Language and Speech

Article

Language *and* Speech
2014, Vol. 57(3) 367–393

DOI: 10.1177/0023830913507694
las.sagepub.com

Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the *Bouba-Kiki* Paradigm

Annette D'Onofrio
Department of Linguistics, Stanford University, USA

Abstract

Sound symbolism is the process by which speakers link phonetic features with meanings non-arbitrarily. For instance, speakers across languages associate non-words with rounded vowels, like *bouba*, with round shapes, and non-words without rounded vowels, like *kiki*, with spiky shapes. Researchers have posited that this link results from a cognitive association between sounds and visual or proprioceptive cues made in their production (e.g. sounds of rounded vowels cue the image of rounded lips, which is mapped to rounded shapes). However, non-words used in previous studies differ from one another along multiple phonetic dimensions, some showing no clear iconic mapping to shape. This study teases apart these features, finding that vowel backness, consonant voicing, and consonant place of articulation each elicit a sound symbolic effect, which is amplified when these dimensions are combined. This investigation also probes object properties that can be involved in sound symbolic association, bringing the "*bouba-kiki*" paradigm, typically involving the use of abstract shapes, into the realm of real-world objects. To shed light on ways that sound symbolism may operate in natural language, this study suggests that future research in this paradigm would benefit from consideration of both more detailed phonetic correlates and more refined object properties.

Keywords

Sound symbolism, phonetic symbolism, cross-modality, embodiment, dimensionality, bouba-kiki

1 Introduction

The arbitrariness of links between sound and meaning has been taken as a basic assumption of linguistic theory (e.g. de Saussure, 1966). However, various studies have established the robust existence of sound symbolism, the phenomenon in which speakers link phonetic features with meanings in a non-arbitrary fashion. Some aspects of sound symbolism have been found cross-linguistically. Within the domain of "synesthetic" sound symbolism, in which a sound comes to cross-modally represent visual, tactile, or proprioceptive properties of objects like shape or size

Corresponding author:
Annette D'Onofrio, Department of Linguistics, Stanford University, 450 Serra Mall, Building 460, Room 127, Stanford, CA 94305-2150, USA.
Email: annetted@stanford.edu

(Hinton, Nicols & Ohala, 1994), a noted finding is the so-called *bouba-kiki* effect. In this paradigm, speakers across a number of languages associate non-words like *maluma* or *bouba* with round shapes, and non-words like *takete* or *kiki* with spiky shapes. Experiments that have produced these results (e.g. Köhler, 1929; Nielsen & Rendall, 2011; Ramachandran & Hubbard, 2001) have uncovered a striking pattern supporting the possibility that this sound symbolic association is cross-linguistic.

The *bouba-kiki* effect was first documented by Köhler (1929), who found that English speakers consistently matched rounded and spiky shapes with the nonsense words *maluma* and *takete*, respectively. Since this initial finding, the same pattern has been replicated with English speakers (Holland & Wertheimer, 1964; Ramachandran & Hubbard, 2001) and with speakers of other languages like Swahili and Kitwongwe (Davis, 1961), Czech (Tarte, 1974) and Swedish (Ahlner & Zlatev, 2010). Only one cross-linguistic exception to this pattern has been documented: Songe speakers in Papau New Guinea were not found to show any consistent preference for matching in either direction (Rogers & Ross, 1975). Most recently, this paradigm has been used to explore cognitive processes underlying sound symbolic phenomena. Ramachandran and Hubbard (2001) have described the cognitive mechanism responsible for these associations to involve some form of cross-modal activation in the brain. In a replication of Köhler's initial study, Ramachandran and Hubbard suggested that cognitive representations of the visual appearance or proprioceptive sensations embodied by the mouth and tongue in sound production are cross-activated by sensory representations of acoustic phonetic percepts. Others have used the *bouba-kiki* paradigm to conjecture that some sound symbolic associations may be pre-linguistic and somehow encoded as "natural" biases. Maurer, Pathman and Mondloch (2006) found that both 2.5 year-old children and adults consistently matched words with rounded vowels to rounded shapes, suggesting that these associations are present early enough in development that sound symbolic correspondences themselves may influence language learning. Other studies (Imai et al., 2011; Ozturk, Krehm & Vouloumanos, 2011) found sensitivity to sound symbolism even earlier in language acquisition, in eleven-month-old and four-month-old infants. These researchers have proposed that the presence of sound-shape correspondences early in development provides evidence that these correspondences are innate, and thus may influence both how language is learned and how language evolves over time.

The cross-linguistic and early developmental nature of sound symbolic phenomena like the *bouba-kiki* effect has led to proposals of a biological basis for some sound symbolic associations. Researchers (e.g. Blakemore & Frith, 2005; Maurer et al., 2006; Ramachandran & Hubbard, 2001) have proposed that the association of mouth shape in the production of a sound and the visual qualities to which that sound is linked can be taken as a biological or anatomical basis for a sound symbolic association. For example, Maurer et al. (2006) described the strong *bouba-kiki* effect in their results as a product of cross-modal activation cued by the sound of a rounded vowel in a nonsense word. This sound is posited to trigger visual or proprioceptive sensation of rounded lip shape that is then mapped onto visual roundedness of an external object. It has been suggested that this process could occur through cognitive mechanisms similar to those that underlie synesthesia (Ramachadran & Hubbard, 2001), the phenomenon in which an individual experiences sensation in a particular modality when a different modality is stimulated (seeing a particular color when hearing a musical note, for example).

Although earlier studies have drawn upon the *bouba-kiki* paradigm to probe cognitive mechanisms and potential explanations for sound symbolism in language, many studies have left details of the paradigm empirically unrefined. Both earlier and more recent experiments have used non-words that differ from one another in a number of phonetic features, any or all of which may contribute to the sound-shape correspondence, without considering the possible effects of each of these features in turn. For example, Köhler's *maluma* and *takete* differed from one another in vowel roundedness and vowel backness, but also in continuant nature of consonants, sonority of consonants, voicing of consonants, and place of articulation of consonants. Ramachandran and Hubbard's (2001) *bouba* and *kiki* similarly differed not only in vowel quality but also in consonant voicing and place. While Maurer et al. (2006) attribute their results largely to the mapping between vowel roundedness and round shapes, all of the nonsense word-pairs used in their experiment consistently differed from one another in a number of other phonetic exponents. Each of their "round" nonsense words (*bamu, bouba, goga, mabuma*) contained consonants that were voiced and either bilabial or velar. Each of their "spiky" nonsense words, (*kuhtay, kaykee, teetay, tuhkeetee*), however, contrasted from the round words not only in their lack of rounded vowels, but also in their consistently voiceless consonants, the presence of alveolar consonants, and the absence of bilabial consonants. While biologically-based theories of sound symbolic phenomena may account for sound-meaning associations between shape and certain types of phonetic features, namely vowel roundedness, it is more difficult to apply such theories to other phonetic features, such as consonant place of articulation or voicing. What iconic link can be held responsible for a symbolic link between, say, a voiced consonant and a rounded shape? Effects of individual phonetic features that have been used previously to contrast auditory stimuli must be investigated for their own sound symbolic potential. If features like consonant voicing are found to be at least partially responsible for the *bouba-kiki* effect, this could point to a cross-modal explanation outside of visual analogy as mediated by visual properties of the mouth in articulation. Instead, such a finding would indicate a conventionalized association between sound and shape that is unlikely to be biologically based.

More recent work in the paradigm has begun to tease apart previously conflated phonetic features, particularly in consonants. Some studies (Ahlner & Zlatev, 2010; Aveyard, 2012; Nielsen & Rendall, 2011; Westbury, 2005) have pointed out that stimuli of prior experiments conflated effects of vowels and consonants. These studies have separated vowel differences from consonant differences to find that features of consonants do, in fact, contribute to sound-shape mappings in the *bouba-kiki* paradigm. Some findings have indicated that effects of consonant differences may even be stronger than effects of vowel differences (Ahlner & Zlatev, 2010; Nielsen & Rendall, 2011). Other studies have also isolated more specific features of consonants, finding that sound symbolic mappings can be effected via consonant voicing, such that voiced consonants are matched with rounded shapes, and voiceless consonants are matched with spiky shapes (Cuskley, Kirby, & Simner, 2010). Consonant manner of articulation also appears to be sound symbolic, such that continuant consonants and sonorant consonants are matched with round shapes, whereas stops and strident consonants are matched with spiky shapes (Aveyard, 2012; Nielson & Rendall, 2011, 2012; Westbury, 2005). Others have delved deeper into particular features of vowels that may contribute to the *bouba-kiki* effect, indicating that even features that may not be phonemically contrastive, such as amplitude, can map onto shape. Parise and Pavani (2011) found that when speakers were asked to produce a token of a single vowel in response to an image of a shape, their productions were consistently louder and had a higher F3 (indicating less lip rounding) when they were presented with a triangle as opposed to a dodecagon. This suggests that shapes with more acute or "sharp" angles map onto vocalizations that are louder and less rounded.

These studies have substantially refined empirical investigations of the *bouba-kiki* phenomenon, demonstrating that aspects of both vowels and consonants can be responsible for the well-documented shape-sound correspondences. However, while the sound symbolic potential of some aspects of consonants, such as manner and sonority, have held particular attention in recent studies, others have remained unaddressed. The distinctions between continuant versus non-continuant (Westbury, 2005) and strident versus sonorant consonant (Nielsen & Rendall, 2011) refer to categories that are quite broad with respect to phonemic inventories in natural languages. Other features, such as consonant place of articulation place of articulation, are also used across languages for phonemic contrast, but this feature has as yet remained unexplored with respect to the *bouba-kiki* phenomenon. This feature may prove particularly interesting in such an investigation, given the ways in which places of articulation of a consonant, like roundedness of a vowel, differentially map to visual cues in production. For example, like lip rounding in producing a rounded vowel, bilabial consonants utilize a gesture with the lips that can be visually apparent to an interlocutor. Furthermore, considering these categories in broad strokes lumps together groups of sounds such as nasal stops, fricatives, and more vowel-like approximants into categories like "continuant," in spite of the fact that the proprioceptive properties of such sounds that may be responsible for a sound symbolic effect can be quite different. Finally, it has been shown that vowels and consonants may work together to strengthen sound symbolic effects: in one study, when a word contained both vowels and consonants that were independently matched with a rounded shape, an individual was more likely to match the word to a rounded shape than if the vowels and consonants operated in incongruent directions (Ahlner & Zlatev, 2010). This finding, coupled with the consistent conflation of particular features of consonants or vowels that repeatedly produce a robust *bouba-kiki* effect, indicates that it may in fact be a combination of a number of phonetic features that gives the *bouba-kiki* paradigm such consistent and seemingly natural force. Considering a single phonetic feature of consonants or vowels does not attend to the ways that particular phonetic features may be used in concert with one another to reinforce a sound symbolic meaning, while considering a single vowel or consonant at the segmental level eliminates phonetic detail that may itself be responsible for the effect.

A different aspect of the *bouba-kiki* paradigm that has in some ways limited its application to natural language and the real visual world is the fact that visual stimuli used in previous studies have been constructed of unfamiliar shapes that nearly always appear as line drawings on a two-dimensional plane. In this domain, the object property of "roundness" or "spikiness" with which a certain word is associated is reduced to a simplified feature of a two-dimensional abstract shape. While these stimuli are useful in isolating a particular contrast in visual qualities, these shapes cannot be easily extended to implicate real-world objects that could be deemed round (like a bowl or plate) or spiky (like a fork or knife), which comprise multiple and more complex visual properties and exist in three-dimensional space. An investigation of the ways in which phonetic features are linked to objects more akin to those that exist in the real world would broaden understandings of the ways that sound symbolism may operate in natural language.

The investigation presented here involves two tasks, each of which refines a different aspect of previous *bouba-kiki* studies. In the first, the abstract shape task, a number of individual phonetic features—voicing of consonants, place of articulation of consonants, and backness/roundness of vowels—are tested for their possible roles in visual-auditory sound symbolic associations found in the *bouba-kiki* paradigm. In order to elucidate which specific cues hearers use when associating nonce words with visual properties, this task examines which individual phonetic features may be responsible for the *bouba-kiki* effect and how these features might work in conjunction with one another to produce or strengthen a link between a word that contains these phonetic features and an object property. The second task, the real-world object task, works in a different vein to explore applications of the *bouba-kiki* paradigm to pictures of objects that a speaker interacts with in the real world. The use of real objects adds facets of meaning to the properties of roundness or spikiness that are otherwise flattened in the use of two-dimensional abstract shapes. Aiming to refine what, exactly, constitutes the visual object properties in the *bouba-kiki* effect, the real-world object task probes whether sound symbolic effects found in *bouba-kiki* tasks still arise when hearers are asked to match nonsense words not with abstract shapes, but with objects that are found in the real world. This task explores ways that more complicated conceptions of this property—roundness in terms of dimensionality, for instance—can interact with sound symbolic association. Taken together, these tasks work to refine both the "sound" and the "shape" aspects of the robust sound-shape correspondence found in previous *bouba-kiki* studies.

2 Abstract shape task

The abstract shape task examines the effects of three phonetic features—backness of vowels, voicing of consonants, and place of articulation of consonants—on the selection of a shape. The experiment closely follows the design of previous *bouba-kiki* type experiments, but it is designed to isolate each phonetic feature's independent effects on shape selection, then allowing for combinations of these features to be analyzed through their interactions. Given the phonetic features found in nonce words of previous experiments, it is hypothesized that words with front/unrounded vowels will be matched with a spiky shape, whereas words with back/rounded vowels will be matched with a rounded shape—the nonsense words *maluma* and *bouba* contain only back vowels, whereas *kiki* contains only front, and unrounded vowels, for example. Furthermore, *maluma* and *bouba* contain only voiced labial consonants, so it is expected that voiced consonants and labial consonants map to the selection of rounded shapes, whereas voiceless consonants and velar consonants (as in *kiki*) will be matched with spiky shapes. Finally, it may be the case that the robustness of the *bouba-kiki* effect comes not solely from a single phonetic correlate like vowel roundedness, but from a grouping of correlates that compositionally amplifies the difference between words that predict selection of a round shape and those that predict selection of a spiky shape.

2.1.1 Methods.

2.1.1.1 Participants. Two hundred participants were recruited and participated online in the abstract shape task via Amazon's Mechanical Turk (www.MTurk.com), a "crowdsourcing" web service. Mechanical Turk provides an interface through which large numbers of workers can perform tasks via the Internet for payment.[1] A number of issues particular to web experiments can arise and must be addressed: avoiding duplicate entries from the same participant, ensuring that participants are listening to sound clips and attending to the task, and controlling for language background. To address these concerns in both tasks presented in this investigation, a number of checks were put into place via Mechanical Turk and the web interface used to conduct the experiment. Workers were instructed to complete the task only once, and they were informed that duplicate entries would be rejected without payment. Each worker is assigned a unique worker identification number via the Mechanical Turk website, and in the case that the task was completed twice or more by the same worker ID, all results from that worker ID were removed prior to analysis. To constrain the participant population to a particular set, the Mechanical Turk interface allows for a task to be restricted only to workers in a given location, which workers report when they create a Mechanical Turk account. To limit the language backgrounds of participants to some extent, all of the tasks presented in this study were restricted to recruit only participants from the United States. Participants were also given a survey following the task, asking for their native language. Participants who did not self-report their native language as English were removed from the dataset. Finally, to ensure that participants were listening to the sound clips presented in the task, each participant was presented with a sound clip of an English word prior to the task, and they were required to type the word into a text box. If the correct word was not entered, workers were not allowed to continue on to the main task. In the survey following the task, participants were also asked which type of listening device they used to perform the task (headphones, earbuds, or speakers). In analysis, type of listening device used had no effect on the results.

Participants were compensated $0.18 for the task, which took an average of 3 min to complete. Those who completed the task in less than 10 s were removed prior to analysis, since it was

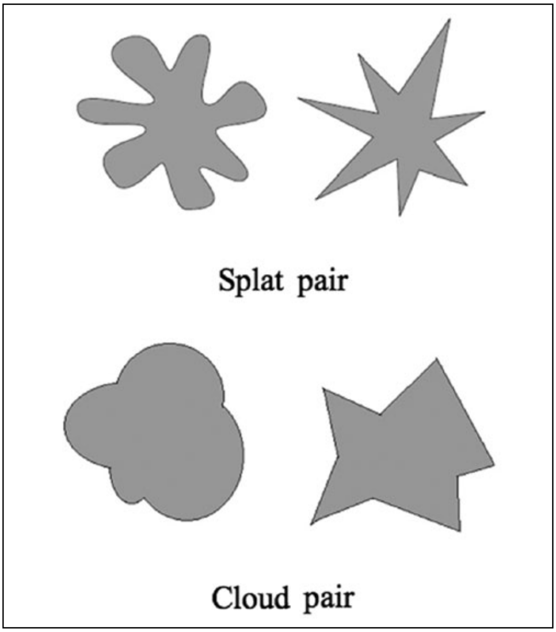

Figure 1. Critical object pairs in the abstract shape task.

determined that properly attending to and completing the task would require at least 10 s for valid answers. In all, data from 170 of the 200 recruited participants were included in the analysis.

2.1.1.2 Visual stimuli. The experiment contained two rounded-spiky shape pairs (or "critical pairs"). These pairs are shown in Figure 1, with what will be referred to as the "splat" pair in the first row, and what will be referred to as the "cloud" pair in the second row. In both pairs, the rounded object is shown on the left, and the spiky object is shown on the right.

The splat pair has been used in previous *bouba-kiki* style studies (Cuskley, Kirby & Simner, 2010). The cloud pair was designed for this study with the intention of including a fuller, more bulbous shape pair that maintains a round versus spiky distinction, while perhaps implementing a different conception of the object property of roundness than the splat pair does. That is, the quality of roundness with respect to the round splat can be interpreted as containing rounded edges (or no sharp angles), whereas the quality of roundness with respect to the round cloud may describe the shape more generally, as approximating a circle. All participants saw both of these pairs. The two objects in each pair were always shown together, but the placement of which object was on the left and which was on the right in each pair was randomized. In order to make the round-versus-spiky pattern less explicit, these two pairs were interspersed with four filler pairs and preceded by one filler "practice" pair. In all, participants saw six pairs of objects after the practice pair, two of which were the critical pairs shown in Figure 1. The practice object pair was accompanied by a filler nonsense word.

2.1.1.3 Auditory stimuli. Each of the six object pairs in the main task was accompanied by one of the auditory stimuli shown in the pool in Table 1.

Each auditory stimulus was disyllabic and contained two matching consonants (voiceless or voiced labial, alveolar or velar stops) and either two different front, unrounded vowels or

Table 1. Pool of martian words in the abstract shape task.

Place	Labial		Alveolar		Velar	
Voicing	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced
Front vowels	/pipe/ /pepi/	/bibe/ /bæbi/	/tite/ /teti/	/dide/ /dedi/	/kike/ /keki/	/gige/ /gegi/
Back vowels	/pupɑ/ /pɑpu/	/bubɑ/ /bɑbu/	/tutɑ/ /tɑtu/	/dudɑ/ /dɑdu/	/kukɑ/ /kɑku/	/gugɑ/ /gɑgu/

two different back vowels, including one rounded vowel /u/. Note that, as a reflex of the English phonemic inventory, in which the only rounded vowels are back vowels, the features of vowel backness and vowel roundedness are necessarily conflated in the abstract shape task: all of the words that contained front vowels contained unrounded vowels only, and every word that contained non-front vowels contained the rounded vowel /u/. While these will be referred to as "front" versus "back" vowels throughout the study for simplicity, this feature is necessarily conflated with vowel roundedness as well. Both possible orderings of vowels were used in the stimuli, as shown in Table 1. The back vowels used were the American English back vowels /u/ and /ɑ/, as these were the vowels used in previous experiments' stimuli. The front vowels used were American English /i/ and /e/, except in the case where the use of these front vowels together would have formed an American English lexical item /bebi/—baby. A different front vowel, /æ/ was used in the place of /e/ in this case to avoid potential confounds that would be introduced by the presence of a recognizable English lexical item in the pool of unfamiliar "martian" words. In the six main trials, participants heard one randomly selected auditory stimulus from each of the six consonant groups (i.e. one /p/ word, one /b/ word, and so on). The word that each participant heard from each consonant group and the order in which these auditory stimuli appeared was counter-balanced. The auditory stimuli were recorded by a male American English speaker in his early thirties in a soundproof booth using a Turner model 2302 microphone (Turner Microphones, Cedar Rapids, IA) with a Rane MS1b preamplifier (Rane Corporation, Mukileto, WA). Recordings were digitized (44.1 kHz, 24 bits) with an Edirol UA-101 USB-Audio interface (Roland Corporation U.S., Los Angeles, CA), recorded into the software program Audacity (Mazzoni & Dannenberg, 2011). Each word was scaled for duration and peak amplitude in Praat acoustical analysis software (Boersma & Weenink, 2011). Since these nonsense words were to be presented to native American English speakers, non-native sounds could have introduced potential confounds in the task. In order to avoid such confounds, the stimuli were pronounced using American English phonotactics—stress fell on the initial syllable of each word, and syllable-initial voiceless consonants were aspirated. It should therefore be noted that references to "voiced" and "voiceless" consonants map to the voiced and voiceless variants of consonants in American English. Thus, throughout this analysis, voicelessness cannot be teased apart from aspiration. This combination of features will be referred to as voiceless for convenience, but the sound symbolic effect and the contrast between these segments and voiced segments could be attributed to either aspiration or the lack of voicing, or both.

2.1.2 Procedure. Participants were directed via Mechanical Turk to a separate web interface through which they performed the task. Participants were told that the task involved martian shapes and martian words. They were told that in each trial they would hear a martian word, and were instructed to click on the martian object that they thought the word referred to. They were given one practice trial followed by six real trials. In each trial, participants were presented with a button labeled "play sound," which they would click on to hear the martian word. Below the button, in the center of the screen, they were presented with two shapes, side-by-side. Once the play button was clicked and

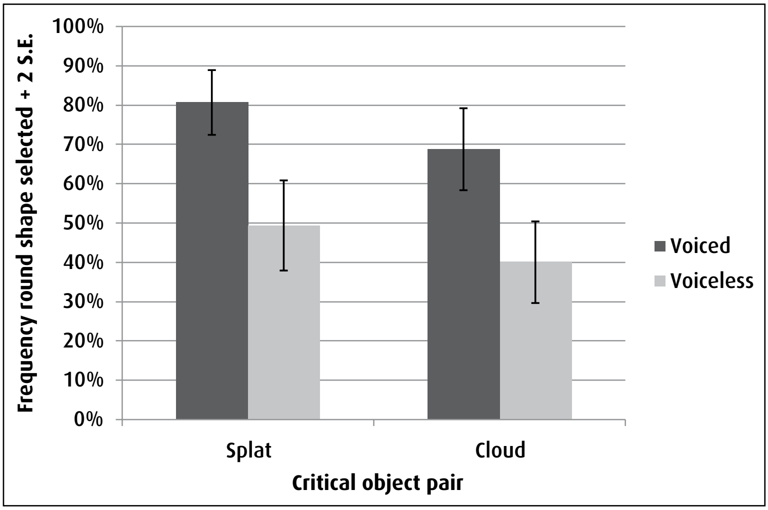

Figure 2. Frequency of round shape selected, by critical object pair and consonant voicing.

the word finished playing, the shape images became clickable. Participants clicked on one of the two shapes in order to advance to the next pair, the use of two shapes giving participants a 50% chance of selecting the expected sound-symbolic association in any given trial. The task was preceded by the sound test and followed by the survey asking for language background and type of listening device. To receive payment, each participant was provided a passcode following their completion of the task, which they entered into the Mechanical Turk interface. Only upon entering a valid passcode did each participant receive payment.

2.1.3 Results. Considered individually, each of the three predicted phonetic features—vowel backness/roundedness, consonant place of articulation, and consonant voicing—had a significant effect on the selection of a rounded or spiky shape in at least one of the two critical object pairs. To evaluate differences in shape selection between variants of the phonetic features, Pearson's chi-square tests were conducted comparing the counts of round shape selected to counts of spiky shape selected between two conditions of a given independent variable (e.g. "voiced" versus "voiceless") (see Appendix 1 for all counts on which these tests were conducted). For differences in shape selection between a phonetic variant and chance, chi-square tests were conducted comparing the counts of round shape selected and counts of spiky shape selected to the expected counts if the results had been at chance (always 50% of the total count). The data show a significant difference by consonant voicing on shape selection for responses in both the splat and the cloud pair (Figure 2).

For each pair, participants who heard a word with voiced consonants in it were more likely to select the round shape in the pair than participants who heard a word with voiceless consonants in it (splat pair c^2 $(1, n = 170) = 18.51, p < 0.001$; cloud pair c^2 $(1, n = 170) = 14.07, p < 0.001$). Since this task involved a choice between two objects, participants had a 50% chance of selecting a given sound-shape correspondence. While words with voiceless consonants did not yield shape responses that were significantly different from chance in either critical object pair, words with voiced consonants did in both pairs, in the expected direction (splat pair c^2 $(1, n = 93) = 34.94, p < 0.001$; cloud pair c^2 $(1, n = 80) = 11.25, p < 0.001$). Not only does the difference in consonant voicing significantly predict selection of a shape type, but also, words with voiced consonants in particular produce likelihoods above chance that a round shape will be selected.

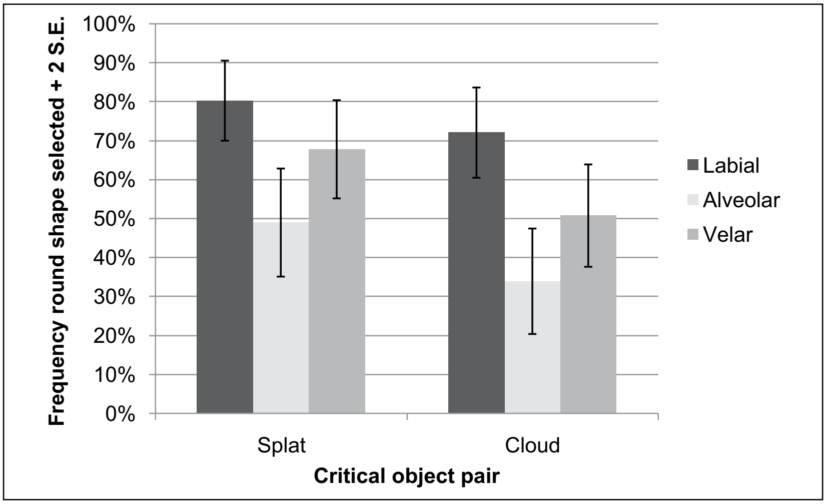

Figure 3. Frequency of round shape selected, by critical pair and consonant place.

When the data for each pair are divided only by consonant place of articulation, they similarly show significant differences for consonant place in each pair (Figure 3).

For both pairs, consonant place of articulation had a significant effect on shape selection (splat pair c^2 (2, n = 170) = 12.52, p = 0.0019; cloud pair c^2 (2, n = 170) = 16.32, $p < 0.001$). The standardized residuals of the chi-square test show that, for the splat pair, the major contributor to the significance of the chi-square was the matching of an alveolar consonant with a spiky shape (z = 2.19, p = 0.029), indicating that the difference in shape selection between alveolars, on the one hand, and velars and labials, on the other, are most responsible for the effect. Alveolars were also a major contributor to the significance of the chi-square for the cloud pair (z = 2.03, p = 0.042), as was the matching of labials to a rounded shape (z = 1.99, p = 0.047), indicating that participants who heard a word with labial consonants in it were significantly more likely to select a round shape over a spiky shape than those who heard a word with alveolar consonants in it. In both pairs, the matching of a round shape with a word with labial consonants was significantly above chance (splat pair c^2 (1, n = 61) = 22.44, $p < 0.001$; cloud pair c^2 (1, n = 61) = 11.95, $p < 0.001$). In the splat pair, the likelihood of a velar being matched with a round shape was also above chance (c^2 (1, n = 56) = 7.14, p = 0.0075), whereas the likelihood of an alveolar being matched with a shape type was not different from chance. In the cloud pair, however, the likelihood of an alveolar being matched with a spiky shape was significantly above chance (c^2 (1, n = 50) = 5.12, p = 0.024), whereas the likelihood of a velar being matched with a shape type was not significantly different from chance.

Finally, the data also show a significant effect of vowel backness, but only for the responses that matched a word with the cloud pair (Figure 4).

When participants matched a word with an object in the cloud pair, those who heard a word with back vowels in it were significantly more likely to select the rounded cloud shape than those who heard a word with front vowels in it (c^2 (1, n = 170) = 18.2828, $p < 0.001$). For the splat pair, the difference between back vowels and front vowels is not significant. In both pairs, words containing back vowels were matched with a round shape type at rates significantly different from chance (splat pair: c^2 (1, n = 96) = 16.67, $p < 0.001$; cloud pair: c^2 (1, n = 88) = 13.14, $p < 0.001$). When participants heard words with front vowels in them, the selection of the round versus the spiky

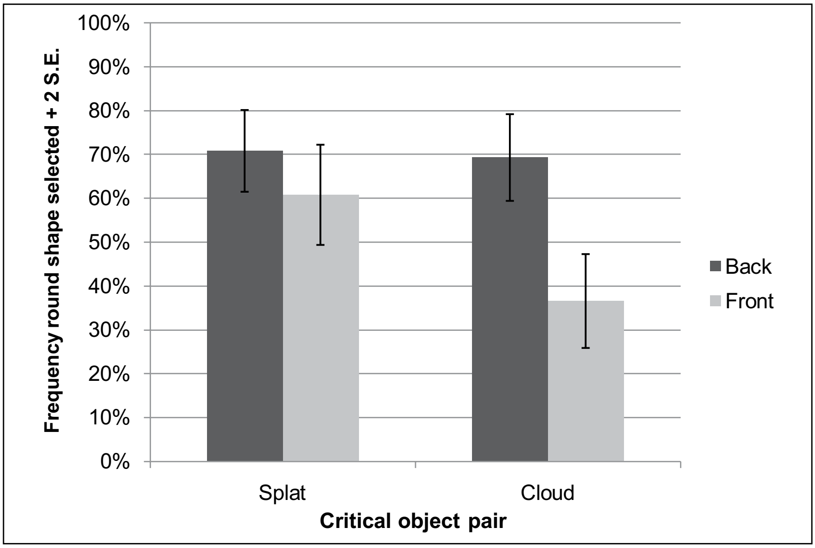

Figure 4. Frequency of round shape selected, by critical pair and vowel backness.

shape was significantly different from chance for the cloud pair (c^2 (1, n = 82) = 5.90, p = 0.015) such that a word with front vowels was more likely than chance to be matched with a spiky cloud, whereas for the splat pair, when participants heard words with front vowels in them, their shape selection was close to chance.

In order to evaluate the ways that these simple factors may interact in the data, a logistic regression model was fitted on the data using the lrm function in the Design package of R statistical software (R Core Team, 2013) on the response set for each of the critical object pairs. The logistic regression model used here tested the effects of backness, voice, place, and all possible interactions among the three as predictors of shape type selected (round versus spiky). Default factor levels in these models were set at backness = back, voicing = voiced, and place = labial. Since round shape was used as the baseline, positive coefficients represent an increase in the likelihood that a spiky shape is selected, and negative coefficients represent a decrease in this likelihood. Two separate models evaluated the effects of these predictors on shape type selected in each of the two critical pairs. Summaries of the predictors for each of the final models are provided in Appendix 1.

In the case of the splat pair, interactions were not found to be significant, so only simple effects were included in the final model. Simple effects of place of articulation (β = 1.53, SE = 0.46, z = 3.33, $p < 0.001$) and voicing (β = 1.42, SE = 0.36, z = 3.89, $p < 0.001$) were the only significant predictors of shape selection for this target pair, as discussed above.

In the logistic regression model fitted on the cloud pair, the simple fixed effect of backness was significant (β = 1.57, SE = 0.39, z = 4.02, $p < 0.001$), as indicated by previous chi-square analysis. However, rather than consonant voicing and consonant place serving as significant predictors of shape type individually, the interaction between the two proved to be the significant predictor of shape selection (voiceless × alveolar: β = 2.77, SE = 0.99, z = 2.79, p = 0.0053; voiceless × velar: β = 2.09, SE = 0.86, z = 2.42, p = 0.015).

The interaction captures the fact that voicing does not influence shape selection in the same manner across all places of articulation, but in fact voicing has a different effect on shape selection depending on whether the consonant is labial, alveolar, or velar. The positive values of these

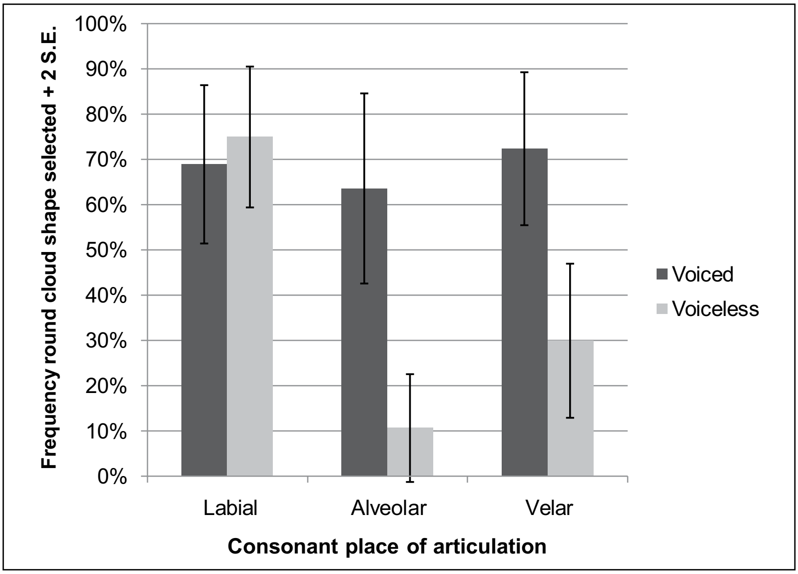

Figure 5. Frequency of round cloud shape selected, by consonant place and voicing.

coefficients indicate that the specified levels of the interaction (voiceless × alveolar and voiceless × velar) increase the likelihood of the selection of a spiky shape, thereby decreasing the likelihood that a round shape will be selected.

When the voiced-voiceless contrast is stratified by place of articulation (Figure 5), it becomes clear that the effect of voicing is quite strong within stimuli that contain alveolar consonants and those that contain velar consonants, but voicing does not have an effect within labial consonants. Furthermore, while voiceless consonants do not independently predict shape selection that was significantly different from chance in the cloud pair, consonants that are both alveolar and voiceless as well as consonants that are both velar and voiceless predict the selection of a spiky shape significantly more often than chance (voiceless alveolars in cloud pair: c^2 (1, n = 28) = 17.29, $p <$ 0.001; voiceless velars in cloud pair: c^2 (1, n = 28) = 4.8, p = 0.028). Thus, at least in the cloud pair, it is not simply the case that all voiceless consonants are matched with spiky shapes, but alveolar and velar voiceless consonants in particular are significantly likely to be paired with spiky shapes, as compared to voiced consonants in either of these places of articulation, and as compared to labial consonants of either voicing level.

2.1.4 Discussion

2.1.4.1 Which sounds are linked with shape selection? This experiment shows that if each of three phonetic features—voicing of consonants, place of articulation of consonants, and backness/roundness of vowels—are considered individually for their influence on shape selection, each in isolation has an effect on the selection of a round versus a spiky shape in at least one of the two critical object pairs. Overall, the circumstances in which results were significantly different from chance involved words containing voiced consonants, labial consonants, and back vowels. All of these features were matched with a rounded shape with a likelihood significantly greater than chance. The other phonetic variants—voiceless consonants, velar and alveolar consonants, and front vowels—were matched with round shapes significantly less frequently than their respective voiced, labial and non-front counterparts were, yet none of these features produced results that predicted the selection of a spiky shape to be significantly different from chance. This indicates that phonetic features that are matched with *rounded shapes* in particular (voiced consonants, labial consonants, and back and/or rounded vowels) may be more responsible for the significant differences between phonetic variants in the *bouba-kiki* effect, a finding that has been suggested previously (Kovic, Plunkett, & Westermann, 2010).

A difference in backness and/or roundness of vowel is a significant predictor of shape selection for the cloud pair, which supports the association of rounded vowels with round shapes, as asserted by previous studies (e.g. Maurer et al., 2006). The distinction between responses to the splat pair versus those to the cloud pair with respect to this feature illuminates a more precise interpretation of roundness that may be associated with vowel backness, roundness, or both. The quality of object roundness as it relates to the two pairs of objects seems to differ conceptually. The round splat is rounded compared to the spiky splat because the round splat contains rounded edges rather than sharp angles. By contrast, the roundness of the round cloud is more general—the shape itself is round in the sense that it is circle-like.

These results with respect to backness and/or roundedness of vowels may bear on whether or not this association is evoked through a mapping from an iconic visual or proprioceptive cue made in the production of a sound (rounded lips) to qualities of objects (rounded shapes). Approximating a circle, the round cloud serves as a better visual match than the round splat to the image of rounded lips, which form a circular shape when producing a rounded vowel. The possibility that an iconic visual or proprioceptive cue of rounded lips is the basis for the *bouba-kiki* effect (e.g. Ramachandran & Hubbard, 2001) may in fact be supported by the findings shown in Figure 2. If we expect that the feature of vowel roundness itself links to visual object properties iconically, an object that more closely matches the circular shape of rounded lips would map to words with rounded vowels significantly more often than to words without rounded vowels, which is what occurs in these data. Future studies might investigate more ways that shapes can be "rounded" in order to further probe the possibility that shapes closely matching a visual image of rounded lips (for example, circular shapes with a hole in the middle) are likely to evoke a sound symbolic association with rounded vowels.

Further support for the possibility of an iconic mapping between visual object properties and sounds as mediated by gestures made in articulation comes from the correlation of labial consonants with rounded shapes. Regardless of voicing, a labial consonant was consistently matched with a rounded shape. Further, unlike the results found for vowel roundedness/backness in this study, the effect of labial place occurs not only across voicing levels, but also across target shape pairs. When producing labial consonants, a gesture is made with the lips that can be perceptible to an interlocutor. Thus, it could be postulated that a lip gesture comes to be associated with curvature in general. More cross-linguistic evidence would be required, however, to conclude that this pattern can be interpreted as biological or innate. In fact, the demarcation between a "natural" or biologically iconic sound-meaning association and a conventionalized one is tenuous, and claims that such an association is innate should be approached with caution. Irvine and Gal's (2000) theory of iconicity points out just this difficulty. Irvine and Gal describe iconization as the process by which ideological representation creates a link between a linguistic feature and a social meaning in a way that attributes necessity or naturalness to such a link, even if that link is in fact conventional. That is, individuals' ideologies may cause them to perceive a certain linguistic feature to iconically reflect a certain social group's essence or biology, even if this association has been created only through cultural or historical factors. Thus, what may seem natural or "essential" has in fact been socially constructed.

Table 2. Percentage round and spiky shapes selected, by voicing of consonants, place of articulation of consonants, and frotness of vowels.

Voicing	Frontness of vowel	Place of stop	Word	Round (%)	Spiky (%)	Total *n*
Voice	Non-front	Velar	*guga/gagu*	91	9	33
Voice	Non-front	Labial	*buba/babu*	82	18	33
Voiceless	Non-front	Labial	*pupa/papu*	80	20	30
Voice	Front	Labial	*bibe/bæbi*	76	24	34
Voice	Front	Velar	*gige/gegi*	68	32	25
Voice	Non-front	Alveolar	*duda/dadu*	67	33	30
Voiceless	Front	Labial	*pipe/pepi*	64	36	25
Voice	Front	Alveolar	*dide/dedi*	55	45	18
Voiceless	Non-front	Velar	*kuka/kaku*	55	45	31
Voiceless	Non-front	Alveolar	*tuta/tatu*	41	59	27
Voiceless	Front	Velar	*kike/keki*	15	85	26
Voiceless	Front	Alveolar	*tite/teti*	7	93	28
			Average %	60	40	340

A different interpretation of the difference between the splat pair and the cloud pair could consider the perception of "weight" as pertains to negative space (i.e. the space around an image's subject). The objects in the splat pair contain more inward angles and thus more negative space than the objects in the cloud pair. This could lead the pairs to be interpreted differently with respect to appearance of weight—the objects in the cloud pair appear heavier than those in the splat pair. This difference cues into a sound symbolic effect termed the "frequency code," a pattern in which low pitch and backness of vowels is associated with large size, heaviness, and darkness, whereas high pitch and frontness of vowels is associated with small size, lightness, and brightness (e.g. Berlin, 1994; Jespersen, 1933; Ohala, 1994; Sapir, 1929). Specifically, speakers associate front vowels with sharpness, while they associate back vowels with softness and heaviness. It is possible, then, that the cue of vowel backness is used to form sound symbolic associations with respect to perceptually heavy or more voluminous objects like those in the cloud pair. The round cloud can more easily be described to have "soft" and "heavy" roundedness in contrast with the round splat, which simply has rounded edges.

2.1.4.2 Compositionality of sound symbolism. Results of the abstract shape task extend beyond simple differences found between variants of individual phonetic features. They indicate that in order to fully understand how nonsense words are mapped to object qualities, it is necessary and fruitful to examine the ways that multiple phonetic features work in concert with one another to predict the selection of a round versus a spiky shape. While voicing alone appears to be a simple predictor of shape type selection, a closer investigation of the interactions between this and other features demonstrates that, at least for the cloud pair, the effect of voicing occurs only within the context of alveolar and velar consonants, and has no effect when it comes to labial consonants.

The idea that phonetic features work together in sound symbolic association is bolstered when the relative strengths of possible feature combinations are considered. Table 2 shows the frequency with which participants selected a round shape based on combinations of the three target phonetic features in the auditory stimulus, in order of the frequency with which the round shape was selected over the spiky shape. Note that in this table, the results for the splat pair and those for the cloud pair are collapsed to provide larger *n*'s by which these frequencies are evaluated.

Words that were matched most frequently with round shapes contain combinations of multiple features that have been shown to independently predict the selection of a rounded shape. Conversely, words that participants matched with round shapes least frequently, and thus spiky shapes most frequently, contain combinations of features that, on their own, predict the selection of a shape that is not significantly different from chance. Although these features may not independently predict the selection of one or the other shape type, when they are combined, they produce a significant association with spiky shape type (voiceless velar and alveolar consonants and front vowels). Further, between these two ends of the spectrum from very frequently matched to a round shape to very infrequently matched to a round shape (and thus frequently matched to a spiky shape), a gradience appears across frequencies of round shape selection by word. That is, different combinations of features show variable selection rates between round and spiky shape. For example, when voiceless labials, which predict the selection of a round shape (Figure 5), are combined in a word with front vowels, which predict the selection of a spiky shape (Figure 4), the frequency with which a round shape was selected (67%) was close to the overall average of round shape selection (60%).

In breaking down these stimuli by feature, it becomes clear that the compositional combination of phonetic features strengthens the effect of a sound symbolic mapping between a word and either a round or spiky object. Considering where this experiment's instantiations of previously used stimuli—/bubɑ/ and /kike/—fall in the breakdown of shape selection frequencies by word, /bubɑ/'s *combination* of features places it at the very-round end of the spectrum, whereas /kike/'s combination of features places it at the very-spiky end of the spectrum. It thus appears that the compositional effect of multiple phonetic features that work in tandem may be responsible for the robustness and strength of the effect found in *bouba-kiki* studies, which have focused on these stimuli in particular. These results demonstrate the value of considering sound-symbolic associations both with respect to individual phonetic features and, crucially, with respect to combinations of features.

3 Real-world object task

The real-world object task brings findings for the abstract shape task in this study and previous *bouba-kiki* type studies into the realm of real-world objects. While the findings of *bouba-kiki* studies reveal interesting patterns that support the presence of sound symbolism in abstract naming, previous studies have largely used unfamiliar shapes that exist only on a two-dimensional plane. This necessarily creates distance between the findings of such studies and objects that exist in the real world, leaving a barrier in understanding between sound symbolic potential on an abstract level and the ways in which sound symbolism may operate in natural language toward naming or learning of labels for real-world objects. Work on sound symbolism in the lexicon has provided evidence for the presence of sound symbolism in natural language itself, as in the naming of plants and animals in Huambisa (Berlin, 1994) and mimetics in Japanese (Hamano, 1998), for example. These studies indicate that sound symbolic patterns can indeed be found in the naming of real-world objects and processes. The real-world object task presented here draws upon this evidence that sound symbolism can and does operate in natural language, using phonetically detailed findings from the abstract shape task to further probe what real-world properties might be embodied on the "shape" end of the sound-shape link. Specifically, the real-world object task is designed to explore how the relationships between features of nonsense words and abstract shapes might be extended to relationships between nonsense words and a class of three-dimensional, real-world objects. Further, this task investigates how considering sound symbolism with respect to objects in the real world might help further identify types of more concrete object properties that can be involved in a sound-symbolic association.

3.1 Participants

Participants were recruited for and participated in the real-world object task online via Amazon's Mechanical Turk: 68 participants for a pilot study to obtain potential items, and 136 participants for the main task. All participants were from the United States and were native English speakers. The assessment of these details and a sound check were performed in the same manner as in the abstract shape task. Participants in both the pilot study and main task were paid $0.25 via Mechanical Turk. The pilot task took an average of 1.5 min to complete, and the main task took an average of 3 min to complete.

3.2 Materials

3.2.1 Visual stimuli. Ten items were used as visual stimuli for the real-world object task, consisting of equally sized images (100 × 100 pixels) of kitchen items. This category was selected to represent a single class of real-world objects that contains a number of members that could be deemed round and a number that could be deemed spiky. In order to obtain relevant kitchen objects that could be evoked by the auditory stimuli used in the main task, a pilot study was conducted in which participants were provided with four auditory stimuli (martian words), each accompanied by an open text box. Participants were instructed to listen to all of the martian words first, then to listen to each word again, and type in the English word for any kitchen item that they thought each martian word might refer to. No pictures were displayed. This resulted in a total of 194 responses from 68 total participants. Due to the open-ended nature of the task in this pilot, participants named many different kitchen items overall, some that could be evaluated as having a salient round or spiky shape quality, and some that would be difficult to place in either category. All kitchen items that were named at least twice in the dataset of the pilot were rated by a panel survey via Mechanical Turk to ascertain which of the kitchen objects could be deemed notably round or spiky. Again, no pictures were shown to the raters. Independent roundness and spikiness ratings were collected for each object name. 20 raters of roundness, and 20 raters of spikiness rated each item as "not at all round/spiky" (scored as 0), "somewhat round/spiky" (scored as 1), or "very round/spiky" (scored as 2). These ratings yielded a number of kitchen objects deemed round or spiky that served to guide the selection of pictures to be used in the main task. Notably, in the open-ended responses provided in this pilot, many more responses were rated as round (n = 12) than as spiky (n = 3). Thus, for the main task, the selection of photos was aimed to prevent this imbalance, the choice of photos including five objects that were deemed round in the open-ended pilot survey (bowl, pan, plate, spoon, ladle) and the three objects that were deemed spiky in the open-ended pilot survey (fork, knife, grater). Two potentially spiky objects (carving fork, tongs) were added. All photos used were presented in the main task in monochrome and were sized to take up as much of the 100 × 100 pixel space as possible, as shown in Figure 6 (spiky objects shown on the top row, round objects on the bottom row).

Since these photos themselves served as stimuli, rather than open-ended lexical items as in the pilot task, ratings of object roundness, spikiness, and size were also obtained for these images. Panels of different raters for roundness (n = 20) and spikiness (n = 20) of the pictures participated via Mechanical Turk. Each rater was compensated $0.25, and the ratings took an average of 2 min to complete. Again, participants were asked to rate each of the 10 pictures as "not at all round/spiky", "somewhat round/spiky", or "very round/spiky," coded as 0, 1, and 2, respectively. The means of these ratings were analyzed to create groupings of objects into roundness and spikiness categories, using a series of Wilcoxon rank-sum tests to compare mean ratings of each of the pictures. The tests ascertained whether or not the ratings of roundness or spikiness of two given

Figure 6. Visual stimuli used in main real-world object task.

objects were significantly different from one another. When two objects were not rated significantly differently, they were placed in the same roundness or spikiness group; when they were, they were placed in separate groups. This analysis yielded four different roundness groups, labeled "not at all round," "round," "more round," and "most round," and three different spikiness groups, labeled "not at all spiky," "spiky," and "most spiky" (details of objects belonging to each group and mean ratings are provided in Appendix 2). Labels for each group were added to represent the gradation from "not at all round/spiky" to "very round/spiky" with respect to the average ratings of the objects. These roundness and spikiness groupings were used in the analysis of the main task. No overlap exists between round and spiky ratings, such that each object's ratings placed it in either the "not at all spiky" group or the "not at all round" group, but never both—an object was rated as being either round, to some extent, *or* spiky, to some extent.

Additionally, size ratings were obtained for all of the visual stimuli used. The fact that the pictures in this task represent objects that are differently sized in the real world brings up a possible sound symbolic confound, particularly given this task's focus on vowel quality. The aforementioned frequency code becomes relevant, as front vowels have been found to symbolize small size, whereas back vowels have been found to symbolize large size (e.g. Ohala, 1994). Since backness of vowels is the major parameter analyzed with respect to the shape link in this task, it is necessary to test whether or not the results could actually be better represented by a sound symbolic link between vowel backness and *size* rather than *shape*. To this end, size ratings were obtained for the stimuli. Size raters participated via Mechanical Turk and were compensated $0.25, the size rating task taking an average of 2 min to complete. Size raters were provided with pictures of all of the objects, presented in random order, and asked to group them into size categories. Five categories, from "smallest" to "largest" were provided, but participants were told that they were not required to use all of the categories, only as many as they deemed necessary to classify all of the objects. They were required to classify every object into a size class. Results were assigned an integer for each size class (0–4) and each object/size category match was normalized to control for the fact that each rater could potentially use a different range of the five sizes or different ends of the size scale provided. Normalization was performed by subtracting the rater's mean size category assignment from each of his or her size category ratings, then dividing by the rater's standard deviation. Analysis of size groupings of the objects were then conducted on normalized size ratings in the same manner as for the round and spiky groupings, using Wilcoxon rank-sum tests on each object's mean size rating to test which objects grouped together and which objects significantly differed in size rating. This yielded five distinct size groupings that were used in analysis of the data to

Table 3. Pool of martian words for the real-world object task.

Condition	Front vowels		Back vowels	
	/i/	/e/	/u/	/ɑ/
1	/pimə/	/pemə/	/pumə/	/pɑmə/
2	/bimə/	/bemə/	/bumə/	/bɑmə/
3	/timə/	/temə/	/tumə/	/tɑmə/
4	/dimə/	/demə/	/dumə/	/dɑmə/
5	/kimə/	/kemə/	/kumə/	/kɑmə/
6	/gimə/	/gemə/	/gumə/	/gɑmə/

determine whether or not object size is, in fact, at play and perhaps a better explanation for the sound symbolic link between these real-world objects and vowel backness.

3.2.2 Auditory stimuli. For this experiment, only the feature of vowel quality was targeted. While consonant-related features were piloted as a separate version of the real-world shape task, results indicated that participants were strongly influenced by lexical interference from the English names of the items, namely the matching of items whose English name contained the same initial consonant as a given nonsense word. In the main task presented here, the initial consonant was held constant by participant and counter-balanced across participants, making only vocalic features relevant. In order to manipulate only the stressed vowel, thereby including only one target vowel from the abstract shape task (/i/, /e/, /u/, and /ɑ/) in a given stimulus, the second syllable of each nonsense word included the unstressed vowel /ə/. Each participant heard four words taking the form /CVmə/, in which only the initial stressed vowel differed between the four words that any one participant heard. That is, one participant would hear all four words that began with /b/, one would hear all four words that began with /p/, and so on, numbered as conditions 1–6 in Table 3.

This ensured that the difference between the object that a participant matched with each word was related only to the difference in the stressed vowel between the words, since, in this particular task, a conflation of a number of target features would make it impossible to tease apart which phonetic feature of a word is being used to formulate responses. Conditions were counter-balanced across participants.

The same male American English speaker who recorded the stimuli in the abstract shape task produced the auditory stimuli, using the same procedure as in the abstract shape task. Each word was scaled for duration and peak amplitude in Praat (Boersma & Weenink, 2011). Again, the stimuli were pronounced using American English phonotactics—stress fell on the initial syllable of each word, and vowels fell within the vowel space for American English vocalic phonemes.

3.3 Procedure

Participants were first presented with a sound check, as outlined in the abstract shape task. Only those who passed the sound check were allowed to proceed to the task. Participants were then presented with four sound clips labeled "Word 1" through "Word 4." They were told that the clips were "martian words," and that "these martians live in a world just like ours, but they speak their own language." Participants were told that each of the four words referred to one of 10 pictured items. Each word contained the same consonant frame, with only the vowel differing by word (Table 3). The order of the words (i.e. which vowel corresponded to Word 1, which to Word 2, and so on) was randomized by participant. All of the visual stimuli were presented together below the sound clips, in two rows (five images per row) in a randomized order. Once the participant had heard each sound clip and seen the items, he or she continued to the next page. The participant was then provided again with the same four sound clips at the top of the screen, still labeled Word 1 through Word 4, which could be listened to as many times as the participant desired. The participant was shown the 10 object images again in a column on the left side of the screen, in a different random order than was used on the introductory screen, accompanied by four empty boxes labeled Word 1 through Word 4 on the right side of the screen. The participant was instructed to listen to each word sound clip again, then to look at all of the object images again to remind themselves of every available object option. They were then asked to drag and drop the image of the object on the left that he or she thought each martian word referred to into its corresponding box for the word on the right. Each participant was instructed to drag exactly one image to each "Word" box before completing the task and was not allowed to complete the task if more than one object was grouped into a single Word box, or if any of the Word boxes remained unfilled. As in the abstract shape task, participants were polled for their native language and the type of listening device they used during the experiment. The type of listening device used did not have an effect on the results. The final dataset included only responses from native English speakers.

3.4 Results

Findings from the real-world object task reinforced the findings of the abstract shape task with respect to the vocalic features that were linked to round and spiky object qualities. However, an experiment that appeals to participants' knowledge of real-world objects necessarily complicates the design of a typical *bouba-kiki* type experiment using abstract shapes, since real-world objects contain multiplex properties in the realm of the visual alone that are not easily reducible to a single salient quality. For example, the possibility that size sound symbolism could, in fact, obscure or override a shape-sound mapping in this task was broached earlier. As such, the influence of object size ratings on the matching of an object with back versus front vowels was tested. Size of an object, as delimited by the five categories determined via size ratings, did not have a significant effect on vowel backness of the stimulus that object was matched with (c^2 (4, n = 544) = 4.27, p = 0.37). A breakdown of vowel backness proportions for each size group (Appendix 2) reinforces that it is extremely unlikely that object size could be responsible for the sound symbolic effects that resulted from this experiment. The directionality of the backness-size mapping shows inconsistency across size classes, and any apparent effects in fact go against the direction predicted by the frequency code. For these objects, larger sizes slightly favor front vowels, whereas the frequency code predicts the opposite. This indicates that either the scaling of picture sizes produced a sound enough control for possible size sound symbolism, or, perhaps, that shape sound symbolism overrode possible size effects in this task.

A further complication that arises in the use of existing real-world objects is that English lexical labels for these objects already exist, the sounds of which could quite possibly influence the martian naming of such objects in this task. In order to examine the ways in which the English word for these items may have influenced categorization in the task, results for the objects that contained as their stressed syllable one of the four target vowels used in the auditory stimuli—/i/, /e/, /u/ or /ɑ/—were assessed in isolation. These objects included those that contained stressed /e/: "grater," "ladle," and "plate," that which contained stressed /u/: "spoon," and that which contained stressed /ɑ/: "tongs." The results for these five objects were broken down by which vowel they were most frequently matched with (Table 4).

For four of the five applicable objects, grater, plate, spoon, and tongs, it appeared that the lexical label of the object likely influenced the matching of that object with a target stimulus vowel. For

Table 4. Objects with lexical items containing a stressed vowel used as a target vowel, by vowel in matched martian auditory stimulus.

Object name	/i/	/e/	/u/	/ɑ/	Total *n*
Grater	4	**9**	3	4	20
Ladle	11	**7**	22	18	58
Plate	13	**21**	7	13	54
Spoon	19	9	**36**	14	78
Tongs	9	8	2	**21**	40

Instances in which target vowel in stimulus corresponded with stressed vowel in lexical label are emboldened.

these four items, the target vowel in the martian word with which the object was most frequently matched was the same as the stressed vowel in the English name for the word. Two of these matches correspond with the expected "sound-shape" link for the object and its vowel's backness (round "spoon" is most commonly matched with the vowel in its lexical label, /u/, which is a back vowel, and spiky "grater" is most commonly matched with the vowel in its lexical label, /e/, which is a front vowel). To prevent the apparent effect of English lexical label from spuriously influencing results, these four objects were removed prior to further analysis. The apparent effect of English lexical label was not found for one relevant object, "ladle," which in fact was matched with its stressed vowel in English, /e/, *least* frequently out of the four target vowels. This object was thus retained in the dataset. The removal of the four lexically-influenced objects eliminates one spikiness category (the "more spiky" category, which contains tongs and grater) and one roundness category (the "round" category, which contains spoon), but maintains the balance between round and spiky objects, leaving three of each (spiky: "carving fork," "fork," "knife,"; round: "pan," "bowl," "ladle"). The proportions of responses matched to words with back vowels, categorized by round grouping of objects and spiky grouping of objects, are shown in Figure 7.

Since every object was a member of one spiky group and one round group, separate chi-square analyses were conducted for the relation of spiky group membership and backness of vowel, and round group membership and backness of vowel. Both spiky group membership and round group membership produced significant results. Overall, round kitchen object responses were matched with words containing back vowels more often than with words containing front vowels (c^2 (2, n = 352) = 14.59, $p < 0.001$). Spiky kitchen object responses were matched with words containing front vowels more often than with words containing back vowels (c^2 (1, n = 352) = 10.05, p = 0.0015). Standardized residuals show that the relation of back vowels and the "more round" group ($z = 2.19$, $p = 0.029$), consisting of the object ladle, was the largest contributor to the effect of roundness group. The relation of front vowels and the "most spiky" group ($z = 1.75$, $p = 0.08$), consisting of carving fork, fork, and knife, was the strongest contributor to the significant effect of spikiness group.

3.5 Discussion: sound-symbolic associations with real-world objects and dimensionality in sound symbolism

The real-world object task, bringing investigations of the *bouba-kiki* effect out of the realm of the abstract, corroborates the findings of the abstract shape task with respect to vowel backness: kitchen objects deemed round were more likely to be matched with nonsense words containing back vowels, while kitchen objects deemed spiky were more likely to be matched with nonsense words containing front vowels. Although bringing this study into the realm of the real world

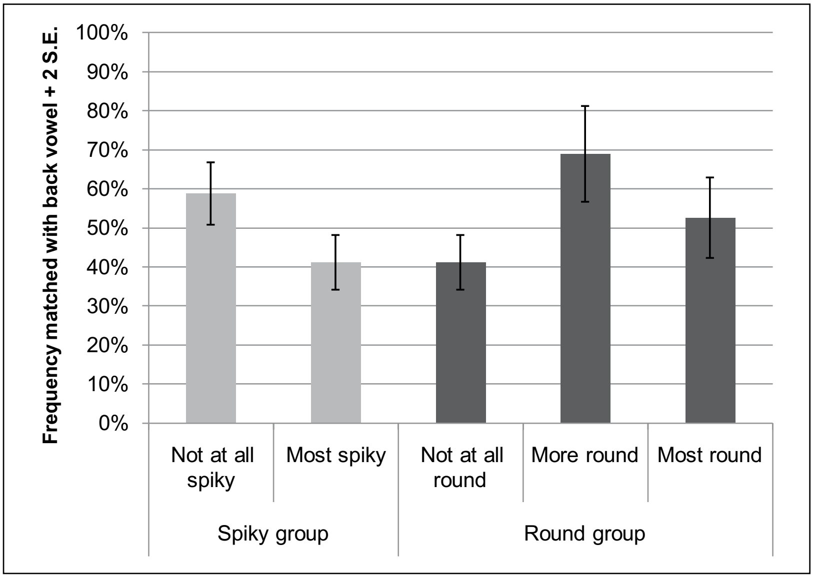

Figure 7. Proportion of kitchen object matches with back vowel stimulus, by spiky/round group membership, lexically-influenced objects removed.

complicates the relationship between sound and shape in some ways (e.g. pre-existing lexical labels, less control over other possible sound symbolic confounds like size), the present study does not allow these issues to preclude the use of such objects, but instead attempts to attend to these complications. This allows for an investigation of this effect that more closely applies to ways in which sound symbolism can operate with respect to objects in the real world, and thus in natural language.

A closer look at real-world objects illuminates more particular and concrete object properties that hearers may cue into when making sound symbolic associations, even within classes of objects that might be described broadly as round or spiky. The term "rounded" or "curved" appeared to serve as a reasonable descriptor for the abstract rounded splat and cloud objects used in abstract shape task. Yet, as shown by the kitchen objects that fit into these categories, a term like roundness alone does not capture the complexity of the property as it is manifested in real-life objects that exist in a three-dimensional world. While five of the objects were rated as somehow embodying roundness as an object quality, more careful consideration of how these objects relate to the property of roundness shows the more complex ways that this idea can be manifested within the set of kitchen items. For example, some of the objects can be considered round in that they are partially spherical ("bowl", "ladle", "spoon"), a three-dimensional conception of roundness. In contrast, some of the objects rated as round are round only two-dimensionally ("plate", "pan"). Since previous *bouba-kiki* studies—including the first experiment in this study—have presented participants with unfamiliar objects on a flat sheet of paper or screen, the role of dimensionality in sound symbolism has not yet been addressed. However, a term like roundness can indicate different qualities with respect to dimension or depth, particularly considering the ways that real-world objects are perceived and represented cognitively. Image-based theories of object recognition support the idea that exemplars of multiple viewpoints of an object are stored to constitute the mental representation of that object (see Palmeri & Gauthier (2004) for a review). Representations of objects could be perceived to be similar based not only on how they appear from a single viewpoint (e.g. a plate and a bowl both appear to be a circle from a particular viewpoint), but in fact how they appear from many different viewpoints. If it is to be claimed that a speaker makes sound symbolic associations between specific sounds and particular object qualities, we might consider the ways in which these representations create within-group distinctions in a single, broad, round category.

In a step toward this end, results from the real-world object task can be analyzed post hoc to begin to probe whether or not an object feature like dimensionality could be involved in mediating symbolic associations between sounds and objects. At a basic level, the group of round kitchen objects can be divided into two classes according to dimensionality of roundness: two of the objects are round only two-dimensionally (plate, pan) and three of the objects contain some sort of three-dimensional roundedness (bowl, ladle, spoon). Since two of these items, plate, and spoon, appeared to be lexically influenced, they were removed from the dataset. However, a comparison can be done between the remaining two-dimensionally round object, pan, the remaining three-dimensionally round objects, ladle and bowl, and the non-round objects. A chi-square test examining the effect of vowel backness on shape type (two-dimensionally round, three-dimensionally round, or not round) shows that this post hoc categorization of objects based on dimensionality is significant (c^2 (2, n = 352) = 14.76, $p < 0.001$), with the "three-dimensionally round" class of objects contributing the most to the effect ($z = 2.24$, $p = 0.025$). This dimensionality grouping yields an object roundness-vowel backness effect that is, in fact, slightly *more* significant than the roundness groups as delimited by ratings. Although this analysis is based on a small number of objects, and the categorization was made post hoc, results indicate that the type of roundness in particular that may be at play in this effect is *three-dimensional*, or spherical roundness, rather than that of two-dimensional circularity or curvature. Furthermore, it should also be noted that of the types of spiky shapes used in this task, the three rated most spiky—knife, fork, and carving fork—could also be characterized as not simply spiky or "angular," but also as physically "sharp." The spiky items that might be deemed less sharp—grater and tongs—were removed from the dataset due to lexical interference, preventing the possibility of a post hoc analysis of spiky categories based on sharpness. However, this feature in particular could also be investigated for its significance in the canonical *bouba-kiki* effect. In fact, previous studies have described the possible correlation of certain consonant sounds in "kiki" with angular shapes as related to physical "sharpness" (Ramachandran & Hubbard, 2001). The use of physical, three-dimensional objects that a participant not only sees, but interacts with, could allow for the distinction between angularity and tactile sharpness to be teased apart in future work.

The use of real-world objects deepens explorations of sound symbolism with respect to objects in the real world, and such investigations are necessary to inform broader claims about the ways that sound symbolism can and does operate in natural language. Studies of sound-shape correspondences in abstract space can offer controls that help disentangle specific phonetic correlates of these associations, as in the abstract shape task of this investigation. However, observing this phenomenon solely in abstract space limits our understanding of which specific object properties listeners use to develop or reinforce a sound-symbolic association with objects that exist in the world and with which speakers interact. Different instantiations of object properties like roundness or spikiness/angularity that are salient in real-world objects (e.g. dimensionality, sharpness), which could be at play in sound symbolism, would remain otherwise opaque in studies that deal only with sound symbolism in a fabricated, flattened world.

4 Conclusions and future directions

These exploratory tasks provide two separate refinements to the investigation of the sound symbolic *bouba-kiki* phenomenon. Examining sound symbolic association with respect to more detailed and linguistically relevant phonetic features on one hand, and more nuanced, real-world object properties on the other, enriches previous findings in sound symbolism in an empirical manner. First, the abstract shape task carefully considers phonetic features involved in round versus spiky sound-shape correspondences, demonstrating that features like place of articulation that have not previously been described to be involved in these correspondences do, in fact, have significant effects on round versus spiky shape selection. In particular, voicing of consonants, place of articulation of consonants, and backness of vowels are significant predictors of whether a participant matched a nonsense word with a rounded abstract shape or a spiky abstract shape. Further, results suggest the robustness of the *bouba-kiki* phenomenon may result from combinations of features that strengthen sound symbolic effects. This investigation of which sounds, or features of sounds, can be symbolic offers a concrete way of discussing the phonetic details of sound symbolic effects. Rather than conflating a number of phonetic features in stimuli, future studies should investigate sound symbolism first in terms of individual features, then combinations of these features, in order to fully understand the linguistic details of the phenomenon.

In a different vein, this study also adds depth to previous work on sound-shape correspondences by moving from a study of these correspondences in an abstract setting into the real world, through the use of a class of real-world objects: kitchen items. A finding of the abstract shape task—that back vowels correlate with round shapes—holds up not only with abstract objects, but also within this group of real-world objects. Further, roundness in terms of dimensionality becomes salient in this study, a facet of object quality that otherwise would not be relevant and thus not taken into consideration in abstract *bouba-kiki* type experiments. The strength of the correlation between back vowels and rounded objects is supported when restricted to those objects that have some sort of three-dimensional roundness, as opposed to objects that are round only two-dimensionally. This suggests that dimensionality is itself a non-auditory percept with which sounds can be associated in a sound symbolic fashion.

Of course, introducing real-world objects into this paradigm poses complications for truly getting at the process of sound symbolism—lexical interference, for example, can be controlled for in some ways, but is unavoidable in a design such as the one presented in the real-world object task. In order to address this issue while retaining the benefits of using complex real-world objects, future experiments may use objects that occur in the real world, but are unfamiliar to participants, such as specialist's tools, or unusually designed objects.[2] The complexity of visual object properties that one real-world object alone can comprise also poses difficulties toward making claims about very specific object properties. However, as this study aims to show, there is explanatory power in investigating the ways that sound symbolism operates with respect to real-world objects. By leaving such studies in the realm of the abstract, researchers limit the reach and depth of their findings. This investigation suggests that the challenges posed by using real-world objects as stimuli should not preclude the use of such objects within studies of sound symbolism; rather, they are issues to be navigated and negotiated in future studies.

Acknowledgements

I would like to thank Eve Clark, John Rickford and especially Penny Eckert for their invaluable comments and guidance on this project. I would also like to thank Isla Flores-Bayer, who collaborated on the first task in this study, and Jason Grafmiller, who recorded stimuli for this study, as well as audiences at the 2012

Linguistic Society of America Annual Meeting and at Stanford University, who provided useful feedback during the development of this project. My thanks goes to Marilyn Vihman, Mutsumi Imai, and particularly Christine Cuskley for their extremely helpful comments on earlier drafts of this paper. I am grateful to Shawn Tice, without whom the web design and execution of these experiments would not have been possible. Finally, my gratitude goes to the research participants whose results are included in this study.

Funding

This research was funded by a graduate student fellowship from the Linguistics Department of Stanford University.

Notes

1. While concerns have been raised in the field regarding the validity of experiments conducted online versus in a laboratory, researchers have found that web experiment methods do, in fact, produce comparable results to those found in a laboratory, and in fact carry a number of advantages (e.g. Birnbaum, 2004; Dandurand, Shultz, & Onishi, 2008; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000), most notably the comparably large number of diverse participants who can complete a task in a relatively short amount of time, for little cost. The reliability of Mechanical Turk itself has been validated for the field of linguistics in particular: in a number of psycholinguistic experiments conducted via Mechanical Turk, results were comparable to results from the same experiment conducted in a laboratory (Schnoebelen & Kuperman, 2010; Snow, O'Connor, Jurafsky, & Ng, 2008). In fact, in two experiments, Mechanical Turk participants were found to produce more consistent trends than participants in a laboratory (Schnoebelen & Kuperman, 2010) in spite of the broader range of demographic categories that Mechanical Turk workers fill in comparison with the typical laboratory participant pool, often primarily comprised of college undergraduates.
2. Many thanks to Christine Cuskley for this insightful suggestion.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*, 298–348.

Aveyard, M. E. (2012). Some consonants sound curvy: Effects of sound symbolism on object recognition. *Memory & Cognition*, *40*, 83–92.

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nicols, & J. Ohala (Eds.), *Sound symbolism* (pp. 76–93). Cambridge, UK: Cambridge University Press.

Birnbaum, M. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, *55*, 803–832.

Blakemore, S., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, *43*, 260–267.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.20. Retrieved from http://www.praat.org/

Cuskley, C., Kirby, S., & Simner, J. (2010). Cross-modality: reviving iconicity in the evolution of language. In A. D. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.) *The evolution of language: Proceedings of the 8th international conference*. London, UK: World Scientific. (Presentation at Evolang8).

Dandurand, F., Shultz, T., & Onishi, K. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, *40*, 428–434.

Davis, R. (1961). The fitness of names to drawings: a cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*, 259–268.

de Saussure, F. (1966). *Course in general linguistics*. New York, NY: McGraw Hill.

Hamano, S. (1998). *The sound symbolic system of Japanese*. Stanford, CA: CSLI Publications; Tokyo, Japan: Kuroshio.

Hinton, L., Nicols, J., & Ohala, J. (Eds.) (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Holland, M., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or 'maluma' and 'takete' revisited. *Perceptual and Motor Skills*, *19*, 111–117.

Imai, M., Asano, M., Arata, M., Okada, H., Kitajo, K., & Thierry, G. (2011). Eleven month-old infants detect sound symbolism: evidence from an ERP study. Poster presented at the 36th Annual Boston University Conference on Language Development. 2011 November 5; Boston, MA.

Irvine, J. T., & Gal, S. (2000). Language ideology and linguistic differentiation. In P. Kroskrity (Ed.), *Regimes of language* (pp. 35–83). Santa Fe, NM: School of American Research Press.

Jespersen, O. (1933). Symbolic value of the vowel *i*. In *Linguistica. Selected papers in English, French and German* (pp. 283–303). Copenhagen, Denmark: Levin & Munksgaard.

Köhler, W. (1929). *Gestalt psychology*. New York, NY: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*, 19–28.

Krantz, J. & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 35–60). New York: Academic Press.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*, 316–322.

Mazzoni, D. & Dannenberg, R. (2000). Audacity (Version 1.3.12.0) [Computer software]. Pittsburgh, PA: Carnegie Mellon University. Available from http://www.audacity.sourceforge.net

McGraw, K., Tew, M., & Williams, J. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science*, *11*, 502–506.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology*, *65*, 115–124.

Nielsen, A., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, *4*, 115–125.

Ohala, J. (1994). The biological bases of sound symbolism. In L. Hinton, J. Nicols, & J. Ohala (Eds.), *Sound symbolism* (pp. 222–236). Cambridge, UK: Cambridge University Press.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2011). Sound-shape correspondences: biased in infancy? Poster presented at the 36th Annual Boston University Conference on Language Development. 2011 November 5; Boston, MA.

Palmeri, T. J., & Gauthier, I. (2004.) Visual object understanding. *Nature Reviews Neuroscience*, *5*, 291–303.

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

R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Rogers, S., & Ross, A. (1975). A cross-cultural test of the maluma-takete phenomenon. *Perception*, *5*, 105–106.

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

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, *43*, 441–464.

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. (2008). Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 254–263) Honolulu, Hawaii: Association for Computational Linguistics.

Tarte, R. D. (1974). Phonetic symbolism in adult native speakers of Czech. *Language and Speech*, *17*, 87–94.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language*, *93*, 10–19.

Appendix 1

Table 5. Counts of round and spiky shape selected in abstract shape task, by critical object pair and consonant voicing.

	Splat pair			Cloud pair		
	Round	Spiky	Total *n*	Round	Spiky	Total *n*
Voiced C	75	18	93	55	25	80
Voiceless C	38	39	77	36	54	90
Total *n*	113	57	170	91	79	170

Table 6. Counts of round and spiky shape selected in abstract shape task, by critical object pair and consonant place.

	Splat pair			Cloud pair		
	Round	Spiky	Total *n*	Round	Spiky	Total *n*
Labial C	49	12	61	44	17	61
Alveolar C	26	27	53	17	33	50
Velar C	38	18	56	30	29	59
Total *n*	113	57	170	91	79	170

Table 7. Counts of round and spiky shape selected in abstract shape task, by critical object pair and vowel backness/roundness.

	Splat pair			Cloud pair		
	Round	Spiky	Total *n*	Round	Spiky	Total *n*
Back V	68	28	96	61	27	88
Front V	45	29	74	30	52	82
Total *n*	113	57	170	91	79	170

Table 8. Splat pair logistic regression model summary in abstract shape task. (n = 170).

Predictor	Estimate	Standard error	z-value	*p*-value
Intercept	–2.4452	0.4653	–5.255	0.000000148
Backness = front	0.6011	0.3697	1.626	0.103991
Voicing = voiceless	1.4152	0.3638	3.890	0.000100
Place = alveolar	1.5298	0.4600	3.326	0.000881
Place = velar	0.6978	0.4640	1.504	0.132573

Table 9. Cloud pair logistic regression model summary in abstract shape task. (n = 170).

Predictor	Estimate	Standard error	z-value	*p*-value
Intercept	–1.73102	0.50095	–3.456	0.000549
Backness = front	1.56896	0.39060	4.017	0.000059
Voicing = voiceless	–0.03999	0.61076	–0.065	0.947797
Place = alveolar	0.45694	0.64209	0.712	0.476680
Place = velar	–0.18648	0.61125	–0.305	0.760302
Interaction = voiceless and alveolar	2.77397	0.99461	2.789	0.005287
Interaction = voiceless and velar	2.09095	0.86285	2.423	0.015380

Table 10. Counts of round and spiky cloud shape selected in abstract shape task, by consonant place and voicing.

Consonant place	Consonant voicing	Round	Spiky	Total *n*
Labial	Voice	20	9	29
	Voiceless	24	8	32
Alveolar	Voice	14	8	22
	Voiceless	3	25	28
Velar	Voice	21	8	29
	Voiceless	9	21	30
	Total *n*	91	79	170

Appendix 2

Table 11. Round and spiky kitchen objects obtained in open-ended pilot survey for real-world object task, average ratings in survey and *n* responses in main experiment.

Round objects	Average roundness rating	*n* responses in open-ended pilot	Spiky objects	Average spikiness rating	*n* responses in open-ended pilot
Bowl	1.95	12	Fork	1.85	31
Mixing bowl	1.85	2	Knife	1.75	41
Pot	1.75	7	Grater	1.45	3
Plate	1.7	17	Total		75
Pan	1.65	10			
Cup	1.65	12			
Straw	1.6	2			
Glass	1.55	3			
Dish	1.55	4			
Ladle	1.5	4			
Kettle	1.45	3			
Spoon	1.3	43			
Total		119			

Table 12. Spikiness and roundness ratings and groupings for visual stimuli used in real-world object main task.

Spikiness group	Pictured object	Mean spikiness rating	Roundness group	Pictured object	Mean roundness rating
Most spiky	Carving fork	1.75	Most round	Plate	1.9
	Fork	1.5		Pan	1.85
	Knife	1.3		Bowl	1.85
Spiky	Grater	0.65	More round	Ladle	1
	Tongs	0.45	Round	Spoon	0.65
Not at all spiky	Spoon	0.05	Not at all round	Tongs	0
	Bowl	0		Grater	0
	Ladle	0		Knife	0
	Pan	0		Carving fork	0
	Plate	0		Fork	0

Table 13. Size ratings and groupings for visual stimuli used in main real-world object task.

Size category	Object	Mean size rating	Normalized mean size rating
Smallest	Fork	0.05263	−1.225
	Spoon	0.05263	−1.225
Small	Carving fork	1.105	−0.372
	Knife	1.211	−0.314
	Tongs	1.263	−0.286
	Ladle	1.474	−0.133
Medium	Grater	1.947	0.203
Large	Plate	2.895	0.943
	Bowl	3.000	1.014
Largest	Pan	3.526	1.398

Table 14. Percentages of objects matched by backness of vowels and size group in main real-world object task.

Size group	Matched with back vowel (%)	Matched with front vowel (%)	Total *n*
Smallest	48.8	51.2	166
Small	54.5	45.5	209
Medium	35.0	65.0	20
Large	47.7	52.3	111
Largest	44.7	55.3	38

Table 15. Counts of objects matched to back and front vowel stimuli in real-world object task, by spikiness group as determined by ratings (items influenced by lexical label removed).

Spikiness group	Back vowels	Front vowels	Total *n*
Not at all spiky	90	63	153
Most spiky	82	117	199
Total *n*	172	180	352

Table 16. Counts of objects matched to back and front vowel stimuli in real-world object task, by roundness group as determined by ratings (items influenced by lexical label removed).

Roundness group	Back vowels	Front vowels	Total *n*
Not at all round	82	117	199
More round	40	18	58
Most round	50	45	95
Total *n*	172	180	352

Table 17. Counts of objects matched to back and front vowel stimuli in real-world object task, by roundness group by dimensionality (items influenced by lexical label removed).

Post hoc dimensionality-based roundness group	Back vowels	Front vowels	Total *n*
Not at all round	82	117	199
2-d round	17	21	38
3-d round	73	42	115
Total *n*	172	180	352