Magnitude sound symbolism influences vowel production

Vainio, Lari

2021-06

Vainio , L 2021 , ' Magnitude sound symbolism influences vowel production ' , Journal of Memory and Language , vol. 118 , 104213 . https://doi.org/10.1016/j.jml.2020.104213

http://hdl.handle.net/10138/353792
https://doi.org/10.1016/j.jml.2020.104213

Magnitude sound symbolism influences vowel production

Vainio, L

Lari Vainio; Email: lari.vainio@helsinki.fi; Tel: +358-2-94129392

1) Helsinki Collegium for Advanced Studies, University of Helsinki, Fabianinkatu 24, Helsinki, Finland.

2) Perception, Action & Cognition Research Group, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Haartmaninkatu 3, Helsinki, Finland

3) Phonetics and speech synthesis research group, Department of Digital Humanities, University of Helsinki, Unioninkatu 40, Helsinki, Finland

Abstract

Segmental properties of speech can convey sound symbolic meaning. This study presents two novel sound-meaning mappings using a choice reaction time paradigm in which participants have to select quickly one of the two vocal response alternatives based on predefined categories of perceptual magnitude. The first study showed that the short distance between perceived objects facilitates the initiation of the vowel [i] production, while long distance facilitates the production of [u] and [æ]. Correspondingly, in the second study, vocal responses produced with [i] and [e] were initiated faster when the stimuli required short vocalizations, while responses produced with [u], [æ] and [y] were faster when the stimuli required long vocalizations. Hence, similar sound-meaning mappings were observed concerning concepts of spatial and temporal length. This suggests that different sound-magnitude effects can be generalized to the common processing of conceptual magnitude. A conceptual magnitude seems to be implicitly and systematically associated with an articulatory response of a specific vowel. The study also suggests that in addition to the vowel openness and backness, the vowel roundness can also associate particular vowels with large magnitudes.

Keywords: speech, sound-meaning mapping, sound symbolism, conceptual magnitude, articulation

1. Introduction

The view of a non-arbitrary relationship between the sound and the meaning of a word, which is most often referred to as sound symbolism, has a long history dating back to Plato’s Socratic dialogue “Cratylus” (see Magnus, 2000). Sound symbolism has been a subject of scientific investigation since the early twentieth century. Researchers such as Jespersen (1922) and Sapir (1929) provided one of the earliest scientific proposals of sound symbolism challenging the dominant thesis, which emphasized the arbitrariness of human language (de Saussure, 1916). Indeed, although the principle of arbitrariness generally holds across the lexicon, exceptions to this principle have been identified since these early observations. Many researchers hold the view that when an individual has to acquire the meaning of thousands of vocabulary items over the years, arbitrariness and non-arbitrariness both bring their own selective advantages (see Dingemanse et al., 2015 for a review). Non-arbitrary relations between the sound and meaning of a word can, for example, improve early language learning (Imai & Kita, 2014; Monaghan et al., 2012), and enhance comprehension of words (Dingemanse et al., 2015).

In many cases of sound symbolism, the term *sound symbolism* can be considered as misleading given that the attribute *symbolism* by definition implies a relation between meaning and sound that is based on convention, while many, if not most, sound symbolism phenomena are not based on convention. In addition, many of the sound symbolism phenomena can be rooted on other underlying properties of speech than *sound* as will be presented below. Anyway, according to Dingemanse et al. (2015), at least two types of non-arbitrariness can be found across natural languages – iconicity and systematicity. Systematicity refers to phonological and prosodic cues such as vowel stress that enable, for example, distinguishing nouns from verbs. In contrast, iconicity refers to relating the form and meaning of words by means of perceptuomotor analogies. In ‘*absolute iconicity’*, sound properties of a word convey meaning by directly imitating the sound of the referent as in onomatopoeia (e.g., *knock, ring, bang*). In contrast, ‘*relative iconicity’* provides a less salient manifestation of sound symbolism as it maps indirect associations or the impression of a specific sound element of a word onto meaning. As a famous example of this relative iconicity, Sapir (1929) showed that 96% of adult participants judged that the pseudo-word *mil* is more likely to stand for an idea of a small object and *mal* for a large object. Since Sapir’s original observation, numerous studies have investigated this sound-size symbolism effect, showing that front vowels ([i] and [e]) are more likely to be associated with small concepts, while back vowels ([u], [o] and [ɑ]) are more likely to be associated with large concepts (Nuckolls, 1999).

Some variance in sound-size symbolism has been observed across natural languages. For instance, front-close vowels are associated with largeness in Bahnar language (Diffloth, 1994). Nevertheless, in general, cross-linguistic studies suggest more similarities in sound-size symbolism across languages than one would expect by chance (Blasi et al., 2016; Shinohara & Kawahara, 2010; Ultan, 1978; Nuckolls, 1999). A consensus is, however, still lacking on explanation of these sound-size symbolism phenomena. What is, for example, so small about *mil* and large about *mal*? Sound-size symbolism has been often explained to reflect automatic matching of articulatory and/or acoustic parameters of speech sounds to physical properties of a referent concept (see Sidhu & Pexman, 2018 for a review). Regarding articulatory explanation, for example, the vocal gesture related to back-open vowels can be assumed to gesturally mimic *largeness* of the referent by enlarging the oral cavity, while front-close vowels mimic *smallness* by making the oral cavity smaller (Sapir, 1929; Ramachandran & Hubbard, 2001). However, it has to be emphasized that these articulatory views of sound-size symbolism are mostly based on anecdotal rather than clear empirical evidence (see for example Ramachandran & Hubbard, 2001).

Regarding acoustic explanations of the sound-size symbolism, small and large sizes can be associated with particular acoustic properties of vowels such as formants F_1 and F_2. It is known that heightened F_1 is a consequence of tongue lowering in vocalization matching it with open vowels, while lowered F_2 is a consequence of tongue backing and lip rounding, as both lengthen the frontal cavity, matching it with back and rounded vowels (Fant, 1960; Ohala & Eukel, 1987). Particularly relevant phenomenon to acoustic accounts of sound-size symbolism is the intrinsic f_0 (fundamental frequency). According to the evidence of intrinsic f_0, close vowels have typically higher f_0 than open vowels across different languages (Whalen & Levitt, 1994). This phenomenon can be observed even in infant babbling (Whalen et al., 1994). Acoustic accounts of sound-size symbolism often assume that vowels that have an intrinsically higher pitch (i.e., front-close vowels) are implicitly associated with smaller objects because smaller things tend to resonate at higher frequencies (see Spence, 2011). This view is closely related to the frequency code hypothesis (Ohala, 1984) emphasizing that relatively lower voice pitch is associated with larger and more dominant individuals.

Furthermore, sound-size symbolism can be attributed to the different intrinsic durations of vowels given that high vowels are typically shorter than low vowels presumably due to the extra time it takes for the jaw to open when producing low vowels (e.g., Lehiste, 1972; Lehnert-LeHouillier, 2007). It could be assumed, for example, that the concept of *largeness* is grounded in sensory experiences based on an intrinsically occurring temporal extension of particular vowels. In line with this, Knoeferle et al. (2017) observed that in addition to F_1 and F_2, vowel duration plays an important role in sound-size symbolism. However, in contrast to the frequency code hypothesis, f_0 was not found to have a role in this phenomenon.

The sound symbolism phenomenon that is semantically close to sound-size symbolism is the effect in which the egocentric distance of a referent is associated with a specific vowel. Most commonly, the high-front vowels [i] and [e] have been associated with meaning referring to proximal distances (i.e., words referring to distances close to an observer such as *here*). In contrast, the back low vowels [o], [u] and [ɑ] have been associated with meaning referring to distal distances (i.e., words referring to distances far from an observed such as *there*) (Tanz, 1971; Woodworth, 1991). Rabaglia et al. (2016) proposed that this distance symbolism can be attributed to vowel frontness/backness so that back vowels have typically longer wavelengths than front vowels. As a consequence, sound of back vowels can travel relatively long distances making them suitable components for words that refer to distal concepts. As such, this explanation is not applicable to sound-size symbolism. Hence, this proposal can be viewed to assume that sound-size and sound-distance symbolism are based on separate sound-meaning mappings: In sound-size symbolism, a specific vowel sound is associated with a specific size (i.e., how small/big something is), while in sound-distance symbolism, a specific vowel is associated with a specific magnitude of egocentric distance (i.e., how near/far something is from the viewer). Nevertheless, it is noteworthy that small/proximal as well as large/distal are semantically related concepts and associated with the same vowels. Hence, it has been debated whether the sound-size and sound-distance mappings present overlapping versions of the same sound-magnitude phenomenon associating abstract *smallness-largeness* concepts with particular vowels, or whether they are indeed based on independent sound-meaning mappings (Traunmüller, 1996).

1.1 The present study

This study explores, for the first time, two types of conceptual magnitude – allocentric distance (in Experiment 1) and temporal length (in Experiment 2) – in the context of sound symbolism using the choice reaction time (CRT) method. The most evidence of sound-magnitude mappings comes from studies based on the interpretation of, for example, pseudo-words or meaningless syllables (e.g., Sapir, 1929; Rabaglia et al., 2016; Knoeferle et al., 2017) rather than focusing on isolated vowels. This is so even though most researchers would agree that sound-magnitude symbolism is attributable to the sound or production of isolated vowels. Furthermore, although some studies have explored how sound symbolism manifests itself in word production (e.g., Laing, 2014; Laing et al., 2017), there is little empirical evidence on whether sound symbolism can affect production of isolated vowels. Therefore, the present study investigates whether processing a magnitude concept influences on producing an isolated vowel as a vocal response.

Furthermore, the most accounts of sound symbolism assume that sound-magnitude mappings are to some extent based on automatically operating connections between conceptual magnitude representations and particular vowels. That is, the sound-magnitude symbolism effects should be observed even when the experimental task does not require linguistic processing of the sound unit, or explicit assigning of magnitude-related meaning to the sound unit. Even so, sound-magnitude symbolism has been mostly explored using paradigms in which participants have to explicitly assign meaning, typically between opposing alternatives (e.g., small-large), to pseudo-words (e.g., Sapir, 1929) or meaningless syllables (e.g., Knoeferle et al., 2017) containing different phonemes. It has been proposed, however, that these kinds of tasks highlight potential associations and may lead participants to become aware of shared aspects between meaning and form of words/syllables that they would not have considered otherwise (e.g., Bentley & Varron, 1933). Given this, and that explicit tasks can reveal different mental processes than implicit tasks (De Houwer & Moors, 2007), it is important to show that sound-magnitude effects can also be robustly observed in implicit tasks. As such, the present study explores whether sound-meaning associations related to conceptual magnitude can *automatically* influence selection of a particular vowel for vocalization responses.

For these purposes, this study uses a CRT paradigm. It is noteworthy that the CRT method used in the current study does not require explicit assigning of any sound symbolic meaning to linguistic alternatives, and hence can be assumed to reveal *automatically* operating linkages between a particular semantic concept (e.g., long distance) and a particular vowel (e.g., [u]). In a prototypical version of the CRT paradigm (see Smith, 1968 for a review), participants are required to select a response out from two response options, as fast and accurately as possible, based on some stimulus property. Additionally, in a typical CRT task, responses are performed manually. As an example of this kind task, in the study of Tucker and Ellis (2001), the participants were presented with images of natural or manufactured objects whose size was congruent with the precision or power grasp, while they were holding the precision and power grip device in their hand. The participants were required to perform quick squeeze-response with the precision or power grip based on natural-manufactured category of the stimuli. In this task, responses were performed particularly rapidly when there was congruency between the object size and the type of grip response. This and other similar stimulus-response effects have been often attributed to direct and automatic stimulus-induced priming of action schema that is congruent with the stimulus property (see Vainio & Ellis, 2020 for a review).

Although CRT paradigms have typically focused on manual responses, there is no reason why it could not be applied as well for articulatory responses. Hence, in the current study, participants are required to discriminate the target stimulus belonging to one of the alternative perceptual categories with opposite meanings (e.g., short vs. long) and select one of the two articulatory response alternatives based on this category. One of the response alternatives is hypothetically associated with one category (e.g., vowel [i] - short distance) and the other response is associated with the other category (e.g., vowel [u] - long distance). It is assumed that responses (i.e., vowel productions) are initiated faster when there is congruency between the perceived category and the required vocal response. Hence, this paradigm allows one to investigate the causal, implicit and direct consequences of processing a particular semantic concept on producing a vocal response that is either congruent or incongruent with the concept.

Finally, the CRT method also allows one to measure acoustic properties of each vocalization. Sensorimotor processing of visual shape-related features has been previously associated with acoustic parameters of the third formant (Parise & Pavani, 2011), whereas sound-magnitude symbolism has been particularly linked to the first two formants instead of the third formant (Knoeferle et al., 2017). In addition, as stated above, magnitude-related sound symbolism phenomena can be linked to acoustic parameters of vowel duration (Knoeferle et al., 2017) and f_0 (Ohala, 1984). However, it has to be emphasized that the previous research that has explored linkages between acoustic properties and magnitude has mostly focused on sound symbolic linking of perceived magnitude information to acoustic properties of perceived vowels (e.g., Knoeferle et al., 2017) instead of observing how magnitude information modulates acoustic properties of produced vowels. The current study measures parameters of duration, f_0, F_1, and F_2 of produced vowels in order to explore whether they provide any explanatory value for potential congruency between a particular vowel and magnitude. Based on the previous accounts, it is hypothesized that larger magnitudes, in comparison to smaller magnitudes, could be associated with longer vocalizations, lower f_0 and F_2 as well as higher F_1.

2. Experiment 1

Experiment 1 investigates the interaction between producing vowels and representing conceptual magnitude by presenting participants with two objects with a short or long distance between them. Participants are required to pronounce one of the two alternative vowels according to the perceived distance. It is predicted that responses are initiated particularly rapidly when the perceived distance is hypothetically congruent (e.g., [i] - short), as opposed to incongruent, with the required vocalization. As such, Experiment 1 investigates whether short and long distances between two visually presented objects (i.e., allocentric distance) can be sound symbolically associated with the same vowels as short and long distances processed in an egocentric context (Tanz, 1971; Woodworth, 1991; Rabaglia et al., 2016). That is, the study explores whether the sound symbolism effect previously associated with a concept of distance can be based on representing a general concept of distance (including egocentric and allocentric distance representations) rather than being based on a specific sound-meaning mapping exclusively related to egocentric distances. The study presents two objects that are vertically or horizontally positioned to each other. If the sound-distance symbolism effect would be similarly observed in relation to distances of vertically and horizontally positioned objects, this outcome would suggest that allocentric distance information can produce the same sound-distance effect, which has been observed previously in relation to egocentric distance information (Tanz, 1971; Woodworth, 1991; Rabaglia et al., 2016). That is, because distance perception of vertically positioned objects can be based on allocentric *and* egocentric aspects of the stimuli. When the distance between two objects is increased in the vertical dimension, one object is further away from the other object as well as from the viewer. In contrast, distance perception related to horizontally positioned objects is based on allocentric aspects of the stimuli.

It is noteworthy that magnitude-related sound symbolism phenomena can be considered to be relatively abstract because, just like in Sapir's (1929) original observation, they are not necessarily referring to a particular perceived environmental property, but rather they can be referring to an abstract idea of a property or concept (e.g., relative size) (Perniss & Vigliocco, 2014). For example, *here* and *there* (i.e., deictic concepts that are sound symbolically associated with particular vowels) can be relatively abstract concepts whose actual meaning varies depending on the context (e.g., *here* can refer to one's hometown or peripersonal space). Indeed, given the abstractness of these sound-meaning phenomena, it has been proposed that, for example, sound-size and sound-distance effects could be versions of the same sound-magnitude phenomenon associating abstract *smallness-largeness* concepts with particular vowels (Traunmüller, 1996). Despite of that, it is generally assumed that the effect between egocentric distance and a particular vowel is largely based on tongue fronting/backing rather than closing/opening (Tanz, 1971; Woodworth, 1991; Rabaglia et al., 2016) because, for example, back vowels have acoustic properties that make them travel relative long distances (Rabaglia et al., 2016). As stated above, this account assumes that sound-size and sound-distance symbolism are based on isolated sound-mapping processes rather than being overlapping versions of the same sound-magnitude phenomenon. In contrast to the sound-distance symbolism effects, the sound-size symbolism effects have been often associated with vowel height (Ramachandran & Hubbard, 2001; Sapir, 1929). Consequently, if the sound-distance effect would be observed to be similarly attributed to vowel height, in addition to vowel backness, this would support the view that sound-size and sound-distance symbolism could be based on overlapping sound-mapping processes. Hence, Experiment 1 studies whether vowel openness, in addition to backness, can associate a vowel with long distances by exploring a hypothetical congruency between the vowel [æ] (unround-front-open) and perceived long distances.

2.1 Methods

2.1.1 Participants

Twenty volunteers naïve to the purposes of the experiment participated in Experiment 1 (20–39 years of age; mean age = 23.2 years; 6 males; 1 female participant was left-handed). The selection of the sample size was based on our previous similar CRT-based investigation of the influence of perceived magnitude cues on vocalizing the vowels [i] and [ɑ], which showed significant effects of vowel-magnitude compatibility using 16 ($\eta_p^2 = .564$) and 18 ($\eta_p^2 = .560$) participants (Vainio et al., 2017). The sample size calculations were carried out by using G*Power software (Faul et al., 2007). All participants were native speakers of Finnish and reported normal hearing and normal or corrected-to-normal vision. Written informed consent was obtained from all participants. The study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Ethical Review Board in the Humanities and Social and Behavioural Sciences at the University of Helsinki.

2.1.2 Apparatus, stimuli, and procedure

Each participant sat in a dimly lit room with his or her head 70 cm in front of a 19″ CRT monitor (screen refresh rate: 85 Hz; screen resolution: 1280 × 1024). A head-mounted microphone was adjusted close to the participant’s mouth. Figure 1 presents the structure of the experiment. At the beginning of each trial, a black fixation cross (1.5° x 1.5°) was presented for 300 ms in the center of the screen. Then, a blank screen was displayed for 1300 ms. After that the reference stimulus was presented for 500 ms. Then, a blank screen was again displayed for 900 ms. After that the target stimulus was presented for 900 ms. Finally, a blank screen was displayed for 1300 ms. All stimuli were presented on a white background.

The stimuli consisted of two identical greyscale drinking glasses (height: 7°; width: 3.8°) that were positioned either vertically or horizontally next to each other. In the reference stimuli, the two glasses were separated from each other by a distance of 2.2° (vertical reference stimuli) or 4.3° (horizontal reference stimuli). In the target stimuli, the distance between the two glasses was either shorter (the distance was 0.9° for the vertical stimuli and 1.4° for the horizontal stimuli) or longer (the distance was 3.8° for the vertical stimuli and 7.1° for the horizontal stimuli) than in the reference stimuli. The horizontal distances were longer than vertical distances because in the pilot studies it was noticed that people find it easier to discriminate modifications in vertical rather than horizontal distances. The same vertical or horizontal arrangement was present in the reference and the target stimuli of each trial. The order of vertical and horizontal stimuli was randomized throughout the experiment.

Participants were required to pronounce a single vowel according to the target stimulus. The vowels that were included to this experiment were [i] (like in *kit*), [u] (like in *soon*) and [æ] (like in *bad*). The experiment was divided into four blocks that were separated by a short break. The order of the blocks was counterbalanced between participants. In one block ([i]/[u]), the participants had to pronounce [i] if the distance was shorter in the target stimulus than in the reference stimulus and [u] if the distance was longer. In the other [i]/[u] block, this vowel-distance mapping was reversed. In one block ([i]/[æ]), the participants had to pronounce [i] if the distance was shorter in the target stimulus than in the reference stimulus and [æ] if the distance was longer. In the other [i]/[æ] block, this vowel-distance mapping was reversed. Hence, the study consisted of two [i]/[u] and [i]/[æ] blocks with opposite vowel-distance mapping conditions. The beginning of each block included a practice session (including the same number of stimulus-response conditions) that lasted for one minute. The block was not started before the participant demonstrated in the practice session that he/she consistently produced responses according to the instructions. Hence, two participants had to perform twice each of the practice session.

---Figure 1 about here---

The participants were instructed to pronounce the vowel as quickly as possible after the onset of the target stimulus. It was emphasized that the vowel should be uttered as a short (e.g., [i]) rather than a long (e.g., [i:]) vowel in a natural talking voice. Finally, each block included 96 no-response trials (48 vertical and 48 horizontal) in which the participants were required to withhold producing any vocal response when the distance between glasses was the same in the reference and the target stimulus. In total, the experiment consisted of 96 no-response trials and 384 response trials [24 repetitions × 4 (block) × 2 (vowel) × 2 (vertical/horizontal)].

The recording levels of the vocal responses were calibrated individually for each participant at the beginning of the experiment. Stimulus presentation and sound recording were done with Presentation® software. The vocal responses were recorded for 1300 ms starting from the onset of the target object. The vocal data were analyzed using Praat v. 5.3.49 (http://www.praat.org). The onsets of the vocalizations were located individually for each trial as the first observable peak in the acoustic signal for the vowel. Similarly, the offsets of the vocalization were located individually for each trial as the observable ending of the acoustic signal. For this task, onsets and offsets were initially located by a highly experienced person who was blind to the condition of each acoustic signal. Later the correctness of these marked locations was double-checked by the author. The spectral components (F_1 and F_2) as well as f_0 were calculated as median values of the middle third of the voiced section. The procedure of locating the onsets and offsets of vocalizations, as well as calculating the acoustic parameters, was carried out in the same way as in my previous studies that similarly investigate reaction times of vocalizations (e.g., Vainio et al., 2015; Vainio et al., 2017; Vainio et al., 2019a).

2.2 Results and Discussion

The no-response trials were removed from the data before carrying out any analysis. The participants incorrectly produced 25 (1.4%) responses in the no-response trials. The data of two participants were removed from the analysis due to technical problems with the recording system. Reaction times were measured from the onset of the target vowel to the onset of the vocalization. Errors [i.e., the participant uttered the wrong speech unit (1.3%), or did not produce any response (1.5%), or the recording finished before the offset of the vocalization (0.3%)] and reaction times faster than 300 ms (0.1%) were excluded from the analysis because it has been shown that responding to a visually presented target takes a minimum of 200-300 ms (Welford, 1980). None of the participants figured out the purpose of the task when a potential purpose was enquired from the participants after the study. The data is available in the following repository (http://dx.doi.org/10.17632/p8gk496r47.2).

The statistical significance of reaction times[1] was tested by using a random intercept model (linear mixed models) that treated Layout (1 = vertical; 2 = horizontal), Vowel-pair (1 = [i]/[u]; 2 = [i]/[æ]), Distance (1 = shortened; 2 = lengthened), and Vowel (1 = [i]; 2 = [u] or [æ]) as fixed within factors and Subject as a random intercept. Errors were not analyzed due to their very low frequencies (Layout: 1 = 0.6%, 2 = 0.7%; Vowel-pair: 1 = 0.5%, 2 = 0.8%;

---

[1] A slight positive skew was observed for reaction times of Experiment 1 (skewness = .866; SE = .03) and 2 (skewness = .936; SE = .023). For the large sample sizes (greater than 300), an absolute skew value larger than 2 may be used as a reference value for determining substantial non-normality. In addition, the analysis carried out for the log transformed data provided the same outcomes as the analysis carried out for the data that was not log transformed.

Distance: 1 = 0.9%, 2 = 0.5%; Vowel: 1 = 0.7%, 2 = 0.6%). The selection of the error covariance structure was based on Schwarz's Bayesian information criterion (BIC). Post hoc comparisons were carried out by using the Bonferroni correction. The analysis was carried out using the SPSS software package (version 25).

After estimating the best fitting error covariance structure (BIC=82279.69), the analysis of vocal reaction times revealed a significant main effect of Layout [F(1,17) = 31.56, p < .001]. Responses were made faster in the horizontal (M = 597 ms) than vertical (M = 614 ms) Layout. The most likely explanation for this is that horizontal distances were longer than vertical distances, and therefore it was easier to discriminate modifications in the horizontal condition. In addition, not surprisingly, the analysis revealed a significant main effect of Distance [F(1,17) = 6.39, p = .022]. Responses were faster when the distance was long (M = 595 ms) rather than short (M = 617 ms). More importantly, there was the two-way interaction of Distance*Vowel [F(1,6588) = 636.35, p < .001]. Responses were initiated faster with Vowel 1 when the distance between objects was shortened (M = 579 ms) rather than lengthened (M = 627 ms) (p < .001, d (Cohen's d) = 0.42), while responses were initiated faster with Vowel 2 when the distance was lengthened (M = 562 ms) rather than shortened (M = 654 ms) (p < .001, d = 0.81). However, the three-way interaction of Vowel-pair*Distance*Vowel [F(1,6610) = 73.55, p < .001] was also significant. As seen in Figure 2, this three-way interaction can be interpreted so that the interaction between Vowel and Distance is significantly larger in the Vowel-pair condition of [i]/[æ] in comparison to the Vowel-pair of [i]/[u].

Regarding the length of vocalization, most importantly, the analysis (BIC=58020.82) revealed a significant interaction between Vowel-pair and Vowel [F(1,6633) = 35.64, p < .001]. In the Vowel-pair condition of [i]/[u], the vowel [i] was produced shorter (M = 127 ms) than the vowel [u] (M = 133 ms) (p = .002; d = 0.19). In the Vowel-pair condition of [i]/[æ], this difference was not significant (p = .795).

The analysis of acoustic parameters (f_0, F_1 and F_2) was carried out using a random intercept model with the same fixed factors and random intercept as in the reaction time analysis. This analysis revealed some potentially interesting significant interactions. For instance, regarding f_0 values, the analysis revealed a significant interaction of Distance*Vowel [F(1,6629) = 40.66, p < .001], showing higher f_0 for Vowel 1 when the Distance was short (M = 195.3 Hz; SE = 11.3) rather than long (M = 194.1 Hz; SE = 11.3) and higher f_0 for Vowel 2 when the Distance was long (M = 191.7 Hz; SE =1 1.3) rather than short (M = 190.2 Hz; SE =11.3). However, the effect size of this and most of the other significant interactions observed in the data did not meet the criteria of having the effect size (Cohen's d) larger than 0.2. In fact, one of the only interactions that met this criteria was the interaction of Vowel-pair*Vowel observed for F_1 [F(1,6400) = 8133.05, p < .001], and F_2 [F(1,6519) = 27393.06, p < .001] values. In general, not surprisingly, this interaction shows that F_1 values are higher for open vowel [æ] that close vowels ([i] and [u]), and that F_2 values are higher for front-close [i] than back-close [u] and open [æ]. Table 1 presents the central outcomes of analysis of acoustic parameters. This results section does not include further reports of analysis of these acoustic parameters because, as stated above, the analysis did not show any other effects that met the minimum criteria of being statistically significant and having the effect size larger than 0.2 but the effects reported in Table 1.

The analysis of acoustic parameters also showed that variability of F_1 and F_2, as signaled in standard deviation (SD), differed between vowels. The mean SD of F_1 was much larger for [æ] (mean SD = 84) than for [i] (mean SD = 24) and [u] (mean SD = 28). Similar, although not as striking, trend was observed in relation to variability of F_2 values ([i] = 79; [u] = 70; [æ] = 96). This is somewhat in line with previous observations showing more variability in these acoustic parameters for open vs. close vowels (e.g., Recasens & Espinosa, 2006). This discrepancy is likely to be based on biomechanical demands in vowel articulation.

In general, the results of Experiment 1 show that the sound-distance symbolism can be observed in reaction times of producing a single vowel according to distance information. In addition to observing this distance symbolism in relation to egocentric distances (Tanz, 1971; Woodworth, 1991; Rabaglia et al., 2016), the present study shows that it can also be observed in relation to allocentric distances. Furthermore, Experiment 1 showed that this effect is based on vowel backing, as signaled in the comparison between the front-close vowels [i] and the back-close vowel [u], and even more so on vowel opening, as signaled in the comparison between the front-close [i] and front-open [æ].

---Figure 2 about here---

3. Experiment 2

The account of common sound-magnitude phenomenon (Traunmüller, 1996) assumes that similar sound-meaning mappings could be observed concerning the concepts of spatial and temporal magnitudes. In order to explore this view, Experiment 2 investigates whether the same vowels that have been previously associated with *small/large* sizes (Sapir, 1929) and distances (Experiment 1 of the present study; Tanz, 1971; Rabaglia et al., 2016) can be also associated with *small/large* temporal durations. For this purpose, Experiment 2 recruits a version of the paradigm previously established to investigate how magnitude information associated with the target stimuli influences selecting short and long vocalizations for a response (Vainio et al., 2019a). The underlying idea behind Vainio et al’ s (2019a) study was based on the fact that, in speech, intentional vowel lengthening denotes an iconic extension in terms of space (size, distance) or time (duration) (e.g., *looong*) (Perniss &

Vigliocco, 2014). In the task used by Vainio et al. (2019a), the participants were presented with a small (1, 2) or large (8, 9) number, and they were required to pronounce the vowel [ɑ] in a short or long form, depending on whether the number was odd or even. The primary finding of the study was that initiation of short vocalizations was facilitated when the number was small, while initiation of long vocalizations was facilitated when the number was large. Correspondingly, research has shown that short and long vocalizations are associated with short and long objects, respectively (Bross, 2018). Similarly to the observation of Vainio et al. (2019a), this also suggests that requirement to pronounce a short or long version of a vowel associates the vocal response with a corresponding semantic magnitude. If it were true that different vowels convey differing sound symbolic meaning of magnitude, on such a basis it would be assumed that requirements to pronounce a particular vowel (e.g., [i]) would automatically benefit selecting a length for the vocalization that is sound symbolically congruent with the vowel (e.g., short). Hence, in order to study the previously presented phenomenon of magnitude-length correspondence (Vainio et al., 2019a) in the context of sound symbolism, participants are initially instructed that one color (e.g., green) requires the short vocal response, while the other color (e.g., blue) requires the long vocal response. Then they are visually presented with one of the alternative vowels in the green or blue color, and they are instructed to pronounce, as fast as possible, the short (e.g., [u]) or long (e.g., [u:]) version of this vowel based on the color. Hence, participants have to process the semantic category of magnitude assigned to the presented color (e.g., short) in order to pronounce a given vowel. It is predicted that vocal reaction times are shortened when the pronounced vowel is congruent rather than incongruent with concept of temporal duration assigned to the stimulus color.

The vowels that are used in this study are [i] (like in *kit*), [e] (like in *dress*), [u] (like in *soon*), [æ] (like in *bad*) and [y] (like in *view*). The study consists of five blocks (1: [i]-[u]; 2: [i]-

[æ]; 3: [e]-[u]; 4: [e]-[æ]; 5: [i]-[y]) that each compare compatibility of two contrasting vowels with short and long vocalizations. Given that sound-magnitude symbolism effects have been most typically linked to vowel backness (Tanz, 1971; Rabaglia et al., 2016) and openness (Sapir, 1929; Ramachandran & Hubbard, 2001), it is predicted that production of [u] and [æ], in comparison to [i] and [e], is facilitated with the long vocalizations. Furthermore, given that the vowels that have been typically linked to long distances are not only back vowels but are also round vowels (e.g., [u] and [o]) (Rabaglia et al., 2016), the [i]-[y] block explores whether vowel rounding can also link a vowel to the long vocalizations. It is hypothesized that production of the front-close-rounded [y], comparison to the front-close-unrounded [i], is facilitated with the long vocalizations.

3.1 Methods

3.1.1 Participants

Twenty-two volunteers naïve to the purpose of the study participated in Experiment 2a (19–44 years of age; mean age = 30.1 years; 7 males; 1 male participant was left-handed), and twenty-two naïve volunteers participated in Experiment 2b (19–44 years of age; mean age = 30.0 years; 5 males; 3 female participants were left-handed). The selection of the sample size was based on our previous similar CRT-based investigation of the influence of processed magnitude cue on selecting between short and long vocalizations, which showed the significant effect of magnitude-length compatibility using 20 ($\eta_p^2 = 0.634$) participants (Vainio et al., 2019a). The sample size calculations were carried out by using G*Power software (Faul et al., 2007). All participants were native speakers of Finnish and reported normal hearing and normal or corrected-to-normal vision. Written informed consent was obtained from all participants. The study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Ethical Review Board in the Humanities and Social and Behavioral Sciences at the University of Helsinki.

3.1.2 Apparatus, stimuli, and procedure

The apparatus, environmental conditions, and calibration were the same as those in Experiment 1. The target stimuli consisted of vowels (consolas font; lowercase; bold; font size: 90) presented at the center of the display. Figure 1 presents the structure of the experiment. At the beginning of each trial, a white fixation cross was presented for 300 ms in the center of the screen. Then, a blank screen was displayed for 500 ms. After that the target vowel was presented in the center of the screen and remained in view for 1300 ms. Finally, a blank screen was displayed for 1300 ms. All stimuli were displayed on a gray background.

Experiment 2a consisted of two blocks that were separated by a short break. The blocks consisted of two different target vowels (Block 1: [i]-[u]; Block 2: [i]-[æ]). Experiment 2b consisted of three blocks that were similarly separated by a short break. These blocks similarly consisted of two different target vowels (Block 1: [e]-[u]; Block 2: [e]-[æ]; Block 3: [i]-[y]). Half of the vowels were presented in green and the other half in blue. The vowels were presented in randomized order within the blocks. In addition, the order of blocks was counterbalanced between the participants. In total, Experiment 2a consisted of 216 trials [27 × 2 (block) × 2 (vowel) × 2 (response length)], and Experiment 2b consisted of 324 trials [27 × 3 (block) × 2 (vowel) × 2 (response length)].

Half of the participants were required to pronounce the vowel as a short (‘*lyhyt’* in Finnish) vowel if it was blue and as a long (‘*pitkä’* in Finnish) vowel if it was green, and the other half of the participants vice versa. It was emphasized that the vowel should be uttered as quickly as possible in their natural talking voice. The beginning of each block included a practice session (including the same number of stimulus-response conditions) that lasted for one minute. All participants demonstrated in the practice session that he/she consistently produced responses according to the instructions. The arrangements of recording vocal responses, locating the onsets and offsets of vocalizations and calculating acoustic parameters were carried out in the same way as in Experiment 1.

At this point, it should be mentioned that Finnish language includes eight different vowels (/ɑ e i o u y æ ø/), and does not include any single letter words. All of these vowels may be short or long. Duration is a distinctive feature for vowels in Finnish leading to words such as *tuli* (fire) and *tuuli* (wind) (Sajavaara & Dufva, 2001). Short vowels, also in unstressed positions, have approximately the same quality as long ones (Heikkinen, 1979). Moreover, in Finnish, letters are (without any exceptions) pronounced as they are written. In other words, every letter of alphabet can be pronounced in one way only. Pronunciation does not depend on a context or surrounding letters.

3.2 Results and Discussion

In Experiment 2a, errors [i.e., the participant uttered the wrong speech unit (1.6%), or did not produce any response (1.2%), or the recording finished before the offset of the vocalization (0.4%)] and reaction times faster than 300 ms (0.2%) were excluded from the reaction time analysis. In Experiment 2b, errors [i.e., the participant uttered the wrong speech unit (1.7%), or did not produce any response (1.2%), or the recording finished before the offset of the vocalization (0.1%)] and reaction times faster than 300 ms (0.0%) were excluded from the analysis. None of the participants figured out the purpose of the task when a potential purpose was enquired from the participants after the study. The data is available in the following repository (http://dx.doi.org/10.17632/p8gk496r47.2).

---Table 1 about here---

The data of Experiments 2a and 2b was analyzed in a single linear mixed model analysis. The statistical significance of reaction time values was tested by using a random intercept model that treated Vowel-pair (1 = [i]-[u]; 2 = [i]-[æ]; 3= [e]-[u]; 4 = [e]-[æ]; 5 = [i]-y]), Vowel (1 = [i]/[i]/[e]/[e]/[i]; 2 = [u]/[æ]/[u]/[æ]/[y]), and Length (1 = short; 2 = long) as fixed within factors and Subject as a random intercept. Errors were not analyzed due to their very low frequencies (Vowel-pair: 1 = 0.9%, 2 = 0.7%; Vowel: 1 = 0.8%, 2 = 0.8%; Length: 1 = 1.2% = 0.6%). The selection of the error covariance structure was based on Schwarz's Bayesian information criterion (BIC). Post hoc comparisons were carried out by using the Bonferroni correction. The analysis was carried out using the SPSS software package (version 25).

After estimating the best fitting error covariance structure (BIC=136346.92), the analysis of vocal reaction times revealed a significant main effect of Length [F(1,39) = 12.64, p = .001]. Long vocalizations were made faster (M = 586) than short vocalizations (M = 598 ms). The two-way interaction of Vowel*Length was significant [F(1,11297) = 209.91, p < .001]. Short vocalizations were initiated faster with Vowel 1 (M = 587 ms) than Vowel 2 (609 ms) (p < .001, d = 0.33), whereas long vocalizations were initiated faster with Vowel 2 (M = 572 ms) than Vowel 1 (M = 599 ms) (p < .001, d = 0.42). The three-way interaction of Vowel-pair*Vowel*Length [F(4,11298) = 2.92, p = .020] was also significant. As presented in Figure 3, the pairwise comparison test showed that, in each Vowel-pair condition, the short vocalizations were initiated significantly faster with Vowel 1 than Vowel 2 (Vowel-pair [i]/[u]: [i] M = 614 ms, [u] M = 632 ms, p = .001, d = 0.20; [i]/[æ]: [i] M = 601 ms, [æ] M = 620 ms, p = .001, d = 0.20; [e]/[u]: [e] M = 576 ms, [u] M = 590 ms, p = .007, d = 0.16[2]; [e]/[æ]: [e] M

---

[2] An effect under 0.2 can be considered trivial.

= 569 ms, [æ] M = 598 ms, p < .001, d = 0.30; [i]/[y]: [i] M = 574 ms, [y] M = 603 ms, p < .001, d = 0.30). In contrast, the long vocalizations were initiated significantly faster with Vowel 2 than Vowel 1 (Vowel-pair [i]/[u]: [i] M = 622 ms, [u] M = 599 ms, p < .001, d = 0.24; [i]/[æ]: [i] M = 612 ms, [æ] M = 583 ms, p < .001, d = 0.30; [e]/[u]: [e] M = 584 ms, [u] M = 561 ms, p < .001, d = 0.24; [e]/[æ]: [e] M = 576 ms, [æ] M = 560 ms, p = .004, d = 0.17[1]; [i]/[y]: [i] M = 600 ms, [y] M = 558 ms, p < .001, d = 0.44).

The previous research has shown associations between particular colors and vowels. For instance, participants have tendency to associate high-front vowels with greenish color (Cuskley et al., 2019; Kim et al., 2018; Moos et al., 2014). In order to explore whether similar color-vowel mappings could be observed in the present data, the data was analyzed so that color-response mapping was included to the analysis as an additional (between-subjects) independent factor. This analysis did not show any significant linkages between color and vowel. The interactions of Vowel*Mapping (p = .479) and Vowel_pair*Vowel*Mapping (p = .087) were not significant.

Regarding the length of vocalization, as expected, the analysis (BIC=120272.84) revealed a significant main effect of Length [*F*(1,42) = 272.28, *p* < .001]. The vocalizations were shorter with short vowels (*M* = 160 ms) than with the long vowels (*M* = 490 ms). This is in line with segmental durations in continuous Finnish speech, where long sounds are, on average, twice as long as the short ones. In addition, the analysis revealed a significant interaction of Vowel-pair*Vowel*Length [*F*(4,11027) = 7.92, *p* < .001]. The pairwise comparison test showed that in the long vocalizations of Vowel-pair conditions of [i]/[u], [e]/[u] and [i]/[y], Vowel 2 was produced significantly longer than Vowel 1 (Vowel-pair [i]/[u]: [i] M = 441 ms, [u] M = 450 ms, p = .016, d = 0.1[1]; Vowel-pair [e]/[u]: [e] M = 514 ms, [u] M = 536 ms, p < .001, d = 0.23; Vowel-pair [i]/[y]: [i] M = 517 ms, [y]: M = 545 ms, p < .001, d = 0.29).

The analysis of acoustic parameters (f_0, F_1 and F_2) was carried out using a random intercept model with the same fixed factors and random intercept as in the reaction time analysis. This analysis revealed some potentially interesting significant interactions. However, similarly to Experiment 1, the effect size of these interactions did not meet the criteria of having the effect size larger than 0.2. In fact, the only interaction that met this criteria was a significant interaction of Vowel-pair*Vowel observed for F_1 [F(4,164) = 4522.59, p < .001], and F_2 [F(4,161) = 12058.97, p < .001] values. In general, not surprisingly, this interaction shows that F_1 values increase as a function of vowel openness, while F_2 increases as a function of vowel frontness. Table 1 presents the central outcomes of analysis of acoustic parameters. This results section does not include further reports of analysis of these acoustic parameters for the same reason as in Experiment 1.

Similarly to Experiment 1, the analysis of acoustic parameters also showed that variability of F_1 and F_2 differs between vowels. The mean SD of F_1 was much larger for [æ] (mean SD = 80) than for other vowels ([i] = 22; [u] = 30; [e] = 40; [y] = 24). Correspondingly, SD values of F_2 were larger for [æ] (mean SD = 110) than for [i] (mean SD = 84) and [u] (mean SD = 66). The variability of F_2 values was almost as large for [e] (mean SD = 103) and [y] (mean SD = 104) as for [æ]. Given that these variability aspects have not been investigated in Finnish, and that there appears to be language-dependent trends in this respect (Recasens & Espinosa, 2006), it is difficult to make any strong statements about these effects. What can be concluded, based on findings of Experiment 1 and Experiment 2, is that the variability of F_1 and F_2 observed in the present study complies with the size range of F_1 and F_2 variability observed in other studies (see for example Nicolaidis, 2003). Secondly, it can be concluded that vowel [æ] is associated with much greater variability, particularly in relation to F_1 values, than other vowels explored in this study, perhaps because it is more open than other vowels included in the study (Recasens & Espinosa, 2006).

These results show that initiation of short vocalizations is facilitated with the unrounded-front vowels [i] and [e], while initiation of long vocalizations is facilitated with the unrounded-front-open vowel [æ], rounded-back-close vowel [u] and rounded-front-close vowel [y]. As such, it seems that the same vowels that are linked to short and long distances in Experiment 1 are also linked to short and long vocalizations, respectively. In addition, the results of the [i]-[y] block show that the vowel rounding associates the vowel with long vocalizations.

---Figure 3 about here---

4. General Discussion

The current study presents two novel effects of sound symbolism. Firstly, Experiment 1 demonstrated for the first time a sound symbolic association between short and long allocentric distances and particular vowels. The participants had to pronounce one of the two response alternatives (e.g., [i] or [u]) depending on whether the distance between the two objects was shortened or lengthened in the target stimulus relative to the reference stimulus. Initiation of the unround-front-close vowel [i] was facilitated when the distance between the two objects was shortened, while initiation of the round-back-close vowel [u] and the unround-front-open vowel [æ] was facilitated when the distance was lengthened. This finding is in line with the previous observations in which [i] is linked to short egocentric distances and [u] is linked to long egocentric distances (Tanz, 1971; Rabaglia et al., 2016). It is likely that both of these distance-related effects are based on representing a general concept of distance rather than independent sound-distance mappings from which one is related to egocentric distance and other is related to allocentric distance. Moreover, the previous research links particularly tongue backness rather than openness with long distances (Tanz, 1971; Woodworth, 1991). However, Experiment 1 showed that the sound-distance effect is significantly larger between a long distance and the front-open vowel [æ] than a long distance and the back-close vowel [u]. This suggests that long distance can be associated solely with the openness dimension of a vowel.

Secondly, Experiment 2 showed that the same vowels that were associated with short and long distances in Experiment 1 can also be associated with short and long vocalizations, respectively. Production of the vowels [i] and [e] was initiated particularly rapidly when the color stimulus called for selecting a short vocalization as a response, while the vowels [u], [æ] and [y] were initiated particularly rapidly when the color stimulus called for selecting a long vocalization as a response. This sound-duration effect is not likely to be caused by consistencies in frequencies of long and short versions of different vowels in Finnish language. That is, because, for example, the same portion of short (22%) and long (20%) vowels are [i]s in Finnish. Roughly, the same applies to the vowels [e] (17% vs. 13%), [u] (10% vs. 12%), [y] (4% vs. 2%) and [æ] (11% vs. 17%) (Häkkinen, 1983; Pääkkönen, 1991). Although there could be a slight trend of, for example, having larger portion of [u]'s and [æ]'s in the long vowels relative to short vowels, the trend seems to go other way around in relation to the vowel [y], yet long vocalizations were also associated with the vowel [y] in the present study.

In addition, the sound-duration effect observed in Experiment 2 is not likely to be caused by different intrinsic durations of vowels because intrinsic vowel durations are typically measured from speech without requiring participants to select between short or long vocalizations intentionally, and because intrinsic vowel durations typically increase as a function of vowel openness (Lehiste, 1970; Lehnert-LeHouillier, 2007). In the present study, the closed vowels [y] and [u] were associated with the long vocalizations, and this was only observed when the performance engaged recalling the meaning of a length cue that was sound symbolically congruent with the required vowel. For these reasons, it is likely that this effect observed in Experiment 2 reflects a genuine sound symbolic association between a concept of duration magnitude and a particular vowel. In the same way as a number magnitude has been shown to bias response selection between short and long vocalizations (Vainio et al., 2019a), the vowel that is required for a vocal response seems to be implicitly associated with a particular concept of magnitude biasing response selection between short and long vocalizations.

As stated in the Introduction, some accounts of sound-magnitude symbolism hold that heightened pitch associates a particular vowel with small magnitudes (Ohala & Eukel, 1987; Ohala, 1984**;** Spence, 2011). The results of the present study do not support this view. As presented in Table 1, increased f_0 values were not systematically associated with the vowels that were congruent with short distances (Experiment 1) or short vocalizations (Experiment 2). Similarly, Knoeferle et al. (2017) investigated how participants associate acoustic properties of perceived vowels with small/large objects and found that f_0 does not seem to have a role in sound-magnitude symbolism. Hence, it seems that sound-magnitude symbolism is not attributable to pitch differences in perceived or produced vowels.

According to an inference about tongue position from the acoustics, F_1 values (as presented in Table 1) support the view that, regarding congruency between large magnitudes and the vowel [æ], vowel openness contributes to the effects of the study. In contrast, regarding the observed congruency between large magnitudes and the vowels [u] and [y], the effects have to be associated with some other parameters of the vowel than openness. This is because both of them are closed vowels with similar F_1 values as their vowel pair [i]. Nor does this parameter that links the vowels [u] and [y] with large magnitudes seem to be $f_{0,}$ as stated above. What might be the parameter that links the vowels [u] and [y] with large magnitudes in Experiments 1 and 2?

It is noteworthy that in Experiment 2, similar effects between long magnitude and a particular vowel were observed with the round-front-close vowel [y] as well as the round-back-close vowel [u], even though according to an inference about tongue position from F_2 values, as seen in Table 1, [u] was considerably more back than [y] (F_2 was 1500 Hz lower for [u] in comparison to [y]). Hence, after questioning the explanatory value of f_0, vowel openness and vowel backness on the linkage between the vowel [y] and long magnitude, what we have left is the vowel roundness. That is, regarding the vowel [y], the vowel roundness seems to be the most salient parameter that differentiates [y] from [i] in the [i]-[y] block of Experiment 2. They are both front-closed vowels, yet only the [y] is rounded vowel.

The view that the vowel roundness can connect the vowel with large magnitudes is in line with the suggestion that lip rounding might associate a particular vowel with distal deixis (Johansson & Zlatev, 2013). What might be the reason for this connection between vowel roundness and the concept of *longness*? One finding of the present study that might shed some light on this issue is that in both studies, the round vowels ([u] in Experiment 1; [u] and [y] in Experiment 2) were produced slightly longer than the unround vowels with which they were paired (i.e., [i] and [e]). It is, for example, possible that participants preferred to produce rounded vowels, in these conditions, as relatively long because rounded vowels are implicitly associated with a conceptual representation of *longness*. Furthermore, gestural explanations could, for example, assume that roundness matches a vowel with a long magnitude because rounding links the vowel iconically to the concept of *longness* by gestural extending of lip protrusion, which lengthens the frontal cavity.

Regarding the rounded-front-close vowel [u], the previous studies have emphasized that it is in particular the vowel backness rather than roundness that matches this vowel to large magnitudes (e.g., Nuckolls, 1999; Thompson & Estes, 2011; Rabaglia et al., 2016). This assumption cannot be rejected based on the results of the current study. However, given that the vowel roundness might be a critical vowel property that connects it with large magnitudes, as stated above, it is likely that the roundness property of the vowel [u], at least, contributes to this sound symbolism effect. Perhaps the relatively large front cavity linked to production of the vowel [u], increased as a function of lip rounding *and* tongue backing, makes this vowel particularly suitable match with large concepts.

As mentioned in the Introduction, many, if not most, stimulus-response effects that have used a CRT method in exploring how perceptual aspects influence responses have attributed the effects to automatic stimulus-induced priming of an action schema (e.g., a particular grasp motor representation) that is congruent with the stimulus property (e.g., size) (Tucker & Ellis, 2001; Vainio & Ellis, 2020). Viewed in this light, one might assume that the present results similarly reflect activation of vowel articulation schemas induced by a perceived magnitude. As an example, in Experiment 1, the facilitated pronunciation of the vowel [æ] in association with long distances might be caused by activation of an articulation schema of the vowel [æ] as a consequence of perceptual and/or conceptual processing of a relatively large distance of the target stimulus. This interpretation of the current findings might be, however, too far-reaching. Nevertheless, at least it could be stated that the results of Experiment 1 showed that processing a concept of distance implicitly engages a vocal response associated with the concept, and consequently biases selection of vocal response in favor of the response alternative that is congruent with the concept.

Together, based on the results of the current study, it can be proposed that the sound-size-, sound-distance-, and sound-duration effects are versions of the same sound symbolism phenomenon anchored on the common conceptual representation of magnitude in terms of space (size, distance) and time (duration). In addition to being semantically related concepts, in all of these effects, *smallness* and *largeness* are associated with the same vowels (i.e., [i]/[e] and [o]/[u]/[ɑ], respectively). This account is further supported by the fact that quick and slow movements, which also are semantically related to concepts of magnitude, are associated with front-close and back-open vowels, respectively (Cuskley, 2013). As such, this view can be linked to “a theory of magnitude” (ATOM) (Walsh, 2003) according to which there is an overlap in sensorimotor processes that encode magnitude information for the metrics of time, space, and quantity. This account holds that overlapping sensorimotor processes, primarily grounded in gestural (manual grasping in particular) sensorimotor processes (e.g., Andres et al., 2004), provide the neural basis for abstracting different kinds of magnitude information, such as number magnitude, speed, length, distance, duration, size, and so forth. The present view holds that in addition to grounding this common magnitude encoding in the manual gestures, as proposed in the ATOM model, the processes responsible for forming articulatory gestures might also play a role in the system that represents common magnitude information. This proposal is compatible with the findings showing connected motor networks between vocal gestures and manual grasping (Vainio, 2019; Vainio et al., 2019b). Hence, these embodied mechanisms that enable representing abstractions of magnitude information in relation to motor processes of manual and articulatory actions manifest themselves in sound-magnitude symbolism phenomena.

In conclusion, specific vowels can be associated with short and long allocentric distances, not just egocentric distances as has been previously presented. In addition, this study argued that not just vowel backness, as previously presented, but vowel openness can also link a particular vowel to a concept of long distance. Furthermore, the study demonstrated that the same vowels associated with short and long distances in Experiment 1 were associated with short and long vocalizations in Experiment 2. In addition to vowel openness, this second study suggested that vowel rounding also seems to associate vowels with the concept of temporal *longness*. Together, these findings reveal that sound-meaning mapping manifests itself in vowel production: the meaning of magnitude can be automatically (i.e., without explicit requirement to assign this meaning to a linguistic unit) associated with an articulatory response of a specific vowel. Finally, it was proposed that all sound symbolism effects that associate *smallness* and *largeness* with particular vowels, regardless of the specific type of magnitude concept (e.g., distance, size, duration, speed), are based on embodied processes that anchor abstracted magnitude concepts in gestural (manual and articulatory) representations.

Acknowledgments

I would like to thank Aleksi Rantala for his contribution to analyzing data, and Aida Lindy, Nelli Frilander, Sara Göransson, Fredrik Lagerstam and Lasse Vatanen for their contribution to planning and carrying out Experiment 1. In addition, I would like to thank Christine Cuskley, James Brand and the anonymous reviewer for their significant contribution for improving quality of the paper during the review process.

References

Andres, M., Davare, M., Pesenti, M., Olivier, E., & Seron, X. (2004). Number magnitude and grip aperture interaction. Neuroreport, 15, 2773-2777.

Bentley, M., & Varon, E. J. (1933). An accessory study of" phonetic symbolism". *The American Journal of Psychology*, *45*(1), 76-86.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. Proceedings of the National Academy of Sciences, 113(39), 10818-10823.

Bross, F. (2018). Cognitive associations between vowel length and object size: A new feature contributing to a bouba/kiki effect. In Proceedings of the Conference on Phonetics & Phonology (pp. 17-20).

Cuskley, C. (2013). Mappings between linguistic sound and motion. Public Journal of Semiotics, 5(1), 39-62.

Cuskley, C., Dingemanse, M., Kirby, S., & Van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: Categorical perception and structure in vowel-color mappings in a large online sample. *Behavior research methods*, *51*(4), 1651-1675.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. *Implicit measures of attitudes: Procedures and controversies*, 179-194.

Diffloth, G. (1994). i: big, a: small. In Sound Symbolism, Hinton, L., Johanna Nichols, J. and Ohala, J.J. (Eds.), pp. 107-14. Cambridge University Press, Cambridge, UK.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. Trends in Cognitive Sciences, 19(10), 603-615.

Fant, G. (1960). I960]: Acoustic Theory of Speech Production. The Hague: Mouton.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Heikkinen, H. (1979). Vowel reduction in the English of Finnish leamers. In J. Lehtonen & K. Sajavaara (Eds.) *Papers in contrastive phonetics,* Jyväskylä Cross-Language Studies 7, Jyväskylä. University of Jyväskylä.

Häkkinen, K. (1983). Suomen kielen äännera- kenteen ominaispiirteistä. Teoksessa A. Hakulinen & P. Leino (eds.), *Nykysuomen rakenne ja kehitys, vol 1* (pp. 39–56). Pieksämäki: Suomalaisen Kirjallisuuden seura.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130298.

Jespersen, O. 1933. “Symbolic value of the vowel i,” In O. Jespersen, Linguistica: Selected papers in English, French and German. Copenhagen: Levin and Munksgaard. 283-303.

Johansson, N., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. Public Journal of Semiotics, 5(1), 3-20.

Kim, H. W., Nam, H., & Kim, C. Y. (2018). [i] is lighter and more greenish than [o]: Intrinsic association between vowel sounds and colors. *Multisensory Research*, *31*(5), 419-437.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. Scientific Reports, 7(1), 1-11.

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language*, *34*(5), 387-405.

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology*, *161*, 32-45.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. The Journal of the Acoustical Society of America, 51(6B), 2018-2024.

Lehnert-LeHouillier, Heike (2007). The Perception of Vowel Quantity: A Cross-Linguistic Investigation. Ph.D. thesis, State University of New York, Buffalo.

Magnus, M. (2000). What's in a Word? Evidence for Phonosemantics (Doctoral dissertation, Norwegian University of Science and Technology).

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5), 1152.

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder. *i-Perception*, *5*(2), 132-142.

Nicolaidis, K. (2003). Acoustic variability of vowels in Greek spontaneous speech. *Proceedings of the 15th international congress of phonetic sciences*, 3221-3224.

Nuckolls, J. B. (1999). The case for sound symbolism. Annual Review of Anthropology, 28(1), 225-252.

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

Ohala, J. J., & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. In honor of Ilse Lehiste, 207-215.

Parise, C. V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. Experimental Brain Research, 214, 373– 380.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130300.

Pääkkönen, M. (1991). A:sta Ö:hön. Suomen kielen yleisyystilastoja. Kielikello, 1.

Rabaglia, C. D., Maglio, S. J., Krehm, M., Seok, J. H., & Trope, Y. (2016). The sound of distance. Cognition, 152, 141-149.

Ramachandran, V.S., & Hubbard, E.M. (2001). Synaesthesia--a window into perception, thought and language. Journal of Consciousness Studies, 8, 3-34.

Recasens, D., & Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech communication*, *48*(6), 645-666.

Sajavaara, K., & Dufva, H. (2001). Finnish-English phonetics and phonology. *International Journal of English Studies*, *1*(1), 241-256.

Sapir, E. (1929). A study in phonetic symbolism. Journal of Experimental Psychology, 12(3), 225.

Saussure, F. D. (1916). Course in general linguistics (trans. Wade Baskin). London: Fontana/Collins, 74.

Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. In *Annual Meeting of the Berkeley Linguistics Society*, 36, 396-410.

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. Psychonomic Bulletin & Review, 25(5), 1619-1643.

Smith, E. E. (1968). Choice reaction time: An analysis of the major theoretical positions. Psychological Bulletin, 69(2), 77.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics, 73(4), 971-995.

Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. Language and Speech, 14(3), 266-276.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly journal of experimental psychology*, *64*(12), 2392-2404.

Traunmüller, H. (1996). Sound symbolism in deictic words. TMH-QPSR, 2, pp. 147-151

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual cognition*, *8*(6), 769-800.

Ultan, Russel. 1978. Size-Sound Symbolism. In J. Greenberg ed., *Universals of Human Language II: Phonology*, 525-568. Stanford: Stanford Univ. Press.

Vainio, L. (2019). Connection between movements of mouth and hand: Perspectives on development and evolution of speech. Neuroscience & Biobehavioral Reviews, 100, 211-223.

Vainio, L., Mustonen, T., & Vainio, M. (2019a). The Influence of Number Magnitude on Vocal Responses. Journal of Motor Behavior, 51(2), 129-140.

Vainio, L., Rantala, A., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2017). Systematic influence of perceived grasp shape on speech production. PloS one, 12(1)

Vainio, L., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2015). Interaction in planning movement direction for articulatory gestures and manual actions. Experimental Brain Research, 233(10), 2951-2959.

Vainio, L., Vainio, M., Lipsanen, J., & Ellis, R. (2019b). The Sound of Grasp Affordances: Influence of Grasp-Related Size of Categorized Objects on Vocalization. Cognitive Science, 43(10).

Vainio, L., & Ellis, R. (2020). Action inhibition and affordances associated with a non-target object: An integrative review. *Neuroscience & Biobehavioral Reviews*, 112, 487-502.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. Trends in Cognitive Sciences, 7(11), 483-488.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of phonetics*, *23*(3), 349-366.

Whalen, D. H., Levitt, A. G., Hsiao, P. L., & Smorodinsky, I. (1995). Intrinsic F 0 of vowels in the babbling of 6-, 9-, and 12-month-old French-and English-learning infants. *The Journal of the Acoustical Society of America*, *97*(4), 2533-2539.

Woodworth, N. L. (1991). Sound symbolism in proximal and distal forms. Linguistics, 29(2), 273-300.

Figure and table captions

Figure 1. Schematic depictions of the temporal structure of Experiments 1 and 2. In Experiment 1, the participants' task was to pronounce one of the two response alternatives (e.g., [i] or [u]) after onset of the target stimulus (b) depending on whether the distance between the two glasses was shortened or lengthened in the target stimulus relative to the reference stimulus (a). In Experiment 2, the participants' task was to pronounce the displayed vowel in a short or long form depending on the color (green, blue) of the vowel.

Figure 2. Box-plot of reaction times (ms = millisecond) for Experiment 1 (horizontal line inside the box = median; cross inside the box = mean; box = 25-75%; whiskers = scores outside the middle 50%). The box-plots show the distributions of RT (reaction time) values as a function of Block (four boxes in left: [i]-[u] block; four boxes in right: [i]-[æ] block), Vowel and Distance. Asterisks indicate statistically significant differences (***p < .001, **p < .01).

Figure 3. Box-plot of reaction times (ms = millisecond) for Experiment 2 (horizontal line inside the box = median; cross inside the box = mean; box = 25-75%; whiskers = scores outside the middle 50%). The box-plots show the distributions of RT (reaction time) values as a function of Block (1: [i]-[u] block; 2: [i]-[æ] block; 3: [e]-[u] block; 4: [e]-[æ] block; 5: [i]-[y] block), Vowel and vocalization Length. Asterisks indicate statistically significant differences (***p < .001, **p < .01).

Table 1. Mean f_0, F_1 and F_2 values of particular vowels in each experimental block based on a linear mixed models analysis described in the results sections of Experiment 1 and 2. Only those differences between mean values of paired vowels are marked with an asterisk that

meet the criteria of being statistically significant and having the effect size (Cohen's d) larger than 0.2. The vowel whose mean value is significantly higher in comparison to the opposite vowel is written in bold.

Figure 1

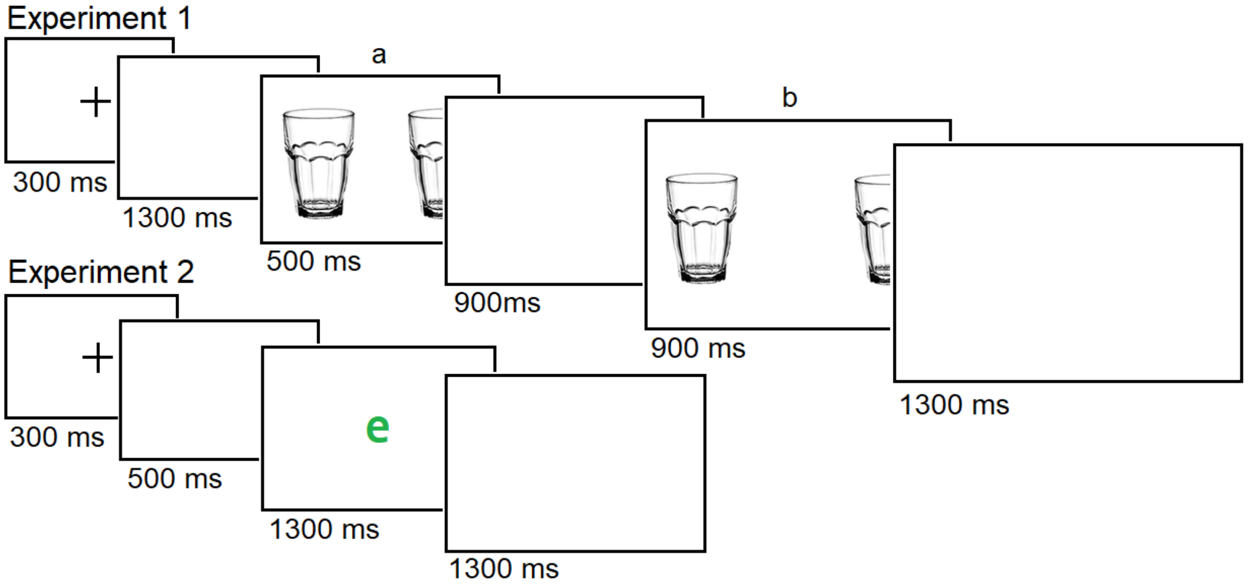

Figure 2

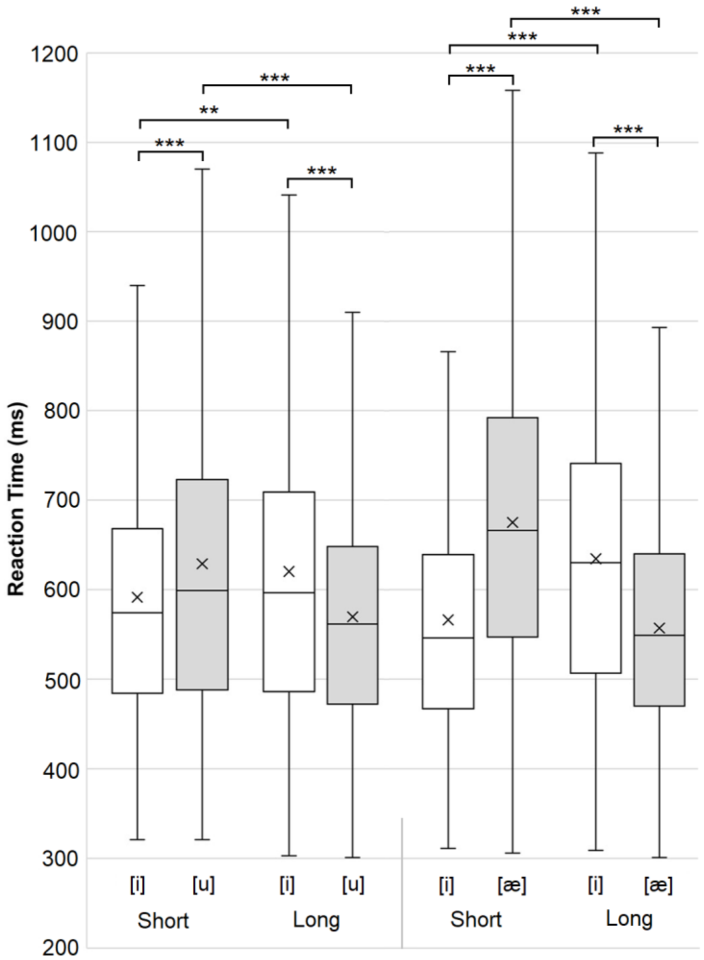

Figure 3

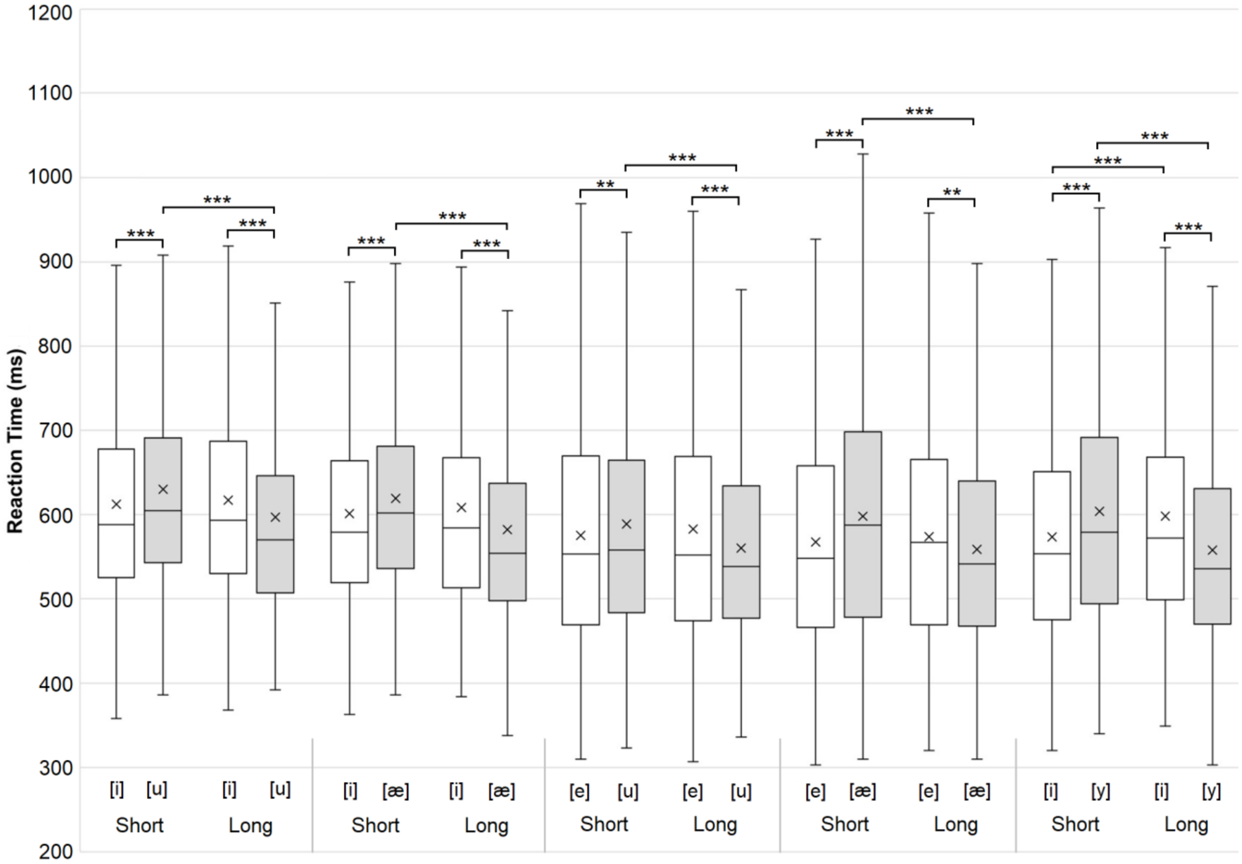

Table 1

Exp	Blocks	f_0 (Hz)	F_1 (Hz)	F_2 (Hz)
1	[i]/[u]	[i] 195 – [u] 194	[i] 311 – [u] 334	**[i] 2842** – [u] 685 *
	[i]/[æ]	[i] 195 – [æ] 188	[i] 304 – **[æ] 618** *	**[i] 2863** – [æ] 1465 *
2	[i]/[u]	[i] 181 – [u] 182	[i] 283 – [u] 313	**[i] 2685** – [u] 627 *
	[i]/[æ]	[i] 178 – [æ] 173	[i] 282 – **[æ] 584** *	**[i] 2698** – [æ] 1482 *
	[e]/[u]	[e] 192 – [u] 199	**[e] 451** – [u] 325 *	**[e] 2486** – [u] 592 *
	[e]/[æ]	[e] 190 – [æ] 188	[e] 440 – **[æ] 676** *	**[e] 2512** – [æ] 1523 *
	[i]/[y]	[i] 199 – [y] 198	[i] 302 – [y] 317	**[i] 2910** – [y] 2142 *