Running head: Sound symbolism and vocalization

Sharp and round shape of seen objects have distinct influences on vowel and consonant articulation

Vainio, L., Tiainen, M., Tiippana, K., Rantala, A., & Vainio, M.

1. *Vainio L (the corresponding author; email: lari.vainio@helsinki.fi; tel: +358-9-29392)

2. *Tiainen M

3. *Tiippana K

4. *Rantala A

5. **Vainio M

* Institute of Behavioural Sciences, Division of Cognitive and Neuropsychology, University of Helsinki, Siltavuorenpenger 5 A, Helsinki PL 9, 00014, Finland

** Institute of Behavioural Sciences, Phonetics and Speech Synthesis Research Group, University of Helsinki, Siltavuorenpenger 5 A, Helsinki PL 9, 00014, Finland

Abstract

The shape and size-related sound symbolism phenomena assume that, for example, the vowel [i] and the consonant [t] are associated with sharp-shaped and small-sized objects whereas [ɑ] and [m] are associated with round and large objects. It has been proposed that these phenomena are mostly based on involvement of articulatory processes in representing shape and size properties of objects. For example, [i] might be associated with sharp and small objects because it is produced by a specific front-close shape of articulators. Nevertheless, very little work has examined whether these object properties indeed have impact on speech sound vocalization. In the present study, the participants were presented with a sharp or round-shaped object in a small or large size. They were required to pronounce one out of two meaningless speech units (e.g., [i] or [ɑ]) according to the size or shape of the object. We investigated how a task-irrelevant object property (e.g., the shape when responses are made according to size) influences reaction times, accuracy, intensity, fundamental frequency, formant 1 and formant 2 of vocalizations. The size did not influence vocal responses but shape did. Specifically, the vowel [i] and consonant [t] were vocalized relatively rapidly when the object was sharp-shaped whereas [u] and [m] were vocalized relatively rapidly when the object was round-shaped. The study supports the view that the shape-related sound symbolism phenomena might reflect mapping of the perceived shape with corresponding articulatory gestures.

Keywords: sound symbolism, perception, articulation, speech, priming, reaction time

Introduction

It might be rightly argued that phonetic sounds of most words do not directly convey information about their actual meaning. In fact, the arbitrary relationship between the sound of a word and its meaning has been usually taken as a fundamental principle of influential studies of language (e.g., Saussure 1916; Newmeyer 1993). However, exceptions of this principle have also been recognized (e.g., Shintel, Nusbaum and Okrent 2006; Ultan 1978). The most famous sound symbolic demonstrations of this relationship between the sounds and meaning associate object shape and size with different speech sounds (Köhler 1929; Sapir 1929). The present study focuses on investigating these two sound symbolic phenomena.

One of the most popular accounts of the sound-shape – also known as the Kiki-Bouba effect – as well as the sound-size correspondence effect proposes that these effects are mostly based on integrated processing of visual shape/size of the object and articulatory representations of the word. For example, Ramachandran and Hubbard (2001) have elaborated this account into a broader synesthetic theory of language evolution assuming that these effects reflect synesthesia-like cross-activation of the motor representations involved in articulating different sounds and the visual representations involved in perception of differently shaped/sized objects. This view does not appear to make a strong distinction between phonetic and articulatory processes but rather adapts views of the motor theories of speech perception (Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967) according to which there is a common interface between articulation and phonetic processing levels in speech production and perception. According to this account, certain speech sounds associated with, for example, the pseudoword '*bouba'* are matched to round shapes because the articulatory production of that word involves rounding of the mouth. Similarly regarding the sound-size correspondence effect, for example, the vowel [ɑ] is matched to large objects because the articulatory production of that vowel involves relatively large opening of the mouth.

This kind of explanation of these sound symbolism phenomena is based on views of embodied cognition according to which perceived objects are contextually represented in distributed

and multi-modal networks consisting of sensory, motor and association cortices (e.g., Barsalou 2008; Binder and Desai 2011). According to this view, it would be plausible to assume that when one has to represent the viewed object, for example, for a categorization task, the shape and size of the object is not represented only at the perceptual level but also at the motor level, partially activating specific motor representations such as the corresponding articulatory representations. As an alternative view, which is not however necessarily mutually exclusive with the articulatory explanation, these sound symbolism phenomena might reflect acquired cross-modal associations between auditory and visual processes in which, for example, *"smooth"* sounds are associated with round shapes whereas "*fractured"* sounds are associated with sharp shapes.

Although the articulatory accounts of these sound symbolism phenomena predict that processing the object properties such as shape and size during preparation of vocalizations should influence vocal responses, vocalization has not been typically involved in the tasks that investigate these phenomena. As a consequence, the present study investigates whether shape and/or size of viewed object could have a systematic influence on vocal responses. Our participants were presented with a sharp or round-shaped object in a small or large size. They were required to pronounce one out of two meaningless speech units (e.g., [i] or [ɑ]) according to the size or shape of the object. Primarily we investigated how a task-irrelevant object property (e.g., a shape when responses are made according to size) influences reaction times (RT) of vocalizations. We predict for example that sharp-edged and/or small objects should prime close and non-round vocalizations (e.g., producing the vowel [i]). We hypothesize that this influences should be observed in relatively fast congruent responses (e.g., vocalization of the vowel [i] when the object is sharp-edged or small) and relatively slow incongruent responses (e.g., vocalization of the syllable [i] when the object is round-edged or large).

In addition to the potential influence of the object shape/size on vocal RTs, we also speculated that these object properties might systematically modulate certain components of voice characteristics. Consequently, the present study also explores whether intensity (i.e., the maximum intensity of the speech sound that is indicated in decibels), fundamental frequency (f_0), formant 1 (F1) or formant 2 (F2) are modulated by the viewed shape or size of the object. The formants reflect

the acoustic resonance of speech, modulated by shapes of the vocal tract. According to the articulatory account, perception of a large object could be assumed to increase the F1 component of voice characteristics as this component reflects opening of a vocal tract during vocalization (Fant 1960); the more the mouth is opened during articulation, the more F1 is increased. Moreover, increase in f_0 can be associated with increased intensity of vocalization (Alku, Vintturi and Vilkman 2002) as well as decreased F1 values (Whalen and Levitt 1995). Consequently, it would be plausible to assume that if F1 is modulated by the size (or shape) of the perceived objects, the vocal characteristics of intensity and f_0 would be also modulated by the size of the perceived object. Furthermore, it might be speculated that F2 would be modulated by the shape of the object so that sharp objects would increase F2 values because F2 reflects frontness position of the tongue in vocalization; the more the tongue is pushed forwards in articulation, the more F2 is increased. This prediction is driven by the fact that the front vowels are usually linked to sharp shapes and the back vowels are linked to round shapes (e.g., Tarte and Barrett 1971). If shape and size-related object properties would be observed to influence some of these vocal characteristics, it would support the view that these objects properties are directly mapped to the articulatory processes which in turn might be a fundamental source of these sound symbolism phenomena.

The sound-shape correspondence phenomena

One famous demonstration of the relationship between the sounds and shape was originally reported by Köhler (1929). He showed that people have a tendency to associate sharp-edged shapes to the pseudoword '*takete*' and round-edged shapes to the pseudoword '*maluma*'. This phenomenon is generally known as the "Bouba-Kiki" effect. The fact that 2.5-year-old children (Maurer, Pathman and Mondloch 2006; Spector and Maurer 2011) and even 4-month-old infants (Ozturk, Krehm and Vouloumanos 2013) make the same sound–shape mappings with task-irrelevant auditory stimuli strongly suggests that the phenomenon operates very automatically, independently of processing graphical features of the written words.

The sounds that are associated with round shapes seem to be back vowels [ɑ] (Maurer et al. 2006), [o] and [u] (Tarte and Barritt 1971, Nielsen and Rendall 2013) as well as voiced consonants [m], [n] and [l] (Westbury 2005; Nielsen and Rendall, 2013; Spector and Maurer 2013). In contrast, those sounds associated with sharp shapes seem to be front vowel [i] (Tarte and Barrett 1971; Spector and Maurer 2013) as well as unvoiced consonants [t] and [p] (Westbury 2005; Nielsen and Rendall 2013). Given that round shapes are associated with unrounded back vowel [ɑ] as well as rounded back vowels [u] and [o] suggest that if the articulatory account of the effect is correct, the backness of the vowel might be at least as important factor in the effect as the roundness factor.

Regardless of the above-mentioned likely vocal outcomes of processing shape information of the stimulus, it is surprising that it has not been properly investigated whether these stimulus properties would really influence vocalizations. The only study that we are aware of that has looked empirically at the influence of an object shape on vocalization has been reported by Parise and Pavani (2011). They used triangles, hexagons and dodecagons as stimuli. Their participants were required to vocalize the vowel [ɑ] as long as the stimulus remained on the screen (2 s) if the object was either triangle or dodecagon (go-trials). It was found that the intensity of the vocalization was modulated by the shape so that is was higher for dodecagons than triangles. In addition, frequency of the third formant (F3) was higher in response to triangles than to dodecagons. Given that lip rounding-spreading during articulation modifies F3 (Lindblom and Sundberg 1971) so that decreased roundness increases F3 values, it might be proposed that increase in sharpness of the stimuli decreased roundness of the articulation in Parise and Pavani's (2011) observation.

The sound-size correspondence phenomena

As another famous example of a sound symbolism phenomenon, across several languages, there is a tendency to use words containing the vowel [i] for smaller objects (e.g., pin or needle) or adjectives that refer to smallness (e.g., diminutive or teeny) and words containing the vowels [ɑ] and [o] for larger objects (e.g., mountain or boat) or adjectives that refer to largeness (e.g., large or enormous) (e.g., Huang, Pratoomraj and Johnson 1969). Sapir (1929) originally demonstrated this

6

link between object size and vowel sound showing that the participants prefer to associate the pseudoword containing an open vowel (e.g., '*mal*') with large objects and the pseudoword containing a closed vowel (e.g., '*mil*') with small objects. Since the original observation the phenomenon has been reported for numerous times (e.g., Tarte and Barritt 1971; Thompson and Estes 2011). This sound-size association has been even observed in 4-month-old infants (Peña, Mehler and Nespor 2011) suggesting that these kinds of sound symbolic mechanisms are innate and may even have a role in language development as they, for example, may be assumed to improve acquiring new vocabulary.

It has been shown that vowels associated with small objects are close and front vowels such as [i] while those associated with large objects are relatively open and back ones such as [ɑ] and [o] (Sapir 1929; Birch and Erickson 1958; Johnson 1967; Tarte and Barritt 1971; Thompson and Estes 2011). Associations between size and different consonants have been much more inconsistent between experiments than those with vowels (Newman 1933; Taylor and Taylor 1962; Greenberg and Jenkins 1966), but the general view is that voiced consonants such as [m] are associated with larger objects than unvoiced ones such as [t] (Newman 1933; Thompson and Estes 2011).

Furthermore, only a couple of studies have investigated whether the size of a viewed object modulates vocalization. Firstly, Parise and Pavani (2011) –the study already mentioned above – did not find any influence of stimulus size on spectral components of the vocalization when their participants were presented with small, middle and large sized stimuli, and they had to vocalize the vowel [ɑ] if the object was either small or large (go-trials). Similarly, Gentilucci and colleagues found that when the participants were vocalizing the syllable [bɑ] and simultaneously observing a small or large object, spectral components were not influenced by the object size (Gentilucci, Campione, Dalla Volta and Bernardis, 2009). Instead, it was found that the size of the viewed object influenced the fundamental frequency (f_0) and intensity of the vocalization when there was a hand grasping these objects in the stimuli. Under these circumstances, both of these parameters increased when observing large objects as compared to small objects. Taken together, it appears that neither an object size nor an object shape – when they are presented alone in absence of a hand grasping the object – is very effective in modulating f_0, F1 or F2.

The present study

The present study investigates the interaction between perception and articulatory motor processes. Behavioural studies that have investigated interplay between perceptual and motor processes have typically used stimulus-response compatibility (SRC) tasks in which the response is executed manually. Perhaps the best known example of SRC task is the so called Simon task (Simon 1969) in which participants are asked to select the hand of response according to, for example, color of the target object that is presented on the left or right of the display. In the Simon effect, participants typically respond faster with the hand that is compatible with the left-right location of the target object, even though the location information is entirely irrelevant to the task. This effect has been often explained by "dual-route" models (e.g. De Jong, Liang and Lauber 1994) which assume that the motor response which spatially corresponds to the location of the target is automatically activated by the stimulus location via a fast direct route. In contrast, the task-relevant feature (e.g. color) and the intentionally selected response to it are processed via a slower indirect route. In incompatible trials, the two routes compete with response selection, resulting in performance costs. This model of the Simon effect is supported by evidence according to which, in the Simon tasks, increased response activation can be observed over hand motor areas that correspond to the stimulus location between 100-300 ms after the stimulus onset (e.g. Cespón, Galdo-Álvarez and Díaz 2013; Leuthold and Schröter 2006).

We have developed a novel SRC task in order to investigate whether systematic connections between perceptual and speech processes could be observed – i.e., whether shape and/or size properties of the viewed object could influence vocalizations. In comparison to the above mentioned traditional SRC tasks, in our task, the manual responses were replaced by vocal responses. The stimuli in all of the present experiments consisted of four objects: two different shapes (sharp-edged/round-edged) that were presented in two different sizes (small/large). One of these objects was presented to the participant in one experimental trial and s/he was required to vocalize - as fast as possible - according to the pre-defined property of the stimulus, one of the two vocal response choices. We were interested in whether the task-irrelevant property of the object

(e.g., shape) influences vocal responses in some systematic manner. As an example, in Experiment 2, the participant was asked to pronounce the syllable [te] if the object is small and the syllable [me] if it is large. In that experiment, we investigated how the shape of the object influences these vocal responses. We assumed that in this task, the object shape (e.g., sharp) would implicitly prime the corresponding articulatory representation (e.g., [te]) resulting, for example, in relatively rapid selection - and consequently also execution - of this corresponding articulatory response. In addition to measuring reactions times, we also investigated whether certain vocal characteristics of the response (intensity, f_0, F1 & F2) are influenced by this task-irrelevant object property.

Experiment 1

The first experiment investigated whether size and/or shape of the viewed object influences vocalizations. The participants were required to vocalize one of the two alternative syllables according to the size (block 1) or shape (block 2) of the target object. The stimuli consisted of one small and one large sharp-edged (i.e., '*kiki*') and round-edged (i.e., '*bouba*') object in both experimental blocks. In the first block, we were particularly interested in how the object shape (i.e., the task-irrelevant property of the stimuli) influences the vocal responses, and in the second block we were particularly interested in how the task-irrelevant object size influences the vocal responses. In this way we assured that the potential congruency effects would reflect implicitly operating priming between the task-irrelevant object property and the vocal response rather than explicit matching of the response type to the object property that is under the investigation.

The syllables that were used as response alternatives were [ti] and [mɑ]. The participants were given enough practice so that they were able to remember properly the response alternatives, and consequently rapidly select the required response immediately after the onset of the object. These syllables were selected because the previous research has associated the consonant [t] (Westbury 2005; Nielsen and Rendall 2013) and the vowel [i] (Maurer et al. 2006; Spector and Maurer 2011) with sharp-edged shapes whereas the consonant [m] (Westbury 2005;

Nielsen and Rendall 2013) and the vowel [ɑ] (Maurer et al. 2006) have been associated with round-edged shapes. Hence, we assumed that the syllable [ti] would be a likely congruence match to the sharp-edged stimuli and the syllable [mɑ] would be likely congruence match to the round-edged stimuli. In addition, the vowel [i] has been previously associated with small objects whereas the vowel [ɑ] has been associated with large objects (e.g., Sapir 1929; Peña et al. 2011). Moreover, the previous research has associated the consonant [m] with the grasp performance that is typically directed towards large (power grip compatible) objects whereas the consonant [t] has been associated with the grasp performance that is typically directed towards small (precision grip compatible) objects (Vainio, Schulman, Tiippana and Vainio 2013; Vainio, Tiainen, Tiippana and Vainio 2014). Hence, we assumed that the syllable [ti] would be likely congruence match to the small stimuli and the syllable [mɑ] would be likely congruence match to the large stimuli.

Taken together, we predicted that participants would articulate the syllable [ti] relatively rapidly if the object is sharp-edged (in block 1) or small (in block 2). In contrast, we predicted that the syllable [mɑ] would be articulated relatively rapidly if the object is round-edged (in block 1) or large (in block 2). In addition, we hypothesized that certain object properties could have systematic influences on certain components of voice characteristics. For instance, it is theoretically plausible to assume that F1 could increase as a function of the object size.

Methods

Participants

Twenty-three naïve volunteers participated in Experiment 1 (17–29 years of age; mean age = 24 years; 4 males). All participants were native speakers of Finnish and had normal or corrected- to-normal vision and were right-handed. We obtained written informed consent from all participants. The study was approved by the Ethical Review Board in Humanities and Social and Behavioural Sciences at the University of Helsinki.

Apparatus, stimuli and procedure

Each participant sat in a dimly lit room with his or her head 60 cm in front of a 19-in. LCD monitor (screen refresh rate: 75 Hz; screen resolution: 1280 x 1024). The head-mounted microphone was adjusted close to the participant mouth. The target stimuli consisted of four different centrally displayed images: small/sharp (vertically 4.1˚ x horizontally 4˚), small/round (3.7˚ x 4˚), large/sharp (8.6˚ x 8.6˚) and large/round (8.1˚ x 8.6˚). The stimuli were standard '*kiki'* and '*bouba'* objects whose sizes were manipulated in order to provide the small and large version of these stimuli. The stimuli were presented in light grey color. A thin black line emphasized the edges of the shapes (see Figure 1).

Each trial started with the presentation of a fixation cross (1˚ x 1˚) for 1000 ms. Then the cross was replaced by an empty white screen, displayed for 1000 ms. Next the target stimulus appeared on the screen for 1000 ms. The participants were required to perform as fast and accurate vocal response as possible to the target. Reaction times were measured from the onset of the target object to the onset of the vocalization. The target stimuli were presented in random order with equal probability. Then the target was replaced by an empty white screen for 1000 ms. Figure 1 illustrates this design. All stimuli were presented on white background.

The experiment was divided into two separate blocks. In one block, the participant was instructed to pronounce either [ti] or [mɑ] according to the target size (size-syllable block) and in other block s/he was instructed to pronounce these syllables according to the target shape (shape-syllable block). Half of the participants started the experiment with the shape-syllable block. Moreover, half of the participants were required to pronounce the syllable [ti] to small (in the size-syllable block) and sharp (in the shape-syllable block) objects, and the syllable [mɑ] to large (in the size-syllable block) and round (in the shape-syllable block) objects (Mapping 1). Half of the participants responded in a reversed mapping condition (Mapping 2). Each block began with practice trials. Each participant was given as much practice as it took to perform the task fluently. In addition, the participant was allowed to have a break twice within each block. In total, the experiment consisted of 320 trials [40 x 2 (response) x 2 (size/shape) x 2 (block)].

The vocal responses were recorded for 2000 ms starting from the onset of the target object. At the beginning of the experiment, the recording levels were calibrated individually using the voice calibration function of Presentation 16.1 software so that the recording levels would match with the natural intensity of the participant's voice. This calibration was carried out before providing written instructions to participants. In the calibration, the participants were required to pronounce the syllables [ti] and [mɑ] one after another approximately once every second. The calibration took around two minutes. Stimulus presentation and sound recording were done with Presentation 16.1 software.

---Figure 1 about here---

Results

The size-syllable block (i.e., how shape influences vocal responses)

Reaction times and Errors: Reaction times were measured from the onset of the target object to the onset of the vocalization. Errors (i.e., the participant uttered the wrong syllable) and RTs more or less than two standard deviations from each participant's condition means were excluded from the reaction time analysis. Of the trials, 1.0% were removed as errors and 5.1% were removed as outliers. The combined removal of errors and outliers left 93.9% of the raw data as correct responses. The condition means of these remaining data were computed for each participant and subjected to a repeated-measures analysis of variance (ANOVA) with the within-subjects variables of Shape (sharp or round) and Response ([ti] or [mɑ]), and the between-subjects variable of Size-syllable mapping (Mapping 1: [ti]-small/[mɑ]-large or Mapping 2: [ti]-large/[mɑ]-small). Post hoc comparisons were performed by means of t-tests applying a Bonferroni correction when appropriate. A partial-eta-squared statistic served as an effect size estimate.

The analysis of percentage of errors did not reveal any significant effects. The analysis of reaction times revealed a main effect of Response, $F(1,21) = 37,298$, $MSE = 35116,685$, $p < .001$, $\eta_p^2 = .640$. The syllable [ma] was pronounced faster ($M = 480$ ms) than the syllable [ti] ($M = 519$ ms).

More importantly, the analysis revealed a significant interaction between Shape and Response, $F(1,21) = 10.59$, $MSE = 2086.16$, $p = .004$, $\eta_p^2 = .335$. The syllable [ti] was pronounced faster when the shape was sharp ($M = 509$ ms) rather than round ($M = 516$ ms), albeit this effect was only marginally significant ($p = .071$) and the syllable [mɑ] was pronounced faster when the shape was round ($M = 471$ ms) rather than sharp ($M = 483$ ms) ($p = .002$). This interaction is presented in Figure 2. The main effect of Size-syllable mapping ($p = .811$) was not significant. The pairwise comparisons test revealed that the effect between Shape and Response was significant in Mapping 1 for the syllables [ti] ($p = .007$) and [mɑ] ($p = .005$). However, in Mapping 2, the effect was missing for both syllables ([ti]: $p = .936$; [mɑ]: $p = .068$).

Voice characteristics: Vocal data was analyzed using Praat v. 5.3.49. Onsets and offsets of the vocalizations were first located individually for each trial. The intensity value was calculated as a peak value of the voiced section. The spectral components (F1 and F2) as well as f_0 were calculated as median values of the middle 1/3 of the voiced section. The values more or less than two standard deviations from each participant's condition means of intensity (0.3%), f_0 (0.5%), F1 (0.6%) and F2 (0.2%) were excluded from the voice characteristic analysis. The analysis of intensity did not reveal any significant main effects or interactions. In addition, the analysis of f_0, F1 and F2 did not reveal any significant main effects or interaction besides the obvious main effects of response (i.e., f_0 and F2 were lower for the vowel [ɑ] than [i] and F1 was lower for the vowel [i] than [ɑ]) that reflect shape of the vocal tract during vocalization of different vowels and the intrinsic f_0 of the vowels (Whalen and Levitt 1995).

---Figure 2 about here---

The shape-syllable block (i.e., how size influences vocal responses)

Reaction times and Errors: Errors and RTs more or less than two standard deviations from each participant's condition means were excluded from the reaction time analysis. Of the trials, 1.0% were removed as errors and 4.7% were removed as outliers. The combined removal of errors and outliers

left 94.3% of the raw data as correct responses. The condition means of these remaining data were computed for each participant and subjected to a repeated-measures analysis of variance (ANOVA) with the within-subjects variables of Size (small or large) and Response ([ti] or [mɑ]), and the between-subjects variable of Shape-syllable mapping (Mapping 1: [ti]-sharp/[mɑ]-round or Mapping 2: [ti]-round/[mɑ]-sharp).

The analysis of percentage of errors did not reveal any significant effects. The only significant effect that was revealed by the analysis of reaction times was a main effects of Response, $F(1,21) = 25.15$, $MSE = 26575.46$, $p < .001$, $\eta_p^2 = .545$. The syllable [mɑ] was pronounced faster ($M = 511$ ms) than the syllable [ti] ($M = 545$ ms). The two-way interaction between Size and Response ($p = .260$) and the three-way interaction between Size, Response and Mapping ($p = .745$) were not significant.

Voice characteristics: The analysis of intensity, f_0, F1 and F2 was carried out in the same way as in the size-syllable block. The values more or less than two standard deviations from each participant's condition means of intensity (0.2%), f0 (0.2%), F1 (0.6%) and F2 (0.3%) were excluded from this analysis. The analysis of intensity did not reveal any significant main effects or interactions. In addition, the analysis of f_0, F1 and F2 did not reveal any significant main effects or interaction besides the obvious main effect of response (i.e., f_0 and F2 were lower for the vowel [ɑ] than [i] and F1 was lower for the vowel [i] than [ɑ]) that reflect shape of the vocal tract during vocalization of different vowels.

Discussion

In the Mapping 1 of Experiment 1, when responses were performed according to the size of the object (in size-syllable block; i.e., the shape was task-irrelevant property of the object), the syllable [mɑ] was pronounced significantly faster when the object was round-edged rather than sharp-edged. An opposite trend was observed in relation to the syllable [ti]. The syllable [ti] was pronounced faster

when the object was sharp-edged rather than round-edged. These effects are in line with the hypothesis that round shapes are associated with the voiced consonant [m] and the back vowel [ɑ] whereas sharp shapes are associated with the unvoiced consonant [t] and the front vowel [i]. The object shape was not observed to influence vocal response times in Mapping 2 when [ti] was produced for large objects and [mɑ] for small objects. In addition, the object size was not observed to produce any congruency effects in reaction times. Finally, the analysis of voice characteristics did not reveal any effects or interactions relevant to our hypotheses.

Experiment 2

The results of Experiment 1 showed that, in Mapping 1 (i.e., [ti] was produced for small objects and [mɑ] for large ones), the object shape influenced vocal responses so that they were performed relatively rapidly when the shape was congruent with the syllable. Experiment 2 investigated whether the vowel and/or consonant content of the vocalizations included in Experiment 1 contributed to the shape-vocalization effect observed in that experiment. That is, the experiment was divided into two blocks. In one block, the participants were required to pronounce the vowel [i] or [ɑ] according to the size of the object. In another block, they were required to pronounce the syllable [te] or [me] according to the size. The consonants were coupled with the vowel [e] because it is difficult to pronounce them alone. The vowel [e] was selected because it is relatively neutral in terms of vocal tract shape. Because the shape-syllable block in Experiment 1 did not reveal any consistent congruency effects in relation to the object size, Experiment 2 focused on investigating how task-irrelevant shape of the target object influences vocal responses. Therefore, the participants were asked to respond according to the size of the object in Experiment 2; that is, the participants were vocalizing the vowels ([i]-[ɑ]) or syllables ([te]-[me]) according to the size of the target object throughout the experiment.

Methods

Participants

Thirty-one naïve volunteers participated in Experiment 2 (20–47 years of age; mean age = 26.5 years; 7 males; 3 left-handers). All participants were native speakers of Finnish and had normal or corrected-to-normal vision. We obtained written informed consent from all participants. The study was approved by the Ethical Review Board in Humanities and Social and Behavioural Sciences at the University of Helsinki.

Apparatus, stimuli and procedure

The apparatus and environmental conditions were the same as those in Experiment 1. The stimuli were mostly similar to the stimuli used in Experiment 1 with an exception that the size difference between the small and large objects was smaller than in Experiment 1. The size difference between small and large objects was made smaller in Experiment 2 in order to make the task somewhat more difficult. It was assumed that more difficult discrimination task would demand increased processing of the stimuli – including the task-irrelevant shape properties of the stimuli – which in turn might boost up the potential congruency effect. The target stimuli consisted of four different centrally displayed images: small/sharp (vertically 5.8˚ x horizontally 5.7˚), small/round (5.7˚ x 5.2˚), large/sharp (8.8˚ x 8.6˚) and large/round (7.8˚ x 8.6˚). The experiment was divided into two separate blocks. In one block, the participants were required to pronounce the vowel [i] or [ɑ] to the size of the object. Half of the participants were asked to pronounce [i] if the object was small and [ɑ] if it was large (size-vowel mapping 1). The size-vowel mapping was reversed (size-vowel mapping 2) for half of the participants. In the other block, the participants were required to pronounce the syllable [te] or [me] to the size of the object. Half of the participants were asked to pronounce [te] if the object was small and [me] if it was large (size-syllable mapping 1). The size-syllable mapping was reversed (size-syllable mapping 2) for half of the participants. The order of these blocks was balanced between the participants. The participants were required to respond as fast as possible, however so that they

would simultaneously make as few errors as possible. Each block began with practice trials. Each participant was given as much practice as it took to perform the task fluently. In addition, the participant was allowed to have a break two times within each block. The procedure was the same to that used in Experiment 1 besides the above mentioned differences. In total, the experiment consisted of 320 trials [40 x 2 (shape) x 2 (response) x 2 (block)].

Results

Reaction times and Errors: The two blocks were analyzed separately. Errors and RTs more or less than two standard deviations from each participant's condition means were excluded from the reaction time analysis. In the vowel block, 0.9% of the data were removed as errors and 4.3% were removed as outliers. The combined removal of errors and outliers left 94.8% of the raw data as correct responses. In the consonant block, 0.9% of the data were removed as errors and 4.1% were removed as outliers. The combined removal of errors and outliers left 95% of the raw data as correct responses. The vowel and consonant blocks were analyzed separately. The condition means of the remaining data were computed for each participant and subjected to a repeated-measures analysis of variance (ANOVA) with the within-subjects variables of Shape (sharp or round) and Response ([i] or [ɑ] in the vowel block, and [te] or [me] in the consonant block 2), and the between-subjects variable of Size-response mapping (Mapping 1: [i]-small/[ɑ]-large in the vowel block or [te]-small/[me]-large in the consonant block; Mapping 2: [i]-large/[ɑ]-small in the vowel block or [te]-large/[me]-small in the consonant block).

The vowel block: The analysis of percentage of errors did not reveal any significant effects. The analysis of reaction times revealed a significant interaction between Shape and Response, $F(1,29)$ = 10.18, MSE = 1627.83, p = .003, η_p^2 = .260. The vowel [i] was produced faster when the shape was sharp (M = 492 ms) rather than round (M = 507 ms) (p < .001). The shape did not influence reaction times of [ɑ] vocalizations (sharp: 507 ms; round: 507 ms) (p = .992). The further analysis

revealed that in Mapping 1, [i] responses were produced faster when the shape was sharp (M = 486 ms) rather than round (M = 505 ms) (p < .001). Similarly, in Mapping 2, [i] responses were produced faster when the shape was sharp (M = 499 ms) rather than round (M = 507 ms) (p = .020). However, [ɑ] responses were not influenced by shape information in either mapping condition (Mapping 1: p = .574; Mapping 2: p = .572)

---Figure 3 about here---

The consonant block: The analysis of percentage of errors did not reveal any significant effects. The analysis of reaction times revealed a significant main effect of response, $F(1,29) = 59.42$, MSE = 48067.71, p < .001, $\eta_p^2 = .672$. Responses were faster when the syllable was [me] (M = 463 ms) rather than [te] (M = 503 ms). The interaction between Shape and Response was also significant, $F(1,29) = 8.81$, $MSE = 2889.03$, $p = .006$, $\eta_p^2 = .233$. The syllable [te] was produced faster when the shape was sharp (M = 496 ms) rather than round (M = 509 ms) (p = .004). In contrast, the syllable [me] was produced faster when the shape was round (M = 460 ms) rather than sharp (M = 466 ms). However, this interaction was not statistically significant (p = .181). The three-way interaction between shape, response and mapping was also significant, $F(1,29) = 22.04$, MSE = 7228.85, p < .001, $\eta_p^2 = .432$. Further analysis showed that the interaction effect between shape and response was evident only in Mapping 1. In that mapping, the participants produced significantly faster [te] responses when the shape was sharp (M = 503 ms) rather than round (M = 529 ms) (p < .001) and faster [me] responses when the shape was round (M = 464 ms) rather than sharp (M = 488 ms) (p < .001). The interaction effect was absent in mapping 2 ([te]: p = .972; [me]: p = .076).

---Figure 4 about here---

Voice characteristics: The analysis of intensity, f_0, F1 and F2 of both of the blocks was carried out in the same way as in Experiment 1. The analysis of intensity did not reveal any significant main effects or interactions. In addition, the analysis of f_0, F1 and F2 did not reveal any significant main effects or

18

interaction besides the obvious main effects of response (i.e., f_0 and F2 were lower for the vowel [ɑ] than [i] and F1 was lower for the vowel [i] than [ɑ]) that reflect shape of the vocal tract during vocalization of different vowels.

Discussion

Experiment 2 suggested that the shape-vocalization correspondence effect observed in Experiment 1 is based on interaction between the sharp shape and articulation of the vowel [i] and the consonant [t] and on interaction between the round shape and articulation of the consonant [m]. The vowel [ɑ] does not seem to contribute to the effect. In addition, similarly to Experiment 1, in Experiment 2, the effect between consonants and object shapes was only observed in Mapping 1 in which [te] was vocalized to small stimuli and [me] to large stimuli. This asymmetry between mappings and the lack of the effect in relation to the vowel [ɑ] are discussed in better more detail in the General Discussion.

Experiment 3

The results of Experiment 2 suggested that articulation of the consonant [t] and the vowel [i] is linked – even when pronounced separately – to a sharp-edged shape whereas articulation of the consonant [m] is linked to a round-edged shape (however, notice that the effect with consonants was only observed in Mapping 1). The results also suggest that the vowel [ɑ] does not contribute to the interaction effect observed in Experiment 1 between the syllable [mɑ] and the round-edged object. That is, the effect associated with that syllable appears to be exclusively linked to the interaction between articulation of the consonant [m] and round shapes. Given that the vowel [ɑ] is not a rounded vowel, this finding is in fact something that could have been expected. In order to validate this proposal, Experiment 3 investigates whether the shape-vocalization congruency effect can be observed between the round vowel [u] and a round shape. The vowel [u] was selected for this experiment because it has been robustly linked to round shapes in the Bouba-kiki effect (e.g., Maurer

et al. 2006). Hence, Experiment 3 replicates the first block of Experiment 2 with the vowels [i] and [u].

Methods

Participants

Twenty-six naïve volunteers participated in Experiment 3 (18–48 years of age; mean age = 24.1 years; 4 males; 1 left-hander). All participants were native speakers of Finnish and had normal or corrected-to-normal vision. We obtained written informed consent from all participants. The study was approved by the Ethical Review Board in Humanities and Social and Behavioural Sciences at the University of Helsinki.

Apparatus, stimuli and procedure

The apparatus, stimuli and environmental conditions were the same as those in Experiment 2. However, in contrast to Experiment 2, in this experiment the participants were required to pronounce the vowel [i] or [u] to the size of the object. Half of the participants were asked to pronounce [i] if the object was small and [u] if it was large (size-vowel mapping 1). The size-vowel mapping was reversed (size-vowel mapping 2) for half of the participants. The participants were required to respond as fast as possible, however so that they would simultaneously make as few errors as possible. Each participant was given as much practice as it took to perform the task fluently. In addition, the participant was allowed to have a break within the study. The procedure was the same to that used in Experiment 2 besides the above mentioned differences. In total, the experiment consisted of 160 trials [40 x 2 (shape) x 2 (response)].

Results

Reaction times and Errors: Errors and RTs more or less than two standard deviations from each participant's condition means were excluded from the reaction time analysis. Of the trials, 1.0% of the data were removed as errors and 4.0% were removed as outliers. The combined removal of errors and outliers left 95.0% of the raw data as correct responses. The condition means of the remaining data were computed for each participant and subjected to a repeated-measures analysis of variance (ANOVA) with the within-subjects variables of Shape (sharp or round) and Response ([i] or [u]), and the between-subjects variable of Size-vowel mapping (Mapping 1: [i]-small/[u]-large or Mapping 2: [i]-large/[u]-small).

The analysis of percentage of errors did not reveal any significant effects. The analysis of reaction times revealed a significant interaction between Shape and Response, $F(1,24) = 29.02$, $MSE = 8199.75$, $p < .001$, $\eta_p^2 = .547$. The vowel [i] was produced faster when the shape was sharp (M = 526 ms) rather than round (M = 547 ms) (p < .001). In contrast, the vowel [u] was produced faster when the shape was round (M = 533 ms) rather than sharp (M = 548 ms) (p = .012). The three-way interaction between Shape, Response and Size-syllable mapping was not significant (p = .506). The pairwise comparisons test revealed that the interaction effect between Shape and Response was significant in Mapping 1 for the vowels [i] (p < .001) and [u] (p = .031). However, in Mapping 2, the effect was only observed for the vowel [i] ([i]: p = .001; [u]: p = .130).

Voice characteristics: The analysis of intensity, f_0, F1 and F2 of both of the blocks was carried out in the same way as in Experiments 1 and 2. The analysis of intensity did not reveal any significant main effects or interactions. In addition, the analysis of f_0, F1 and F2 did not reveal any significant main effects or interaction besides the obvious main effect of response (i.e., f_0 and F2 were lower for the vowel [u] than [i] and F1 was lower for the vowel [i] than [u]) that reflect shape of the vocal tract during vocalization of different vowels.

---Figure 5 about here---

Discussion

The results of Experiment 3 again replicated the interaction effect between the vowel [i] and sharp-edged shapes. In addition, the interaction effect was also observed between the vowel [u] and round-edged shapes even though similarly to the previous experiments it again seemed that the Mapping 1 provided better conditions for the effect to be observed.

General Discussion

Although influential sound symbolism theories assume that shape and size properties of viewed objects resonate with articulatory representations resulting in well-known sound symbolism phenomena of sound-size and sound-shape correspondence effects (e.g., Ramachandran and Hubbard 2001; Oberman and Ramachandran 2008), very little work has examined whether these object properties indeed influence vocalizations. Hence, the current study investigated influence of shape and size on vocalizations. In Mapping 1 of Experiment 1 ([ti] was performed for small objects and [mɑ] for large ones), the [mɑ] vocalizations were relatively faster when the target object was round-edged and the [ti] vocalizations were relatively faster when the object was sharp-edged. The results of Experiments 2 and 3 showed that the effect observed in Experiment 1 can be replicated when vowels are produced alone without any consonants. In Mappings 1 and 2 of Experiments 2 and 3, the vowel [i] was associated with relatively rapid vocal responses when the object was sharp-edged. In contrast, [u] was associated with relatively rapid vocal responses, in Mapping 1 of Experiment 3 (i.e., [i] was performed for small objects and [u] for large ones), when the object was round-edged. Moreover, the effect was also replicated, in Mapping 1 of Experiment 2 (i.e., [te] was performed for small objects and [me] for large ones), when only the consonant content of the

22

vocalized speech units were changing. That is, sharp-edged objects were associated with relatively fast [te] vocalizations whereas round-edged objects were associated with relatively fast [me] vocalizations.

The shape-vocalization correspondence effect was observed most robustly in the mapping condition in which the participants were required to produce those vocal responses to small objects that were anticipated to be congruent with the sharp shape and those vocal responses to large objects that were anticipated to be congruent with the round shape (i.e., Mapping 1). In the reversed mapping condition (i.e., Mapping 2) the effect was mostly missing. The only shape-vocalization effect that was observed in Mapping 2 (Experiment 3 and the block 1 of Experiment 2) was linked to articulation of the vowel [i]. That is, articulation of the consonants [t] and [m] were facilitated by the shape of objects (sharp and round, respectively) only when [t] was produced for small objects and [m] for large objects. Similarly, articulation of the vowel [u] was facilitated by the sharp-shape of the object when it was produced to large objects. Unfortunately, the current data do not provide a straightforward explanation for why Mapping 1 offers better circumstances for the shape-vocalization correspondence effect to be observed than Mapping 2. One could speculate that the participants preferred to articulate [t] and [i] when the object was small and [m] and [u] when it was large (i.e., Mapping 1), and in the reversed mapping condition there was interference for not being able to respond according to this tendency. Consequently, this interference might have impaired any visual priming effects that the objects had on articulation. However, this explanation was not fully supported by the data because such an interference could have been assumed to lead to relatively slow overall responses in Mapping 2, which was not observed in the study. In general, the responses were not performed significantly faster in Mapping 1 than in Mapping 2. Hence, this and other potential explanations for the above mentioned mapping phenomenon require further investigation. Nevertheless, it needs to be emphasized that the mapping phenomenon does not invalidate the shape-vocalization effect observed mostly in Mapping 1 that the shape has a significant influence on reaction times of vocal responses.

The shape-vocalization correspondence effect was not observed with the vowel [ɑ] in Experiment 2 (the vowel block). This suggests that the correspondence effect observed in

Experiment 1 with the syllable [mɑ] was exclusively based on the consonant content of the syllable. This interpretation was supported by the fact that articulation of the consonant [m] was primed by round shaped objects in Experiment 2 (the consonant block). However, the lack of any interaction between articulation of the vowel [ɑ] and round shapes is not that surprising given that the vowel [ɑ] is not rounded. Moreover, the fact that the sharp objects were clearly associated with the vowel [i] but not with the vowel [ɑ] in Experiment 2, which are both in fact unrounded vowels, suggests that the shape-vocalization effect can be observed between sharp objects and unrounded-front vowels and between round objects and round-back vowels. That is, the effect can be only observed if the roundness as well as frontness criteria are both fulfilled. Indeed, given that articulation of the [u], which is close-back-rounded vowel, was primed by round objects in Experiment 3, it supports the view that sharp visual shapes are associated with vocalization of front-unrounded vowels (i.e., [i]) whereas round shapes are associated with vocalization of back-round vowels. Perhaps roundness is not a sufficient articulatory feature on which the shape-related sound symbolism phenomenon is grounded but rather the phenomenon is - to some extent - also based on frontness dimension of articulation in addition to roundness.

Object size did not produce any effects in reaction times or voice characteristics. This effect was absent in all experiments of the current study. This finding is in line with the previous observations, which also failed to show any interaction between object size and vocalization (Parise and Pavani 2011; Gentilucci et al. 2009). Nevertheless, our initial assumption on which the present study was based on was that size-related sound symbolism phenomena (Sapir 1929) might rely on interaction between the object size and openness of the articulation. The current study clearly does not support this hypothesis. In Experiments 1 and 2, open vocalizations ([ɑ]) were not produced faster than close vocalizations ([i]) when the objects were large. In addition, F1 values of voice characteristics that are assumed to reflect openness of the vocal tract during vocalization (Fant 1960) were not influenced by object size. Absence of these effects in addition to the previous failures to observe any interaction between object size and openness of vocalization (Parise and Pavani 2011; Gentilucci et al. 2009) suggest that size-related sound symbolism phenomena (Sapir 1929) might mostly operate within some other representative medium than sensorimotor processes related to

articulation. In contrast, the fact that an object's shape has been observed to automatically modulate spectral components of vocal responses (Parise and Pavani 2011) and that the present study showed an influence of an object's shape on vocal responses suggests that the articulatory account works better for the sound-shape congruency phenomenon.

We assume that the shape-vocalization effect was observed in the current study because visual shape of the object (e.g., sharp shape) implicitly primes the corresponding articulatory representation stored in the motor system (e.g., the articulatory representation for producing the vowel [i] or the consonant [t]) leading to relatively rapid vocal responses when the shape matches with the required vocal response. The fact that the effect was exclusively observed in vocal reaction times rather than shape-triggered modulations of voice characteristics suggests that the perceived shape causes relatively coarse biases in selection of the articulatory gesture prototype required for the response rather than fine shape-driven modulations of the selected gesture. Nevertheless, our view concerning the effects observed in the present study is mostly similar with the articulation account of the shape-related sound symbolism phenomena previously proposed by Ramachandran and Hubbard (2001) and Oberman and Ramachandran (2008). We similarly assume that an object's shape resonates with articulatory representations that are congruent with the shape resulting in these sound symbolism effects. That is, for example, round object is associated with the pseudoword *bouba* because articulation of that word requires relatively round shaping of a mouth.

The shape-vocalization correspondence effect is to some extent similar with the vision-speech effect observed by Kerzel and Bekkering (2000). They showed that vocal reaction times related to the articulation of the syllables [bɑ] and [dɑ] are influenced by viewing a mouth movement – in absence of speech sound – that is producing either [bɑ] or [dɑ] vocalization. Vocal responses are made faster when they are compatible rather than incompatible with the visual speech stimuli. This finding was proposed to support the motor theory of speech perception (Liberman et al. 1967) according to which the articulatory motor system is recruited for perceiving auditory and visual speech. That is, the viewed speech-related mouth shapes are directly mapped to corresponding articulatory gestures. In line with this view, more recent evidence shows that people have tendency

to simulate observed articulatory mouth movements in their own articulatory system (e.g., Watkins, Strafella and Paus 2003). Of course it has to be highlighted that if the present finding is indeed based on similar sensorimotor mapping processes between visual mouth shapes and articulation as the effect observed by Kerzel and Bekkering, the visuo-speech associations observed in the present study must operate under representations that employ coarse articulation-related abstractions drawn from different shapes. That is, for example, the 'kiki' stimulus does not really look like a mouth that is shaped to produce the syllable [ti] but rather might be assumed to present some sort of abstraction of that mouth shape. Indeed, recent evidence suggests that people have strong implicit tendency to extract visual features, potentially related to articulation, from relatively abstract stimuli, and automatically integrate these features with speech processes (Sweeny et al. 2012; D'Ausilio, Bartoli, Maffongelli and Fadiga 2014). Hence, the driving mechanism behind the shape-vocalization correspondence effect might be the capability to extract generalized sensorimotor primitives from relatively abstract stimulus properties and automatically map them to corresponding articulatory prototypes in the speech system.

In conclusion, our results demonstrate that the shape of '*kiki-bouba'* stimuli influences vowel and consonant vocalizations, providing a new behavioural method to investigate how different speech sounds could be associated with different object properties. In support of the articulatory account of the Bouba-kiki effect, the production of the vowel [i] and the consonant [t] was faster when viewing sharp-edged objects. In contrast, the production of the vowel [u] and the consonant [m] was faster when viewing round-edged objects. Although these effects were mostly observed in Mapping 1 condition, we propose that this shape-vocalization correspondence effect is likely to reflect biasing influence of object shape on selecting the articulatory gesture prototype required for the response rather than fine shape-driven modulations of the selected gesture. That is because the effect was observed in vocal reaction times rather than in voice spectral components. In addition to revealing several novel findings regarding interaction between perceptual and speech processes, this study provides evidence for supporting the articulatory account of shape-related sound symbolism phenomena.

Compliance with Ethical Standards

Conflict of Interest: The corresponding author (Lari Vainio) declares that he has no conflict of interest. In addition, all of the co-authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Alku, P., Vintturi, J., & Vilkman, E. (2002). Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation. *Speech Communication, 38*, 321-334.

Barsalou, L.W. (2008). Grounded action. *Annual Review of Psychology, 59*, 617-645.

Birch, D., & Erickson, M. (1958). Phonetic symbolism with respect to three dimensions from the semantic differential. *Journal of General Psychology, 58*, 291–297.

Binder, J.R., & Desai, R.H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*, 527-536.

Cespón, J., Galdo-Álvarez, S., & Díaz, F. (2013). Similarities and differences between interference from stimulus position and from direction of an arrow: behavioral and event-related potential measures. *International Journal of Psychophysiology*, 90, 180-189.

D'Ausilio, A., Bartoli, E., Maffongelli, L., Berry, J.J., & Fadiga, L. (2014). Vision of tongue movements bias auditory speech perception. *Neuropsychologia, 63*, 85–91.

De Jong, R., Liang, C.C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-responce correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 731-750.

Fant, C. (1960). *Acoustic Theory of Speech production*. Mouton's, Gravenhague.

Gentilucci, M., Campione, G.C., Dalla Volta, R., & Bernardis, P. (2009). The observation of manual grasp actions affects the control of speech: a combined behavioral and Transcranial Magnetic Stimulation study. *Neuropsychologia, 47*, 3190-202

Greenberg, J.H., & Jenkins, J.J. (1966). Studies in the psychological correlates of the sound system of American English III and IV. *Word, 22*, 207-242.

Huang, Y.-H., Pratoomraj, S., & Johnson, R. C. (1969). Universal magnitude symbolism. *Journal of Verbal Learning and Verbal Behavior, 8*, 155–156.

Johnson, R.C. (1967). Magnitude symbolism of English words. *Journal of Verbal Learning and Verbal Behavior, 6*, 508–511.

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 634-647.

Köhler, W. (1929). Gestalt psychology. New York: Liveright Publishing Corporation.

Liberman, M., Cooper, F.S., Shankweiler, D.S., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Leuthold, H., & Schröter, H. (2006). Electrophysiological evidence for response priming and conflict regulation in the auditory Simon task. *Brain Research, 1097*, 167-180.

Lindblom, B.E.F., & Sundberg, JE.F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America, 50*, 1166–1179.

Maurer, D., Pathman, T., & Mondloch, C.J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science, 9*, 316-22.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology, 45*, 53-75.

Newmeyer, F.J. (1993). Iconicity and generative grammar. *Language, 68*, 756-796.

Nielsen, A.K., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology, 67*, 153-63.

Oberman, L.M. & Ramachandran, V.S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience, 3*, 348-355.

Ozturk, O., Krehm, M., & Vouloumanos, A.J. (2013). Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. *Experimental Child Psychology, 114*, 173-86.

Parise, C.V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research, 214*, 373-80.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science, 22*, 1419-21.

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

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 239-255.

Saussure, F. (1916). *Cours de linguistique générale*, ed. C. Bally and A. Sechehaye, Lausanne and Paris.

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

Simon, J.R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81*, 174-176.

Spector, F., & Maurer, D. (2011). The colors of the alphabet: Naturally-biased associations between shape and color. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 484-495.

Spector, F., & Maurer, D. (2013). Early sound symbolism for vowel sounds. *i-Perception, 4,* 239–241.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition, 124*, 194-200.

Thompson, P.D., & Estes, Z. (2011).Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology, 64,* 2392-2404.

Tarte, R.D., & Barritt, L.S. (1971). Phonetic symbolism in adult native speakers of English: three studies. *Language and Speech, 14*, 158-168.

Taylor, I. K., & Taylor, M. M. (1962). Phonetic symbolism in four unrelated languages. Can. J. Psychol., 16, 344–356.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds.), Universals of human language, Vol. 2: Phonology. Stanford, CA: Stanford University Press. 527-568.

Vainio, L., Schulman, M., Tiippana, K., & Vainio, M. (2013). Effect of syllable articulation on precision and power grip performance. *PLoS One 8*(1):e53061.

Vainio, L., Tiainen, M., Tiippana, K., & Vainio, M. (2014). Shared processing of planning articulatory gestures and grasping. *Experimental Brain Research, 232*, 2359-68.

Watkins, K.E., Strafella, A.P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*, 989-94.

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain & Language, 93*, 10-9.

Whalen, D. H. & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics 17*, 193-203.

Figure captions

Figure 1. The illustration of the design (a) and the stimuli (b) used in Experiments 1, 2 and 3. The target stimuli consisted of round and sharp objects that were presented in small or large size. More detailed description of the stimuli is provided in Methods section.

Figure 2. Influence of object shape on vocalization (Experiment 1): The mean reaction times (RTs) as a function of the syllable and the object shape in Mappings 1 ([ti] was responded to small objects and [mɑ] for large objects) and 2 ([ti] was responded to large objects and [mɑ] for small objects).

Figure 3. Influence of object shape on vocalization (the vowel block of Experiment 2): The mean reaction times (RTs) as a function of the vowel and the object shape in Mappings 1 ([i] was responded to small objects and [ɑ] for large objects) and 2 ([i] was responded to large objects and [ɑ] for small objects).

Figure 4. Influence of object shape on vocalization (the consonant block of Experiment 2): The mean reaction times (RTs) as a function of the syllable and the object shape in Mappings 1 ([te] was responded to small objects and [me] for large objects) and 2 ([te] responded to large objects and [mɑ] for small objects).

Figure 5. Influence of object shape on vocalization (Experiment 3): The mean reaction times (RTs) as a function of the vowel and the object shape in Mappings 1 ([i] was responded to small objects and [u] for large objects) and 2 ([i] responded to large objects and [u] for small objects).
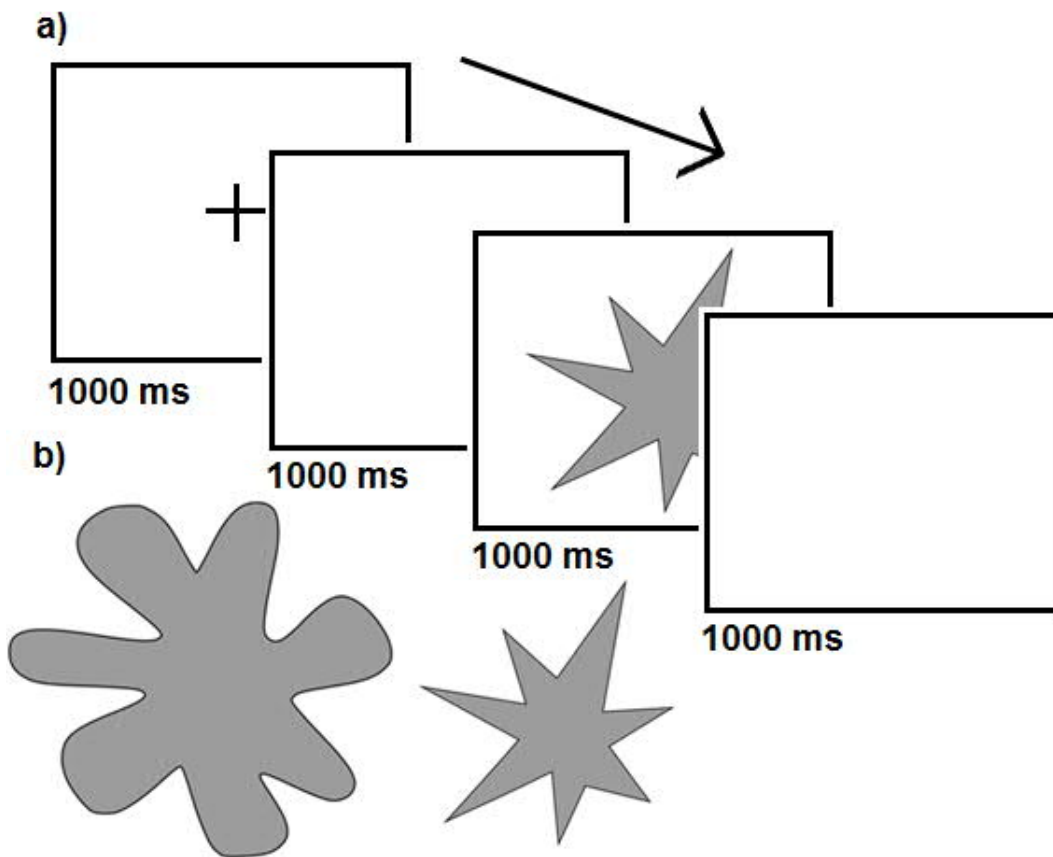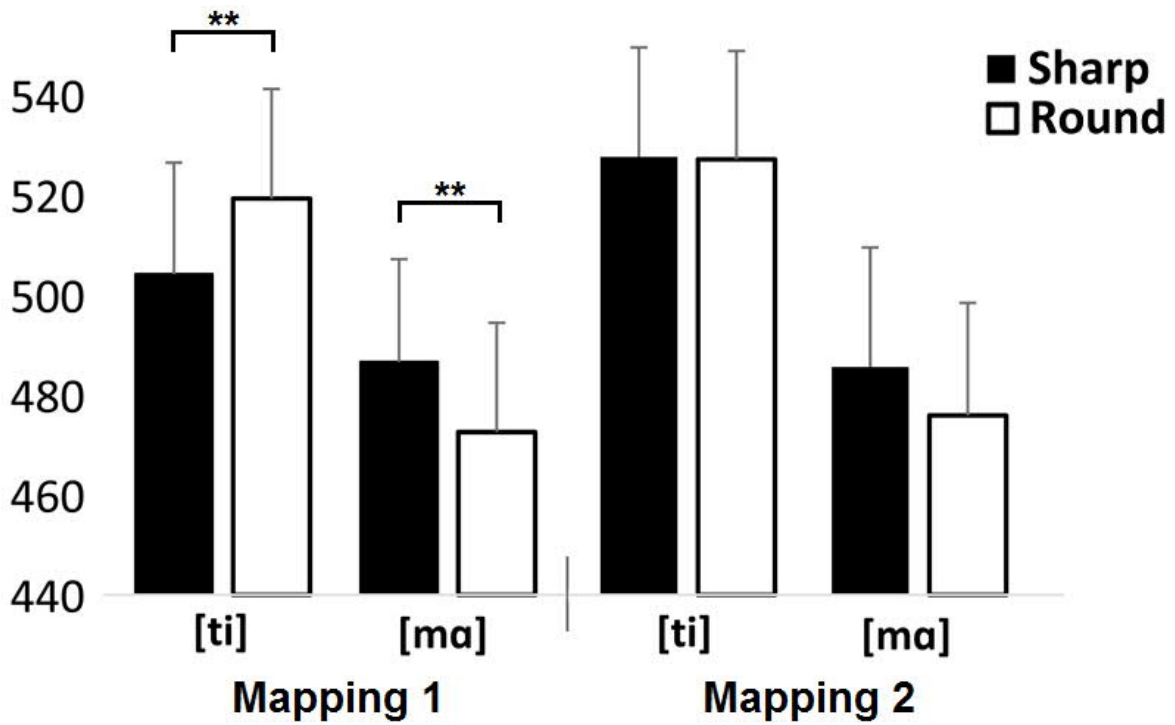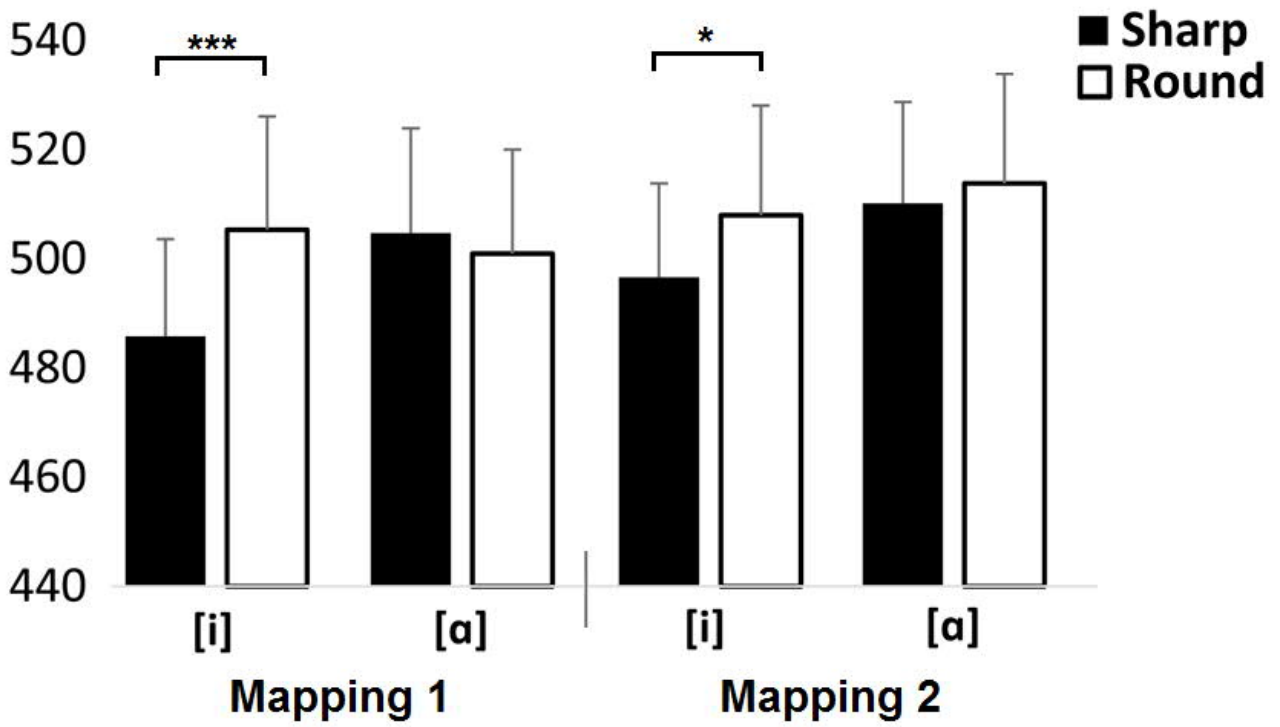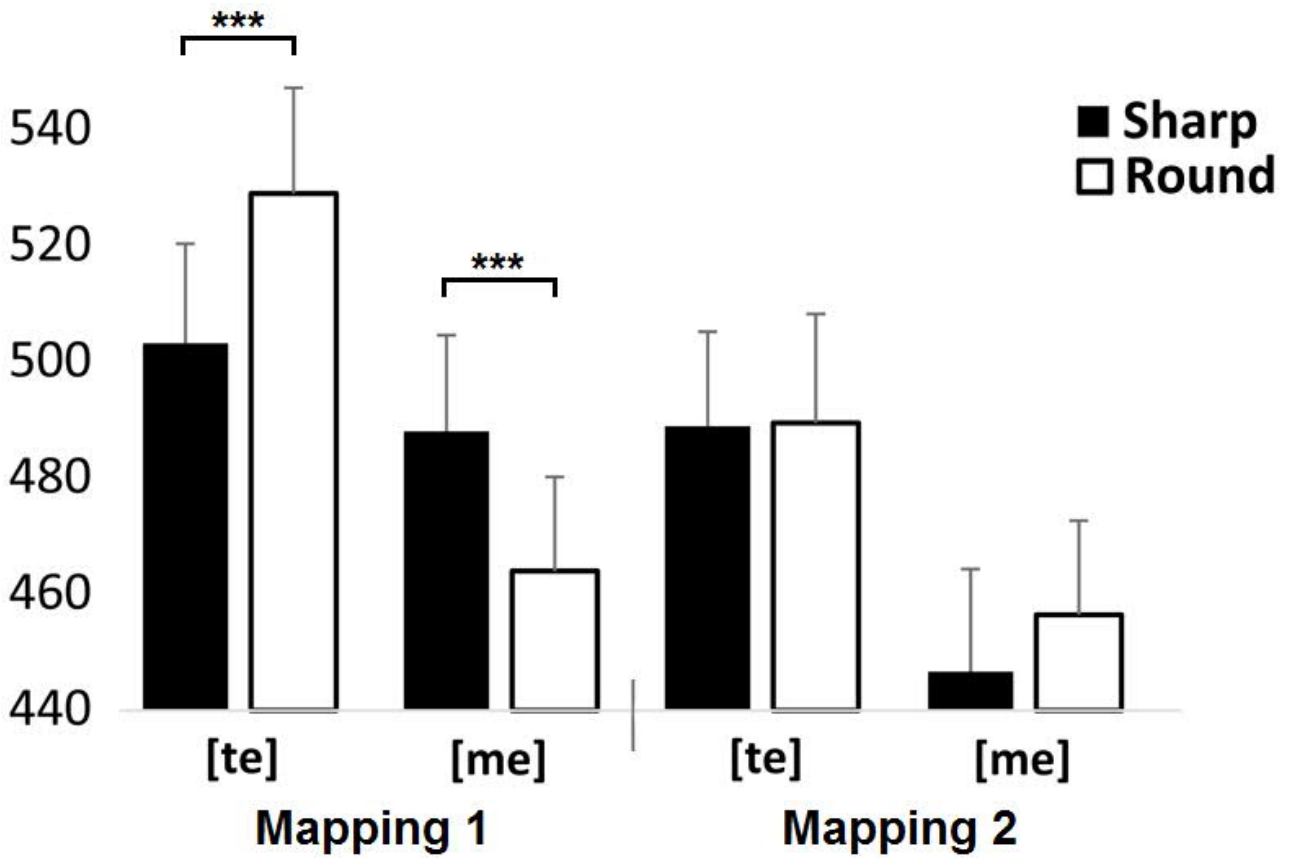
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5