RESEARCH REPORT

What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages

MARK DINGEMANSE
Max Planck Institute for Psycholinguistics

WILL SCHUERMAN
Max Planck Institute for Psycholinguistics, and *Universiteit Utrecht*

EVA REINISCH
Ludwig Maximilian University Munich

SYLVIA TUFVESSON
Max Planck Institute for Psycholinguistics

HOLGER MITTERER
Max Planck Institute for Psycholinguistics, and *University of Malta*

Sound symbolism is a phenomenon with broad relevance to the study of language and mind, but there has been a disconnect between its investigations in linguistics and psychology. This study tests the sound-symbolic potential of ideophones—words described as iconic—in an experimental task that improves over prior work in terms of ecological validity and experimental control. We presented 203 ideophones from five languages to eighty-two Dutch listeners in a binary-choice task, in four versions: original recording, full diphone resynthesis, segments-only resynthesis, and prosody-only resynthesis. Listeners guessed the meaning of all four versions above chance, confirming the iconicity of ideophones and showing the viability of speech synthesis as a way of controlling for segmental and suprasegmental properties in experimental studies of sound symbolism. The success rate was more modest than prior studies using pseudowords like *bouba*/*kiki*, implying that assumptions based on such words cannot simply be transferred to natural languages. Prosody and segments together drive the effect: neither alone is sufficient, showing that segments and prosody work together as cues supporting iconic interpretations. The findings cast doubt on attempts to ascribe iconic meanings to segments alone and support a view of ideophones as words that combine arbitrariness and iconicity. We discuss the implications for theory and methods in the empirical study of sound symbolism and iconicity.*

Keywords: iconicity, sound symbolism, ideophones, pseudowords, prosody

1. INTRODUCTION. The existence of an arbitrary relation between a word and its meaning can be seen as an important design principle of language (Hockett 1960). However, it does not have exclusive reign, as shown by various types of regularities in form-meaning associations in natural languages. Such regularities include well-known iconic uses of the vowel space in mapping size (Ohala 1984) and distance (Woodworth 1991, Johansson & Zlatev 2013), but also iconic patterns found in phonaesthemes (Kwon & Round 2014) and in IDEOPHONES, vivid sensory words found in many of the world's languages (Vigliocco & Kita 2006, Dingemanse 2012).

There appears to be a tendency to either underplay or exaggerate the significance of iconicity in the study of language and mind. In linguistics, for instance, the number of 'pictorial, imitative or iconic words' has been declared to be 'vanishingly small' (Newmeyer 1992, relying on Whitney 1874), a claim that is hard to reconcile with the typological prevalence and numerical preponderance of ideophones (also known as 'expressives' or 'mimetics') in languages as diverse as Japanese, Korean, Zulu, Gbaya, and Quechua (Voeltz & Kilian-Hatz 2001). But the iconicity of ideophones may sometimes

* Thanks to Felix Ameka, Jiyoun Choi, Sho Tsuji, and Makiko Sadakata for help with the selection and recording of Ewe, Korean, and Japanese ideophones. This research was funded by the Language Comprehension and Language & Cognition departments of the Max Planck Institute for Psycholinguistics, and by an NWO Veni grant (016.154.087) to MD.

have been overclaimed. They have been characterized as 'words or phrases that do the work of representation by phonetic means' (Tedlock 1999:118), and in line with this, some descriptions of Japanese ideophones have attempted to pin down the iconic meanings of individual phonemes; for example, /p/ is proposed to indicate the 'abrupt movement of a tensely stretched small surface' (Hamano 1986:114; cf. Imai et al. 2008). We dub this the STRONG ICONICITY ASSUMPTION. If it were literally true, readers should be able to tell aspects of the meaning of the Zuni ideophone *ch'uk'i-* or the Japanese ideophone *pin* right away.[1]

A similar all-or-nothing approach has characterized work in psychology: while some theories of psycholinguistics have the assumption of arbitrariness built in, leaving little room for nonarbitrary associations of form and meaning (Levelt et al. 1999, Friederici 2002), other work has heralded sound symbolism as a stepping stone in language evolution (Ramachandran & Hubbard 2001) and as a crucial factor supporting language development (Maurer et al. 2006). Key to the latter are findings based on pseudowords like *bouba* and *kiki* presented in binary forced-choice paradigms. How such findings can be linked to the role of iconicity in natural languages is an open question.

In this article we present evidence that some sound-symbolic correspondences between form and meaning found in natural languages can be recognized as such by speakers of other languages. Our results highlight the nature of ideophones as iconic words and the importance of prosody in inviting and supporting sound-symbolic interpretations. However, our results also show that segmental information alone is insufficient to suggest meaning, that overall performance is relatively modest compared to prior work, and that the potential for iconic associations differs across semantic domains. Showing more modest and differentiated effects than previous work using pseudowords, our findings motivate a critical reappraisal of what sound symbolism can and cannot do in natural language.

1.1. THE EXPERIMENTAL STUDY OF SOUND SYMBOLISM. Sound symbolism is a phenomenon whose experimental investigation started around the same time in linguistics (Sapir 1929) and psychology (Köhler 1929)—twin roots that affirm the relevance of sound symbolism to the study of both language and mind. The early experimental studies found a wide following in psychology, where scholars continue to document cross-culturally consistent nonarbitrary associations between sound and meaning. Best known is the association of pseudowords like *maluma* or *bouba* with rounded shapes, and the contrasting association of *takete* and *kiki* with angled shapes (Köhler 1929, Brown & Nuttall 1959, Davis 1961, Ramachandran & Hubbard 2001, Maurer et al. 2006, Bremner et al. 2013, Lockwood & Dingemanse 2015). Recent studies have also shown that verb meanings in the domain of motion are learned more easily for novel words that are construed on the template of Japanese ideophones than for novel arbitrary words (Imai et al. 2008, Kantartzis et al. 2011). Only in rare cases have real words been presented to listeners who do not know the languages from which the words are taken (Slobin 1968, Kunihira 1971, Nygaard, Cook, & Namy 2009, Lockwood et al. 2016).

These experiments all report either above-chance recognition at first try or better performance in a word-learning task if a (pseudo)word is coupled with its appropriate rather than an alternative meaning. As Table 1 shows, participants are quite comfortably above the chance level of 50%, with an average success rate of 74% in pseudoword studies.

[1] According to Tedlock (1999:118), *ch'uk'i-* 'evokes sounds like that of eye popping out of its socket'. According to Hamano (1986:115), *pin* refers to 'the plucking of a guitar string' or 'the stretching of a thread'.

Using signal-detection theory to translate this into an effect-size measure (mean difference divided by the standard deviation) provides a relatively large effect of $d' = 1.3$.

STUDY	SUCCESS RATE IN %
Ramachandran & Hubbard 2001	95
Imai et al. 2008, exp. 1b/exp. 2/exp. 3	64/70/82
Kantartzis et al. 2011	70
Maurer et al. 2006, adults/three-year-olds	82/70
Monaghan et al. 2012	55
AVERAGE	74%

TABLE 1. Success rates in guessing pseudoword meanings in previous studies.

Findings like these have given rise to far-reaching claims about the potential relevance of sound symbolism to language evolution, language structure, and language learning. According to some studies, the *bouba-kiki* phenomenon provides a 'vital clue for understanding the origins of proto-language' (Ramachandran & Hubbard 2001:19), and it 'can indeed facilitate the learning of languages in which rounder objects tend to be labelled with rounded vowels' (Maurer et al. 2006:321). Before such conclusions can be drawn, however, it is worthwhile to consider their empirical grounding more closely. One reason for caution is that success rates vary quite strongly across studies. Ramachandran and Hubbard (2001), with informally described methods, report 95% correct choices on first encounter. Monaghan and colleagues (2012), with an admirably detailed description of methods and materials, report only 55% correct on first encounter. Such discrepancies imply that matters of experimental design deserve particular scrutiny.

First, there is a question of ecological validity. The stimuli used in binary forced-choice tasks tend to be designed for maximal contrast and minimal complexity. The words are contrived forms that usually bear no direct relation to sound-symbolic forms found in natural languages, and the meanings tend to be mutually exclusive antonyms that differ on one dimension only (e.g. rounded vs. pointy). While some degree of simplification is essential in order to achieve experimental control and isolate key variables, a legitimate question arises as to how behavior in a task like this can be linked to the facts of natural languages. Our communicative needs go far beyond distinguishing round from pointy objects, and our linguistic systems are shaped by a myriad of competing motivations (Dingemanse et al. 2015). Moreover, in everyday language use we seldom encounter single-purpose words with antonymic alternative interpretations; instead, words have multiple layers of meaning and are interpreted in context. This suggests that the forms and functions of sound symbolism in natural languages may be different from the laboratory versions studied so far.

Second, prosody is a potential confounding factor that is not controlled for in many studies. Experimental control should begin at the point of recording, where speakers should be unaware of the experimental design to avoid unconscious biases (Rosenthal & Fode 1963). Additionally, controlling for the prosodic properties of auditory stimuli is critical, because prosody has known effects on sound-symbolic interpretations. For instance, performance in a forced-choice task is enhanced by an expressive tone of voice (Kunihira 1971), and prosody alone can successfully cue semantic distinctions like *strong-weak* or *hot-cold* (Nygaard, Herold, & Namy 2009), even in novel-word learning (Reinisch et al. 2013). Tellingly, investigations of sound symbolism have rarely considered the role of prosody, instead tending to attribute the effect to phonetic aspects of vowels and consonants alone (as in Maurer et al. 2006, Imai et al. 2008, Aveyard 2012, Nielsen & Rendall 2013). This means that many current accounts of sound symbolism may overestimate the role of segments and underestimate the role of prosody.

1.2. CURRENT STUDY. Addressing these challenges requires methodological innovations that introduce a greater degree of ecological validity while keeping experimental control over key factors. To achieve this goal, in this study we use the existing diversity in ideophone systems of languages around the world as a natural laboratory, and we use speech synthesis to create controlled versions of auditory stimuli. We chose five languages described as having a major lexical class of words that show recurring iconic associations between form and meaning: Japanese, Korean, Semai, Siwu, and Ewe (Table 2). Different linguistic traditions have used slightly different descriptive terms for these words, but there is a broad consensus in linguistic typology that these represent essentially the same phenomenon, for which we use the cover term 'ideophones' (Samarin 1970, Diffloth 1972, Nuckolls 1999, Voeltz & Kilian-Hatz 2001, Dingemanse 2012).[2] If sound symbolism plays a role in natural languages, this is one of the places where we should be able to find it.

LANGUAGE	PHYLUM	DESCRIPTIVE CATEGORY
Japanese	Isolate	'mimetics' (Kita 1997, Akita 2009)
Korean	Isolate	'ideophones' (Martin 1962, Lee 1992)
Semai	Mon-Khmer, Austro-Asiatic	'expressives' (Diffloth 1976, Tufvesson 2011)
Siwu	Na-Togo, Niger-Congo	'ideophones' (Dingemanse 2012)
Ewe	Kwa, Niger-Congo	'ideophones' (Westermann 1907, Ameka 2001)

TABLE 2. Languages used in the experiment.

We compiled a collection of ideophones by sampling from five semantic domains in each of the languages: Sound, Motion, Texture, Shape, and Visual appearance. This introduces greater ecological validity in two ways: the forms are not contrived but based on existing words, and the meanings represent a cross-section of the actual semantic distinctions made by ideophonic words in natural languages. Prior work has focused on associations across modalities (as in *bouba*/*kiki*, where sound is used to depict aspects of shape). Our stimuli include both associations within modality (as in the Sound domain, where sound depicts sound) and associations across modalities (in the Motion, Texture, Shape, and Visual appearance domains, where sound is used to depict aspects of other sensory experiences). We hypothesize that Sound ideophones will be easier to guess because the iconic associations are within the same modality, so we will able to use that as a baseline test, while the other domains make comparison with prior work possible.

In order to exercise greater control over stimulus design and to tease apart the role of segmental versus suprasegmental properties, we use diphone resynthesis to generate stimuli. Each word was prepared in four versions: (i) a native speaker's utterance (original recording), (ii) a diphone resynthesis replicating the segments, pitch, and amplitude contour of the original utterance (full resynthesis), (iii) a resynthesis using typical Dutch phoneme durations and a flat intonation contour, retaining only the segmental information (phones-only), and (iv) a resynthesis using the original phoneme durations and amplitude contour, but synthesized with different phones, retaining only the

[2] Our use of one cover term does not imply that these words show exactly the same grammatical behavior in all languages; as is well known from linguistic typology, languages may differ in the details of how they implement similar types of words. In terms of Haspelmath's (2010) useful distinction, we use 'ideophone' as a comparative concept, recording the descriptive categories in Table 2.

prosodic information (prosody-only). As a yardstick for the sound-symbolic potential of these words, we tested how well native speakers of Dutch unfamiliar with any of the languages could guess the meanings of these words from two alternatives.

A key aspect of this kind of two-alternative forced-choice task is the question of how to select the foil. For a given ideophone, the foils were translations of other ideophones from the same language and from the same semantic category. For example, for the Ewe ideophone *falefale* 'thin, as paper', we selected foils from the other words within the Ewe 'shape' category such as *legbee* 'long form' or *goroo* 'spherical'. This improves the ecological validity of the forced-choice task by presenting participants with a choice between two nonantonymic alternatives encoded in the language, rather than between perfectly antonymic alternatives. The same items served as target and foils in different trials. This minimizes the possible influence of particular meanings or translations on the results.

The following comparisons are critical. A first question is whether ideophones show sound-symbolic potential at all; this would be shown if naive listeners are able to choose the correct meaning more often than chance would permit. A second question is whether all ideophones are sound-symbolic to the same extent; we test this by comparing performances across different semantic domains. One prediction is that associations within modality (e.g. sound-sound) are more transparent than associations across modalities (e.g. sound-motion). Third, to show that the resynthesis method can capture relevant phonetic and prosodic detail, the full resynthesis should lead to performance similar to that with the original recordings. Fourth, if this is the case, the phones-only and prosody-only stimuli will reveal the extent to which segmental and prosodic information contribute to sound-symbolic interpretations. If (as is often claimed) sound symbolism can be attributed mainly to segmental properties, the phones-only version should lead to relatively better performance than the prosody-only version. A final question is whether the success rate with stimuli taken from actual languages is comparable to success rates reported in earlier studies using pseudowords; this will provide a way to judge the relation between the present findings and prior studies, and will allow a more realistic estimate of the effect size of real-world sound symbolism.

2. METHOD.

2.1. PARTICIPANTS. A total of eighty-five participants from the Max Planck Institute's participant pool participated in the study. They were native speakers of Dutch sampled from the student population in Nijmegen, The Netherlands. The data from five participants who were tested was not included in the analysis. Three were excluded because they indicated in a postexperiment questionnaire that they had followed some classes on Japanese, and two were excluded because they did not perform the task properly, since their reaction times indicated that they responded before reading the translations. The remaining eighty participants had no knowledge of any of the languages used in this study.

2.2. MATERIALS. We strove to select eight ideophones from each language for each semantic category. Due to constraints of the ideophone inventories of the languages, the final number for each cell of the design varied between five and eleven; the total number of ideophones used in the experiment is 203 (see Table 3). Visual appearance included examples of color terms and meanings such as 'transparent' or 'hanging in clusters'. Motion terms included motion of humans and animals (e.g. 'walking hurriedly') as well as motion of inanimate objects (e.g. 'rolling ball'). Shape terms included generic terms (such as 'round' or 'spherical') as well as description of specific forms (such as ‘skinny’ or ‘fat’). Sound terms included human and animal sounds (e.g. ‘dog barking’) as well as environmental sounds (e.g. ‘sound of a running creek’). Texture terms included terms of surface structure (e.g. ‘bumpy’) and terms of consistency (e.g. ‘brittle’). A complete list of all ideophones used and their different versions can be found in the online supplementary materials.[3] Ideophones were selected by linguists on the basis of native speaker input or with the help of linguistically trained native speakers, who were asked to think of typical ideophones in their language for these categories. For each of the 203 ideophones, a Dutch translation was prepared with the help of a native speaker of the language or a linguist studying the language.

LANGUAGE	CATEGORY					TOTAL
	VISUAL APPEARANCE	MOTION	SHAPE	SOUND	TEXTURE	
Japanese	5	8	9	11	10	43
Korean	9	10	8	10	9	46
Semai	6	8	8	9	7	38
Siwu	7	6	10	5	6	34
Ewe	7	11	8	8	8	42
TOTAL	34	43	43	43	40	203

TABLE 3. Number of ideophones per language and category used in the experiment.

As noted above, four different versions of each ideophone were prepared using the original recordings and speech resynthesis: (i) the original recording, (ii) a full resynthesis, (iii) a phones-only resynthesis retaining only segmental properties, and (iv) a prosody-only resynthesis retaining the prosodic properties of the original.

For the ORIGINAL RECORDINGS, the selected ideophones were either taken directly from field recordings (for Semai and Siwu) or recorded by a native speaker (for Ewe, Japanese, and Korean) in a sound-attenuated booth. The field recordings were treated for noise reduction using Adobe Soundbooth. Using the original recordings as a basis, we generated resynthesized versions with a diphone synthesizer (MBROLA; Dutoit et al. 1996) using the voice database NL2, from a female speaker. The diphone synthesizer takes as input a text file with phonetic labels, their durations, and the pitch contour. To generate these files, all recordings were segmented on a phone-by-phone basis and the segment intervals were labeled with the closest Dutch phonemes.

The resulting labels and their respective durations formed the basis for the FULL RESYNTHESIS together with the original pitch contour (which, where necessary, was transposed to the pitch range of a typical female speaker, c. 200 Hz). The full resynthesis file was then modified to generate two further versions of the stimuli. A PHONES-ONLY version substituted the observed durations with typical phoneme durations as observed in Dutch (Klabbers & Van Santen 2000) along with a flat intonation contour. A PROSODY-ONLY version was generated by using the original phoneme durations and pitch contour but substituting phonemes, in an attempt to break any potential nonarbitrary links between segments and meanings. For vowels, the substitute phones were selected by rotating the vowel triangle 180° counterclockwise (/i/ → /a/, /a/ → /u/, /u/ → /i/). Consonant phones

[3] The online supplementary materials can be accessed at http://muse.jhu.edu/article/619540. The materials consist of the following. S1 is a list of the ideophones, along with their source languages, semantic domains, and translations. S2 is a zip archive containing the auditory stimuli used in the experiment (also available separately at http://www.holgermitterer.eu/ideoPhonesSamples.html). S3 is a zip archive containing the MBROLA .pho files and rules used in creating the synthesized versions of the stimuli.

were substituted first by place of articulation, then by manner of articulation, then by voicing, according to the following mapping: place: alveolar → labial, labial → velar, velar → alveolar; manner: stop → nasal, fricative → liquid, nasal → glide, liquid → stop, glide → fricative, affricate → glide; voice: voiced ↔ voiceless (Table 4 exemplifies the substitution rules for Japanese). If application of manner of articulation would lead to substitution with a phoneme not present in the phonological system of a given language, then the closest existing phone was chosen. If application of voicing would generate an illicit phone, voicing changes were not applied. Thus, Japanese *gorogoro* 'rolling' became *nebenebe*, Korean *cholchol* 'sound of a small creek' became *wepwep*, Semai *grahɒːc* 'rough, uneven' became *pnɤneːm*, Siwu *pumbuluː* 'fat belly' became *ninipiː*, and Ewe *tyatya* 'walk rapidly' became *wuwu*.

The output of the synthesizer was given the amplitude contour of the original utterance using the intensity manipulation option in Praat (Boersma & Weenink 2015). To achieve this, the signal was first multiplied by the inverse of its own amplitude contour (to achieve a flat intensity contour) and then overlaid with the contour of the original utterance. To make sure every item sounded like a possible word, small adjustments were made for segment-specific intensity contours (e.g. the closure-burst intensity contour of a stop was not overlaid on a nasal) and phone durations (e.g. adding 10 ms to the duration of a converted /p/ in order to avoid it being perceived as /b/).

PLACE OF ARTICULATION		MANNER OF ARTICULATION		VOICING	
FROM	TO	FROM	TO	FROM	TO
alveolar	labial	stop	nasal	voiced	voiceless
labial	velar	fricative	liquid	voiceless	voiced
velar	alveolar	nasal	glide		
		liquid	stop		
		glide	fricative		
		affricate	glide		

TABLE 4. Phoneme substitution rules for Japanese.

2.3. PROCEDURE. Participants sat in a sound-attenuated booth in front of a computer screen and a two-button response box. Written instructions explained the procedure by stating that the participant would hear words from five different languages, which were sound-symbolic for the speakers of those languages in the sense that their form suggested something about their meaning. Participants were asked to guess the correct translation from the two options that would be presented on the right and left sides of the screen by pressing the right or left button of the response box.

In each trial, participants heard a stimulus word, and 1350 ms later, the two translations appeared on the screen. After one second the stimulus was repeated. After a button had been pressed, the chosen translation was moved slightly higher on the screen while the other translation disappeared, indicating that the response had been registered.

Each participant heard only one version of each ideophone, for a total of 203 trials. Across participants, a randomization procedure counterbalanced the number of times each version of a given ideophone was presented as a target and a foil. Clusters of four participants saw the same two response options on the screen and heard the same target words, but each participant heard a different audio version (original, resynthesized, phones-only, prosody-only). The next four participants saw and heard the same targets but now combined with a different foil, and so forth, until all possible within-category combinations were run through. Each participant heard approximately one quarter of the words in each audio condition. For each trial, the position of the target (right or left side of the screen) was allocated randomly. Randomization was done offline, and eighty different lists were prepared for the experiment.

2.4. Design and analysis. The independent variables were Semantic category (five levels: Color/Visual appearance, Motion, Shape, Sound, and Texture) and Audio version (four levels: Original, Full resynthesis, Phones-only, and Prosody-only).[4] The dependent variable was the proportion of correct responses. For statistical analyses, we used linear mixed-effects models with a logistic linking function, using the lme4 package (Bates et al. 2011) in R (R Core Team 2011). In addition to the two independent variables implemented as fixed effects, including an interaction term, participant and item were used as random effects. Unless otherwise noted, the models made use of a maximal random-effect structure; that is, all possible random slopes of fixed effects over participants (Semantic category and Audio version) and items (Audio version only, since Semantic category is between items) were included in the models (Barr et al. 2013).

The use of linear mixed-effects models with a logistic linking function has two advantages for this experimental design. First, we deal with the categorical nature of the dependent variable (using ANOVAs with untransformed proportions would be problematic; see Dixon 2008, Jaeger 2008). Second, the predicted values are logOdds. A value of zero in logOdds is equivalent to 50% correct, that is, chance performance in our two-alternative forced-choice task. A positive and significant intercept value can therefore be interpreted as above-chance performance.

As our factors had more than two levels, we used model comparisons to evaluate the overall significance of each factor.[5] A model with both factors and their interaction included was compared to a model that differed only in the absence of one factor (starting with the elimination of the interaction term). All random slopes of the fixed factors were retained in the simpler model. The two models were then compared with a log-likelihood test. This test is similar to the test of a main effect in an ANOVA. A significant effect in the model comparison indicates that the more complex model (i.e. the one with one more factor included) allows a significantly better prediction of the data.

3. Results. Figure 1 shows the mean proportion of correct responses across the different categories and audio versions. A first finding is that the overall success rate is above chance for most ideophones in most conditions—though in some cases only barely so (Fig. 1, top panel). Only ideophones referring to sound were recognized with greater than 65% accuracy, and only when presented with segmental and prosodic information. Moreover, while the full resynthesis led to similar performance as the original recordings (c. 57% correct on average, c. 55% excluding Sound), the stimuli with no particular prosodic information (phones-only) or with different segments (prosody-only) led to a clear cost in accuracy (c. 53% correct on average, 4% less than for the full resynthesis).

[4] Language was not an independent variable since we had no hypothesis connected to source language. However, we tested whether effects differed between languages to exclude the possibility that the languages for which original recordings were made under field conditions differed from the other languages. This was not the case ($\chi^2(4) = 3.36$, $p > 0.1$).

[5] The summary of a linear mixed-effects model contains regression weights and their respective significance levels. The significance (or absence thereof) of regression weights does not necessarily mean that a factor with more than two levels as a whole is significant. A significant overall effect of a factor is possible even if no single regression weight is significant, and the presence of a significant regression weight does not mean the factor overall has to be significant.

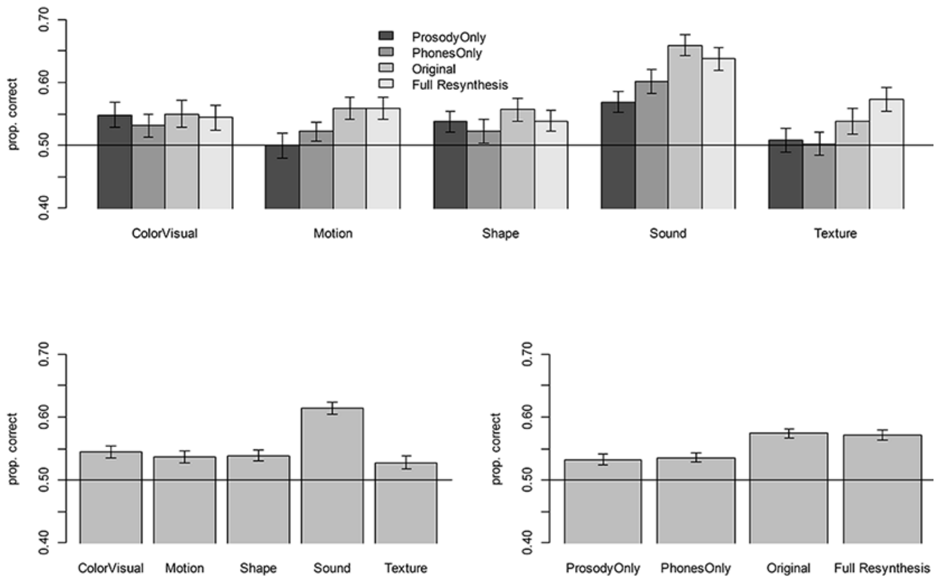

FIGURE 1. Mean proportions of correct responses in each combination of Semantic category and Audio version (upper panel); aggregated over Semantic category (lower left panel), and aggregated over Audio version (lower right panel). Error bars represent subject-based standard errors, and the horizontal line reflects chance performance at 50% correct responses.

We first tested whether the interaction of Semantic category and Audio version influenced the proportion of correct responses, that is, whether differences between audio versions were comparable across the different levels of Semantic category. A comparison of the models with and without the interaction between Semantic category and Audio version was not significant ($\chi^2(12) = 15.1$, $p > 0.1$),[6] suggesting a better fit of the simpler model without the interaction. As the two factors appeared to be independent, we analyzed the main effects of Semantic category and Audio version separately.

3.1. SEMANTIC CATEGORY. For the independent variable Semantic category (Fig. 1, bottom left), a model comparison with and without this fixed effect revealed a significant difference ($\chi^2(4) = 12.9$, $p = 0.011$). This shows that performance differed significantly across the different Semantic categories. Table 5 shows the regression weights for the linear mixed-effects model with Semantic category as a fixed effect. In this model, the level Color/Visual appearance is mapped on the intercept and a regression weight is calculated for the difference between the intercept level and all other levels of this factor. The analysis revealed a significant intercept, which shows that performance for items in the category Color/Visual appearance was significantly above chance. Comparing the other categories to this intercept, only the regression weight for Sound was significant, indicating that ideophones in this category lead to significantly better performance than ideophones in the Visual appearance category, as predicted. To test whether the performance for each semantic category by itself was above chance, we additionally ran intercept-only models with the data from each one of the levels of the factor Semantic category. All analyses showed significant intercepts (Shape: $z = 0.21$, $p = 0.03$;

[6] Unlike the other models below, this model comparison was based on an anticonservative random-intercept-only structure for the random effects, since a model with a maximal random-effect structure did not converge.

Color/Visual appearance: $z = 0.19$, $p = 0.03$; Motion: $z = 0.27$, $p = 0.01$; Sound: $z = 0.66$, $p = 0.001$; Texture: $z = 0.22$, $p = 0.01$), showing above-chance performance for ideophones in each of the categories.

	INTERCEPT	MOTION	SHAPE	SOUND	TEXTURE
β	0.19	−0.03	−0.02	0.31	−0.07
z	2.04*	−0.23	−0.18	2.53*	−0.54

TABLE 5. Regression weights for the analysis of the fixed effect Semantic category, dummy coded with Color/Visual appearance mapped on the intercept. Note: * = $p < 0.05$.

3.2. AUDIO VERSION. For the independent variable Audio version (Fig. 1, bottom right), the analysis focused on cross-modal ideophones (i.e. ideophones in the domains of Visual appearance, Motion, Shape, and Texture) in order to achieve comparability with previous studies, which also looked at associations across modalities. A model comparison indicated a better fit for the model that included a main effect of Audio version than a model with only an intercept term ($\chi^2(3) = 9.4$, $p = 0.024$). To further explore this effect, we devised three linearly independent contrasts (Table 6). The first contrast compares the two full versions (original and full resynthesis) with the two reduced versions (phones-only and prosody-only). The two remaining contrasts are comparisons within the full and reduced versions (original vs. full resynthesis and phones-only vs. prosody-only). Table 6 shows the contrast coding and the outcome of the analyses. The intercept represents the overall mean with above-chance performance, since the three contrasts all add up to zero. The regression weights show that there was a clear difference between the full and the reduced stimuli (contrast 1) but no differences for the other contrasts, neither between the original and the full resynthesis (contrast 2) nor between the prosody-only and the phones-only stimuli (contrast 3).

To test whether performance for the two types of reduced stimuli was still significantly above chance (observed means: 52.2% for prosody-only and 51.8% for phones-only), a linear mixed-effects model with only an intercept was run using the data from just these two conditions. The intercept was not different from zero ($b_{PhonesOnly} = 0.08$, $z = 1.41$, $p = 0.15$; $b_{ProsodyOnly} = 0.10$, $z = 1.70$, $p = 0.09$), showing that participants did not perform significantly above chance for cross-modal ideophones in the prosody-only and phones-only conditions. That is, both segmental and suprasegmental properties appear to contribute to sound symbolism, but neither of them alone is sufficient to trigger above-chance performance across all domains. Of interest is the fact that ideophones in the Color/Visual appearance and Shape domains did show above-chance performance at c. 55% in the prosody-only condition (Fig. 1). This demonstrates the potential of prosody to convey (disambiguating) information in these semantic domains. To rule out the possibility that the poor recognition in the other domains was simply due to the unintelligibility of the resynthesis, we tested whether performance was above chance for the Sound ideophones, which was the case ($b_{PhonesOnly} = 0.46$, $z = 3.3$, $p < 0.001$; $b_{ProsodyOnly} = 0.29$, $z = 2.04$, $p = 0.04$). This shows above-chance performance for ideophones imitative of sound even in the prosody-only and phones-only versions.

4. DISCUSSION. Our findings show that people can correctly guess aspects of the meaning of ideophones from languages they do not speak at an above-chance rate, clearly demonstrating the existence of sound symbolism in the ideophone systems of natural languages and affirming the viability of the methodological approach taken here. At the same time, the results are more differentiated and attenuated than much prior work has suggested: the overall effect is considerably weaker than in most studies using contrived stimuli, performance is not uniform across semantic domains, and the

REGRESSION WEIGHT	CONTRAST WEIGHTS PER LEVEL					
	ORIGINAL UTTERANCE	FULL RESYNTHESIS	PHONES ONLY	PROSODY ONLY	β	*z*
intercept	—	—	—	—	0.16	5.82**
contrast1	0.5	0.5	−0.5	−0.5	−0.14	3.76**
contrast2	1	−1	0	0	−0.01	0.33
contrast3	0	0	1	−1	0.02	0.21

TABLE 6. Regression weights for the analysis of the fixed effect Audio version, coded as three linearly independent contrasts. Note: ** = $p < 0.001$.

results owe as much to prosodic implementation as to segmental information. We discuss the implications of these points in turn.

4.1. REAL-WORLD SOUND SYMBOLISM. We have shown that naive listeners can glean some information from ideophones that allows them to choose the correct meaning from between two alternatives more often than chance would permit. This is an experimental demonstration that ideophones across the five languages have iconic properties, as suggested in the descriptive literature. However, while the overall success rate is statistically distinguishable from chance performance, it is lower than many previous experiments that used pseudowords, and also lower than certain claims about the iconicity of ideophones would lead one to believe. Our findings disprove the strong iconicity assumption: the idea that ideophone forms are direct phonetic representations of meaning (Tedlock 1999). Instead, they support a more moderate view of ideophones as words that combine a significant degree of arbitrariness with WEAK ICONICITY (Lyons 1977): form is not enough to fully predict meaning, but given both, we can see iconic correspondences between them by means of perceptual analogies. Weak iconicity in this sense is common in language use. For instance, the iconic gestures that often accompany speech show a similar type of weak iconicity: they iconically represent aspects of the meaning conveyed by the verbal material, but their interpretation is scaffolded to an important degree by context and common ground. It has been argued that the interpretation of ideophones is likewise contextually dependent (Samarin 1967, Childs 1994), something rendered likely by our finding that the interpretation of ideophones in isolation is relatively difficult for naive participants. So the meanings of ideophones, even if reinforced by iconic properties and supported by prosody, are conventionalized and enriched by context just like other words.

Our study shows that when we use form and meaning contrasts found in natural languages, sound-symbolic effects are attenuated, which suggests a more nuanced view of the role of sound symbolism in language evolution and language learning. The far-reaching claim that sound symbolism provides a 'vital clue for understanding the origins of proto-language' (Ramachandran & Hubbard 2001:19) was based on a *bouba*/*kiki* task with a 95% success rate. In contrast, we find a much lower success rate and differentiated performance across semantic domains: as predicted, sound-sound associations are easiest to guess, and associations that cross modalities are significantly harder (though still above chance). This indicates that iconicity may be helpful in some semantic domains, but limited in others, and points to constraints on the possible uses of sound symbolism in languages today and by extension in a putative vocal protolanguage.

Recent work reviewing iconic patterns in spoken and signed languages proposes to elevate iconicity to a 'general property' of language, on a par with the 'design feature' of arbitrariness (Perniss et al. 2010; see also Hockett 1960). While a reevaluation of the assumption of arbitrariness is important, Hockett's framing may detract from the reality of languages as complex systems that are not designed but evolved—systems that form not static sets of symbols but dynamic resources for making meaning. Iconicity may be profitably characterized as an AFFORDANCE of language in interaction (Gibson 1977), shifting the perspective from the properties of the language system to the potential for action inherent in it. This also points to a fruitful avenue of research on how the iconic affordances of different semiotic modalities shape and constrain the role of iconicity in the lexicons of spoken versus signed languages (Dingemanse et al. 2015).

4.2. THE ROLES OF SEGMENTS AND PROSODY IN SOUND SYMBOLISM. A key finding of our study is that both segments and prosody contribute to the effect of sound symbolism, and that neither alone is sufficient to drive the effect. This adds to the body of converging evidence suggesting that iconic associations between form and meaning are not to be sought in single phonemes and their supposed meanings, but in structural correspondences that recur across words, and that involve both segmental and suprasegmental information (Nuckolls 1996, Tufvesson 2011, Emmorey 2014). The finding that prosody is as important as segmental information is in line with research showing the role of prosody in guiding iconic interpretations in word learning (e.g. Kunihira 1971, Nygaard, Cook, & Namy 2009) and in spontaneous vocalizations (Shintel et al. 2006). One way in which prosody can help is by disambiguating potential associations. So phonemes may not be directly associated with a particular meaning, but the prosodic implementation of a phoneme may bias listeners toward one interpretation over another. Stop consonants, for instance, may be interpreted quite differently when produced with a long closure and a strong release burst than when produced with a short closure and a relatively weak release burst.

That segments and prosody both provide independent contributions to iconicity is a reflection of how ideophones are used in natural language. Here, there are even more cues, because ideophones are not spoken like ordinary words but delivered as performances, with prosodic foregrounding setting them apart from other speech and iconic gestures adding to the performance (Kunene 2001, Dingemanse 2012). The semiotic package as a whole is what contributes to the iconic interpretation of the words. On this view, trying to pinpoint the iconic meaning of single segments is like taking single dots from a pointillist painting and asking what they represent. The performative nature of ideophones may help explain native speakers' reports that they feel these words are strongly iconic (Nuckolls 1996, Kita 1997). Essentially, speakers have a lifetime of experience with performed ideophones in which various factors—segments, prosody, gesture—help to build a representation of the words that emphasizes their iconic potential. Cross-modal associations acquired over the course of experience (Mitterer & Jesse 2010) may account for intuitions about iconicity that could turn out to be language-specific, in addition to the universal iconicity we have examined here.

Our finding that prosody is important raises concerns regarding the interpretability of previous studies of sound symbolism, which usually attempt to attribute effects to segmental properties alone yet do not mention how the stimuli were recorded and what their suprasegmental properties were. Variations in the prosodic implementation of stimulus presentation may partly explain the striking differences in success rates between studies, ranging from 95% (Ramachandran & Hubbard 2001) to only 55% (Monaghan et al. 2012). Monaghan and colleagues (2012) deal with the matter of suprasegmental properties in an exemplary way. They not only report that the speaker was unaware of the purpose of the recordings, but also present two pages worth of phonetic measurements, showing that suprasegmental properties were well controlled. Yet the list of possible properties to control for is enormous; for instance, there may have still have been relevant differences in voice quality (jitter, shimmer, breathiness) or other subtle phonetic features that are hard to control for in a recording booth or in data from speech corpora. Here, our stimulus generation method may provide a useful complementary method.

4.3. Methodological contributions and future questions. An important finding of our study is that the results of the resynthesized stimuli are not statistically distinguishable from the original recordings. This means that speech resynthesis is a viable method for controlled stimulus generation. The tools used in our study—MBROLA (Dutoit et al. 1996) for diphone synthesis and Praat (Boersma & Weenink 2015) for post-processing—are easy to use, scriptable, and freely available, making speech resynthesis a promising way to achieve systematic control over suprasegmental properties in experimental studies of sound symbolism. Here, we have used it to approach the ecological validity of natural recordings while creating carefully controlled alternate versions. Many other uses are possible, for instance, generating series of forms that systematically differ in articulatory features in order to test sound-symbolic oppositions found in natural languages or proposed in the literature.

Other contributions lie in the experimental design of our study. While we use a two-alternative forced-choice paradigm like many prior studies, two important changes bring our design closer to ecological validity and stack the deck against inflated effect sizes: (i) we use only form-meaning pairings attested in natural languages, and (ii) the choice in the task is not between perfect antonyms but between lexicalized meanings from the same semantic domain. Whereas prior studies may have overestimated the sound-symbolic potential of speech due to both the nature of the stimuli and the nature of the task, the design of our study renders it likely that the effect we found, however modest, is a genuine effect that represents the degree to which segments and prosody together can serve as cues to meaning.

Our results suggest a number of directions for future work. We have tested participants with a linguistic background in which sound symbolism appears to be less prevalent, at least as a coherent lexical class (Nuckolls 2004). Are speakers of languages with many ideophones more prone to make sound-symbolic associations? Which iconic associations found in ideophones are language-specific, and which transcend languages? These questions could be addressed by running similar experiments in a wider range of languages, and by investigating the iconic associations of particular (sets of) forms and meanings in more detail. Crosslinguistic comparative studies may also bring to light other pervasive iconic associations and oppositions beyond the well-worn *bouba*/*kiki*-type, providing novel cross-modal associations to probe experimentally.

We have shown that the success rate (at least in a forced-choice task) is significant yet modest, suggesting that iconic associations may be better thought of as cues influencing interpretations rather than directly encoding meaning (cf. Nielsen & Rendall 2012 for related evidence from a learning task). Iterated learning offers one way to study how such biases may shape and constrain linguistic items over the course of cultural evolution (Tamariz & Kirby 2015). Further questions relate to mechanisms. Iconicity can be thought of as grounded in perceptual analogies by means of structure-mapping (Gentner 1983, Tufvesson 2011, Emmorey 2014). Which structures are mapped? Is it just the spectral structure of speech, or do articulatory gestures also play a role? This could be investigated with an experimental manipulation in which participants do not just hear the word (as in the current study) but also pronounce it, which may boost sensitivity to iconic associations (Oda 2000). Finally, further research is needed with regard to the possible functions of sound symbolism in language learning and communication (Perniss & Vigliocco 2014).

5. CONCLUSIONS. There has long been a disconnect between experimental work on sound symbolism and linguistic work on ideophones and like phenomena. Our study is among the first to bridge this gap, showing that (i) naive listeners can correctly guess the meanings of ideophones from natural languages with a success rate above chance; (ii) effect sizes are more modest than assumed, suggesting that experimental findings based on pseudowords cannot be automatically translated into claims about natural languages; (iii) prosodic implementation is just as important as segmental information in supporting iconic interpretations; and (iv) speech synthesis offers a viable way to achieve experimental control in the study of sound symbolism.

Our results call for a stronger empirical grounding of research on the role of sound symbolism in natural language. Stimulus design can benefit from taking into account lexically relevant semantic distinctions, and the suprasegmental aspects of stimulus presentation are a nontrivial factor in sound-symbolic interpretations. Accounts of the role of sound symbolism in language evolution and language learning will gain ecological validity by taking into account the forms and functions of sound symbolism in natural languages. When empirical grasp and experimental control go hand in hand, we will better understand what sound symbolism can and cannot do.

REFERENCES

AKITA, KIMI. 2009. Phonosemantic evidence for the mimetic stratum in the Japanese lexicon. *Berkeley Linguistics Society* 34.1–12. DOI: 10.3765/bls.v34i1.3552.

AMEKA, FELIX K. 2001. Ideophones and the nature of the adjective word class in Ewe. In Voeltz & Kilian-Hatz, 25–48.

AVEYARD, MARK E. 2012. Some consonants sound curvy: Effects of sound symbolism on object recognition. *Memory & Cognition* 40.83–92. DOI: 10.3758/s13421-011-0139-3.

BARR, DALE J.; ROGER LEVY; CHRISTOPH SCHEEPERS; and HARRY J. TILY. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–78. DOI: 10.1016/j.jml.2012.11.001.

BATES, DOUGLAS; MARTIN MAECHLER; and BEN BOLKER. 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375. Online: http://cran.r-project.org/package=lme4.

BOERSMA, PAUL, and DAVID WEENINK. 2015. Praat: Doing phonetics by computer. Online: http://www.praat.org.

BREMNER, ANDREW J.; SERGE CAPAROS; JULES DAVIDOFF; JAN DE FOCKERT; KARINA J. LINNELL; and CHARLES SPENCE. 2013. 'Bouba' and 'Kiki' in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition* 126.165–72. DOI: 10.1016/j.cognition.2012.09.007.

BROWN, ROGER, and RONALD NUTTALL. 1959. Method in phonetic symbolism experiments. *Journal of Abnormal and Social Psychology* 59.441–45. DOI: 10.1037/h0045274.

CHILDS, G. TUCKER. 1994. African ideophones. *Sound symbolism*, ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala, 178–204. Cambridge: Cambridge University Press.

DAVIS, R. 1961. The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology* 52.259–68. DOI: 10.1111/j.2044-8295.1961.tb00788.x.

DIFFLOTH, GÉRARD. 1972. Notes on expressive meaning. *Chicago Linguistic Society* 8.440–47.

DIFFLOTH, GÉRARD. 1976. Expressives in Semai. *Oceanic Linguistics Special Publications* 13.249–64. Online: http://www.jstor.org/stable/20019159.

DINGEMANSE, MARK. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6.654–72. DOI: 10.1002/lnc3.361.

DINGEMANSE, MARK; DAMIÁN E. BLASI; GARY LUPYAN; MORTEN H. CHRISTIANSEN; and PADRAIC MONAGHAN. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences* 19.603–15. DOI: 10.1016/j.tics.2015.07.013.

DIXON, PETER. 2008. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 59.447–56. DOI: 10.1016/j.jml.2007.11.004.

Dutoit, Thierry; Vincent Pagel; Nicolas Pierret; François Bataille; and Olivier Van der Vrecken. 1996. The MBROLA project: Towards a set of high quality speech synthesizers free of use for non-commercial purposes. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 96)*, vol. 3, 1393–96. DOI: 10.1109/ICSLP.1996.607874.

Emmorey, Karen. 2014. Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369:20130301. DOI: 10.1098/rstb.2013.0301.

Friederici, Angela D. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6.78–84. DOI: 10.1016/S1364-6613(00)01839-8.

Gentner, Dedre S. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7.155–70. DOI: 10.1016/S0364-0213(83)80009-3.

Gibson, James J. 1977. The theory of affordances. *Perceiving, acting, and knowing: Toward an ecological psychology*, ed. by Robert Shaw and John Bransford, 67–82. Hillsdale, NJ: Lawrence Erlbaum.

Hamano, Shoko Saito. 1986. *The sound-symbolic system of Japanese (ideophones, onomatopoeia, expressives, iconicity)*. Gainesville: University of Florida dissertation.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86.663–87. DOI: 10.1353/lan.2010.0021.

Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203.89–96. Online: http://www.scientificamerican.com/article/the-origin-of-speech/.

Imai, Mutsumi; Sotaro Kita; Miho Nagumo; and Hiroyuki Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109.54–65. DOI: 10.1016/j.cognition.2008.07.015.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.434–46. DOI: 10.1016/j.jml.2007.11.007.

Johansson, Niklas, and Jordan Zlatev. 2013. Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. *Public Journal of Semiotics* 5.3–20. Online: http://journals.lub.lu.se/ojs/index.php/pjos/article/view/9668.

Kantartzis, Katerina; Mutsumi Imai; and Sotaro Kita. 2011. Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science* 35.575–86. DOI: 10.1111/j.1551-6709.2010.01169.x.

Kita, Sotaro. 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35.379–415. DOI: 10.1515/ling.1997.35.2.379.

Klabbers, Esther, and Jan Van Santen. 2000. Predicting segmental duration for Dutch using sums-of-products approach. *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, vol. 3, 670–73. Online: http://www.isca-speech.org/archive/icslp_2000/i00_3670.html.

Köhler, Wolfgang. 1929. *Gestalt psychology*. New York: Liveright.

Kunene, Daniel P. 2001. Speaking the act: The ideophone as a linguistic rebel. In Voeltz & Kilian-Hatz, 183–91.

Kunihira, Shirou. 1971. Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior* 10.427–29. DOI: 10.1016/S0022-5371(71)80042-7.

Kwon, Nahyun, and Erich R. Round. 2014. Phonaesthemes in morphological theory. *Morphology* 25.1–27. DOI: 10.1007/s11525-014-9250-z.

Lee, Jin-Seong. 1992. *Phonology and sound symbolism of Korean ideophones*. Bloomington: Indiana University dissertation.

Levelt, Willem J. M.; Ardi Roelofs; and Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–38. DOI: 10.1017/S0140525X99001776.

Lockwood, Gwilym, and Mark Dingemanse. 2015. Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology* 6:01246. DOI: 10.3389/fpsyg.2015.01246.

Lockwood, Gwilym; Mark Dingemanse; and Peter Hagoort. 2016. Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. DOI: 10.1037/xlm0000235.

Lyons, John. 1977. *Semantics*, vol. 2. Cambridge: Cambridge University Press.

Martin, Samuel E. 1962. Phonetic symbolism in Korean. *American studies in Altaic linguistics*, ed. by Nicholas Poppe, 177–89. Bloomington: Indiana University Publications.

Maurer, Daphne; Thanujeni Pathman; and Catherine J. Mondloch. 2006. The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science* 9.316–22. DOI: 10.1111/j.1467-7687.2006.00495.x.

Mitterer, Holger, and Alexandra Jesse. 2010. Correlation versus causation in multisensory perception. *Psychonomic Bulletin & Review* 17.329–34. DOI: 10.3758/PBR.17.3.329.

Monaghan, Padraic; Karen Mattock; and Peter Walker. 2012. The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38.1152–64. DOI: 10.1037/a0027747.

Newmeyer, Frederick J. 1992. Iconicity and generative grammar. *Language* 68.756–96. DOI: 10.2307/416852.

Nielsen, Alan K. S., and Drew Rendall. 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* 4.115–25. DOI: 10.1515/langcog-2012-0007.

Nielsen, Alan K. S., and Drew Rendall. 2013. Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 67.153–63. DOI: 10.1037/a0030553.

Nuckolls, Janis B. 1996. *Sounds like life: Sound-symbolic grammar, performance, and cognition in Pastaza Quechua*. New York: Oxford University Press.

Nuckolls, Janis B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28.225–52. DOI: 10.1146/annurev.anthro.28.1.225.

Nuckolls, Janis B. 2004. To be or not to be ideophonically impoverished. *Proceedings of the eleventh annual Symposium about Language and Society—Austin (SALSA)* (Texas linguistic forum 47), 131–42. Online: http://salsa-tx.org/proceedings/2003/nuckolls.pdf.

Nygaard, Lynne C.; Allison E. Cook; and Laura L. Namy. 2009. Sound to meaning correspondences facilitate word learning. *Cognition* 112.181–86. DOI: 10.1016/j.cognition.2009.04.001.

Nygaard, Lynne C.; Debora S. Herold; and Laura L. Namy. 2009. The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science* 33.127–46. DOI: 10.1111/j.1551-6709.2008.01007.x.

Oda, Hiromi. 2000. *An embodied semantic mechanism for mimetic words in Japanese*. Bloomington: Indiana University dissertation.

Ohala, John J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41.1–16. DOI: 10.1159/000261706.

Perniss, Pamela; Robin L. Thompson; and Gabriella Vigliocco. 2010. Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1:227. DOI: 10.3389/fpsyg.2010.00227.

Perniss, Pamela, and Gabriella Vigliocco. 2014. The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369:20130300. DOI: 10.1098/rstb.2013.0300.

R Core Team. 2011. R: A language and environment for statistical computing. Version 2.14. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org/.

Ramachandran, V. S., and Edward M. Hubbard. 2001. Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies* 8.3–34.

Reinisch, Eva; Alexandra Jesse; and Lynne C. Nygaard. 2013. Tone of voice guides word learning in informative referential contexts. *The Quarterly Journal of Experimental Psychology* 66.1227–40. DOI: 10.1080/17470218.2012.736525.

Rosenthal, Robert, and Kermit L. Fode. 1963. The effect of experimenter bias on the performance of the albino rat. *Behavioral Science* 8.183–89. DOI: 10.1002/bs.3830080302.

Samarin, William J. 1967. Determining the meaning of ideophones. *Journal of West African Languages* 4.35–41. Online: http://main.journalofwestafricanlanguages.org/index.php/downloads/download/11-volume0402/63-determining-the-meanings-of-ideophones.

SAMARIN, WILLIAM J. 1970. Inventory and choice in expressive language. *Word* 26.153–69. DOI: 10.1080/00437956.1970.11435590.

SAPIR, EDWARD. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12.225–39. DOI: 10.1037/h0070931.

SHINTEL, HADAS; HOWARD C. NUSBAUM; and ARIKA OKRENT. 2006. Analog acoustic expression in speech communication. *Journal of Memory and Language* 55.167–77. DOI: 10.1016/j.jml.2006.03.002.

SLOBIN, DAN I. 1968. Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology* 10.301–5. DOI: 10.1037/h0026575.

TAMARIZ, MÓNICA, and SIMON KIRBY. 2015. Culture: Copying, compression, and conventionality. *Cognitive Science* 39.171–83. DOI: 10.1111/cogs.12144.

TEDLOCK, DENNIS. 1999. Ideophone. *Journal of Linguistic Anthropology* 9.118–20. DOI: 10.1525/jlin.1999.9.1-2.118.

TUFVESSON, SYLVIA. 2011. Analogy-making in the Semai sensory world. *The Senses and Society* 6.86–95. DOI: 10.2752/174589311X12893982233876.

VIGLIOCCO, GABRIELLA, and SOTARO KITA. 2006. Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes* 21.790–816. DOI: 10.1080/016909600824070.

VOELTZ, F. K. ERHARD, and CHRISTA KILIAN-HATZ (eds.) 2001. *Ideophones*. (Typological studies in language 44.) Amsterdam: John Benjamins.

WESTERMANN, DIEDRICH HERMANN. 1907. *Grammatik der Ewe-Sprache*. Berlin: Dietrich Reimer.

WHITNEY, WILLIAM DWIGHT. 1874. Physei or thesei—natural or conventional? *Transactions of the American Philological Association* 6.95–116.

WOODWORTH, NANCY L. 1991. Sound symbolism in proximal and distal forms. *Linguistics* 29.273–300. DOI: 10.1515/ling.1991.29.2.273.

[mark.dingemanse@mpi.nl]
[will.schuerman@mpi.nl]
[evarei@phonetik.uni-muenchen.de]
[sylvia.tufvesson@mpi.nl]
[holger.mitterer@um.edu.mt]

[Received 1 July 2015;
accepted 19 February 2016]