may shed light on the issue of how phonetic considerations, both in terms of articulation and acoustics, may influence the linguistic knowledge that speakers of many languages have (Kawahara 2020). Third, there is also a very influential idea, proposed and developed by Imai and Kita (2014), that sound symbolism contributes to language-acquisition processes to a non-trivial degree. Words with sounds that in some way or another represent their meanings are easier to learn, and such words may potentially help bootstrap the initial phase of language-acquisition processes. The role of sound symbolism has been demonstrated both for first-language acquisition (Maurer et al. 2006) as well as for second-language acquisition (Nygaard et al. 2009).

Fourth, there is conjecture that iconic vocalizations may have been the source of the origin of human languages (Perlman and Lupyan 2018). For example, a study by Perlman and Lupyan (2018) shows that English speakers can produce iconic vocalizations to distinguish 30 different meanings that can be inferred by other native speakers of English to a level that is much higher than mere chance level. A follow-up study by Ćwiek et al. (2021) shows that the same vocalization can be understood by people from 25 different linguistic and cultural backgrounds. These studies raise the possibility that iconic vocalizations, which are likely to be precedents of sound–symbolic patterns, entailed a pre-modern form of human languages (see, also, related proposals by Berlin 2006 and Haiman 2018).

For the reasons briefly reviewed above, it is safe to say that sound symbolism now forms a central topic for general linguistic inquiry. Bodo Winter, in their Abralin talk, goes so far as to argue that iconicity, of which sound symbolism is an example, is one of the design features of human languages, contrary to what Hockett argued in 1959[1]. Reflecting this research trend, there are now a handful of review articles on this topic, each written from a slightly different perspective. See, for example, Dingemanse et al. (2015) and Lockwood and Dingemanse (2015). Kawahara (2020), which is itself a review article on sound symbolism, contains a list of the other overview articles written as of 2020. The above-mentioned talk by Bodo Winter also offers some review of the state-of-the-art research in this domain.

Positional Effects in Sound Symbolism

With this theoretical background in mind, we zoom in on a particular sound–meaning relationship to address a question that is understudied in the literature. The sound–symbolic meanings that we explore in this paper are that obstruents tend to be judged as hard, whereas sonorants tend to be judged as soft (Kawahara et al. 2005; Sidhu et al. 2022; Uemura 1965; Uno et al. 2022). Obstruents include plosives/stops (e.g., [t]), fricatives (e.g., [s]), and affricates (e.g., [ts]) and are defined as sounds for which the intraoral air pressure arises so greatly that spontaneous vocal-fold vibration is impossible (Chomsky and Halle 1968). Sonorants, on the other hand, include nasals (e.g., [n]), liquids (e.g., [r]), and glides (e.g., [j]), for which the air flow is not too obstructed in the oral cavity for spontaneous voicing to naturally occur. It has been shown by some previous work that obstruents tend to be judged to be harder than sonorants. Kawahara et al. (2005), expanding upon an old observation by Uemura (1965), used an experimental method and found that Japanese speakers judge nonce words with obstruents to be harder than those with sonorants (see also Kumagai et al. 2022 and Uno et al. 2022 for related sound–symbolic connections in the context of expressions of food texture). A recent study by Sidhu et al. (2022) also identified a systematic relationship between sonorants and softness by native speakers of English. We make use of this well-established sound–symbolic relationship to address a particular issue that we believe is under-studied in the literature.

The issue that we would like to address in the paper is positional effect (Kawahara 2020): psycholinguistically speaking, segments in different word positions have different degrees of psycholinguistic prominence, such that it would not be surprising if sounds in particular word positions may have more salient sound–symbolic meanings. In particular, segments in the word-initial position play an eminent role in lexical access (see e.g., Hawkins and Cutler 1988 for an overview). For example, when speakers recall only a part of a word (the "tip of the tongue" phenomenon), the most successful cue to recall the entire word is the initial syllables (Browman 1978). In a phoneme-monitoring task in which listeners are asked to detect whether or not there was a particular phone, response times are faster for segments in word-initial syllables than for segments in non-word-initial syllables (Mattys and Samuel 2000). Segments in initial syllables are, therefore, psycholinguistically more important than segments in other syllables.

Moreover, as we look at the phonological systems of many languages, segments in word-initial positions are privileged in several ways (e.g., see Beckman (1998) for a

review). For example, languages tend to allow more segment types in initial syllables than in non-initial syllables. Initial syllables often trigger vowel-harmony patterns, determining the overall vowel quality of the whole words. Initial syllables are also known to attract segments and other suprasegmental properties, such as tone and stress. Therefore, given the psycholinguistic and phonological privileges of initial segments, it would not be surprising if word-initial segments exhibit salient sound–symbolic meanings. Some studies have explored this question in the context of sound symbolism, but more case studies seem to be warranted, a gap that we would like to address.

Kawahara et al. (2008) deployed a sound–symbolic connection between voiced obstruents and dirtiness, which is known to hold rather robustly in Japanese, and demonstrated that voiced obstruents evoked more dirty images in the word-initial position than in the non-word initial position. Adelman et al. (2018) analyzed about 37,000 words from five different languages (English, Spanish, Dutch, German, and Polish) and demonstrated that word-initial sounds are a good predictor of the valence values of these words. On the other hand, Haynie et al. (2014) found sound–symbolic values to be salient in prominent positions (segments at word edges). McGregor (1996) explored the sound–symbolic values of root-final consonants in Gooniyandi. It, thus, seems important to us to pin down more precisely the positional effects on sound symbolism with more case studies. Given the robust body of evidence reviewed above that segments in initial positions are both psychologically and linguistically privileged, our initial expectation was that sound–symbolic values are stronger in initial syllables than in non-initial syllables.

2. Experiment 1

2.1. Method

In this experiment, native speakers of English and Japanese were presented with disyllabic nonce words and were asked to judge how hard they sound. In some words, obstruents appear in the initial syllables, whereas in other words, obstruents appear in the non-initial syllables. The general prediction we initially had was that nonce words should be judged harder when obstruents appear in the initial syllables.

Following the spirit of the open-science initiative in linguistics (see Winter 2019), the raw data and the R markdown files are made available at the following Open Science Framework repository[2]. This folder contains the relevant data for both Experiment 1 and Experiment 2.

2.2. Stimuli

For both English and Japanese, the stimuli were disyllabic nonce words. In one condition, the nonce words started with obstruents, and the second syllables had sonorant onsets. In the other condition, the nonce words started with sonorants, and the second syllables had obstruent onsets. The lists of the stimuli appear in Table 1 (those for English speakers) and Table 2 (those for Japanese speakers)[3]. For the stimuli in Japanese in Table 2, we added a word-final <n> (usually transcribed as a uvular nasal, [N]) so that all the stimuli would be perceived as nonce words; otherwise, it was very difficult to find a set of disyllabic sequences that would be considered as non-existing words by Japanese speakers. In contrast, we did not add a final consonant for the English stimuli because that would make the final syllable heavy and, hence, more likely to receive stress. Stress, however, would make final syllables more prominent than otherwise, and that would introduce a layer of confound, which we wished to avoid. The choice of obstruents was the same across the two languages; the choice of sonorants differed slightly between the two languages in that <r> is used for English, whereas <j> is used for Japanese. This was necessitated by the need to make all the stimuli nonce words. Within each language, the vowel quality was controlled between the two conditions.

Table 1. The list of stimuli for English speakers (Experiment 1).

Obs-Son	Son-Obs
pami	mapi
pini	nipi
puri	rupi
tami	mati
tini	niti
turi	ruti
kami	maki
kini	niki
kuri	ruki

Table 2. The list of stimuli for Japanese speakers (Experiment 1). Note that <j> represents a palatal glide, not a post-alveolar voiced affricate. <n> is usually transcribed as a uvular nasal [N].

Obs-Son	Son-Obs
pamon	mapon
panon	napon
pajon	japon
toman	motan
tonan	notan
tojan	jotan
kumun	mukun
kunun	nukun
kujun	jukun

2.3. Participants and Procedure

A total of 68 native speakers of English and 109 native speakers of Japanese participated in this study. The study was administered online using SurveyMonkey. The participants all read through a consent form approved by the authors' institution. The consent form for English speakers made it clear that only data from native speakers would be analyzed; in addition, a post-experimental question asked whether they are a native speaker of English. All Japanese speakers were from regions in or near Tokyo. The task of the experiment was described as follows:

> Imagine you are working for a food company. Your team is going to produce some new brands of snacks to be exported to other countries. The snacks are ready for production, but you have not decided on their names. So, your boss asks your team to come up with several names that may match the images of the snacks you are producing. There are several examples of names already, and your present task is to judge the images evoked by each name. Now, you are going to judge how hard or soft each name sounds.

The stimuli were presented as katakana orthography to Japanese speakers, as this is the standard orthographic system used to represent nonce words. The stimuli for English speakers were presented orthographically as well, although both groups of speakers were encouraged to read the stimuli in their minds before making a judgment, with an accent (high pitch for Japanese and stress for English) on the initial syllables rather than on the final syllables[4]. English speakers were also told that the final <i> is not to be read as a diphthong (i.e., [ai]) but rather as a monophthong ([i]), as that would make the final syllable heavy. The order of the stimulus items was randomized for each participant by SurveyMonkey.

Before the main trials, the participants were provided with one practical trial, with <yusi> for English speakers and <hasin> for Japanese speakers. The participants were presented with each nonce word per trial and were asked to judge how hard each nonce word sounds using a five-point Likert scale in which higher values were harder. The experiment usually took about 15 min.

2.4. Statistics

Since the responses were obtained using a Likert scale, we fit a mixed-effects ordinal logistic regression with participants and items included as random intercepts. The independent variables are the type of initial segments (obstruents vs. sonorants) and language (English vs. Japanese); random slopes for speakers were also coded for the initial segments. If the initial segments evoked stronger sound–symbolic meanings, and if obstruents were associated with the image of hardness, obstruent-initial stimuli (those in the left columns of Tables 1 and 2) should be judged as harder. If this pattern were to hold both in English and Japanese, the interaction term should not be substantial.

2.5. Results

Figure 1 summarizes the results in the form of violin plots. Each transparent dot represents an average hardness rating from each participant. Each violin represents a normalized probability distribution of these averages. The thick red dots represent the grand average for each condition. The left facet represents the results from English speakers and the right facet those from Japanese speakers. Within each facet, the left violin shows the ratings of the obstruent-initial items, and the right violin shows the ratings of the sonorant-initial items.
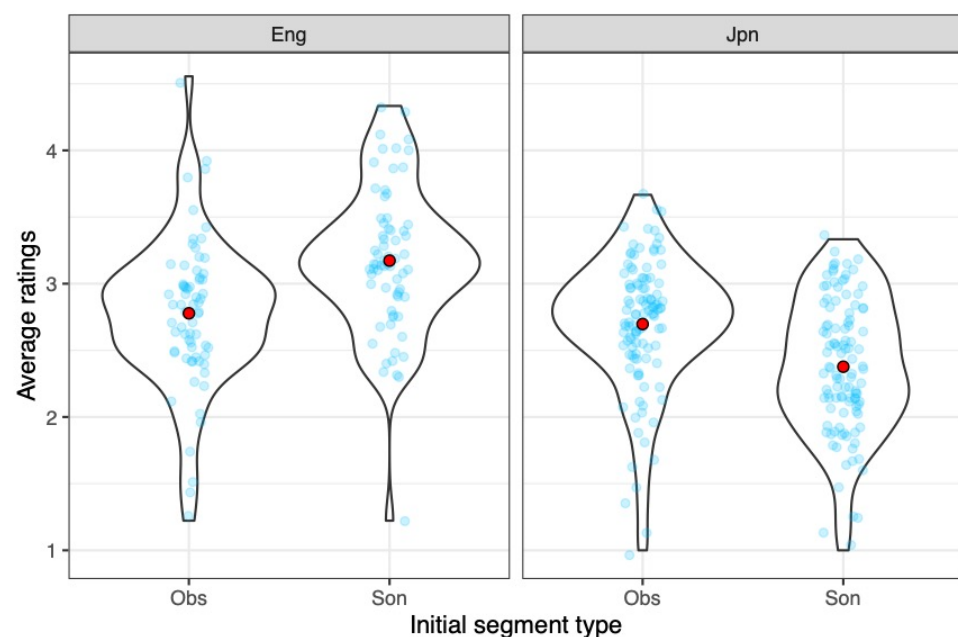

Figure 1. The results of Experiment 1. The x-axis labels are the type of initial segments. The plots are violin plots showing the probability distributions of the averaged responses from each speaker. Each transparent blue dot represents an average response from each speaker. The solid red dots represent the grand averages.

The results from the Japanese speakers are as expected from the results of previous studies. To the extent that obstruents evoke harder images, and to the extent that initial segments are more important than non-initial segments (i.e., onset consonants in non-initial syllables), obstruent-initial words should be judged as harder. The English pattern seems to indicate the opposite pattern, however. Indeed, the interaction term of the ordinal-regression model shows a significant effect (*beta* = -1.31, $z = -2.94$, $p < 0.01$). The main effect of the initial consonant type was also significant in that sonorant-initial items generally tended to be judged as harder at the baseline level (*beta* = 0.76, $z = 2.38$, $p < 0.05$). The language difference was not significant at the baseline level (*beta* = -0.15, $z = -0.44$, n.s.).

2.6. Discussion

The pattern of the Japanese participants accords well with our initial prediction. Obstruents evoke hard images; if word-initial segments (i.e., onset consonants in word-initial syllables) play a prominent role, then obstruent-initial words should be judged as harder. The pattern exhibited by the English speakers, alternatively, is puzzling. At this point, there are two possibilities to explain this unexpected pattern observed in Experiment 1: either (1) English speakers may find sonorants to be harder than obstruents, or (2) English speakers may find onset consonants in final syllables to sound symbolically more important than those in initial syllables. The next experiment was designed to tease apart these two possibilities.

Whichever explanation is on the right track, we find this cross-linguistic difference (i.e., the significant interaction term in the statistical model) to be intriguing. We based our hypothesis on the assumption that the salience of sound–symbolic values has its root in the psycholinguistic prominence of initial segments (onset consonants in initial syllables in this case), which should hold regardless of whether the target language is English and Japanese. This cross-linguistic difference, therefore, raises many new questions—to the extent that there are cross-linguistic differences in positional effects in sound symbolism, how different can languages be? What is the source of positional effects in sound symbolism? Can they be independent of the psycholinguistic prominence of particular syllables? These questions are highly intriguing but addressing them is beyond the scope of the paper.

3. Experiment 2

To tease apart the two possibilities discussed above, we compared disyllabic nonce words that consist of only obstruent onsets with those consisting of only sonorant onsets. If the first possibility—that sonorants tend to be judged as harder by English speakers—is on the right track, then the second set of stimuli should be judged as harder.

3.1. Stimuli

The list of stimuli for Experiment 2 is listed in Table 3. Nine items were included in each condition. Note that the second syllables, unlike those in Experiment 1, are heavy syllables, but we judged this to be non-problematic because, in this experiment, we were not interested in the difference between the first and second syllables[5].

Table 3. The list of stimuli for Experiment 2 (English speakers only).

Obs-Obs	Son-Son
popoi	momoi
pepei	memei
papai	mamai
totoi	nonoi
tetei	nenei
tatai	nanai
kokoi	loloi
kekei	lelei
kakai	lalai

3.2. Participants and Procedure

A total of 44 native speakers of English completed the online experiment administered using SurveyMonkey. The other details of the experiment are identical to those of Experiment 1, except that the practical trial used <yasoi>.

3.3. Results

Figure 2 shows the results, again in the form of violin plots. We can observe that the items with obstruents were judged to be harder than those with sonorants and that this difference was statistically robust (*beta* = −1.57, *z* = −4.41, *p* < 0.001), a result that is

compatible with a recent study by Sidhu et al. (2022). This result indicates that it is not the case and that English speakers find sonorants to be harder than obstruents.
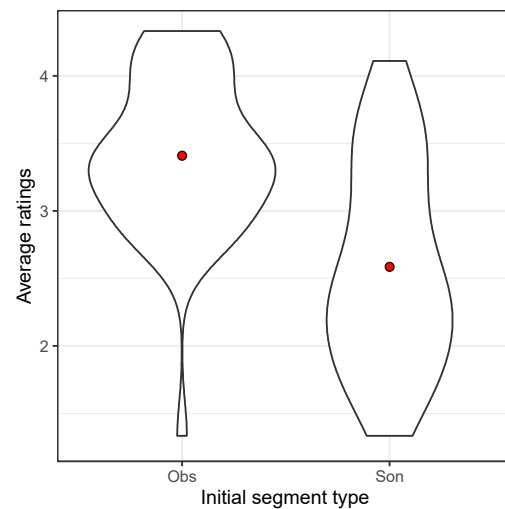

Figure 2. The results of Experiment 2. The x-axis labels are the type of consonants. The plots are violin plots and show the probability distributions of the averaged responses from each speaker. Each transparent blue dot represents an average response from each speaker. The solid red dots represent the grand averages.

3.4. Discussion

Recall that Experiment 1 left two possibilities open: either English speakers tend to find sonorants as hard, or English speakers find onset consonants in the last syllables to be more important than those in the initial syllables. Experiment 2 shows that the first possibility is not a viable explanation of Experiment 1, as English speakers do find obstruents to be harder than sonorants. Taken together, then, we are left with the conclusion that for English speakers, the onset consonants in the final syllables are more important than those in the initial syllables, at least in the context of judging the hardness of nonce words.

One limitation of this study is that since we used disyllabic words, we were not able to tease apart the difference between final syllables and non-final medial syllables. Whether or not we observed a tripartite distinction between the initial, final, and medial syllables is an interesting topic that is worthy of further experimentation. Another limitation of the study is the fact that we used orthography to present the stimuli, and it would be interesting to replicate the current experiments with auditory stimuli only.

4. General Discussion

As reviewed in the introduction, there seems to be no doubt that exploring sound-symbolism connections in natural language offers important insights into the nature of the linguistic systems of human languages, and, as such, we are witnessing a surprisingly fast-growing body of studies on this topic. We attempted to contribute to this enterprise by addressing a gap that is under-studied in the literature—a positional effect. Since we know from previous psycholinguistic studies that segments at different word positions have different degrees of importance in terms of lexical access, it would not have been surprising to observe a positional effect in sound symbolism.

The current study took the sound–symbolic relationship between obstruents and hardness as its empirical target. Experiment 1 found that Japanese speakers judged obstruent–sonorant sequences to be harder than sonorant–obstruent sequences, which is just as predicted, if Japanese speakers find consonants in initial syllables to be more important when deciding on the sound–symbolic values of nonce words. However, Experiment 1, unexpectedly, found the opposite results for English speakers. Experiment 2 confirmed that English speakers find obstruents to be harder than sonorants. Taken together, we conclude

that English speakers find onset consonants in the final (or non-initial) syllables to be more important when calculating sound–symbolic values. This result was rather surprising, given that we know that initial syllables are psycholinguistically prominent and also given that we ensured that the participants placed stress on the initial syllables rather than on the final syllables. As puzzling as it is, we find this cross-linguistic difference to be highly interesting (and novel, to the best of our knowledge).

While we can only offer some post hoc speculative remarks regarding the behavior of English speakers, which went against our initial expectation, one possible interpretation of this result is due to recency effects (Gupta 2005; Gupta et al. 2005)—sounds that occur more recently stay more clearly in the memory. If English speakers were more susceptible to recency effects than the psycholinguistic prominence of the initial syllables, that may explain the behavior of English speakers. We fully admit, however, that this hypothesis is merely speculative and raises a new question of why English speakers, and not Japanese speakers, were sensitive to recency effects.

Another admittedly speculative hypothesis is that in some sense, final syllables are prominent in English. For example, in hypocoristic-truncation patterns, the final syllables, rather than the initial syllables, can survive truncation, as in <Beth> from <Elizabeth> and <Bert> from <Albert>. Syllabic sonorants, as in those words such as <kitten> and <little>, are found in word-final positions but not in word-initial positions. However, the phonological strengths of the final syllables in English are not well-established in the literature, to the best of our knowledge, and this explanation too is post hoc. We would like to make it clear here that whatever explanation we may come up with is going to be post hoc, which has to be tested with a new experiment.

Our results open up possibilities for future experimentation. First, we should extend the coverage of target languages beyond English and Japanese. A lesson that arises from the current study is that we may be able to find a language in which final syllables are phonologically strong. If the second hypothesis entertained above is on the right track, the prediction is that they should behave like English speakers. Second, we should explore the positional effects in sound symbolism by making use of different sound–symbolic patterns; for example, we could deploy other well-studied sound–symbolic relationships, such as those between obstruents and angularity; sonorants and roundedness; [a] and largeness; [i] and smallness; [o] and roundness, etc. (see Westbury et al. 2018 for a comprehensive review). At this point, we have no reasons to expect that different sound–symbolic patterns should show different patterns of positional effects, but this question can only be answered through actual empirical studies. Finally, we can explore whether a finer distinction (e.g., initial vs. medial vs. final) holds in terms of the force of sound–symbolic values. It would also be important to explore potential task effects on the found pattern; for example, it would be interesting to test whether the same patterns would hold when the stimuli are presented purely auditorily. While we acknowledge that many questions are left open, we hope to situate the current study as a non-trivial stepping stone for future work on positional effects in sound symbolism.

Author Contributions: Conceptualization, K.S. and R.U.; methodology, K.S.; investigation, K.S. and R.U.; data curation, K.S and R.U; writing—original draft preparation, K.S.; writing—review and editing, K.S. and R.U; project administration, K.S.; funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics Committee of Tokyo University of Agriculture and Technology (approval No. 210806-0281, 15 September 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Notes

1　https://www.youtube.com/watch?v=R1ETw21oCGE&t=6s (accessed on 4 July 2022).

2　https://osf.io/s2ngc/?view_only=930c5db2193942f9a7fe62c8e8f53137X (accessed on 4 July 2022).

3　An anonoymous reviewer asked if we should be concerned that <tini> may resemble existing words such as <tiny> or <teeny>. Since the current experiment was about hardness judgment rather than size judgment, we believe that it is not too much of a concern.

4　An anonymous reviewer asked if non-linguistically informed Japanese participants were able to follow this instruction, as most Japanese speakers are not conciously aware of how the pitch-accent system in their language works. Nevertheless, this is not too much of a concern, as it is very unlikely that nonce words receive an accent on the final syllables in Japanese (see Kubozono 2006). In case of English, disyllabic nouns are usually pronounced with a stress on the initial syllables. We made this clear to the participants by presenting an example, <water>, before the practice session. Therefore, we assume that English-speaking participants could easily put a stress on the initial syllables.

5　By making the second syllables heavy, we also hoped to direct the participants' attention to vowel quality differences in the final syllables, although we have no way of knowing whether this actually worked. This, however, is tangential to the main question that we would like to address in this experiment.

References

Adelman, James S., Zachary Estes, and Martina Cossu. 2018. Emotional Sound Symbolism: Languages Rapidly Signal Valence via Phonemes. *Cognition* 175: 122–30. [CrossRef] [PubMed]

Alderete, John, and Alexei Kochetov. 2017. Integrating Sound Symbolism with Core Grammar: The Case of Expressive Palatalization. *Language* 93: 731–66. [CrossRef]

Beckman, Jill. 1998. Positional Faithfulness. Doctoral dissertation, University of Massachusetts, Amherst, MA, USA.

Berlin, Brent. 2006. The First Congress of Ethonozoological Nomenclature. *Journal of Royal Anthropological Institution* 12: 23–44. [CrossRef]

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler, and Morten H. Christiansen. 2016. Sound-meaning Association Biases Evidenced across Thousands of Languages. *Proceedings of the National Academy of Sciences* 113: 10818–23. [CrossRef] [PubMed]

Browman, Catherine. 1978. *Tip of the Tongue and Slip of the Ear: Implications for Language Processing*. UCLA Working Papers in Phonetics 42. Los Angeles: University of California.

Chomsky, Noam, and Moris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Ćwiek, Aleksandra, Susanne Fuchs, Christoph Draxler, Eva Liina Asu, Dan Dediu, Katri Hiovain, Shigeto Kawahara, Sofia Koutalidis, Manfred Krifka, Pärtel Lippus, and et al. 2021. Novel Vocalizations Are Understood across Cultures. *Scientific Reports* 11: 10108. [CrossRef]

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen, and Padraic Monaghan. 2015. Arbitrariness, Iconicity and Systematicity in Language. *Trends in Cognitive Sciences* 19: 603–15. [CrossRef]

Gupta, Prahlad. 2005. Primacy and Recency in Nonword Repetition. *Memory* 13: 318–24. [CrossRef]

Gupta, Prahlad, John Lipinski, Brandon Abbs, and Po-Han Lin. 2005. Serial Position Effects in Nonword Repetition. *Journal of Memory and Language* 53: 141–62. [CrossRef]

Haiman, John. 2018. *Ideophones and the Evolution of Language*. Cambridge: Cambridge University Press.

Harris, Roy, and Talbot J. Taylor. 1989. *Landmark in Linguistic Thoughts*. New York and London: Routledge.

Hawkins, John, and Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In *Explaining Language Universals*. Edited by John A. Hawkins. Oxford: Basil Blackwell, pp. 280–317.

Haynie, Hannah, Claire Bowern, and Hannah LaPalombara. 2014. Sound Symbolism in the Languages of Australia. *PLoS ONE* 9: e92852. [CrossRef]

Hinton, Leane, Johanna Nichols, and John J. Ohala. 1994. *Sound Symbolism*. Cambridge: Cambridge University Press.

Hockett, Charles. 1959. Animal "Languages" and Human Language. *Human Biology* 31: 32–39.

Imai, Mutsumi, and Sotaro Kita. 2014. The Sound Symbolism Bootstrapping Hypothesis for Language Acquisition and Language Evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130298. [CrossRef] [PubMed]

Kawahara, Shigeto. 2020. Sound Symbolism and Theoretical Phonology. *Language and Linguistic Compass* 9: 168–82.

Kawahara, Shigeto, Yoshihiro Matsunaka, Akira Nakayama, and Kazuko Shinohara. 2005. An Experimental Case Study of Sound Symbolism in Japanese. Paper presented at New Directions in Cognitive Linguistics, Brighton, UK, October 23–25.

Kawahara, Shigeto, Kazuko Shinohara, and Yumi Uchimoto. 2008. A Positional Effect in Sound Symbolism: An Experimental Study. *Proceedings of the Japan Cognitive Linguistics Association* 8: 417–27.

Kubozono, Haruo. 2006. Where Does Loanword Prosody Come From? A Case Study of Japanese Loanword Accent. *Lingua* 116: 1140–70. [CrossRef]

Kumagai, Gakuji, Ryoko Uno, and Kazuko Shinohara. 2022. The Sound-symbolic Effects of Consonants on Food texture: An Experimental Study of Snack Names in Japanese. In *The Language of Food in Japanese: Cognitive Perspectives and Beyond*. Edited by Kiyoko Toratani. Amsterdam: John Benjamins, pp. 78–110.

Lockwood, Gwilym, and Mark Dingemanse. 2015. Iconicity in the Lab: A Review of Behavioral, Developmental, and Neuroimaging Research into Sound-Symbolism. *Frontiers in Psychology* 6: 1246. [CrossRef]

Mattys, Sven, and Arthur Samuel. 2000. Implications of Stress-pattern Differences in Spoken-word Recognition. *Journal of Memory and Language* 42: 571–96. [CrossRef]

Maurer, Daphne, Thanujeni Pathman, and Catherine J. Mondloch. 2006. The Shape of Boubas: Sound-shape Correspondences in Toddlers and Adults. *Developmental Science* 9: 316–22. [CrossRef]

McGregor, William. 1996. Sound Symbolism in Gooniyandi, a Language of Western Australia. *Word* 47: 339–64. [CrossRef]

Nielsen, Alan, and Mark Dingemanse. 2020. Iconicity in Word Learning and Beyond: A Critical Review. *Language and Speech* 64: 52–72. [CrossRef]

Nygaard, Lynne C., Alison E. Cook, and Laura L. Namy. 2009. Sound to Meaning Correspondence Facilitates Word Learning. *Cognition* 112: 181–86. [CrossRef]

Ohala, John J. 1994. The Frequency Code Underlies the Sound Symbolic Use of Voice Pitch. In *Sound Symbolism*. Edited by Leane Hinton, Johanna Nichols and John J. Ohala. Cambridge: Cambridge University Press, pp. 325–47.

Perlman, Marcus, and Gary Lupyan. 2018. People Can Create Iconic Vocalizations to Communicate Various Meanings To Naïve Listeners. *Scientific Reports* 8: 2634. [CrossRef] [PubMed]

Sapir, Edward. 1929. A Study in Phonetic Symbolism. *Journal of Experimental Psychology* 12: 225–39. [CrossRef]

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.

Sidhu, David M., Gabriella Vigliocco, and Penny M. Pexman. 2022. Higher Order Factors of Sound Symbolism. *Journal of Memory and Language* 125: 104323. [CrossRef]

Styles, Suzy J., and Lauren Gawne. 2017. When Does Maluma/Takete Fail? Two Key Failures and a Meta-Analysis Suggest that Phonology and Phonotactics Matter. *i-Perception* 8: 2041669517724807. [CrossRef] [PubMed]

Uemura, Yukio. 1965. Onsei-no hyoushousei-ni tsuite [On the Symbolic Aspects of Sounds]. In *Gengo Seikatsu*. Tokyo: Honami Shuppan, pp. 66–70.

Ultan, Russell. 1978. Size-sound Symbolism. In *Universals of Human Language II: Phonology*. Edited by Joseph Greenberg. Stanford: Stanford University Press, pp. 525–68.

Uno, Ryoko, Fumiyuki Kobayashi, Kazuko Shinohara, and Sachiko Odake. 2022. Analysis of the Use of Japanese Mimetics in the Eating and Imagined eating of Rice Crackers. In *The Language of Food in Japanese: Cognitive Perspectives and Beyond*. Edited by Kiyoko Toratani. Amsterdam: John Benjamins, pp. 55–77.

Westbury, Chris, Geoff Hollis, David M. Sidhu, and Penny M. Pexman. 2018. Weighing Up the Evidence for Sound Symbolism: Distributional Properties Predict Cue Strength. *Journal of Memory and Language* 99: 122–50. [CrossRef]

Winter, Bodo. 2019. *Statistics for Linguists: An Introduction Using R*. New York and London: Routledge.

Winter, Bodo, and Marcus Perlman. 2021. Size Sound Symbolism in the English Lexicon. *Glossa: A Journal of General Linguistics* 6: 79. [CrossRef]