Title
A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN
1069-7977

Authors
Kawahara, Shigeto
Shinohara, Kazuko

Publication Date
2012

Peer reviewed

A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation

Shigeto Kawahara (kawahara@rci.rutgers.edu)
Linguistics Department,
Rutgers University, New Brunswick NJ 08901, USA

Kazuko Shinohara (k-shino@cc.tuat.ac.jp)
Institute of Engineering, Tokyo University of Agriculture & Technology,
2-24-16 Nakacho, Koganei, Tokyo 184-8588, JAPAN

Abstract

The current project is a case study–and an extension–of the traditional investigation into sound symbolism (Hinton et al., 1994). Several studies have shown that certain sounds evoke images of particular shapes; for example, oral stop consonants are often associated with angular shapes, whereas sonorants (nasals, liquids, and glides) are associated with round shapes (Berlin, 2006; Köhler, 1947). Berlin (2006) attributes these associations to the similarities between abrupt acoustic amplitude modulation of stop consonants and abrupt change of the directions of lines, i.e., abrupt visual changes. In this study, we extend the stop-angular sound symbolic relation to the domain of emotions. Stops not only evoke the images of angular shapes, but are also associated with emotions that involve abrupt onsets. We further show that angular shapes themselves are associated with such emotions. Our three experiments thus establish a tripartite trans-modal symbolic relationship among three domains of cognition (sounds, shapes, and emotions). As an additional general implication, we argue that our experimental results support acoustic, rather than articulatory, bases of sound symbolism.

Keywords: sound symbolism; stop consonants; sonorants; angularity; emotions; abrupt modulations; modularity; synesthesia

Introduction

A prevalent assumption in modern linguistics is the autonomy of semantics (meanings) from sounds. One of the Saussurian principles of languages suggests that there is no inherent connection between meanings and sounds. For example, there is no inherent reason why what we call [k^hæt] is called in such a way. In fact, different languages call that animal by different names. Therefore, the argument goes, the sound-meaning relationship must be arbitrary. Saussure in fact raises this principle of arbitrariness as a first principle of natural languages (de Saussure, 1960).

However, evidence for an opposing view—that there is some inherent connection between sounds and meanings—is also available, and proponents of this view date back to at least Cratylus in Plato (Harris & Taylor, 1989). Especially since the seminal experimental work by Sapir (1929), a substantial body of experimental work shows that sounds are often associated with particular meanings. This relationship, often referred to as sound symbolism, is usually not absolute, and comes with some lexical exceptions. For example, Sapir found that English speakers tend to associate [a] (back and low vowels) with big objects and [i] (high front vowels) with small objects; however, there are lexical items such as *big* that go against this trend. Since Sapir's work, the tendency to associate lower and backer vowels with bigger objects have been found in many other languages (e.g. Shinohara & Kawahara (2012) among many others).

In short, the previous work on sound symbolism in natural languages has established that we can at least identify stochastic tendencies toward some connection between sounds and meanings (despite the fact that some lexical items may not strictly follow such connections) (Hinton et al., 1994).

There is thus little doubt, albeit perhaps in non-categorical ways, that there are some associations between sounds and meanings. Moreover, these associations between sounds and meanings tend to make phonetic sense. For example, [a] is often considered to be larger than [i] in many different languages, and this association arises because [a] has wider opening of the mouth than [i] (Sapir, 1929; Shinohara & Kawahara, 2012). Alternatively, viewed from a (psycho)acoustic perspective, [a] involves lower second resonant frequency (F2) than [i], which would imply a larger resonator (Ohala, 1994). See also Shinohara & Kawahara (2012) for the comparison of these two theories of sound symbolism.

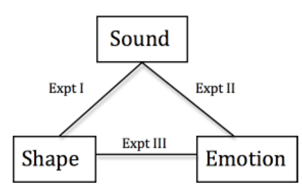

Figure 1: The roadmap of this paper.

In the current project, we focus on the meanings of oral stop consonants (like [p, t, k]), which acoustically involve abrupt amplitude changes, as opposed to sonorants, which involve gradual amplitude changes (nasals, liquids, glides, like [n, l, r, j]) (see section 2 for the illustration of these phonetic concepts.). We show that these acoustic characteristics of stops are associated with angular shapes by native speakers of English (Experiment I). Furthermore, Experiment II shows that the acoustic characteristics of stops are also associated with emotion types with abrupt onsets, as opposed to those that involve gradual onsets. The final experiment goes beyond the traditional sound symbolism studies and shows that there is a direct connection between the emotion types with abrupt onsets and angular shapes. The last two experiments were motivated by a recent finding with Japanese speakers that particular types of emotions can be associated with particular sounds and shapes (Shinohara et al., 2011). Overall, our experiments show a tripartite relationship between three domains of cognition (auditory sounds, visual shapes, and emotion types). The overall conclusion and the roadmap of this paper are illustrated in Figure 1.

Background: Phonetics of stops and sonorants

Oral stops are those sounds that are made with complete oral occlusion (and without the leakage of air through the nasal cavity), which results in rise of the intraoral air pressure (Ohala, 1983). The acoustic consequence of the rise in intraoral air pressure, upon the release of the stop occlusion, is abrupt bursts. Stop consonants thus involve a burst with abrupt amplitude changes after the release of the oral closure. Figure 2 shows a waveform of the 0.05 sec (=50 ms) interval centering around a stop burst of [t]. It represents amplitude changes on the y-axis across time on the x-axis. The transition from a closure phase (oral occlusion) into a burst is rather abrupt, and the burst itself involves abrupt amplitude changes.

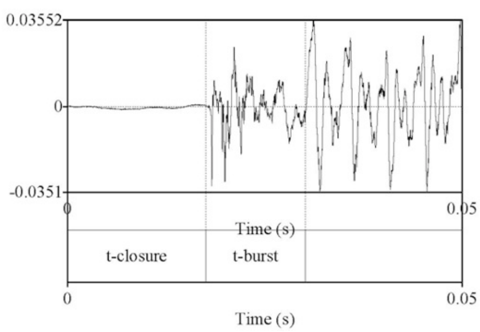

Figure 2: A waveform illustrating a closure phase and a burst of [t]. It shows amplitude changes within a 0.05 sec interval centering around a stop burst. A silent closure is followed by a burst with an abrupt onset.

Sonorants, on the other hand, include a class of sounds consisting of nasals ([m, n]), liquids ([l, r]), and glides ([w, j]) ([j] is the sound that is often represented with "y" in English orthography, as in *young*). In contrast to obstruents, sonorants do not involve rise in the intraoral air pressure because their aperture is wide enough to allow spontaneous vibration of vocal folds (Chomsky & Halle, 1968). Sonorants are thus characterized by energies with gradual changes, and their boundaries with respect to surrounding vowels are gradual. Figure 3 illustrates 0.05 sec intervals of transitions from a vowel to a nasal [n] and to a glide [j]. As observed in the figure, the transitions from the vowel to sonorants are blurry, and the sonorants themselves are characterized by gradual amplitude changes.

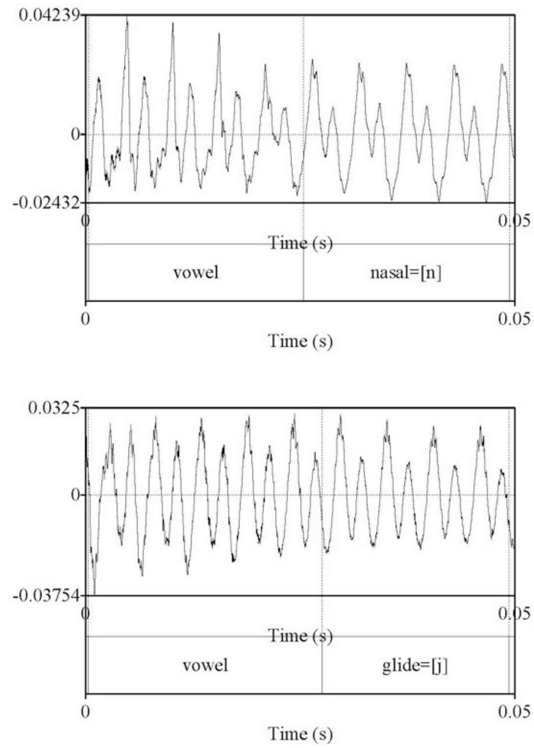

Figure 3: Wave forms of [n] and [j], illustrating 0.05 sec intervals including transitions from the preceding vowels to the sonorants. The boundaries between the vowels and the sonorants are blurry. The consonants themselves involve energies with gradual change.

Experiment I: Sounds-Shapes

As observed in Figures 2 and 3, stops acoustically involve abrupt amplitude changes whereas sonorants involve gradual amplitude changes. Previous experiments have shown that speakers map these acoustic characteristics to the visual domain, considering stops to be more angular than sonorants. Köhler presented two types of figures, one with an angular shape and one with a round shape (see Figure 4), with two sound stimuli, *takete* and *maluma*. The result was that people often associate *takete* with the angular shape and *maluma* with the round shape. Berlin combined this observation and Ohala's theory of sound symbolism (Ohala, 1994) to investigate animal nomenclature patterns (Berlin, 2006). He suggests [p. 34] that "[a]n angular, sharp, long-legged, streamlined bodied rail ought to show a preference for voiceless consonants, especially voiceless stops, while the rounded, short-legged tinamou should not favour these sounds." (see also Ramachandran & Hubbard (2001) and footnote 2 for other sound distinctions that may yield the images of angularity and roundness.) Experiment I replicates these findings by testing whether English speakers map the acoustic characteristics of stops and those of sonorants to a visual domain. For this purpose, we auditorily presented stimuli with stops and those with sonorants together with pairs of angular shapes and round shapes, and asked them to match each stimulus sound with either an angular shape or a round shape.

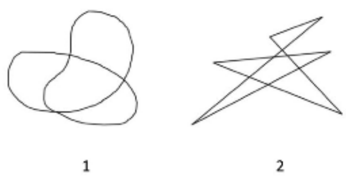

Figure 4: Our reproduction of a pair of shapes used by Köhler (1929/1947). The pair was also used in Experiments I and III.

Method

Experiment I expanded on the previous results (Köhler, 1947; *et seq*.), and tested the connection between stops and angular shapes. The experiment used many stimulus pairs to test the generality of this connection. Furthermore, to avoid the effects of orthography, we used auditory stimuli.

Stimuli The stimuli were all disyllabic CVCV nonce words (i.e. nonexisting words in English). In one condition, both syllables contained stop onsets; in the other condition, both syllables contained sonorant onsets.[1] The vowel quality was controlled between the two conditions: the first vowels were [a, e, ɪ, o, u], and the second vowels were [ə, i] (10 vowel combinations). Two items were included for each vowel combination. The stimulus list is provided in Table 1.

Two native English speakers (one female, one male) pronounced all the stimuli three times in a sound-attenuated booth, at the sampling frequency of 44.1k Hz. To control for potential effects of F0 contour on the listeners' judgments about the images of the stimuli's shapes, the recorded tokens were acoustically resynthesized with a uniform falling contour from the first vowel to the second vowel. For the female speaker, F0 of the first syllable was adjusted to 300 Hz, and F0 of the second vowel to 200 Hz, with linear interpolation in between. For the male speaker, the two F0 parameters used were 150 Hz and 100 Hz, again with linear interpolation. Also, to control for the potential effects of amplitude, peak amplitude of all the stimulus files was modified to 0.7 by using Praat (Boersma & Weenink, 1999-2012). Together with 40 nonce words consisting of stops and 40 nonce words consisting of sonorants, seven different pairs of shapes, each pair consisting of an angular shape and a round shape, were prepared, as exemplified in Figure 5 (the experiment also included the pair of shapes similar to Köhler's, shown in Figure 4). The experiment thus had a total of 560 stimuli (80 auditory stimuli * 7 figure pairs).

[1] The current stimuli do not include fricatives, which also involve frication with abrupt amplitude changes. Testing the visual images associated with fricatives awaits further experimentation.

Table 1: The stimulus list for Experiments I and II.

	Stop condition	Sonorant condition
a-ə	[tagə] [bakə]	[jamə] [ralə]
e-ə	[depə] [tekə]	[wejə] [rewə]
ɪ-ə	[kɪbə] [tɪbə]	[jɪmə] [wɪjə]
o-ə	[dokə] [dopə]	[jorə] [nojə]
u-ə	[dukə] [pukə]	[munə] [mujə]
a-i	[kabi] [tadi]	[maji] [jawi]
e-i	[tegi] [tepi]	[reni] [jewi]
ɪ-i	[tɪpi] [tɪgi]	[jɪni] [nɪwi]
o-i	[boki] [pobi]	[joli] [woji]
u-i	[buki] [gugi]	[wuni] [luri]

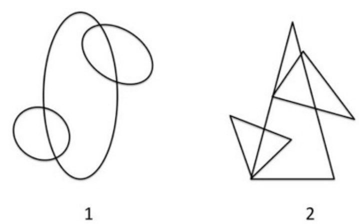

Figure 5: A sample pair of visual cues used in Experiments I and III.

Procedure For each trial, the participants were presented with a pair of objects, one angular and one round, immediately followed by a stimulus sound. They were then asked to choose a shape that better matched each auditory stimulus. The maximum time for the participant to respond to each trial was 3000 ms; if they did not respond within this time limit, that trial was skipped. The inter-trial interval was 250 ms. The visual and audio stimuli were presented using Superlab ver. 4.0 (Cedrus). All the participants wore high quality headphones (Sennheiser HD 280 Pro), and registered their responses using an RB-730 response box (Cedrus).

The experiment started with a practice block in order for the participants to familiarize themselves with the procedure. To avoid loss of attention due to exhaustion, the main session was organized into two blocks. The first block contained the combination of all the auditory stimuli with four pairs of shapes; the second block presented the rest of the three visual pairs. The two blocks were separated by a self-timed break. The order of trials within each block was randomized per each participant by Superlab.

Participants Seventeen native speakers of English participated in the experiment in a sound attenuated room. They were all undergraduate students at Rutgers University, and received extra credit for linguistics classes.

Statistics Since the responses were categorical (angular or not), a logistic linear mixed model regression was run in which the dependent variable was the angular response and the independent variables were the difference between stops and sonorants as a fixed factor and subject as a random factor. All statistical analyses in this paper were performed using R.

Results and discussion

Figure 6 presents the percentages of angular responses for the stop condition and the sonorant condition.[2] As observed in Figure 6, English listeners associated angular shapes much more frequently with the stop stimuli than with the sonorant stimuli. This difference between the two conditions is statistically significant (z = 35.00, $p < .001$). We thus conclude that stops, which involve bursts with abrupt amplitude changes, are associated with angular shapes, and that sonorants, which involve energies with gradual changes, are associated with round shapes.[3]

Experiment II: Sounds-Emotions

Experiment II extended the results of Experiment I to the domain of emotions. We tested whether acoustic abrupt changes of stop consonants are projected on the domain of emotion types. We tested a pair of two emotions like "shocked" and "sad", the former of which involves an abrupt onset; i.e. those types of emotions that start abruptly. The prediction is that, if there is a trans-modal relationship between sounds and emotion types, stops are more likely to be associated with emotions with abrupt onsets (Shinohara et al., 2011).

[2] A signal detection analysis would have been an alternative, which would tease apart sensitivity from bias (Macmillan and Creelman, 2005). We report percent correct analyses throughout this paper for the ease of interpretation.

[3] One remaining question to be addressed in future research is the effect of place of articulation and voicing. Ramachandran & Hubbard (2001) show that [kiki] is considered to be more angular than [bouba]—both of these nonce words contain stops. It seems that, in addition to the differences in vowels, [k] is more angular than [b]. Furthermore, this difference seems to arise because [b] is both labial and voiced. Labial sounds may be associated with round images because they involve movement of lips; voiced stops may be more round than voiceless stops, because their bursts are usually weaker, and voiced stops involve voicing during closure (intervocalically), which consists of gradual amplitude changes (Berlin, 2006; Ohala, 1983). Testing the effect of place of articulation and voicing (and vowels, for that matter) requires future experimentation. See also Jakobson (1978) for discussion on the effect of the acute/grave distinction on the images of sharpness.

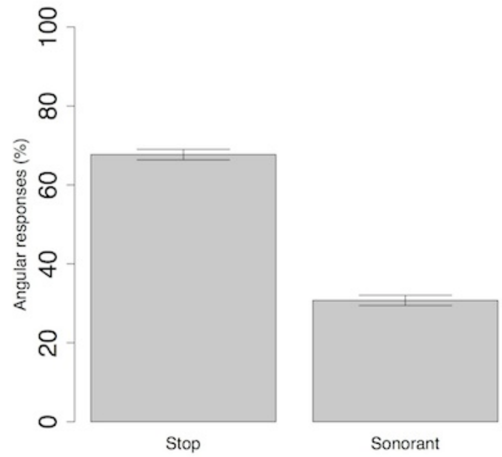

Figure 6: The percentages of angular responses in Experiment I. The error bars are 95% confidence intervals.

Method

The auditory stimuli were identical to those used in Experiment I. For emotion types, we used a pair of negative emotions ("shocked" vs. "sad") and a pair of non-negative emotions ("surprised" vs. "happy").

In this experiment, the participants were told that they would be hearing words of a language that they do not know. After each auditory stimulus was presented, they were presented with one of the two pairs of words in English orthography ("shocked" vs. "sad" or "surprised" vs. "happy"). They were instructed to choose the meaning that better matches the auditory stimuli. Other details of the experiment were identical to Experiment I, except that there was no short break because the experiment was much shorter (80 auditory stimuli * 4 emotions = 320 stimuli). Experiment II was conducted right after Experiment I after a short break with the same participants (seventeen native speakers of English). A logistic regression was run on abrupt responses, with the difference between stops and sonorants and the two types of pairs (negative vs. non-negative) and their interaction as fixed factors and subject as a random factor.

Results and discussion

Figure 7 shows "abrupt responses" ("shocked" for the negative pair and "surprised" for the non-negative pair) for each condition. We again observe that English listeners associated stops with those emotions with abrupt onsets. The difference between stops and sonorants was significant ($z = -7.80$, $p < .001$). The difference between the two types of pairs was also significant (z = 2.13, $p < .05$), because abrupt responses were generally higher for the negative pair of emotions. The interaction term, however, was not significant ($z = -1.53$, *n.s.*), showing the difference between the stop condition and the sonorant condition was consistent between the two pairs. Experiment II thus shows that stops are not only associated with angular shapes, but also with emotion types that involve abrupt onsets. This finding, together with Shinohara et al. (2011), as far as we know, adds a new type of sound symbolism to the sound symbolism literature.

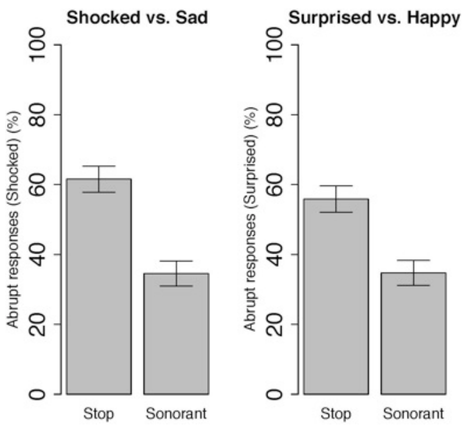

Figure 7: The percentages of abrupt responses in Experiment II.

Experiment III: Shapes-Emotions

The final experiment tested the relationship between emotion types and shapes, the question being whether those emotion types with abrupt onsets are associated with angular shapes. The previous two experiments established that these two are both associated with stops, and the final experiment addressed whether emotions and shapes are directly related. The prediction from the previous two experiments is that English speakers would associate angular shapes with those emotions that involve abrupt onsets.

Method

The stimuli were 16 pairs of angular and round shapes, as exemplified in Figure 5.[4] In this experiment, the participants were instructed to be an assistant of Steven Spielberg, a film-director. They were told that in his new movie, the setting is an extraterrestrial planet where people communicate using visual symbols rather than sounds. The participants were presented with a pair of visual cues, one with a round shape and the other with an angular shape (see Figure 5), and were asked to choose which one better matches a particular meaning ("shocked" or "sad", "surprised" or "happy"). Experiment III was conducted as an online questionnaire survey, as it did not involve auditory stimuli. The experiment was created and distributed using surveymonkey, and the participants were recruited on Psychology on the Net, an online forum for psychology experiments. 37 native speakers of English voluntarily participated in the experiment. No compensation was offered for this experiment.

[4] Since Experiment III tested the combination of only two pairs of emotions (see below), it allowed us to use more pairs of visual stimuli than Experiment II.

Results and discussion

Figure 8 represents the percentages of how often the angular shapes were associated with each emotion in Experiment III.

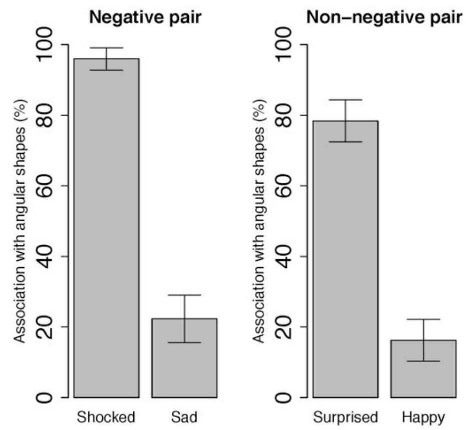

Figure 8: The percentages of how often the angular shapes were associated with each emotion in Experiment III.

We observe that the predictions are borne out: angular shapes were associated more frequently with "shocked" and "surprised" than with "sad" and "happy", i.e. those emotions that involve abrupt onsets. Statistically, the difference between the two types of emotions (those with abrupt onsets vs. those without) was significant ($z = 9.57$, $p < .001$). There was no overall difference between the negative pair ("shocked" vs. "sad") and the non-negative pair ("surprised" vs. "happy") in terms of angular responses ($z = -1.21$, *n.s.*). The interaction was significant ($z = -2.66$, $p < .01$), because the difference between the two emotions was more pronounced in the negative pair than in the non-negative pair. Since the interaction was significant, simple analyses were run separately for the negative pair and non-negative pair. They revealed that the difference within each pair was significant ($z = 9.50$, $p < .001$ for the negative pair, and $z = 9.35$, $p < .001$ for the non-negative pair).

General discussion

To summarize, Experiment I has shown that English speakers associate oral stop consonants with angular shapes, supporting the previous work on this sound symbolic relationship. Experiment II shows that stops are also associated with emotion types that involve abrupt onsets. This sound-emotion connection, to the best of our knowledge, is new and adds a new instance of a sound symbolic relation to the literature. More generally, the results from these two experiments lend further support to the existence of sound symbolism, a general idea that there are particular sound-meaning relationships. Experiment III, going beyond traditional sound symbolic studies, has shown that angular shapes are associated with those types of emotions that involve abrupt onsets. Taken together, our three experiments establish a tripartite trans-modal symbolic relationship between three domains of cognition (auditory sounds, visual shapes, and emotions), among those that involve abrupt modulation, as summarized in Figure 1.

In addition to establishing this tripartite trans-modal relationship among stops, angular shapes, and emotions with abrupt onsets, we suspect that the results of Experiment I have one implication for a general debate about sound symbolism: the debate concerning whether sound symbolism is based on articulation or acoustics. It seems plausible to assume that the image of angular shapes comes from the bursts of stops; i.e. it makes acoustic sense. Acoustically, the stop bursts with abrupt amplitude changes look "spiky" if we track the amplitude changes of stop bursts across time, as we illustrated in Figure 2. By contrast, if we track amplitude changes of sonorants across time, they look "roundish", as in Figure 3. The association between stops and angular shapes and the one between sonorants and round shapes can be considered as projection of the acoustic characteristics of sounds to the visual domain.

On the other hand, an articulation-based explanation of the current results seem difficult, because there is nothing in the articulation of stops that is angular. In fact, the only superlaryngeal articulatory difference between [t] and [n] is opening of velum in [n], and it is not immediately clear why the opening of the velum can be associated with round shapes.[5]

Independent of our speculation about the basis of sound symbolic relationships, our results show that characteristics of sounds can be projected onto the domain of emotions. We have further shown in Experiment III that such a trans-modal relationship directly holds between the domain of visions and emotions. We hope that further research will address how different modalities of cognitions are linked to one another. We suspect that this line of research may go in tandem with general research on synesthesia (Ramachandran & Hubbard, 2001), as our results show that there may be tighter relationships between different modalities of cognitions than previously assumed. Finally, to strength our claim about the trans-modal relationship, it would be desirable to test the claim with more instances of emotions, and with a wider range of experimental paradigms. We leave this task for future research.

[5] It is of course possible that an articulatory basis exists for the stop-angular sound symbolic relationship, which we are simply unaware of. However, a follow-up study of the current study provided further evidence for the acoustic basis of the angularity image. The follow-up study presented English listeners with non-speech sounds like sine waves and square waves. They were told that these sounds were used by an extraterrestrial language, and asked to judge whether sine waves and square waves mean "abrupt" or "gradual". The results show that square waves, which acoustically involve more abrupt change, are indeed judged to mean "abrupt" more often. The result shows that English speakers can map abrupt acoustic change of non-speech to semantic meaning of abruptness, which is presumably the basis of the results of Experiment I. Since the non-speech sounds used in this follow-up experiment are unlike human sounds, whose articulatory origin cannot even be speculated by listeners, the results support the acoustic basis of the images of abruptness.

Acknowledgments

We thank the research assistants at the Rutgers Phonetics Laboratory for their assistance with this project. For comments on earlier draft of this paper, we thank Kimi Akita, Young Ah Do, Shinsuke Nakajima, Seunghun Lee and two anonymous reviewers.

References

Boersma, P. & Weenink, D. (1999–2012). Praat: Doing phonetics by computer. Software.

Berlin, B. (2006). The first congress of ethonozoological nomenclature. *Journal of Royal Anthropological Institution*, 23–44.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

de Saussure, F. (1916/1960). *Cours de linguistique générale[Course in general linguistics]*. Paris: Payot.

Harris, R. & Taylor, T. J. (1989). *Landmark in linguistic thoughts*. London & New York: Routledge.

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

Jakobson, R. (1978). *Six Lectures on Sound and Meaning*. Cambridge : MIT Press.

Köhler, W. (1929/1947). *Gestalt Psychology*. New York: Liveright.

Macmillan, N. & Creelman, D. (2005). *Detection Theory: A User's Guide. 2nd Edition*. Mahwah: Lawrence Erlbaum Associate Publishers.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P., (Ed.), *The Production of Speech* (pp. 189-216). New York: Springer-Verlag.

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

Ramachandran, V. & Hubbard, E. M. (2001). Synesthesia–a window into perception, thought, and language. *Journal of Consciousness Studies*, 8(12):3–34.

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

Shinohara, K. & Kawahara, S. (2012). A cross-linguistic study of sound symbolism: The images of size. In *Proceedings of BLS 36*. Berkeley: BLS

Shinohara, K., Natsume, F., & Matsunaka, Y. (2011). Sound-shape-emotion iconicity in visual psychomimes in Japanese. A talk presented at the Eighth International Symposium on Iconicity in Language and Literature, Linnaeus University, Vaxjo, Sweden.