Canadian Journal of Experimental Psychology
2011, Vol. 65, No. 2, 115–124

1196-1961/11/$12.00 DOI: 10.1037/a0022268

The Sound of Round: Evaluating the Sound-Symbolic Role of Consonants in the Classic *Takete-Maluma* Phenomenon

Alan Nielsen and Drew Rendall
University of Lethbridge

Köhler (1929) famously reported a bias in people's matching of nonsense words to novel object shapes, pointing to possible naïve expectations about language structure. The bias has been attributed to synesthesia-like coactivation of motor or somatosensory areas involved in vowel articulation and visual areas involved in perceiving object shape (Ramachandran & Hubbard, 2001). We report two experiments testing an alternative that emphasizes consonants and natural semantic distinctions flowing from the auditory perceptual quality of salient acoustic differences among them. Our experiments replicated previous studies using similar word and image materials but included additional conditions swapping the consonant and vowel contents of words; using novel, randomly generated words and images; and presenting words either visually or aurally. In both experiments, subjects' image-matching responses showed evidence of tracking the consonant content of words. We discuss the possibility that vowels and consonants both play a role and consider some methodological factors that might influence their relative effects.

Keywords: sound symbolism, language structure, bouba-kiki, name-shape biases

Debate about the potential naturalness of words dates to at least Socratic times. In Plato's *Cratylus* dialogue, Socrates argues for the natural relationship between the structure of words and the things they denote (i.e., their meaning), while his interlocutor Hermogenes argues that the relationship is purely conventional. The naturalist and conventionalist positions have been debated many times since, but following Saussure (1949) the dominant view in contemporary linguistics is that *the form of the sign is arbitrary* and established by social convention.

This conclusion is not seriously disputed for the vast majority of words, but many possible exceptions have also been noted. Such exceptions are often considered together under the banner of *sound-symbolism* to indicate that the physical form of words (or of higher-order language constructions) might sometimes bear a non-arbitrary connection to the meanings instantiated (reviewed in Hinton et al., 1994). Among the best known examples is the phenomenon of vowel-specific marking of size diminution and augmentation where real or nonsense words containing high, front vowels (e.g., *bit*, *chip*, *teeny*) are judged more natural for denoting small size, while words containing low, back vowels (e.g., *block*, *hunk*, *chunk*) are judged more natural for denoting large size (Sapir, 1929).

This effect has been traced to broader sound-size relationships obligatorily embodied in the physics of sound production generally and more specifically in the physics of voice production in humans and other animals (Morton, 1994; Ohala, 1994). For example, compared to small individuals, larger individuals also often have larger sound production apparatuses which naturally produce sounds of lower frequency. This basic relationship between body size and sound frequency holds across a variety of animal species and modes of sound production. It is important that it also applies among primate species, including humans, where sound production is primarily voice-based. In humans and other primates, the vocal folds of the larynx and the vocal-tract cavities that resonate sounds generated by the larynx are both generally larger in larger-bodied individuals. This pattern yields lower frequency voices (with both lower pitch and lower resonances) in, for example, adults compared to children and in men compared to women (reviewed in Fitch & Hauser, 1995; Fitch, 2000; Rendall, Kollias, Ney & Lloyd, 2005; Ghazanfar & Rendall, 2008).

Consistent evidence that sound-symbolic relationships like this are actually manifest in the structure of real words is often lacking, controversial, or complicated (e.g., Bentley & Varon, 1933; Ultan, 1978; Diffloth, 1994). As a result, proposals of sound-symbolism are often viewed with skepticism and treated as marginalia in linguistics and psycholinguistics. However, the consistency of people's behavior in experimental studies of sound-symbolism, even if they involve interpreting nonsense word material, suggests that people might operate with inherent perceptual-cognitive biases that yield naïve expectations about language structure. Hence, it is important to investigate these possibilities further because perceptual-cognitive biases of this sort could serve to facilitate normative language processing, or even language acquisition, if the biases are commensurate with real language constructions. Alternatively, such biases could serve to frustrate or impede language processing or acquisition if they run counter to the relationships instantiated in real languages. Either way, the biases could

Alan Nielsen and Drew Rendall, Behaviour and Evolution Research Group, Department of Psychology, University of Lethbridge.

We thank Doug Mewhort and Daphne Maurer for comments that significantly improved the work. We thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for grant support.

Correspondence concerning this article should be addressed to Drew Rendall, Department of Psychology, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada. E-mail: d.rendall@uleth.ca

influence peoples' language facilities and they must, therefore, feature in the development of a fuller understanding of the psychology of language.

The Takete-Maluma Phenomenon

Another paradigm example of potential naïve language structure expectations involves the "takete-maluma" phenomenon. This phenomenon was famously described by Köhler (1929, 1947) who reported that subjects reliably matched nonsense words such as *takete* and *maluma* to images of unfamiliar *jagged* and *curved* objects, respectively, suggesting that people might indeed have some naïve expectations about the kinds of words best suited to name certain kinds of objects. Similar effects have been reported in various additional studies and these studies have also extended the observations to include other kinds of images and other nonsense words, such as *bouba* and *kiki*.

For example, Davis (1961) found that participants preferentially matched the nonsense word *takete* with a jagged image and *uloomu* with a curved image, and this was true for both English and Swahili speakers. Westbury (2005) tested the effect using a more indirect, selective interference design and found that participants were quicker to recognize and classify nonwords (by comparison to real words) that were presented visually in text form if the consonant class of the nonwords matched the shape of the frame within which the words were presented. Specifically, nonwords presented within curved frames were recognized more quickly if they contained continuant consonants (e.g., /L/ and /m/) and nonwords presented within jagged frames were recognized more quickly if they contained stop consonants (e.g., /t/ and /k/). In another test, Tarte found that speakers of both English (Tarte & Barritt, 1971) and Czech (Tarte, 1974) consistently matched words containing the vowel /ee/ (as in *beat*) to figures that were triangular in shape and they matched words containing the vowel /oo/ (as in *rule*) to figures that were elliptical.

Despite results like these, there have been lingering doubts about the robustness of the matching biases involved. For example, there have been concerns about the limited number and variety of word and image stimuli used in many experiments. There have also been concerns that inadvertent selection biases by experimenters might account for some of the effects reported making them effectively circular (reviewed in Westbury, 2005). At the same time, explanations for the source of the *taketa-maluma* phenomenon have also varied considerably and added to the ambiguities. Thus, some studies trace the effect to the consonant content of words (e.g., Westbury, 2005), while others trace it to the vowel content (e.g., Tarte & Barritt, 1971; Tarte, 1974; Ramachandran & Hubbard, 2001; Maurer, Pathman, & Mondloch, 2006).

Recently, Ramachandran and Hubbard (2001) offered a comprehensive neuropyschological account of the phenomenon. They proposed that the object-shape: word-form matching bias reflects synesthesia-like coactivation of the motor or somatosensory areas involved in articulating different sounds and the visual areas associated with the perception of differently shaped objects. They argued that such cross-modal linkages yield the natural matching of, for example, the visual percept of a curved, or rounded, object and the motor representation involved in the round-mouth articulation of the /oo/ vowel sound in the word *bouba*. Ramachandran and Hubbard (2001) elaborated this proposal into a broader synaesthetic theory of language origins and consciousness.

Maurer et al. (2006) extended this synaesthetic proposal in experiments conducted on young children (2.5 years old, sometimes referred to as toddlers). Despite relatively little language experience, these children showed the same matching tendencies as adults. Maurer et al. attributed the effect in infants to synesthesia-like, cross-modal linkages that they propose are a general feature of neural organization in infants. Many of these cross-modal linkages are thought to be gradually pruned during development to yield the relative independence of different sensory systems that characterize most adults, but the cross-modal linkages are argued to be especially functional in infancy for language learning (Maurer & Mondloch, 2004; Mondloch & Maurer, 2004).

This recent emphasis on vowels and on potential cross-modal neural activity in accounts of the *taketa-maluma* phenomenon is compelling. However, so too is the possibility that the consonant content of words might play a role (Westbury, 2005). Thus, there are clear spectral density and attack differences between /k/ and /m/ that make /k/ a relatively harsh (or strident) consonant and /m/ a relatively mellifluous (or sonorant) consonant. These basic differences in spectral structure might naturally tend to imply or conjure "*harsh, fractured, or jagged*" constructs on the one hand and "*smooth connected, rounded*" constructs on the other, and these effects might characterize other strident and sonorant consonants.

Part of our impetus for reevaluating the latter possibility stems from consistent *affective-semantic* relationships observed widely in the communication systems of animals (Rendall & Owren, 2009). Across many primates and other animal species, harsh, noisy and punctuate (i.e., strident) sounds are associated with situations of high arousal and often also hostility and aggression. In contrast, smoother more harmonic (i.e., sonorant) sounds are associated with situations of lower arousal and also positive affiliation and contact (reviewed in Morton, 1977; Owren & Rendall, 1997, 2001; Rendall & Owren, 2009).

For example, in aggressive confrontations with predators or rival group members, many primate species produce very loud calls, often referred to by investigators as "barks," "screeches," or "screams." As the latter names for these calls imply, the calls tend to have sharp (abrupt) onsets and to be composed primarily of unstructured, broad-band noise. These calls tend to induce reflexive reactions in listeners preparatory to fight-or-flight type responses based on fairly direct stimulation of autonomic processes regulating whole-body arousal. In contrast, when calmly foraging or seeking close contact with social companions, the same primates produce calls with more gradual onsets and structured tonal or harmonic frequency spectra that tend to be referred to as "coos" or "peeps." These kinds of calls tend not to induce the same sort of dramatic autonomic responses in listeners (reviewed in Owren & Rendall, 2001; Rendall & Owren, 2009).

These and other vocal-affect relationships are continuous with effects also known in human voice production. For example, young infants variously produce either sonorant-type coos or relatively strident loud and abrasive screeches and screams in situations reflecting, respectively, relative comfort or contentment versus hunger or distress. And calls of these two types tend to induce very different affective and behavioral responses in caretakers and other adult listeners. Similarly, adult voicing patterns can be relatively punctuate, harsh and noisy in situations of high arousal (whether related to fear or anger) compared to during normative relaxed speech (reviewed in Rendall, 2003; Bachorowski & Owren, 2008).

Taken together, there appear to be some very broad relationships between strident and sonorant sounds and the contexts they mediate that have different social and behavioral salience for listeners. It is possible that these relationships might extend, at least in a limited way, to certain consonant sounds of language that exemplify this same strident-sonorant distinction and that thus inherit some natural semantic potential.

To test this hypothesis, we undertook a set of experiments which were organized by the following logic: 1. Our first objective was simply to replicate the matching biases reported in previous studies of the takete-maluma phenomenon, utilizing the same word and image materials and the same testing methods used previously. This seemed a critical first step to be certain we were dealing with the same phenomenon. 2. A second major objective was to address concerns raised about stimulus materials used in previous studies and possible biases in their selection and construction. To address this issue, we developed new methods for generating random image and word materials that eliminate subjectivity in the selection of experimental stimuli. We used these randomly generated images and words to extend previous work on the takete-maluma phenomenon and evaluate its robustness. 3. The third major objective was to test the extent to which the matching biases involved might be influenced specifically by the consonant content of the words, according to the strident-sonorant distinction just outlined.

Experiment 1: Replicating and Extending the Taketa-Maluma Phenomenon

The first step in research was an attempt to replicate the word-form: object-shape matching patterns reported in previous studies and extend past research on the influence that consonants in particular might play. To do this, our first experiment replicated the design of previous studies of the takete-maluma phenomenon and used similar word and image materials. Thus, participants were shown line drawings of an unfamiliar jagged object and an unfamiliar curved object and were given a choice of two nonsense words to match to them. At the same time, our experiment included two additional conditions designed to address methodological concerns raised about previous studies and to specifically test the role that consonants might play in the word-matching bias. One condition involved simply swapping the consonant and vowel contents of the original word stimuli to test whether participants' matching performance would track the previous vowel content or the consonant content. The other condition involved using an entirely new set of randomly generated images and words in an attempt to eliminate potential biases in experimenters' selection and construction of stimuli.

Methods

Participants

Participants were 24 undergraduate students (13 female, 11 male) who were enrolled in introductory psychology courses at the University of Lethbridge and received partial course credit for their participation.

Word and Image Stimuli

In Condition 1 of this experiment, which was the direct replication condition, the word and image stimuli came directly from those used in previous studies by Köhler (1947) and Maurer et al. (2006). The word and image stimuli used are shown in Table 1 and Figure 1, respectively. Some of the stimuli used in Maurer et al. (2006) involved a mixed media presentation that allowed for haptic interaction with the objects. To improve comparability to previous studies, including the original work by Köhler, we used the typical two-dimensional graphical representations of these objects.

For Condition 2, we created a second set of word pairs by swapping the consonant content of the original word pairs used in Condition 1. For example, the classic word pair used originally by Köhler (1947) contrasted *takete* and *maluma*. In Condition 2, this word pair became *maleme* and *takuta* (see Table 1 for other examples). This manipulation was undertaken to test the effect that consonants might have using exactly the same word materials and methods that were used in previous studies. If previous interpretations were correct that the vowel content of words is primarily responsible for the preferential matching effects reported, then swapping the consonant content of the words should not alter the result; subjects' matching responses should remain the same. However, if the consonant content of words has an effect, then subjects' matching responses should often change and follow the consonant (not vowel) content of the original words.

For Condition 3, we created an entirely new set of words to more systematically test the potential role of consonants and to minimize additional possible confounds raised previously. For example, to construct words, we selected as strident consonants /t/, /k/, and /p/ and as sonorant consonants /L/, /m/, and /n/. The former are established stop-consonants in which voicing (i.e., vibration of the vocal folds) follows release of the consonant, creating a brief pulse of aspirated turbulent noise prior to the onset of stable vocal fold vibration for the following vowel sound. The latter sonorants are continuant consonants which are produced simultaneously with the vocal-fold vibration involved in the following vowel sound which makes the consonant component of the sound relatively vowel-like in structure. This selection of consonants allowed us to test the proposed affective-semantic distinction between relatively noisy, strident sounds and relatively mellifluous, sonorant sounds. This selection of consonants also minimized potential orthographic confounds in which the visual form of the letters in the words resembles the jagged or curved form of the object shapes. This has been proposed to account for some previous results when for example words containing the letter /K/ are matched to jagged objects while words containing the letter /B/ are matched to rounded objects. However, such orthographic effects should not have been a factor in our experiment, because all three sonorant consonants, which are predicted to be matched to curved object shapes, are actually relatively jagged in their capitalized form (/L/, /M/, /N/) and one remains so in lower-case form (/L/). At the same time, one of the strident consonants, which are proposed to be matched to jagged object shapes, is patently curved in both upper- and lower-case forms (/P/, /p/). As a result, the orthographic

Table 1
Word Stimuli Used in Experiment 1 and Their Vowel- and Consonant-Type Classifications and Predicted Image-Type Associations[1]

Condition 1 (Word pairs from Köhler and Maurer et al.)		Condition 2 (Consonant-swapped word pairs from Condition 1)		Condition 3 (Randomly generated word pairs)	
Kuhtay (U:J) (St:J)	Baamoo (R:C) (So:C)	Buhmay (U:J) (So:C)	Kaatoo (R:C) (St:J)	Paka (St:J) Lana (So:C)	Peka (St:J) Mela (So:C)
Tuhkeetee (U:J) (St:J)	Maaboomaa (R:C) (So:C)	Muhbeemee (U:J) (So:C)	Taakootaa (R:C) (St:J)	Kepa (St:J) Muna (So:C)	Tatu (St:J) Lalu (So:C)
Kaykee (U:J) (St:J)	Boobaa (R:C) (So:C)	Baybee (U:J) (So:C)	Kookaa (R:C) (St:J)	Tuka (St:J) Nema (So:C)	Keka (St:J) Lulu (So:C)
Teetay (U:J) (St:J)	Gogaa (R:C) (So:C)	Geegay (U:J) (So:C)	Totaa (R:C) (St:J)	Kapa (St:J) Munu (So:C)	Kuta (St:J) Luna (So:C)
Takete (U:J) (St:J)	Maluma (R:C) (So:C)	Maleme (U:J) (So:C)	Takuta (R:C) (St:J)	Tapa (St:J) Namu (So:C)	Petu (St:J) Mana (So:C)

Note. Note that, in Condition 1, the vowel and consonant classifications are fully confounded and thus make exactly the same predictions about the image-type associations for the words. In Condition 2, this confound is broken by swapping the consonant content of words within pairs and the predicted image-type associations for the words are now different based on their vowel and consonant content. In Condition 3, the confound is eliminated by using randomly generated words in which the vowel content is balanced across word pairs. R = Rounded vowel; U = Unrounded vowel; St = strident consonant; So = Sonorant consonant; J = jagged image; C = Curved image.
[1]Letters in brackets beside each word indicate the word's vowel-type classification from previous studies, their consonant-type classification from the current study, and the image type associations that are predicted to follow from these classifications.

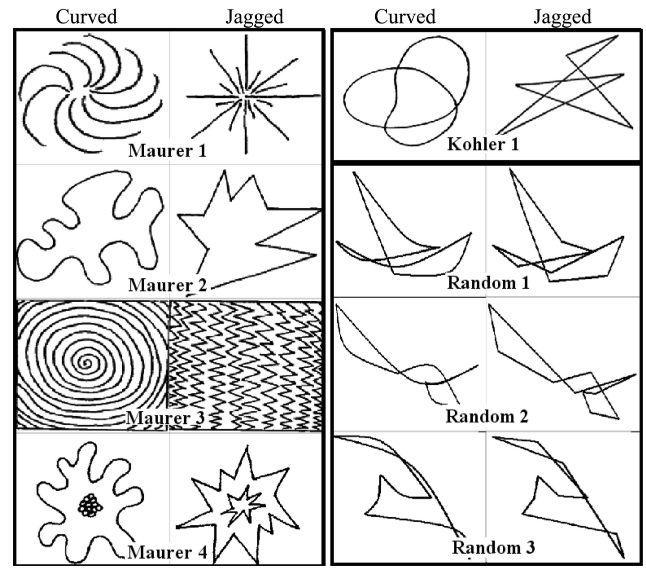

Figure 1. Image stimuli used in Experiment 1. Side-by-side comparisons of curved and jagged image pairs from previous studies by Maurer et al. (2006) and Köhler (1947) that were used in Conditions 1 and 2, and a sample of the new, randomly generated curved and jagged images used in Condition 3. Note that the latter curved and jagged images are very subtly different, varying only in the curvature of the lines linking calculus points.

representation of the consonants used in this experiment often worked against the predictions of our hypothesis.

For vowels, we chose not to use /o/ and /i/ because the orthographic form of these two vowels could be too obviously and canonically associated with rounded and jagged images, respectively. We used instead the vowels /a/, /e/, and /u/ and always in their lower-case forms. Hence, the orthographic forms of these vowels were all similarly curved or rounded and would therefore introduce no systematic bias.

To avoid any possible subjectivity in the construction of words from this limited alphabet, we created word stimuli using Gammadyne Random Word Generator . Words were constrained to be four letters long in a two-syllable, consonant-vowel-consonant-vowel (cVcV) sequence, with the first letter capitalized. An additional constraint was that the letter /e/ could not appear in word-final position because that construction alters the pronunciation of the preceding vowel sound and can create a single-syllable word out of a two-syllable one (e.g., *Paku* vs. *Pake*). The result was a database of randomly constructed, two syllable words that were either entirely strident (e.g., *Taku*) or entirely sonorant (e.g., *Nemu*) by consonant content (for a complete listing of word stimuli used in Condition 3, see Table 1).

For Condition 3, we also created an entirely new set of randomly generated curved and jagged images. To accomplish this, we developed a radially constrained mathematical formula which populated a field of finite size (Figure 2-A) with either five or 10 sequentially ordered, randomly generated, calculus points. The location of these points was determined by creating vectors which departed from the centre of a field at random magnitudes and directions. These points were numbered in the order that they were created, and sequentially labeled points had a second vector cal-

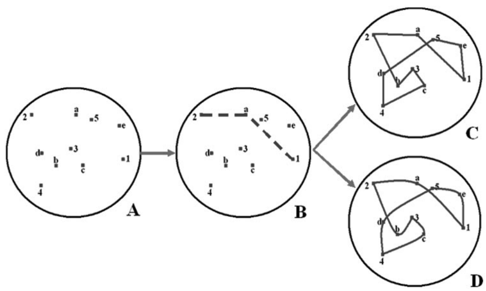

Figure 2. Schematic illustrating the process used to create curved and jagged versions of the same image based on a set of randomly generated calculus points.

culated from the midpoint of the line connecting them, giving a calculus point for the drawing of simple curves. Individual line segments were defined by their two original points and the generated curvature point (Figure 2-B). They were then joined with either straight (Figure 2-C) or curved (Figure 2-D) lines. Sequential line segments were then connected to one another. The resulting curved and jagged images were identical to one another except for the curvature of the lines used to draw them, which often yielded only very subtle differences between them (see Figure 1). Most important, this method also standardized the size of images and so avoided any possible size-related confounds in subjects' matching responses.

Together, these precautions for Condition 3 addressed common concerns about the limited variety of past word and image stimuli and potential biases in experimenters' selection of them.

Experimental Design

The experiment was conducted on computer via a graphical interface created using Runtime Revolution, Version 2.8.1. In each trial, the participant was shown two images side-by-side (one curved, one jagged) with two words below. Participants were asked to label the images with the words. It is important that, as in previous studies, labeling one image with a given word automatically led to the assignment of the remaining word to the other image. Participants were free to change their labeling. They advanced to the next trial by clicking a button on-screen.

The experiment involved 20 of these forced-choice trials in three conditions: Condition 1 consisted of five trials using the original word and image pairs from previous experiments; Condition 2 consisted of five trials using the original image pairs and the original word pairs in which the consonant content of the words had been swapped; Condition 3 involved 10 trials using the new set of random images and random words. The coupling of word and image pairs, and their placement on screen, in each trial was randomized within condition and the ordering of trials belonging to the different conditions was randomized within and across subjects.

Data Analysis

Matching responses for experimental trials were scored for each subject using two different schemes. In the first scheme, subjects' matching performance was scored according to the vowel content of words based on criteria used in previous studies. Hence, a correct score was defined as matching the curved image with the word containing vowels labeled in previous studies as rounded (/ah/, /oh/, /oo/), or matching the jagged image with the word containing vowels labeled in previous studies as unrounded (/uh/, /ay/, /ee/). In the second scheme, subjects' matching performance was scored according to the consonant content of words as per our alternative hypothesis emphasizing the acoustic quality of consonants. Here, a correct score was defined as matching the curved image with the word containing sonorant consonants (/L/, /m/, /n/) or matching the jagged image with the word containing strident consonants (/t/, /k/, /p/). Because some previous studies were not focused on differences in consonant quality, some of the word pairs we used that replicated past word pairings were complicated in that they contained only strident consonants. To address this issue, we categorized the different words within such pairs according to differences in their relative sonority using established differences in the voice-onset-time and spectral density of their consonants (see Table 1). Subjects correct matching scores were averaged across trials within each condition, and individual subject averages were tested for their deviation from chance (50% correct) using one-sample *t* tests.

Results

In Condition 1, subjects' responses to the original stimuli used by Köhler (1947) and by Maurer et al. (2006) were scored based on both the vowel and consonant matching schemes and the original findings of these authors were replicated (see Figure 3). Participants chose correctly 82% of the time when trials were scored using the vowel matching scheme, $t(23) = 8.35$, $p < .01$. Participants also chose correctly 82% of the time when trials were scored using the consonant matching scheme, $t(23) = 8.35$, $p < .01$. Note that this pattern of outcomes, in which subjects could show correct matching of original stimuli according to both the vowel content of words and the consonant content of words, is possible because the word stimuli used in previous studies in-

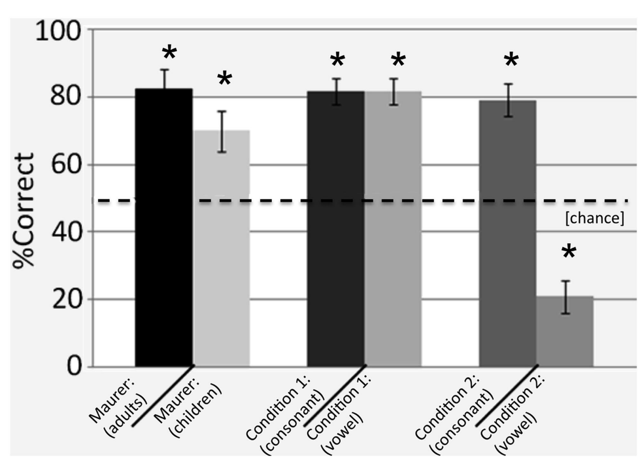

Figure 3. Mean (±SEM) correct matching performance by adults and children in Maurer et al. (2006) compared to subjects' correct matching performance in Conditions 1 and 2 of Experiment 1 when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance, 50% ($p < .01$).

volved inadvertent consonant-vowel associations. That is, vowels that were labeled in previous studies as rounded appeared in words containing sonorant consonants and vowels previously labeled as nonrounded appeared in words containing strident consonants. This confound is actually apparent in Table 1 which shows the vowel- and consonant-type classifications of the original word stimuli and their predicted jagged or curved object-type associations. It is apparent that, in Condition 1, the rounded and unrounded vowel-type categories and the strident and sonorant consonant-type categories are perfectly confounded and thus make the same object-type association predictions.

In Condition 2, this confound is broken because the consonant content of words was swapped within pairs. Hence, the vowel and consonant content of the words now make different predictions about object-type associations. In this condition, correct scores based on consonant content remained significant at 79%, $t(23) = 5.97$, $p < .01$, but correct scores based on the vowel matching scheme dropped to 21%, $t(23) = -5.97$, $p < .01$, confirming that subjects were tracking consonant content.

Condition 3 used new and randomly generated word and image materials and was a test specifically of the possible influence of different consonant types on subjects' matching choices. Correct scores based on the vowel scheme could not be calculated then because our word generation method controlled for and balanced vowel content across word types which therefore did not preserve the rounded-unrounded vowel distinction used in previous studies. In this condition, according to the distinction between strident and sonorant consonants, participants chose correctly 81% of the time, $t(23) = 7.71$, $p < .01$; Figure 4.

Finally, when the data for Conditions 1 and 2 were collapsed and analyzed together, the consonant content of word stimuli significantly predicted participant choices overall, 80%: $t(23) = 7.57$, $p < .01$, while performance based on vowel content was not different from chance, 51%; $t(23) = 0.85$, $p = .20$; Figure 4.

Discussion

Results of Experiment 1 replicated the word form:object shape matching patterns reported in previous studies. In Condition 1,

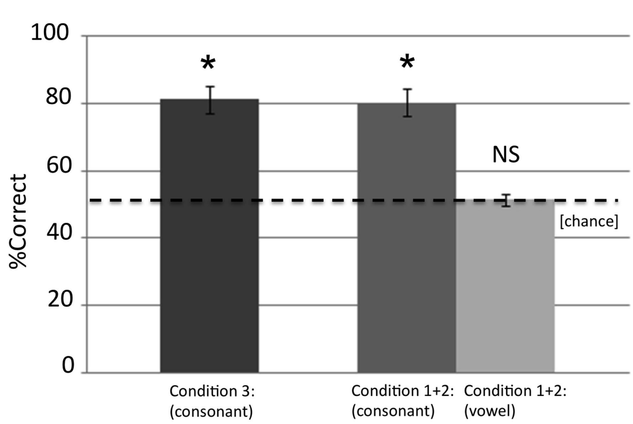

Figure 4. Mean (±SEM) correct matching performance in Experiment 1 in Condition 3 and in Conditions 1 and 2 combined when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance ($p < .01$) except where noted (NS).

using similar methods and materials as previous studies, subjects showed the same matching bias reported in those studies. For example, our subjects consistently matched *baamoo* with the curved image and *kuhtay* with the jagged image, as in previous work. However, in Condition 2, when the consonant content of words was simply swapped within pairs, subjects changed their selections. Subjects still showed a matching bias but it followed the consonant rather than vowel content of the words. For example, the word *kaatoo* was now matched to the jagged image and *buhmay* was matched to the curved image. This consonant-based matching bias was confirmed in Condition 3 using randomly generated words and images. In this condition, subjects matched words containing strident consonants to jagged images and they matched words containing sonorant consonants to curved images even though the differences between curved and jagged image forms in this condition were extremely subtle.

It is important to emphasize that our results are unlikely to be attributable to orthographic confounds. Our selection of vowels specifically omitted /o/ and /i/ that have obvious visual associative value with curved and jagged images, respectively, and used instead only /a/, /e/ and /u/ in their lowercase forms which are all relatively curved in visual appearance. Furthermore, by using consonants in sonorant and strident categories that had both jagged and curved visual form polymorphisms, depending on their capitalization, we reduced the possibility of any systematic orthographic biases in the visual form of consonants.

At the same time, there are some potential limitations in our experiment and in the inferences that can be drawn from it. First, our experiment followed previous studies in using a procedure in which both types of images and both categories of words were presented simultaneously for subjects to match. This design is intuitive but risks making the matching task relatively transparent to subjects by allowing them to directly compare both types of images and both categories of words on each trial.

In addition, the simultaneous presentation procedure makes it impossible to discover whether the observed strident-jagged and sonorant-curved matching outcomes truly reflect two biases (and choices) or really only one, because, on every trial, the subjects second "choice" is dictated automatically by which image and word they first match. In other words, with the simultaneous presentation design used, a one-sided matching bias would yield the same evidence as a two-sided matching bias. This ambiguity does not mean that there is, in fact, no underlying matching bias, only that we cannot be sure on the basis of our results whether and to what degree it involves both a strident-word:jagged-image association and a sonorant-word:curved-image association as opposed to only one of these. The very same problem plagues previous studies that have used this experimental design.

Finally, although our results highlighting the potential role of consonants in subjects' matching preferences were unlikely attributable to orthographic effects, we cannot definitively conclude that it was necessarily the acoustic quality of the different consonant types that mediated subjects' performance as hypothesized because, strictly speaking, the words were presented only visually in this experiment and not aurally. It is possible that when subjects saw the words on-screen, they tended to pronounce the words to themselves softly or subvocally, but we cannot be sure of this.

As a result, we undertook a second experiment to address these issues.

Experiment 2: Auditory Test of the Role of Consonant Quality in the Taketa-Maluma Phenomenon

Experiment 2 paralleled the design of Experiment 1 but involved two methodological changes. One change involved moving to sequential rather than simultaneous presentation of image materials. Thus, on each trial subjects saw only a single object image and subjects were given a choice of two words to match to that object (i.e., a two-alternative forced choice design). This change eliminated the opportunity for direct comparison between curved and jagged images on each trial and made the matching task less transparent. The second change involved shifting the word-dimension of the matching task into the auditory domain where the hypothesized effects arising from the acoustic qualities of different consonants could be tested more definitively. That is, words were presented aurally on each trial rather than visually in text form.

Methods

Participants

Participants were 88 undergraduate students (75 female, 13 male) who were enrolled in introductory psychology courses at the University of Lethbridge and received partial course credit for their participation.

Image and Word Stimuli

Experimental stimuli once again derived from previous studies by Köhler (1947) and by Maurer et al. (2006). In this experiment, we also included stimuli used in a study by Westbury (2005). In total, 20 images (10 curved, 10 jagged) and 42 pairs of words from these previous studies were used as stimuli in the present experiment. Of the original word pairs, 21 preserved the original vowel and consonant content and 21 involved pairs in which the consonant content of words in a pair had been swapped as in Condition 2 of Experiment 1. Once again, we supplemented this sample of materials with a set of random images and words that we generated ourselves according to procedures described for Experiment 1. In total, we used 22 images (11 curved, 11 jagged) from the random image generation technique and 42 pairs of words from the random word generator. In generating words, we again used the stop-consonants /t/, /k/, and /p/ as strident consonants and the continuants /L/, /m/, and /n/ as sonorant consonants. We used only the vowels /a/, /e/, and /u/ and retained the restriction that the vowel /e/ could not appear in word-final position.

To facilitate aural presentation of word material in this experiment and standardize the pronunciation of all words, we used a commercial text-to-speech program (Swifttalker) to generate spoken versions of each word from their associated text forms. Spoken versions of each word were generated in each of two realistic-sounding synthetic voices, one male (David) and the other female (Diane). These synthetic voices were developed using the parameters and standard phoneme pronunciations for American English without any strong regional accent. Hence, the pronunciation of individual word tokens did not vary between David and Diane. For word stimuli derived from previous studies, the original pairing of words was left intact (e.g., *Bouba-Kiki*). For new word stimuli, word pairs were constructed to control and balance vowel and consonant content.

Experimental Design

The experiment was conducted on computer via a graphical interface created using Runtime Revolution, Version 2.9. On each trial, the participant was shown a single image, and two words were played sequentially via headphones. Participants were then asked to choose which word best matched the image using a button labeled either <*word 1*> or <*word 2*>. Participants could replay the words by clicking a button on-screen.

The experiment involved 84 of these forced-choice trials: each of the 42 images was presented twice, each time coupled with a different word pair one of which involved words derived from previous studies and one of which involved our new randomly generated words. The coupling of particular word pairs and images was randomized. Which word in a pair was played first and the sex of the synthesized speaker were counterbalanced across trials.

Data Analysis

To preserve the comparability of the results of the two experiments, the analysis of data for Experiment 2 paralleled that for Experiment 1. Thus, subjects' matching responses in experimental trials were once again scored based on both the vowel content and the consonant content of words as per Experiment 1. Subjects' correct matching scores according to these two schemes were averaged across trials within each condition, and individual subject averages were tested for their deviation from chance (50% correct) using one-sample *t* tests.

Results

Condition 1 of this experiment once again replicated the matching patterns reported in previous studies and in Condition 1 of our own first experiment (see Figure 5). Using the original word and image materials, participants chose correctly 60% of the time based on the vowel matching scheme, $t(87) = 6.24$, $p < .01$. Participants also necessarily chose correctly 60% of the time using the consonant matching scheme, $t(87) = 6.24$, $p < .01$ because the

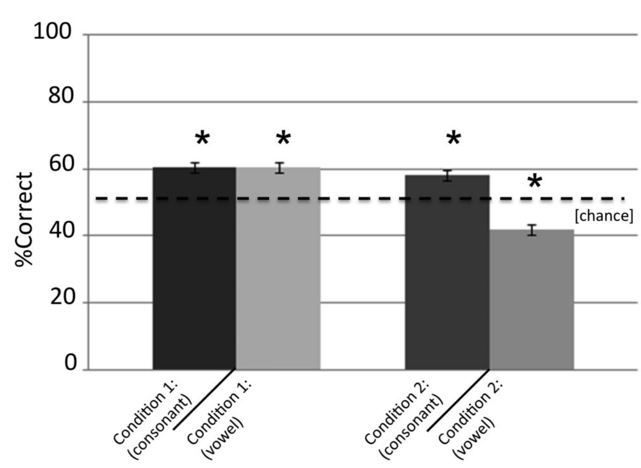

Figure 5. Mean (±SEM) correct matching performance in Conditions 1 and 2 of Experiment 2 when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance, 50% ($p < .01$).

vowel and consonant content of the original word pairs used in this condition were perfectly confounded.

In Condition 2, in which the consonant content of the word stimuli used in previous studies was simply swapped, subjects' performance based on consonant content once again remained significant at 58%, $t(87) = 4.72$, $p < .01$, but their performance based on the vowel matching scheme dropped below chance, 42%, $t(87) = -4.72$, $p < .01$, Figure 5.

In Condition 3, using new and randomly generated word and image materials, participants chose correctly 59% of the time based on consonant content, $t(87) = 8.05$, $p < .01$; Figure 6. Finally, when the data for Conditions 1 and 2 were collapsed and analyzed together, the consonant content of word stimuli significantly predicted participant choices overall, 59%: $t(87) = 6.95$, $p < .01$, while vowel content did not, 51%; $t(87) = 1.28$, $p = .10$; Figure 6.

Overall, subjects' matching performance based on the consonant content of words was higher for jagged images (63% correct) than for curved images, 54% correct: $t(86) = 4.66$, $p < .01$. Matching performance based on consonant content was also better for the more stylized and differentiated jagged and curved images used in previous studies (63% correct) than for the randomly generated jagged and curved images that we developed and that were only subtly different from each other, 55% correct: $t(86) = 4.89$, $p < .01$.

Discussion

Results of Experiment 2 corroborated those of Experiment 1 in showing a bias to match unfamiliar jagged and curved images with nonsense words based on differences in their consonant content. In both experiments, words containing strident consonants were preferentially matched with jagged images, and words containing sonorant consonants were matched with curved images. Results of Experiment 2, which involved only aural presentation of word stimuli, suggest that these consonant-based effects were mediated by differences in the acoustic quality of strident and sonorant consonants as hypothesized.

Results of Experiment 2 also tended to confirm some methodological concerns about previous studies and our own Experiment 1. These revolved around using simultaneous presentation of both types of words and images on each trial which allows subjects to make direct comparisons among them. This approach raises questions about the relative transparency of the overall task and also about the possibly one-sided nature of the matching bias needed to account for previous findings. We attempted to address both issues in Experiment 2 by moving to a procedure that involved sequential presentation of experimental stimuli. Consistent with this change in experimental procedure, subjects' correct matching performance also dropped substantially to an average of 58% in Experiment 2 from nearly 80% in Experiment 1. We attribute this drop in performance to the change in experimental procedure which made the task more difficult (less transparent), but it is possible that it also reflects other differences between the two experiments related to the processing of word materials presented either visually or aurally.

At the same time, subjects' responses in Experiment 2 were found to involve an asymmetry in matching bias: the association of strident words with jagged images was stronger than the association of sonorant words with curved images.

The latter matching asymmetry pertained only to trials involving our newly generated images, however, and reflected low correct matching scores specifically for the curved images. In reexamining images, it is clear that, in general, our curved images are less curved and more jagged than are the stylized types of curved images used in previous studies. This results from the fact that, in our image generation technique, the curved and jagged forms of images were generated from exactly the same set of randomly generated calculus points. This precaution was undertaken purposefully as an attempt to reduce concerns about inadvertent experimenter biases creeping into the choice of image materials used previously. However, in some cases, our "unbiased" figure generation technique necessarily yielded intersecting lines that resulted in acute interior angles in the final form of the curved version of each image (see Figure 1 for examples). As a consequence, and in the context of the larger image set they were exposed to, subjects might have viewed many of our curved images as relatively jagged.

If true, the overall asymmetry in correct matching of jagged and curved images that we report might not reflect a real asymmetry in subjects' word-form:object-shape matching biases so much as an additional image-related methodological factor to be cognizant of in future work as discussed in more detail below.

General Discussion

We report two experiments testing the classic Takete-Maluma phenomenon. In both experiments, subjects preferentially matched nonsense words to unfamiliar jagged and curved images based on the consonant content of words. It is important that the effects held for classic image and word material used originally by Köhler and by other investigators since, as well as for a new set of randomly generated images and words created specifically to circumvent concerns about previous biases in the selection of stimulus materials. The latter outcomes with controlled and randomized words and randomized images in which jagged and curved image forms were only subtly different from each other helps to buttress the conclusion that the word-form:object-shape matching bias is a bona fide effect, which has often been questioned. They also

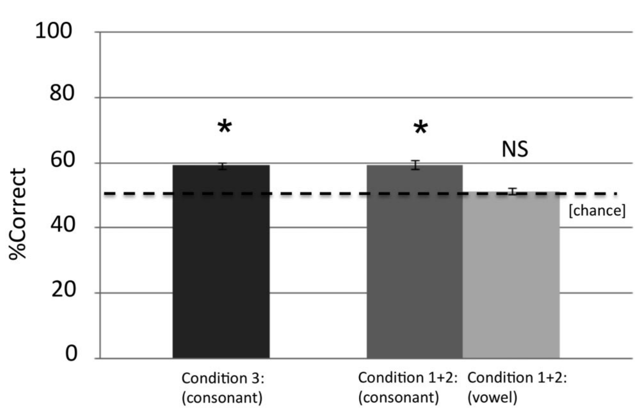

Figure 6. Mean (±SEM) correct matching performance in Experiment 2 in Condition 3 and in Conditions 1 and 2 combined when scored by either the vowel or the consonant content of words. Effects shown are significantly different from chance ($p < .01$) except where noted (NS).

suggest that the observed matching biases are not leashed to a specific set of experimenter-selected images, nor do they depend on dramatic differences in image curviness or jaggedness, which has also been suspected. At the same time, the results highlight that the details of object form are certainly important. For example, we found that subjects' correct performance with our curved images was not as strong as it was with our jagged images, likely because the mathematics of our random figure generation technique necessarily produced curved images containing some acute angles that gave them an element of jaggedness. This finding opens the door to additional future work testing how subjects' matching biases are sensitive to such specific details of visual object form. In this respect, the novel figure generation techniques introduced here offer considerable promise because they allow a virtually limitless number, range, and variety of controlled images to be generated for future testing.

Overall, our results are consistent with some previous findings highlighting the role of consonants in the Takete-Maluma phenomenon (e.g., Westbury, 2005). They also revealed that some of the results obtained in earlier studies that were previously attributed to the vowel content of words reflected a nonarbitrary coupling of consonants and vowels within words. When this inadvertent consonant-vowel association was broken in our experiments, we found that subjects' matching choices tracked the consonant content of words more than they did the vowels. This consonant-based effect was demonstrated to hold whether words were presented visually or aurally. This pattern of findings is consistent with our proposal that the consonant-based effect is likely mediated by differences in the auditory perceptual quality that different consonants have owing to variation in their pattern of spectral density and attack which make them either relatively harsh sounding (strident) or relatively smooth sounding (sonorant). And, as outlined in the introduction, these effects may be continuous with vocal-affective relationships observed more widely in the communication systems of many animal species which involve consistently differentiated use of harsh, strident vocalizations versus smooth, sonorant vocalizations in social behavioral contexts that portend different affective consequences for listeners (reviewed in Owren & Rendall, 1997, 2001).

On the surface of it, this consonant-based account of subjects' matching biases appears to weaken the proposal that previous effects were due to the vowel content of words and to synesthesia-like linkages between the visual percepts of object shape and motor (or somatosensory) activity associated with articulation of different vowel sounds. However, the two accounts need not be mutually exclusive. Thus, although consonants appeared to trump vowels in influencing subjects' performance in our experiments overall, it is likely that the vowel content of words accounted for some of the variation in subjects' responses. In addition, our experiments used only a limited selection of consonants. It is therefore possible that, in the context of different consonants, vowels might play a greater role in matching biases (e.g., Tarte & Barritt, 1971; Tarte, 1974; Ramachandran & Hubbard, 2001; Maurer et al., 2006) and that, when they do, their effects might be mediated by cross-modal linkages of exactly the sort proposed (Ramachandran & Hubbard, 2001; Maurer et al., 2006). It is also possible that the relative influence of vowels and consonants might vary depending on whether word stimuli are presented either visually or aurally, or maybe more pertinently on whether they are processed either visually or aurally by subjects no matter how they are canonically presented to subjects.

Furthermore, while the consonant account that we emphasize does not require a synesthetic linkage between visual and somatosensory systems per se, it does nevertheless point to some type of communication across, or integration among, sensory domains in as much the auditory quality of consonant sounds is proposed to interact with the visual system to produce regularities in the matching of auditory and visual percepts.

It is important also to note that the consonant-based effects observed here do not mean necessarily that strident consonants actually conjure specific semantic constructs like "*harsh, jagged, fractured*", or that sonorant consonants conjure the opposite semantic constructs ("*smooth, connected, rounded*"). Indeed, strident and sonorant consonants might have no very specific semantic associations in any absolute sense. Instead, the consonant effects might be rooted in processes that are very low-level, peripheral, and presemantic (Westbury, 2005).

Nevertheless, whatever their proximate reality, the effects involved could yield naïve expectations about language structure at some level, either preprepared or experientially driven, and these are intriguing to consider for their, as yet, relatively unstudied influences on language processing and learning (e.g., Maurer et al., 2006). For example, an increasing body of work confirms the importance of object shape in childrens' early word learning (reviewed in Samuelson & Smith, 2005). Although not yet systematically studied, it is possible that this process could be facilitated further, or, alternatively, that it could be hampered to some degree, depending on the extent to which the form of real object words is either commensurate with, or contradicts, any such naïve expectations about language structure.

Résumé

Köhler (1929) a rendu célèbre un biais dans la façon d'associer des mots dépourvus de sens à des nouvelles formes d'objet, démontrant la présence possible d'attentes naïves à propos de la structure du langage. Le biais a été imputé à une coactivation de type synesthésique des aires motrices ou somatosensorielles impliquées dans l'articulation des voyelles et des aires visuelles impliquées dans la perception des formes (Ramachandran & Hubbard, 2001). Nous rapportons deux expériences visant à tester une variante mettant l'accent sur les consonnes et les distinctions sémantiques naturelles émanant de la qualité des perceptions auditives de différences acoustiques saillantes entre elles. Nos expériences visaient à reproduire des études antérieures en utilisant des mots et des images semblables mais en incluant des conditions additionnelles afin de substituer les voyelles et les consonnes des mots; utiliser des mots et images nouveaux, générés aléatoirement; présenter les mots visuellement ou auditivement. Dans les deux expériences, les réponses des sujets à la tâche d'association d'images ont indiqué des signes de repérage du contenu en consonnes des mots. Nous discutons de la possibilité que les voyelles et les consonnes jouent toutes deux un rôle et considérons quelques facteurs méthodologiques pouvant influencer leurs effets relatifs.

Mots-clés : symbolisme des sons, structure du langage, bouba-kiki, biais nom-forme

References

Bachorowski, J.-A., & Owren, M. J. (2008). Vocal expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *The handbook of emotion* (3rd ed., pp. 196–210). New York: Guilford Press.

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

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

Diffloth, G. (1994). i:big, a:small. In L.Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 107–114). Cambridge: Cambridge University Press.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Science, 4,* 258–267.

Fitch, W. T., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology, 37,* 179–190.

Ghazanfar, A. A., & Rendall, D. (2008). The evolution of human vocal production. *Current Biology, 18,* 457–460.

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

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

Köhler, W. (1947). *Gestalt psychology* (2nd ed.). New York: Liveright.

Maurer, D., & Mondloch, C. (2004). Neonatal synesthesia: A reevaluation. In L. Robertson & N. Sagiv (Eds.), *Attention on synesthesia: Cognition, development and neuroscience* (pp. 193–213). Oxford: Oxford University 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.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, and Behavioral Neuroscience, 4,* 133–136.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some birds and mammal sounds. *American Naturalist, 111,* 855–869.

Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge: Cambridge University Press.

Ohala, J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 325–347) Cambridge: Cambridge University Press.

Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate signaling. In D. H. Owings, M. D. Beecher, & N. S. Thompson (Eds), *Perspectives in ethology* (Vol. 12: *Communication,* pp. 299–346). New York: Plenum Press.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology, 10,* 58–71.

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

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America, 113,* 3390–3402.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117,* 944–955.

Rendall, D., & Owren, M. J. (2009). Vocalizations as tools for influencing the affect and behavior of others. In: S. M. Brudzynski (Ed.), *Handbook of mammalian vocalization* (pp. 177–185). New York: Elsevier Press.

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: Spontaneous naming and attention to shape. *Developmental Science, 8,* 182–198.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12,* 225–239.

Saussure, Ferdinand de. (1983). *Course in general linguistics,* Charles Bally and Albert Sechehaye Eds. Translated by Roy Harris. La Salle, IL: Open Court.

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

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

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

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

Received January 18, 2010
Accepted November 1, 2010 ■