Sara Marie List April 10, 2014

The sound of shape: Functional and neural correlates of sound to shape mapping in natural language

by

Sara Marie List

Lynne C. Nygaard, PhD

Adviser

Neuroscience and Behavioral Biology Program

Lynne C. Nygaard, PhD

Adviser

Krish Sathian, MD, PhD

Committee Member

Laura Otis, PhD

Committee Member

Marjorie Pak, PhD

Committee Member

2014

The sound of shape:

Functional and neural correlates of sound to shape mapping in natural language

By

Sara Marie List

Lynne C. Nygaard, PhD

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Neuroscience and Behavioral Biology Program

2014

Abstract

The sound of shape:

Functional and neural correlates of sound to shape mapping in natural language

By Sara Marie List

Linguistic theory has historically assumed that the sound structure of words and the physical attributes of word referents are arbitrarily assigned. Recent behavioral, linguistic, and neuroscience studies have challenged this assumption through the demonstration of sound symbolism, or non-arbitrary correspondences between sound and meaning. Research, including preliminary fMRI studies, indicates that sound symbolism may result from cross-modal interactions within the brain.

This study attempts to pinpoint specific sound-to-meaning mappings through analysis of associations between the phonetic and acoustic attributes of systematically varied nonsense words with two opposite physical traits, rounded and pointed. Using these identified sound-to-meaning mappings, a speeded classification task was developed to examine the functional effects of sound symbolic correlations. Certain phonemic traits were consistently rated as either pointed or rounded across tasks. Pairing nonsense words with these traits with congruent and incongruent visual stimuli did not significantly affect time to identify the visual stimulus, although learning may have affected reaction times. This research provides insights into the nature of linguistic reference as well as into the structures that produce and interpret meaning in natural language.

Keywords: sound symbolism, arbitrariness, cross-modal correspondence, grounded cognition

The sound of shape:

Functional and neural correlates of sound to shape mapping in natural language

By

Sara Marie List

Lynne C. Nygaard, PhD

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Neuroscience and Behavioral Biology Program

2014

Acknowledgements

I would like to thank Dr. Lynne Nygaard, Jee Young Kim, and Kelly McCormick for their help and guidance throughout this project. Thanks also to all members of the Ny-Laab, especially and including Speech and Perception Lab Manager Kate Coblin. Thanks to members of Language Group for thoughtful insights into data presentation and design. For their suggestions and feedback, thanks to my honors committee members, Dr. Krish Sathian, Dr. Laura Otis, and Dr. Marjorie Pak.

Funding for this project provided by the NIH funded Computational Neuroscience Training Grant Fellowship and Emory University Psychology Department.

Table of Contents

Introduction……………………………………………………………………………..…1

Project Goal and Hypothesis …………………………………………………..………9

Experiment 1 ……………………………………………………………….…...…10

Method ………………………………………………………...…………………...10

Participants ……………………………………...…………………..…………...10

Materials……………………………..…………..……………………………..…11

Procedure………………………………..……...……………………………..…13

Results & Discussion………………………...……………………………………..…14

Force-choice Task………………………..………………………………15

Likert-type rating Tasks………………….……………………...………………17

Experiment 2 ……………………………………………………………..………...…21

Method ………………………………………………………………...…………...21

Participants ……………………………………...……………………..…………...21

Materials………………………...………………………………………...…………..…21

Procedure………………………...………………………………………..…………..…23

Results & Discussion……………………...…………………………………….…..…24

General Discussion……………………...………………………………...………………..…27

Future Directions & Conclusions………………...……………….……………...………..…35

References...…………………………………………………....……………….……………..…37

Appendix 1: *Nonword Auditory Stimuli for Experiment 1*…...……………….……………..…42

Appendix 2: *Auditory Stimuli for Experiment 2* ……………...…………….…….……………..…50

Figure 1: *Representation of vowels in articulatory space based on height and backness indicating vowels used in nonsense words* ……………………………………………………….……13

Figure 2: *Round and pointy stimuli for Speeded Classification Task*……………………………22

Figure 3: *Mean response time for each group condition in Experiment 2*………………...…….26

Table 1: *Proportion pointed judgments by vowel and consonant class in the forced choice task of Experiment 1*……………………………………………………………………….….16

Table 2: *Mean Likert ratings for the pointed scale (1=not pointed, 7=very pointed) and for the rounded scale (1=not rounded, 7=very rounded) by vowel and consonant class in Experiment 1*…………………………………………………………………………………..18

Table 3: *Mean percent of trials excluded based on response time and standard error*……...….24

Table 4: *Mean proportion correct and standard error as a function of visual stimulus and congruency condition*……………………………………………………………………….25

Introduction

A major tenet on which the characterization of human language rests is that words are arbitrarily assigned to meaning. Scholars as early as John Locke (1715) have insisted that the association between a word and the corresponding meaning is “perfectly arbitrary, not the consequence of a natural connexion [sic],” (p. 393) with the argument that if language were not arbitrary, humans would have developed a single common language in which the sounds of words would necessarily resemble properties of their referents. de Saussure (1959) proposed that language is the link between a conceptual understanding, which is known as the signified, and a sound-image, or the signifier. For example, language links the concept of what a tree is, the signified, with the written or spoken word “tree”, the signifier. An image or concept of the tree can be expressed through communication of that word or sign. These linguistic signs were assumed to bear no connection or resemblance to the signified or referent. There is no necessary characteristic of the signified determining properties of the word to represent that concept.

If words did have a non-arbitrary connection to the concepts conveyed, similar types of words and sound structures would convey equivalent meanings across different languages. However, similar concepts often do not share similar sound structure or even directly translate from one language to another. The Spanish words “conocer” (to know or be familiar with, especially people or skills) and “saber”(to know facts or information) have meaning encompassed by a single infinitive in English “to know”. Because English uses a single linguistic form that encompasses multiple meanings in another language, there must not be a rigid connection between sound structure and meaning. Additionally, in Spanish, though both words broadly translate into the English meaning “to know” the words used to represent this concept are structurally different, indicating that there is not a one to one mapping between a word and its meaning, even within a single language. If there were a necessary connection between words and their referents, these types of differences across languages and within a single language would not arise. This arbitrary property of language allows for a productive system of reference with an infinite number of signs to represent the link between signified and signifier (Gasser, 2004; Monaghan, Christiansen, & Fitneva, 2011; de Saussure, 1959).

Recently, modern linguists have begun to question this assumption of complete arbitrariness (Monaghan & Christiansen, 2006). Examples of potentially non-arbitrary pairings in natural languages include onomatopoeia and Japanese mimetics (e.g. “yoboyobo” meaning to become wobbly legged, usually with age), which contain characteristic sounds that suggest their meaning (Hamano, 1998). In addition, behavioral paradigms, such as the bouba/kiki phenomenon, illustrate that humans as young as three consistently indicate a correspondence between the sounds in a word (e.g., bouba) and the physical shape to which that word is assigned (e.g. bouba = round, blob; Köhler, 1929; Maurer, Pathman, & Mondloch, 2006). These relationships extend beyond sound and shape. Sapir (1929) investigated the relationships between object labels and size of the object using two tables of differing sizes as stimuli. Participants were asked to choose one label, either *mal* or *mil*, for each object. Nonwords were used as stimuli so that the participants could not use any preconceived biases from their knowledge of or practice with language to influence their decisions, and participants associated the nonword *mil* with the small table and *mal* with the larger table. Sapir noted that judgments of the object’s size were based on the vowel used, with /a/ from *mal* acting as a cue for a large object. The sound symbolic associations made with nonwords suggest that sound symbolic properties are not constrained solely by previous language experience.

These sound symbolic associations translate across languages, and this phenomenon is known as cross-linguistic sound symbolism. Cross-linguistic sound symbolism refers to the ability of humans to associate certain speech sounds with certain physical phenomena, regardless of whether these associations are novel, originating from a language in which the person is unfamiliar, or nonsensical, in that the sounds and correlated meanings are contrived. For instance, native speakers of English can more quickly and accurately identify actual Japanese word/English meaning pairs than arbitrarily assigned word-meaning pairs, which suggests that there exist systematic links between semantic meaning and spoken word structure to which learners across languages are sensitive (Nygaard, Cook, & Namy, 2009). In addition, Monaghan and Christiansen (2006), using corpus analyses have found that sound to meaning correspondences in language may be both systematically as well as arbitrarily organized and that these non-arbitrary mappings apply across different languages, including English, Japanese, and French.

Sound symbolic mappings have been found to influence language processing and learning as well. Findings of Nygaard, Cook, and Namy (2009) with Japanese antonym pairs support the possibility that non-arbitrary correspondences between sound and meaning may aid language acquisition. In addition, Kantartzis, Imai, and Kita (2011) studied English-speaking children listening to Japanese mimetic-like stimuli to investigate whether the child's ability to infer word meaning from the sound structure of the words was cross-linguistic in that these sound to meaning mappings held across languages. When the stimuli were sound symbolic, children were more accurately able to infer word meaning, which supports the idea that listeners were cross-linguistically sensitive to and exhibited world learning benefits due to sound symbolism (Imai, Kita, Nagumo, & Okada, 2008). Even proponents of the arbitrariness assumption, such as Gasser (2004), concede that iconic language that violates the principle of arbitrariness may be advantageous in situations when the word set is exceptionally small, such as during acquisition of first words, or when the space of possible distinctions among signified concepts is large.

One possible interpretation of the sound symbolic properties of language is that certain sounds are culturally enforced to encourage an association with a given physical meaning. Traditional theorists posit that language meaning is represented separately from perception, action, and emotion in the brain. Additionally, conceptual content is perhaps encoded into abstract symbols with no relation to memories of the sensorimotor conditions that formed these concepts (Landauer & Dumais, 1997). For example, Bedny & Caramazza (2011), from their examination of the perception of action verbs in an fMRI experiment, suggest that understanding words does not require simulation of sensory-motor processes and that modality-independent neural circuits play a greater role in language comprehension, though perception may contribute at a higher-level processing stage.

Another explanation of the origins of non-arbitrary correspondences in natural language is that neural correlates may connect the processing of sensory information from the physical world to the processing of conceptual information in language. Ramachandran and Hubbard (2001) do not dispute that sound symbolic associations are culturally influenced but posit that there is a neural basis to sound symbolism as well that is based in an extension of sensory-motor synesthesia. According to this theory, sound symbolism could arise from modality-specific representations in the brain, and Barsalou (2008) describes these modality specific representations in the linguistic domain as language and situated simulation, or LASS. These modality-specific representations could arise from associations between any number of an object's physical properties and either aspects of articulation, such as the shape of the lips upon uttering the word, or the sound of the word that represents that object. For example, because the phoneme /o/ is produced with rounded lips, a synesthetic association between the production of that sound and the round shape of a balloon may be neurally represented (Ramachandran & Hubbard 2001).

An expanding body of evidence suggests that words referring to different visual properties of objects are encoded by regions in the brain that either overlap or lie adjacent to regions that are active during the perception of those properties (Kemmerer 2009; Barsalou 2008). This theory of grounded cognition suggests that meaning in human language is grounded, or influenced by, the sensorimotor processes that allow for the perception of said meaning (Pezzulo et al., 2013). If semantics are not amodal, grounded cognition may explain why systematic sound to meaning mappings are found. This theory allows for the possibility that the sound structure of language may have a non-arbitrary property of sound-to-meaning mapping that is cortically represented, rather than being simply culturally influenced. For instance, if semantics in language are partially derived from the experience of physical sensation, perhaps certain sensory experiences, such as seeing a ball, are represented through a mental simulation of a ball in the brain that includes what the ball feels like, looks like, and sounds like when bounced. This mental simulation of the roundness of a ball could then translate into the physical production of sound with rounded lips during expression. In this manner, grounded cognition can account for a systematic organization of sound-symbolic properties through mental simulation of sensory experience (Barsalou, 2008).

Another idea of sound symbolic mechanisms suggests that listeners do not simply mentally simulate to make sound symbolic associations under the framework of grounded cognition, but may go a step further and actually experience the symbolic cues in multiple modalities, such as vision and hearing. Cross-sensory mappings have been studied extensively in individuals with synesthesia, who report experiences such as "seeing" textures or "hearing" numbers, and Ramachandran and Hubbard (2001) have developed a theory of the evolution of language based on studying synesthetes as well as examining evidence from cross-sensory correspondences in other domains, such as pitch and brightness (Marks, 1987; Melara & Marks, 1990a; Melara & Marks, 1990b; Melara & O'Brian, 1987). For example, high-pitched tones are consistently judged to correspond with bright hues, and these associations are referred to as synesthetic, indicating a joining of the senses, because they reflect a presumed connection among attributes from different sensory modalities (Melara & O'Brian, 1987). Ramachandran and Hubbard use these synesthetic correspondences to propose that language arose from sensory-to-motor synesthesia, such that certain lip and tongue movements represented in motor areas may be systematically mapped to certain phonemic and acoustic perceptions mapped in auditory areas of the brain.

In order to investigate the role of cross-modal correspondence in sound symbolism, neuroimaging studies have challenged the assumption of arbitrariness in recent years by examining the neural correlates of cross-linguistic sound symbolism and the possibility that this phenomenon is cross-modally mapped in the brain (Pirog Revill, Namy, Defife, & Nygaard, 2013). For sound symbolic correspondences in particular, fMRI studies by Peiffer-Smadja (2010) have indicated a difference in processing of matching and mismatching shape-pseudoword pairs (bouba/kiki effect) within the lateral occipital cortex. Kovic, Plunkett, and Westermann (2010) found that participants were faster to identify novel objects with sound symbolic rather than arbitrary labels and that this behavioral finding was correlated with early negative EEG-waveforms within 200 ms of object presentation. These findings indicated a potential neural mechanism linking auditory and visual features that supports sound symbolic associations (Kovic et al., 2010). Others have implicated multisensory integration areas in the temporal and parietal lobes, such as the left superior parietal cortex (SPC), the superior temporal cortex (STC), or the superior temporal gyrus (STG), in cross-modal correspondences related to sound symbolism (Arata, Okuda, J., Okuda, H. & Matsuda, 2010; Pirog Revill et al., 2013). These regions may be implicated in functions that allow perceptual attributes of the referent to be correlated with word structure.

Moreover, these synesthetic connections may be reflected within the structure of language in the form of cross-linguistic sound symbolism. The existence of cross-linguistic sound symbolism has been documented extensively in behavioral paradigms, and neural correlates of these findings have been suggested in preliminary neuroimaging studies (Kovic et al., 2010; Pirog Revill et al., 2013). However, few studies have examined what aspects of the sound structure of language evoke associations with particular visual attributes (Nuckolls, 1999). For example, it is unclear, whether it is sensitivity to the overall acoustic form of particular linguistic segments, the manner in which speech sounds are produced (e.g., lip rounding), or a combination of these factors that underlie these associations with meaning (Ramachandran & Hubbard, 2001; Spence 2011). In fact, Nielsen and Rendall (2011) attempted to describe word attributes that contribute to sound symbolic properties using nonsense words and found that both vowels and consonants have an effect on whether or not a nonsense word is non-arbitrarily paired with one visual stimulus or another. In his work examining the sound symbolic associations of nonwords, Sapir (1929) proposed that either certain vowels have greater acoustic volume than others in auditory space or the large spatial relationships between articulatory apparatus during sound production (i.e., tongue position and resonance cavity) is symbolic of a larger reference.

Further studies have attempted to determine whether there may be a processing speed advantage to sound symbolic associations (Parise & Spence, 2009; Parise & Spence, 2012). Parise and Spence (2012) found a stronger coupling of synesthetically matched vs. mismatched stimuli using an implicit association task and provided evidence that sound symbolic congruency can facilitate sensory-perceptual processing. Synesthetic crossmodal correspondences therefore appear to support crosstalk, at least between visual and auditory domains. In a similar finding, Lupyan & Spivey (2010) found that hearing a letter label lowered visual letter detection thresholds, suggesting an influence of auditory labeling on low-level visual processing. Cross-modal interactions in other dimensions, such as pitch and size or pitch and brightness, can also affect response times in speeded classification tasks, indicating that these associations are not limited to language and may further suggest that cross-modal interactions have a neural basis through cross-sensory mappings (Melara & Marks, 1990a; Melara & O’Brian, 1987).

This study attempts to expand on previous work in order to detail the phonetic properties that enhance sound symbolic associations (Nielsen & Rendall, 2011) and to determine whether these associations may yield functional advantage in a speeded classification task (Parise & Spence, 2008). If sound symbolic phonemic properties are found to enhance visual processing of stimuli, perhaps a neural facilitation mechanism based on cross-modal correspondences is engaged during sound symbolic processing that is used to provide information during sensory perception. By identifying the meaningful associations evoked by phonetic, acoustic, or articulatory properties of speech and the functional significance of such mappings, the current study is designed to examine potential sensory aspects of language processing.

Project Goal & Hypothesis

In Experiment 1, in order to determine which speech sounds correspond with particular meanings, an analysis of which phonemes within nonsense words are associated with certain basic visual sensory-perceptual properties was completed. Possible correlations between certain phonemes and judgments of word meaning, rounded or pointed, were analyzed. Speech sounds that contain noise, such as fricatives and affricates (e.g. a nonword pronounced “feesee” or “sehfih”) may be associated with the attribute pointed. This hypothesis is consistent with the findings of Nielsen and Rendall (2011) that obstruent consonants, such as affricates, fricatives, and especially stops (e.g. “tehkay”) were used most often when participants used nonsense words to label jagged object images.

Additionally, if sound symbolism is grounded in synesthetic sensory-motor associations (Ramachandran & Hubbard, 2001), nonsense words containing rounded vowels (e.g. “mohloo” or “kohtoo”) may be more likely to be associated with the meaning rounded due to lip rounding that physically mimics a round shape. Ramachandran and Hubbard (2001) have suggested that cross-modal neural activations account for links between the phonological properties of words and their referents and posit that sound to meaning mappings may arise via auditory sound patterns during vocalization and visual properties of the referent.

Using information about the correspondences between certain speech sounds and visual shapes, in Experiment 2, a behavioral task was developed to determine if these associations have functional significance for language learning and processing. Language users were asked to respond to auditory-visual stimulus pairings in which visual objects (rounded or pointed) were presented with nonsense words that had sound structures which either matched or mismatched visual object shape. This task was designed to determine if the sound to meaning mappings influence visual shape processing as suggested by previous work (Parise & Spence, 2008; Lupyan & Spivey, 2010; Lupyan & Ward, 2013). If cross-modal correspondences underlie sensitivity to sound symbolism, then the sound structure of the nonsense words should influence how quickly participants can identify the visual object shape (Melara & O'Brian, 1987). When nonwords that sound "pointy" are paired with a pointy visual stimulus, judgments of the visual stimulus should be made more quickly than when these same "pointy" nonwords are paired with a round visual stimulus. When the sound symbolic properties of the nonwords match the visual shape properties, then identification should be facilitated. When nonwords mismatch visual object shape, interference between the sound to meaning associations and the visual sensory input is expected.

Experiment 1

Participants' sound-to-meaning assignments were analyzed to determine if there were specific phonetic attributes of the nonsense words that were associated with the probability that a subject would choose one meaning over the other. Certain speech sounds may be associated with the meaning of rounded or pointed based on cross-modal correspondences that are grounded in the manner of production of these sounds. For instance, rounded vowels may be associated with the meaning rounded due to the rounded configuration of the lips during production (e.g. "sohfoh"; Ramachandran & Hubbard, 2001).

Method

Participants. A total of 64 Emory University undergraduates between the ages of 18 and 30 received course credit for their participation. Thirty-four participated in the forced-choice task and 30 in the Likert-type rating tasks. All participants were native English speakers and reported no known hearing, speech, or language disorders.

Materials.

570 nonsense words were constructed using only sounds and combinations of sounds that occur in the English language. Each nonword consisted of two syllables of the form consonant, vowel, consonant, vowel (CVCV). Figure 1 depicts the vowels that were used to create the set of nonwords.

Vowels varied in height and backness. Vowel height refers to the vertical height of the tongue within the lower jaw, which makes the oral cavity more narrow or wide (Reetz & Jongman, 2009). For example, the word “bit”, the word “bait”, and the word “bet” consist of a high vowel, a mid vowel, and a low vowel, respectively. Vowel backness refers to the arching of the tongue during an utterance, which may be toward the front, center, or back of the oral cavity (Reetz & Jongman, 2009). For example, “got” contains a back vowel and “get” contains a front vowel. Roundedness refers to the position of the lips during an utterance (Reetz & Jongman, 2009). For example, /o/ in “loop” is a rounded vowel. For this set of nonwords, roundedness is correlated with backness so that all front vowels were unrounded (e.g "bait") and back vowels were rounded (e.g. “boot”). Each of these vowel sounds can be characterized by all three of these categories. For instance, the nonword “tehkay” contains unrounded, front, mid vowels and the nonword “mohloo” contains a vowel in the first position (V1) that is round, back, and mid and a vowel in the second position (V2) that is rounded, back, and high.

Each nonword also contained consonants varying in manner of articulation (sonorants, affricates/fricatives, or stops) and voicing (either voiced or unvoiced). Voicing refers to whether the vocal cords are engaged during the production of the sound. A voiced consonant includes the sound of the letter /g/ in “gab” versus an unvoiced consonant like /k/ in “kite”. Sonorant consonant segments include sounds that are produced without complete obstruction of the vocal tract. An example of a nonword with sonorant consonants is “looloh”. Stops, affricates, and fricatives are occasionally grouped together because all involve an obstruction of the airflow. Stops refer to the complete block or stop of oral flow of air for a brief moment, such as in the nonword “pehtay”. Fricatives are produced by the narrow passage of air through the place of articulation (e.g. the lips, palate, etc) to create turbulence of air and a hissing sound, such as in the nonword “soosoh”. Affricates involve the combination of a stop and fricative, such as in the nonword “chuhsoh” (Reetz & Jongman, 2009).

Nonword stimuli had no repeated syllables, so nonwords such as “kiki” were excluded. In addition, because three of the vowel sounds recognized by the International Phonetic Alphabet (IPA), /I/ /ɛ/ and /ʊ/, do not occur at the end of words in English, the stimuli only included these vowels in the first vowel position (V1) of a nonsense word but not the second position (V2). Also, different combinations of consonant or vowel type within each nonword were excluded. Thus, nonwords contained only front or back vowels or only sonorants or obstruents, but not a combination of front and back vowels or sonorants and obstruents (See Appendix 1). Any combinations that resulted in a real English word, as rated by four trained listeners, were excluded as well.

Each nonsense word was recorded by a native English speaker with neutral intonation. All nonwords had a mean duration of 450 ms and a standard deviation of approximately 50 ms. Nonsense words were read from a computer screen and recorded in a sound-attenuated room using a Zoom 2 Cardioid microphone and directly digitized onto a Dell computer at a 44.1 sampling rate. Each nonsense word was then segmented into separate files and re-digitized at a 22.050 kHz sampling rate and amplitude normalized using PRAAT (Boersma & Weenink, 2012), a speech analysis software. Each nonword was checked by four raters for accurate pronunciation of the intended segments and mispronounced nonwords were rerecorded, edited, and rechecked by the same four raters.

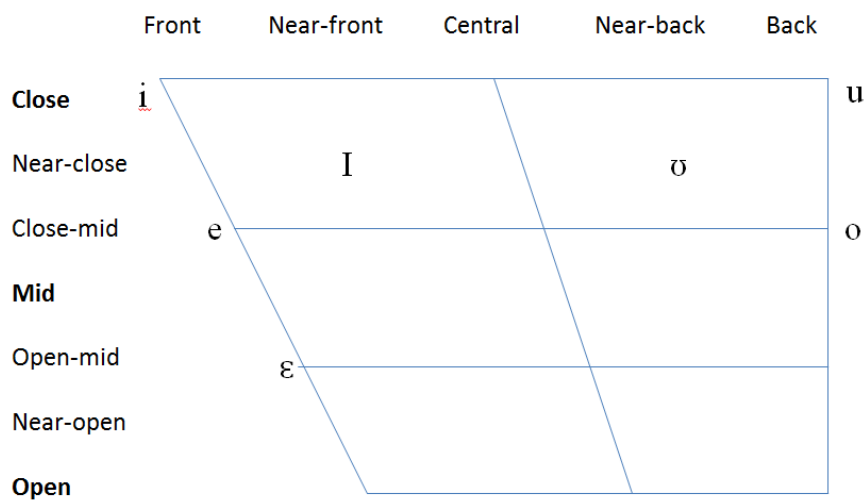

Figure 1. Representation of vowels in articulatory space based on height and backness indicating vowels used in nonsense words. The back vowels were also rounded.

Procedure.

Forced-choice task. The 570 nonsense words were presented through headphones, and participants were asked to assign a label of rounded or pointed to each nonword in a forced-choice task (i.e. choice of either rounded or pointed). Participants were instructed that they would hear a nonsense word and were asked to guess the meaning of the nonword, rounded or pointed. Nonwords were presented auditorily over Beyerdynamic DT100 headphones at approximately 75db SPL. On each trial, the nonsense word was presented auditorily, while the choices of rounded and pointed appeared on the screen. The mapping of "rounded" and "pointed" response options to response box keys (Psychology Software Tools Model 200A serial response box) was counterbalanced across conditions, such that half of the participants (n = 17) chose the right-most button to indicate "rounded" on the right side of the screen and the left-most button to indicate "pointed". The other half of participants (n = 17) were asked to respond with the left button to indicate “rounded” and right for “pointed”. The stimuli were presented in random order.

Likert-type rating tasks. Two Likert-type rating tasks were administered to separate groups of listeners. Participants completed each task in which they were presented the nonwords as described for the forced-choice task but were asked to rate each nonword on a Likert-type scale from 1-7. One set of participants (n = 15) rated each nonword for degree of pointiness with 1 representing not pointed and 7 representing very pointed, and the other group of participants (n = 15) rated each nonword for degree of roundedness with 1 representing not rounded and 7 representing very rounded.

Results and Discussion

Data analysis for the judgment tasks focused on identifying correspondences between specific properties of the systematically varied nonsense words and meaning assignment. First, trials with response times greater than 2 standard deviations from the mean response time for each subject were excluded. The features of the nonsense words were coded and used to predict behavioral data in order to determine if listener judgments were influenced by the specific phonetic information in each nonword. In order to delineate the role that phonetic segments play in the variability of participants’ judgments of rounded or pointed, multiple regression analyses were conducted for each task. These analyses served to assess those phonetic features that co-vary for each of the meaning domains (rounded or pointed). The objective of this analysis was to uncover which phonetic categories are the best predictors of listeners’ judgments of rounded and pointed and the extent to which these associations can be attributed to each phonetic feature or category. The behavioral ratings (e.g., proportion of pointed responses) were regressed on the coded phonological features (e.g., consonant manner). Results are reported as significant at the $p < 0.05$ and $p < 0.01$ levels.

For the forced-choice task, the average proportion of pointed responses averaged across subjects for each of the 570 nonsense words was regressed on manner of consonant articulation, consonant voicing, vowel roundedness (which co-varied with vowel backness), first position vowel height, and second position vowel height. For the Likert-rating tasks, the average Likert score across subjects for each nonsense word was regressed on the same variables.

These predictor variables were chosen because each is representative of the phonological categories used to describe consonants and vowels. For example, for the feature of consonant voicing, a consonant must either be voiced or unvoiced and cannot be both, so using the variable proportion of voiced consonants also then reflects the proportion of unvoiced consonants. Phonological categories in which there existed multiple values, such as manner of consonant articulation, were instead divided into two broad categories (i.e. sonorants and obstruents) for the purposes of this analysis.

Forced-choice task.

As illustrated in Table 1, nonwords with obstruents , including both stops and affricates/fricatives, were more likely to be rated pointed than rounded by listeners, and nonwords with sonorants were more likely to be assigned the label of rounded, with the largest observed mean difference of 0.3224 (stops - sonorants). Nonwords with unvoiced consonants were more likely to be judged as pointed than rounded by listeners, and nonwords with voiced consonants were more likely to be assigned the label of rounded. Nonwords with unrounded, front vowels were more likely to be judged as pointed than rounded by listeners, and nonwords with rounded, back vowels were more likely to be assigned the label of rounded. Nonwords with mid vowels in the first vowel position were slightly more likely to be judged as pointed than rounded by listeners, and nonwords with high vowels in the first vowel position were marginally more likely to be assigned the label of rounded, with the smallest observed mean difference of 0.0137 (V1 mid – V1 high). Nonwords with mid vowels in the second vowel position were slightly more likely to be judged as pointed than rounded by listeners, and nonwords with high vowels in the second vowel position were marginally more likely to be assigned the label of rounded.

Class	**Factor**	**Mean Proportion Pointed**
Consonant Manner	Sonorants	0.2693
	Obstruents	0.5749
	Stops	0.5917
	Affricates/Fricatives	0.5579
Consonant Voicing	Voiced	0.4163
	Unvoiced	0.6583
Vowel Place of Articulation	Unrounded/front	0.6177
	Rounded/back	0.3696
First Vowel Height (V1)	High	0.5075
	Mid	0.5212
Second Vowel Height (V2)	High	0.5286
	Mid	0.4978

Table 1: Proportion pointed judgments by vowel and consonant class in the forced choice task of Experiment 1

In the multiple regression analysis, the predictor variables accounted for a significant proportion of the variance for the proportion of pointed judgments ($R^2 = .692$, $F(5, 564) = 253.741$, $p < .001$). Of the variables examined, the presence of obstruents (ß= - .391, $t(564) = -15.283$, $p < .001$) and the presence of unvoiced consonants in the nonwords (ß = -.356,

$t(564) = -13.915$, $p < .001$) were significant predictors of the proportion of pointed judgments. The presence of unrounded, front vowels ($ß=.552$, $t(564) = 23.201$, $p < .001$), the presence of high vowels in the first position ($ß = 0.0821$, $t(564) = 3.434$, $p < .001$), and the presence of high vowels in the second position ($ß=.0768$, $t(564)=3.268$, $p =0.001$) were also significant predictors of the proportion of pointed judgments.

Likert-type rating tasks.

To evaluate correspondences between the phonological features and listener judgments, mean Likert responses for each nonword averaged across subjects are reported for each segmental class (e.g. consonant manner, consonant voicing, consonant place of articulation, first vowel height, and second vowel height) for both the pointed Likert-type rating task and the rounded Likert-rating task.

As illustrated in Table 2, in the pointed Likert-type rating task, nonwords with obstruents, including both stops and affricates/fricatives, were more likely to be rated as more pointed, and nonwords with sonorants were more likely to be rated less pointed, with the largest observed mean difference of 1.0041 (stops - sonorants). Nonwords with unvoiced consonants were more likely to be rated as more pointed, and nonwords with voiced consonants were more likely to be rated as less pointed. Nonwords with unrounded, front vowels rated more pointed than nonwords with rounded, back vowels. Nonwords with mid vowels in the first vowel position were slightly more likely to be rated pointed than nonwords with high vowels in the first vowel position. Nonwords with mid vowels in the second vowel position were not rated more pointed than nonwords with high vowels in the second vowel position.

Also shown in Table 2, in the rounded Likert-type rating task, nonwords with sonorants were rated more rounded than nonwords with obstruents, with the largest difference of 0.7602 between sonorants and stops. Nonwords with voiced consonants were more likely to be rated more rounded than nonwords with unvoiced consonants. Nonwords with rounded, back vowels were more likely to be rated more rounded than nonwords with unrounded, front vowels. Nonwords with high vowels in the first vowel position were slightly more likely to be rated more rounded than nonwords with mid vowels in the first vowel position. Nonwords with high vowels in the second vowel position were not rated more rounded than nonwords with mid vowels in the second vowel position.

Class	**Factor**	**Average Likert Rating (Very Pointed = 7)**	**Average Likert Rating (Very Rounded = 7)**
Consonant Manner	Sonorants	3.3585	4.1374
	Obstruents	4.1667	3.3879
	Stops	4.3626	3.3772
	Affricates/Fricatives	3.9708	3.3985
Voicing	Voiced	3.8259	3.7522
	Unvoiced	4.2737	3.2161
Vowel Place of Articulation	Unrounded/front	4.2085	2.9543
	Rounded/back	3.7253	4.3400
First Vowel Height (V1)	High	4.0473	3.6014
	Mid	3.9469	3.4503
Second Vowel Height (V2)	High	3.9979	3.5078
	Mid	4.0122	3.5677

Table 2. Mean Likert ratings for the pointed scale (1=not pointed, 7=very pointed) and for the rounded scale (1=not rounded, 7=very rounded) by vowel and consonant class in Experiment 1.

In the multiple regression analysis for the pointed Likert scale, the predictor variables accounted for a significant proportion of the variance in mean pointed ratings ($R^2 = .445$, $F(5, 564) = 90.577$, $p < .001$). Of the variables examined, the presence of obstruents (ß= - .447, t(564) = -13.014, $p < .001$) and the presence of unvoiced consonants in the nonwords (ß = -.161, t(564) = -4.698, $p < .001$) were significant predictors of higher Likert ratings, toward very pointed judgments. The presence of unrounded, front vowels (ß=.398, t(564) = 12.470, $p < .001$) and of high vowels in the first position in the nonwords (ß = 0.162, t(564) = 5.045, $p < .001$) were also significant predictors of higher Likert ratings, toward very pointed judgments. The presence of high vowels in the second position was not a significant predictor (ß=.00194, t(564)=0.0615, p = 0.951).

In the multiple regression analysis for the rounded Likert scale, the predictor variables accounted for a significant proportion of the variance in mean rounded ratings ($R^2 = .769$, F(5, 564) = 375.957, $p < .001$). Of the variables examined, the presence of sonorants (ß= .265, t(564) = 11.951, $p < .001$) and the presence of voiced consonants in the nonwords (ß = .193, t(564) = 8.704, $p < .001$) were significant predictors of higher Likert ratings, toward very rounded judgments. The presence of rounded, back vowels (ß= - .797, t(564) = - 38.680, $p < .001$), the presence of mid vowels in the first position (ß = - .0683, t(564) = - 3.298, p =.001), and the presence of mid vowels in the second position in the nonwords (ß= - .0416, t(564)= - 2.044 , p = 0.041) were also significant predictors of higher Likert ratings, toward very rounded judgments.

The findings indicate that particular phonemic properties of nonsense words were significantly associated with the assignment of higher rounded or pointed ratings on a seven-point Likert scale, and with the choice of rounded or pointed meanings in a forced-choice task. For instance, rounded vowels (e.g., “moomoh” or “fohsoo”) were consistently associated with the rounded meaning across tasks and unrounded vowels (e.g., “fihsee” or “kihpee”) were associated with the pointed meaning.

Across judgment tasks, listeners appeared to use a combination of phonemic cues when assessing word meaning in the dimension of rounded or pointed. Words that had obstruents, unvoiced consonants, unrounded, front vowels, and high vowels in positions one and two of the nonword were more likely to be judged pointed in the forced choice task. In contrast, words that had sonorants, voiced consonants, rounded, back vowels, and mid vowels in positions one and two of the nonsense word were more likely to be judged rounded in the forced choice task. In the pointed Likert-type rating task, nonwords that had obstruents, unvoiced consonants, unrounded, front vowels, and a high vowel in position one were rated more pointed, or higher, on the seven-point Likert scale. In the rounded Likert-type rating task, nonwords that had sonorants, voiced consonants, rounded, back vowels, and mid vowels in positions one and two were rated more rounded, or higher, on the seven-point Likert scale.

The regression equation for the pointed Likert rating task was not significant for the variable of second position vowel height, suggesting a less strong relationship between the phonological feature of vowel height and listener judgments of rounded and pointed meaning, at least in the second syllable of these nonsense words. Findings that were consistent across all judgment tasks were that nonwords with obstruents, unvoiced consonants, unrounded, front vowels, and a high vowel in position one were more likely to be judged as pointed or more pointed on the Likert scale (e.g. “tihkay” or “kehtay”). Additionally, nonwords with sonorants, voiced consonants, rounded, back vowels, and mid vowels in positions one and two were more likely to be judged as rounded or more rounded on the Likert scale (e.g. “mohmoo” or “lohloo”).

Experiment 2

To determine if these sound symbolic correspondences have functional significance and may be related to cross-modal mappings, a speeded classification task was designed utilizing nonwords of varying "pointedness" and "roundedness", based on ratings in Experiment 1. Cross-modal interference and facilitation was assessed by the extent to which the properties of the auditory stimuli influenced visual shape judgments. Interference was defined as occurring when the auditory nonsense words slowed visual shape identification and facilitation was defined as occurring when the auditory nonsense words speeded visual shape identification judgments. How quickly participants responded to the shape stimuli was expected to depend on whether the sound structure of the nonsense words matched or mismatched properties of the visual stimulus.

Method

Participants. A total of 18 Emory University undergraduates between the ages of 18 and 30 received either course credit or payment for their participation. All participants were native English speakers, reported no known hearing, speech, or language disorders. Two participants had participated in one of the tasks in Experiment 1.

Materials.

Visual stimuli. In a speeded classification task, participants were asked to judge one of two selected visual stimuli in the presence of auditory stimuli chosen from the set of nonsense words used in Experiment 1. The visual stimuli, one round and one pointy, were chosen from ratings obtained from a separate set of 34 participants. Participants rated sets of round and pointy figure shapes varying in number of protuberances and depth of protuberances on a seven-point Likert-type scale from 1= not round or pointy to 7= very round or very pointy. The shape that was rated as most pointy was used for the pointy visual stimulus with an average pointy rating of 6.85 and an average round rating of 1.59. The shape that was rated as most round was used for the round visual stimulus with an average pointy rating of 1.15 and an average round rating of 6.06 (Figure 2).

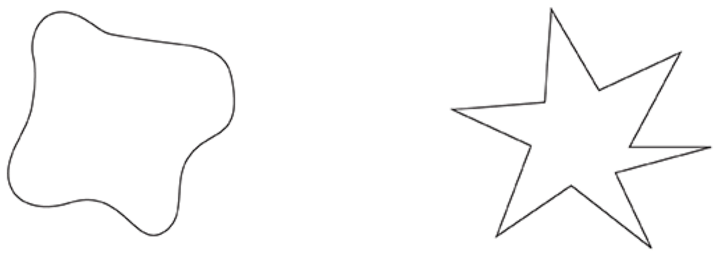

Figure 2. Round and pointy stimuli for Speeded Classification Task

Auditory stimuli. Phonemic content of the auditory stimuli matched, mismatched, or was neutral compared to the physical properties or shape of the visual stimulus. Nonwords were chosen from the set of stimuli used in Experiment 1 to construct five sets of auditory stimuli. Participants' ratings of rounded and pointed in Experiment 1 were compiled to assign all 570 nonsense words on a scale from round to pointy (or a scale of 0 - 1). Five nonwords were chosen in each of five categories including most round (0.10 ± 0.05), less round (0.30 ± 0.05), neutral (0.50 ± 0.05), less pointy (0.70 ± 0.05), and most pointy (0.90 ± 0.05). These sets of nonword stimuli were paired with the two visual stimuli to make five conditions ranging from very congruent to very incongruent. Matching nonword stimuli were either strongly congruent or weakly congruent. Mismatching nonword stimuli were either strongly incongruent or weakly incongruent. Neutral stimuli were neither congruent nor incongruent. In order to create these five conditions, nonwords, five of each type, rated as very rounded, moderately rounded, neutral, moderately pointed, and highly pointed were paired with each shape. For example, for highly congruent trials, a nonword consistently assessed to be rounded across conditions in Experiment

1, such as “kehtay,” was paired with the round visual stimulus. In the highly incongruent condition, the same word “kehtay” would alternatively be paired with the pointy visual stimulus. Nonwords that were just as likely to be assessed rounded as pointed, such as “sehfih”, were paired with either the round or pointy visual stimulus and considered neither congruent nor incongruent. A silent control condition presenting the visual stimulus without auditory cue was also included (See Appendix 2).

Procedure

Twenty-five nonsense words were chosen, as well as five silent trials, and each of these thirty auditory stimuli were paired once with the pointy visual stimulus and once with the round visual stimulus randomly within blocks of sixty trials. These sixty trials were repeated for a total of eight blocks and trials were randomized within each block. Congruency assignments for trials pairing the visual stimulus with either the most pointed or most rounded auditory stimulus were characterized as strongly congruent or strongly incongruent. Trials pairing the visual stimulus with either a less pointed or less rounded auditory stimulus were characterized as weakly congruent or weakly incongruent. Trials pairing the visual stimulus with a neutral auditory stimulus were characterized as neutral. For example, if a less rounded word like “nihmih” were paired with the pointy visual stimulus, that trial would be weakly incongruent.

Before the testing phase, participants were given three auditory stimuli not used during testing and instructed to adjust the volume on their headphones to an adequate level after each stimulus. Following this practice phase, the testing phase began, and on each trial, participants were instructed to focus on a fixation cross before presentation of either stimulus. The fixation cross was presented for 1.5s and then the auditory stimulus was presented for an average duration of 450ms. The visual stimulus was presented 75ms from the end of the auditory stimulus for 75ms, simultaneous with the end of the auditory stimulus. Response choices of "round" and "pointy" were presented on the computer screen at the onset of the visual stimulus. The choices of round or pointy remained on the screen until participants made a response, and the next trial began after 50ms. Response times for each trial were measured from the onset of the visual stimulus presentation.

Results and Discussion

Analysis of the behavioral learning task involved standard mixed effects models using items and subjects as random variables and conditions of interest as fixed variables. Data analysis for the judgment tasks focused on identifying whether congruency between the auditory and visual stimuli would influence visual object identification. Trials with response times greater than 2.5 standard deviations from the mean response time for each subject were excluded (See Table 3). In addition, trials in which the visual stimulus was incorrectly judged were excluded (See Table 4). The congruency level of each trial was coded, and listener response times were averaged across items within conditions.

Visual Stimulus	**Congruent**			**Incongruent**			**Neutral**		
Round		Mean %	SEM		Mean %	SEM		Mean %	SEM
	Strongly	2.0833	0.0493	Strongly	2.2222	0.0441	Neither	1.1111	0.0322
	Weakly	0.8333	0.0185	Weakly	1.6667	0.0496	Silent	4.1667	0.0458
Pointy		Mean %	SEM		Mean %	SEM		Mean %	SEM
	Strongly	1.8056	0.0369	Strongly	2.0833	0.0365	Neither	1.1111	0.0419
	Weakly	0.9722	0.0286	Weakly	1.2500	0.0379	Silent	3.4722	0.0328

Table 3. Mean percent of trials excluded based on response time and standard error

Visual Stimulus	**Congruent**			**Incongruent**			**Neutral**		
Round		Mean	SEM		Mean	SEM		Mean	SEM
	Strongly	.9819	0.1074	Strongly	.9819	0.1094	Neither	.9819	0.1088
	Weakly	.9847	0.1102	Weakly	.9903	0.1099	Silent	.9694	0.1092
Pointy		Mean	SEM		Mean	SEM		Mean	SEM
	Strongly	.9667	0.1081	Strongly	.9722	0.1088	Neither	.9722	0.1093
	Weakly	.9667	0.1044	Weakly	.9764	0.1085	Silent	.9806	0.1100

Table 4. Mean proportion correct and standard error as a function of visual stimulus and congruency condition

Participant response times were recorded and categorized according to the auditory and visual stimulus classifications derived from Experiment 1. For example, the trials including a most rounded auditory stimulus paired with the round visual stimulus as well as the trials including a most pointed auditory stimulus paired with the pointy visual stimulus were categorized as strongly congruent. To evaluate the influence of congruency on participant response time, mean response time and standard error for each condition (strongly congruent, weakly congruent, neutral, weakly incongruent, strongly incongruent, and silent control) was calculated. Results are reported as significant at the $p < 0.05$ and $p < 0.01$ levels.

In order to evaluate the role that sound symbolic nonwords had on visual perception, response times across each level of congruency were compared. As can be seen in Figure 3 below, silent trials were significantly slower than all other categories (one-way ANOVA, $F(5,13) = 4.765$, partial $\eta 2 = .002$, $p < .001$). Individual t-tests with a Bonferroni correction confirmed results of the ANOVA ($t(13) = 2.74$, $p < .005$). No other difference among congruency levels was significant. However, observable trends indicate that participants responded somewhat faster for strongly congruent trials than weakly congruent, neither, and strongly incongruent trials. Contrary to the hypothesis, participants responded slightly faster in the visual judgment task during weakly incongruent trials than during weakly congruent trials and almost equally fast during weakly incongruent and strongly congruent trials.

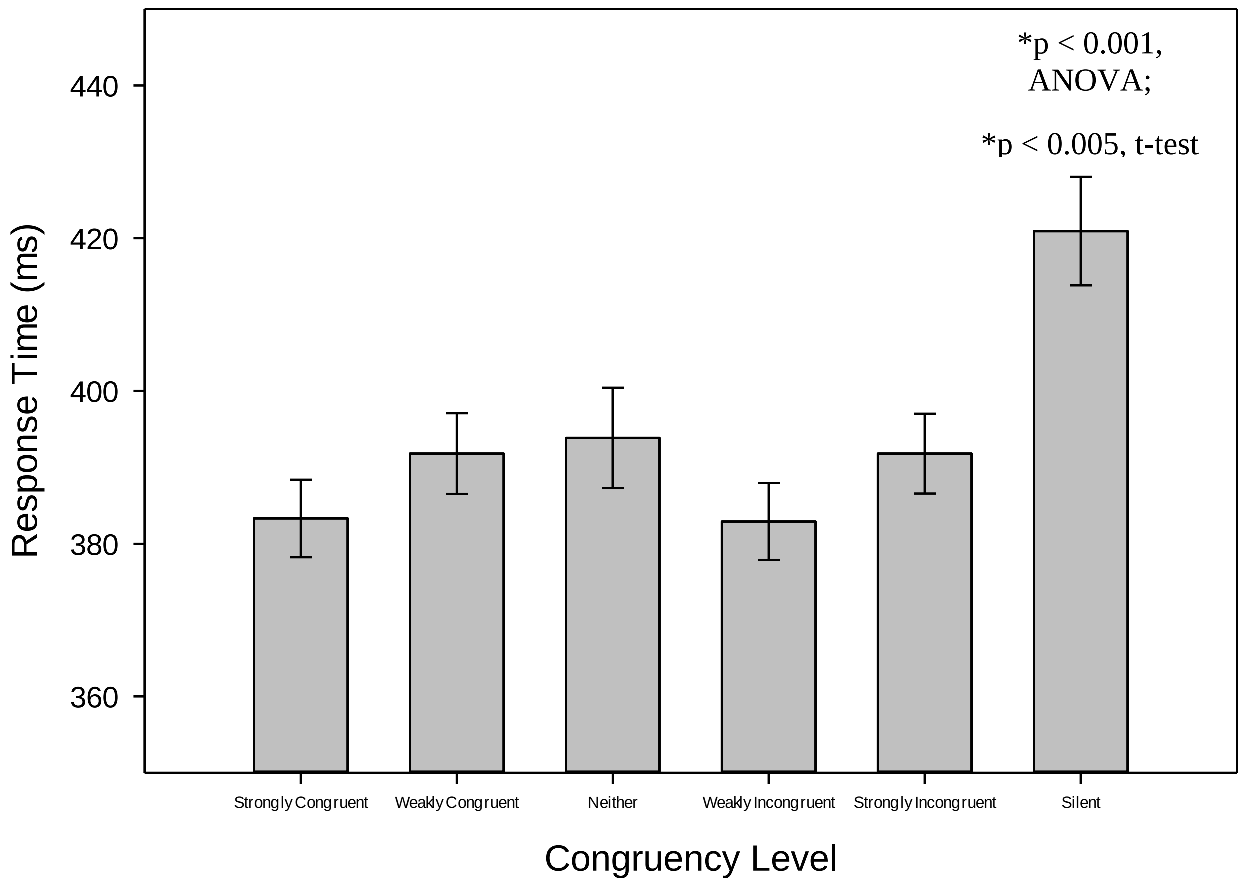

Figure 3. Mean response time for each group condition in Experiment 2.

Findings indicate that the level of congruency between a sound symbolic auditory stimulus and a visual shape did not affect visual object identification judgments in this task. Contrary to the hypothesis, only response times during silent control trials were significantly different than the other conditions, suggesting that interference or facilitation of visual judgment from the auditory properties of sound symbolic stimuli does not occur in this priming task.

General Discussion

In Experiment 1, participants rated systematically varied nonwords as rounded or pointed in Likert-type rating and forced-choice tasks. In order to determine which speech sounds correspond with particular meanings, phonemes within the nonwords were analyzed for consistent associations with rounded or pointed, two basic visual sensory-perceptual properties. Based on work by Nielsen and Rendall (2011), speech sounds that contain noise, such as fricative and affricates (e.g. “feesee” or “sehfih”) were expected to be associated with the attribute pointed. Based on the theory that sound symbolism is grounded in synesthetic sensory-motor associations (Ramachandran & Hubbard, 2001), nonsense words containing rounded vowels (e.g. “mohloo” or “kohtoo”) were expected to be associated with the meaning rounded due to lip rounding that physically mimics a round shape. Sound to meaning mappings may arise via auditory sound patterns during vocalization and visual properties of the referent, and cross-modal neural activations may account for links between the phonological properties of words and their referents (Ramachandran & Hubbard , 2001).

In Experiment 2, a behavioral task was developed to determine if the associations between certain speech sounds and visual shapes found in Experiment 1 have functional significance for language learning and processing. The speeded classification task was designed to determine if the sound to meaning mappings influence visual shape processing as suggested by previous work (Parise & Spence, 2008; Lupyan & Spivey, 2010). Participants responded to auditory-visual stimulus pairings in which visual objects (round or pointy) were presented with nonsense words containing sound structures that either matched or mismatched visual object shape. If cross-modal correspondences underlie sensitivity to sound symbolism, then the sound structure of the nonsense words was expected to influence how quickly participants could identify the visual object shape (Melara & O’Brian, 1987). When nonwords that sound "pointed" were paired with a pointy visual stimulus, judgments of the visual stimulus were expected be made more quickly than when these same "pointed" nonwords were paired with a round visual stimulus. When the sound symbolic properties of the nonwords matched the visual shape properties, then identification was expected to be facilitated, and when nonwords mismatched visual object shape, interference between the sound to meaning associations and the visual sensory input was expected, leading to slower reaction times.

Although different phonemic features played a role in listener judgments for rounded and pointed in Experiment 1, some features, such as consonant manner, were more predictive than others, such as vowel height. However, all phonemic segments analyzed influenced listeners' judgments of nonsense word meaning, rounded or pointed. Nonwords judged to mean rounded contained sonorant consonants, voiced consonants, rounded, back vowels, and mid vowels in position one more often than those nonwords judged to mean pointed. In contrast, nonwords judged to mean pointed had obstruent consonants, unvoiced consonants, unrounded, front vowels, and high vowels in position one more often than those nonwords judged to mean rounded. From the results of Experiment 2, those nonwords judged to sound very rounded or very pointed when paired with corresponding round and pointy visual stimuli did not have the predicted effect on speed of visual perception and identification. Only response times during control silent trials were significantly different than any other condition, suggesting that interference or facilitation of visual judgments through the auditory perception of sound symbolic stimuli did not necessarily occur in this priming task.

The findings from Experiment 1 are consistent with suggested non-arbitrary associations between sound structure and perceptual properties in language (Köhler, 1929; Maurer, et al., 2006). The results of Experiment 1 extend the findings of Nielsen and Rendall (2011) who found that participants labeled curved object images with novel words containing relatively more sonorant consonants and rounded vowels, and labeled jagged object images with novel words containing relatively more stop consonants and unrounded vowels. Although certain properties of vowels were found to affect rounded/pointed judgments in Experiment 1, such as vowel roundedness, vowel height was in some cases not a significant predictor of these judgments, in particular the vowel height in position two in the CVCV structure of the nonsense words. Perhaps listeners were judging meaning before the auditory stimulus was fully presented, which may partially explain why second position vowel height did not have a consistent effect across tasks.

While Experiment 1 is consistent with other findings (Maurer & Mondloch, 2006; Nielsen & Rendall, 2011), this study focused only on native English speakers for whom English was the only language spoken in the home. Further investigation with participants of a different or with varying language backgrounds would enhance the idea that these sound symbolic properties are not a function of previous language experience and are in fact cross-linguistic phenomena as Kantartzis et al. (2011) found with Japanese word-learning in English-speaking children or as Pirog Revill et al. (2013) found in English speakers assigning meaning to foreign words. The possibility remains that participants mapped common or universal associations between sound and meaning onto nonwords during the judgment task. Alternatively, participants may have used phonemic characteristics found in English words that signify rounded or pointed, respectively, to infer the meaning of the nonword auditory stimuli. Characteristics of English language structure may explain the consistent mappings between nonword speech segments and the visual property of rounded or pointed.

Although evidence supports the idea that certain words are phonetically associated with semantic meaning (Köhler, 1929; Maurer et al., 2006; Nygaard et al., 2009), the specific characteristics of word structure that may correspond to multiple physical dimensions, other than rounded and pointed, are not yet known. Expanding the participant judgments from rounded and pointed to other binary dimensions for these nonwords would further explore the consistency of sound symbolic associations across visual dimensions, such as big and small (Sapir, 1929), or even across other modalities, such as the spatial perception of near and far.

The findings in Experiment 2 are inconsistent with previous cross-modal priming research investigating the effects of the auditory modality on speed of judgment in the visual modality (Marks & O'Brian, 1987; Melara & Marks 1990a; Melara & Marks 1990b; Parise & Spence, 2008). Earlier findings have demonstrated that auditory input, and particularly sound symbolic input, may influence speeded visual classification tasks and that this facilitation or interference of response may be due to the congruency of cross-modal correspondences (Marks, 1987; Melara & Marks, 1990a; Parise & Spence, 2008). However, with the exception of weakly incongruent trials, trends aligned with Parise & Spence's (2008) implicit association task (IAT) that found stronger coupling of synesthetically matched over mismatched stimuli, but the relatively small sample size for Experiment 2 (n = 18) may need to be increased in order to yield significant results. One particular drawback of both the IAT and speeded classification tasks is that learning across repeated stimulus blocks may also have affected the pattern of findings (Parise and Spence, 2008). For IAT tasks, there are frequent response assignment changes that can affect results, but learning could also have influenced the current findings in the speeded classification task. For instance, participants may have become habituated to the extraneous auditory stimulus with increasing block number, diminishing interference or facilitation effects.

Alternatively, participants may have been distracted by the additional information provided by the auditory stimulus while trying to learn the task in earlier blocks, causing them to ignore extraneous information and affecting response times in earlier blocks versus later blocks.

Additionally, the task implemented here differed from previous tasks (Parise & Spence, 2008; Melara & O'Brian, 1987) and was subject to limitation based on design. The speeded classification task done in Experiment 2 differs from the implicit association task (IAT) in that for the IAT, stimuli are presented alone as opposed to simultaneously in the speeded classification task. Also, Experiment 2 utilized twenty-five nonsense words and silent controls as auditory stimuli paired with two visual stimuli compared to the IAT paradigm that used two auditory and two visual stimuli. Parise and Spence (2008) also note several disadvantages to the speeded classification task. Because two stimuli in different modalities (e.g. auditory and visual) have been presented simultaneously, or nearly so, in each trial, any stimulus-dependent modulation of response times may in fact reflect a failure of selective attention. Participants may be unable to fully focus attention on the task and most relevant stimuli while ignoring the distracting stimuli enough to perform quickly and accurately.

By investigating the associations made between nonsense words and the opposing physical traits of rounded and pointed, this study pinpoints specific sound-to-meaning mappings and characterizes these mappings through phonetic analyses. Although linguistic as well as cognitive theories have supported the idea that language is purely arbitrary (de Saussure, 1959), this study challenges the idea that language and semantic meaning are confined to the amodal and symbolic realm (Farmer, Christiansen, & Monaghan, 2006). Studies have recently indicated that the conceptual understanding represented by language is linked or related to the sensorimotor memories that formed physical concepts, such as spatial relationships, visual properties, and actions or motion (Kemmerer, 2009; Kovic et al., 2010). Even studies supporting separate meaning and perceptual cortical areas find that the two systems are spatially close, indicating an "anterior shift" that allows for potential overlap (Bedny & Caramazza, 2011; Kemmerer, 2009).

The possibility of neural correlates which connect sensory processes and influence sound symbolism has implications for the evolution of language. Christiansen and Chater (2008) introduce a theory of language evolution in which language is another function of the brain that is subject to change based on the brain's evolution (Barsalou, 2008; Ramachandran & Hubbard, 2001). Kovic et al. (2010) found a sensitivity to sound symbolic associations in an EEG experiment and suggest that sound symbolism may result from a general integration of auditory and visual features within the brain. Barsalou (2008) suggests that language may be cognitively grounded within frameworks of experience, spatial relations, and the senses through mental simulation. Ramachandran and Hubbard further conjecture that some language, including sound symbolic language, employs cross-modal correspondences informed by sensory processing in different modalities. If language is shaped by the brain, as Christiansen and Chater (2008) posit, and sound symbolism results from associations made across modalities, either through mental simulation or perception, then the current study implies that language may have evolved through sound symbolic vocalizations as the result of cross-modal connections between auditory and other sensory perception areas within the brain (Melara & Marks, 1990b; Spence, 2011).

While the current findings may add to conjecture over the evolution of language, sound symbolic properties also have functionality in modern use. For instance, examining lexical processing through the phenomenon of sound symbolism may yield specific insight that can explain how language is used to comprehend the natural world, especially in understanding and enjoying poetic language and in advertising and marketing. Examples of abstraction abound in literature, especially within poetry. The poet Malea Powell uses the physical attributes of a kiwi to describe the less concrete concept of perplexity by elaborating that "sometimes *perplexity* has exactly the *color* and *taste* of kiwi eaten with the skin on, a *fuzziness* that *hangs* on your teeth until you spit it out and pretend it's a perfectly normal thing to do" (Powell, 2012, emphasis added). Here Powell describes a concept that has no direct physical embodiment and invokes the senses of color, taste, and somatosensory cues (i.e. texture and weight) to evoke the idea of perplexity in the reader. Perhaps the connection between physical attributes and an abstract concept lies within the theory of grounded cognition and the idea that conceptual bodies within the mind are supported by sensory processes (Barsalou, 2008; Pezzulo et al., 2013). Based on the current and previous studies (Nielsen & Rendall, 2011), these connections may be reflected within the structure of languages in the form of cross-linguistic sound symbolism and be represented cortically through cross-sensory integration. Sound symbolism can be used to understand and create more evocative poetry or prose using the sounds of words rather than solely relying on their meanings to encourage mental simulation, and in fact, non-arbitrary connections between sound and meaning may be the result of this mental simulation (Barsalou, 2008; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008).

Experiment 1 supports the possibility that non-arbitrary and systematic associations between words and the visual properties of their referents inform communicative strategies in natural language (Gasser, 2004). If these non-arbitrary associations are due to connections between the various perceptual systems in the brain, sound symbolism may allow for greater sensory simulation as well as a depth of abstraction in expression and communication not possible under the arbitrariness assumption (Barsalou, 2008; Gasser, 2004). In addition, these cross-modal associations may confer an advantage to communication that pure arbitrary word-referent assignment does not in daily use. Everyday examples of sound symbolic advantage can be seen through the advertising industry, and studies have been conducted indicating that marketers can enhance certain perceptions of consumer products by ensuring that the sound symbolism of the brand name and sound or shape symbolism within the labeling or packaging establishes the desired product-related sensory associations (Shrum, Lowrey, Luna, Lerman, & Liu, 2012; Spence, 2012; Spence & Deroy, 2013).

Sound symbolism may operate on an unconscious level, allowing consumers to make judgments and associations with products or brands based on these implicit connections and without overtly evaluating the information presented (Spence, 2012). Shrum et al. (2012) found that these sound symbolic properties hold across languages for participants who were bilingual in French, Spanish, or Chinese and in English. Participants preferred words in which there was congruency between the product qualities and the sound symbolism of the label. These results held when participants completed the study in their first or second language and regardless of second-language proficiency or whether the Chinese language representations were in logographic or alphabetic form. Understanding how particular sound segments are associated with particular physical dimensions can have an influence on how advertisers plan campaigns for new products, encouraging buyers through their implicit associations with sound. However, these findings are not only significant in the marketing world for brand development, but also provide evidence for the idea that certain sound symbolic associations may arise from cross-modal mapping within the brain rather than developing through cultural exposure to a given language (Kantartzis, Imai, & Kita, 2011; Shrum et al., 2012).

Although Experiment 2 did not yield conclusive evidence over the existence of cross-modal correspondences, Experiment 1 revealed systematic associations between phonemic content and meaning judgment. These findings indicate that language may consist of an aggregate of both arbitrary and non-arbitrary signifiers, which yields optimal communicative conditions (Christiansen & Chater, 2008). If language is shaped by the brain and the neural firing patterns that drive it (Christiansen & Chater, 2008), then these non-arbitrary pairings imply that language emerged through a complex combination of culture and learning biases that has become implicit in the structural framework of language.

Future Directions & Conclusions

The neural correlates of sensory processing and potential cross-modal correspondences that may influence sound symbolism during language processing are not yet known. A paradigm for investigating the neural mechanisms underlying sound symbolism may be developed from the current experiments into an fMRI imaging study. Such as study would delve into the neural framework of language comprehension similar to work done by Pirog Revill et al. (2013) examining the neural correlates of sound symbolism in foreign words.

If sound symbolism is simply based on cultural or conventional pairings, specific sound to meaning associations will depend on the language background of the individual, and an fMRI will reveal less consistent overlap between language areas and perceptual areas for a given task across subjects of varying language experience. However, if sound to meaning mappings are grounded in perceptual-motor systems (Arata et al., 2010), sound symbolic associations will not differ as dramatically across language background within participant groups, and fMRI will indicate cross modal correspondences between auditory areas responsible for processing linguistic form and visual or other perceptual processing areas represented in the brain. Imaging data can then be compared to behavioral task performance in order explore the association of specific patterns of neural activation and connectivity with sensitivity to sound symbolism. Additionally, the fMRI experiment may reveal active areas such as the SPC, the SLF, and the angular gyrus, which are associated with cross-modal perceptual integration (Pirog Revill et al., 2013). The methodological approach of fMRI and behavioral tasks may both yield insight into the nature of language as well as the neural structures that produce and interpret language. (Simmons et al., 2008; Peiffer-Smadja 2010; Pirog Revill et al., 2013)

Sound symbolism, the non-arbitrary connections between sound structure and meaning, has implications for exploring the human conceptual framework that involves dynamic interaction between the physical world and conceptual understanding. The theory of grounded cognition posits that conceptual and physical representations are not separate but rather interacting systems that are cortically integrated in the brain. This experiment explored the validity of that idea in the context of general cross-modal correspondences and their relevance to sound symbolism, finding systematic mappings between the physical property of rounded or pointed and the phonemic aspects of nonwords. Further exploration into the neural correlates of the associations made during sound symbolic interpretation may yield insight into cortical mechanisms of language comprehension as well as provide clues toward understanding language evolution.

References

Arata, M., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2010). Gesture in language: How sound symbolic words are processed in the brain. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 1374-11379.

Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59(1), 617-645. doi: 10.1146/annurev.psych.59.103006.093639

Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: the case from action verbs. *Annuals of the New York Academy of Sciences*. 1224(1), 81-95. doi: 10.1111/j.1749-6632.2011.06013.x

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 489-558. doi: 10.1017/S0140525X08004998

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). From the cover: Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203-12208.

Hamano, S. (1998). *The sound-symbolic system of Japanese*. Stanford, CA: CSLI Publications.

Hubbard, E. M., & Ramachandran, V. (2005). Neurocognitive mechanisms of synesthesia. *Neuron*, *48*(3), 509-520. doi: 10.1016/j.neuron.2005.10.012

Gasser, M. (2004). The origins of arbitrariness in language.*Proceedings of the Cognitive Science Society Conference*. (pp.434-439). Hillsdale, NJ: LEA.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109,* 54-65.

Kantartzis, K., Imai, M., and Kita, S. (2011). Japanese sound-symbolism facilitates verb learning in English speaking children. *Cognitive Science, 35*, 575–586.

Kemmerer, D. (2009). How words capture visual experience: The perspective from Cognitive Neuroscience. In B. Malt & P. Wolff (Eds.), Words and the mind: How Words Capture Human Experience (pp. 287 – 327). New York: Oxford University press.

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

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114(1),* 19-28. doi: 10.1016/j.cognition.2009.08.016

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. In Psychological Review, 104(2), 211–240.

Locke, J. (1715). Book III On words: Chapter I Of words or language in general. In *An essay concerning human understanding*. London: Printed for J. Churchill; and Samuel Manship.

Lupyan, G., & Spivey, M. J. (2010). Making the invisible visible: Verbal but not visual cues enhance visual detection (J. Lauwereyns, Ed.). *PLoS ONE*, 5(7), E11452. doi: 10.1371/journal.pone.0011452

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196-14201. doi: 10.1073/pnas.1303312110

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316-322. doi: 10.1111/j.1467-7687.2006.00495.x

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. Journal of Experimental Psychology: Human Perception and Performance, 13(3), 384.

Melara, R. D., & Marks, L. E. (1990a). Dimensional interactions in language processing: investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 539-554. doi: 10.1037//0278-7393.16.4.539

Melara, R. D., & Marks, L. E. (1990b). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition,* 18(5), 477-495. doi: 10.3758/BF03198481

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4), 323-336. doi: 10.1037/0096-3445.116.4.323

Monaghan, P., & Christiansen, M. H. (2006). Why form-meaning mappings are not entirely arbitrary in language. *Cognitive Science Journal*, 1838-1843. Retrieved September 21, 2013, from http://csjarchive.cogsci.rpi.edu

Monaghan, P., Christiansen, M.H. & Fitneva, S.A. (2011). The Arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General* (140)3, 325-347. doi: 10.1037/a0022924

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 65(2), 115-124. doi: 10.1037/a0022268

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*,*28*(1), 225-252. doi: 10.1146/annurev.anthro.28.1.225

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181-186. doi: 10.1016/j.cognition.2009.04.001

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes . *PLoS ONE, 4(5)*, E5664. doi: 10.1371/journal.pone.0005664

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, 220(3-4), 319-333. doi: 10.1007/s00221-012-3140-6

Peiffer-Smadja, N. (2010). *Exploring the bouba/kiki effect: A behavioral and fMRI study* (Unpublished master's thesis). Université Paris Descartes.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. Frontiers in Psychology, 3(612), 1-11. doi: 10.3389/fpsyg.2012.00612

Pirog Revill, K., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2013). *Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI*. Manuscript submitted for publication, Emory University.

Powell, M. (2012). A meditation partially composed in a D.C. coffeehouse because there isn't anything better to do in this city of dead white fathers... In Q. Driskill, D. H. Justice, D.

Miranda, L. Tatonetti, I. Allegra, & L. E. Cruz (Authors), Sovereign erotics a collection of two-spirit literature (pp. 84-85). Tucson: University of Arizona Press.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia - A window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3-34.

Reetz, H., & Jongman, A. (2009). Phonetics: Transcription, production, acoustics, and perception. Chichester, U.K.: Blackwell.

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

de Saussure, F. D. (1959). Chapter 2. In Course in general linguistics. (pp. 76-89). New York: Philosophical Library.

Shrum, L., Lowrey, T., Luna, D., Lerman, D., & Liu, M. (2012). Sound symbolism effects across languages: Implications for global brand names. *International Journal of Research in Marketing,* 29(3), 275-279. doi: 10.1016/j.ijresmar.2012.03.002

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). FMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris*, 102(1-3), 106-119. doi: 10.1016/j.jphysparis.2008.03.014

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics*, *73,* 971-995.

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology,* 22(1), 37-54.

Spence, C., & Deroy, O. (2013). On why music changes what (we think) we taste. *I-Perception,* 4(2), 137-140. doi: 10.1068/i0577ic

Appendix 1 Nonword Auditory Stimuli for Experiment 1

Sonorant	Voiced	Unrounded	Front		Sonorant	Voiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
l	e	l	i		l	o	l	u
l	ɛ	l	e		l	u	l	o
l	ɛ	l	i		l	ʊ	l	o
l	i	l	e		l	ʊ	l	u
l	I	l	e		l	o	m	o
l	e	m	e		l	o	m	u
l	e	m	i		l	u	m	o
l	ɛ	m	e		l	u	m	u
l	ɛ	m	i		l	ʊ	m	o
l	i	m	e		l	ʊ	m	u
l	I	m	e		l	o	n	o
l	i	m	i		l	o	n	u
l	I	m	i		l	u	n	o
l	e	n	e		l	u	n	u
l	e	n	i		l	ʊ	n	o
l	ɛ	n	e		l	ʊ	n	u
l	i	n	e		m	o	l	o
l	I	n	e		m	o	l	u
l	i	n	i		m	u	l	o
l	I	n	i		m	u	l	u
m	e	l	i		m	ʊ	l	o
m	ɛ	l	e		m	ʊ	l	u
m	ɛ	l	i		m	o	m	u
m	i	l	e		m	u	m	o
m	I	l	e		m	ʊ	m	o
m	i	l	i		m	ʊ	m	u
m	e	m	i		m	o	n	o
m	ɛ	m	e		m	o	n	u
m	ɛ	m	i		m	u	n	o
m	i	m	e		m	u	n	u
m	I	m	e		m	ʊ	n	o
m	I	m	i		m	ʊ	n	u
m	e	n	e		n	o	l	o
m	e	n	i		n	o	l	u
m	ɛ	n	e		n	u	l	o
m	i	n	e		n	u	l	u
m	I	n	e		n	ʊ	l	o
n	e	l	e		n	ʊ	l	u

Sonorant	Voiced	Unrounded	Front		Sonorant	Voiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
n	e	l	i		n	o	m	o
n	ɛ	l	e		n	o	m	u
n	i	l	e		n	u	m	o
n	I	l	e		n	u	m	u
n	i	l	i		n	ʊ	m	o
n	I	l	i		n	ʊ	m	u
n	e	m	e		n	o	n	u
n	e	m	i		n	u	n	o
n	ɛ	m	e		n	ʊ	n	o
n	ɛ	m	i		n	ʊ	n	u
n	i	m	e					
n	I	m	e					
n	i	m	i					
n	I	m	i					
n	e	n	i					
n	ɛ	n	e					
n	ɛ	n	i					
n	i	n	e					
n	I	n	e					

Stops	Voiced	Unrounded	Front		Stops	Voiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
b	ɛ	b	e		b	o	b	u
b	ɛ	b	i		b	u	b	o
b	i	b	e		b	ʊ	b	o
b	I	b	e		b	ʊ	b	u
b	I	b	i		b	o	d	o
b	e	d	e		b	o	d	u
b	e	d	i		b	u	d	o
b	ɛ	d	e		b	u	d	u
b	ɛ	d	i		b	ʊ	d	o
b	i	d	e		b	ʊ	d	u
b	e	g	e		b	o	g	o
b	e	g	i		b	o	g	u
b	ɛ	g	e		b	u	g	o
b	ɛ	g	i		b	u	g	u
b	i	g	e		b	ʊ	g	o
b	I	g	e		b	ʊ	g	u
b	i	g	i		d	o	b	o

Stops	**Voiced**	**Unrounded**	**Front**		**Stops**	**Voiced**	**Rounded**	**Back**
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
d	e	b	e		d	o	b	u
d	e	b	i		d	u	b	o
d	ɛ	b	e		d	u	b	u
d	i	b	e		d	ʊ	b	o
d	I	b	e		d	ʊ	b	u
d	i	b	i		d	o	d	u
d	I	b	i		d	u	d	o
d	e	d	i		d	ʊ	d	o
d	ɛ	d	e		d	ʊ	d	u
d	ɛ	d	i		d	o	g	o
d	i	d	e		d	o	g	u
d	I	d	e		d	u	g	o
d	e	g	e		d	u	g	u
d	e	g	i		d	ʊ	g	o
d	ɛ	g	e		d	ʊ	g	u
d	ɛ	g	i		g	o	b	o
d	i	g	e		g	o	b	u
d	I	g	e		g	u	b	o
d	i	g	i		g	u	b	u
d	I	g	i		g	ʊ	b	o
g	e	b	e		g	ʊ	b	u
g	e	b	i		g	o	d	o
g	ɛ	b	e		g	o	d	u
g	ɛ	b	i		g	u	d	o
g	i	b	e		g	u	d	u
g	I	b	e		g	ʊ	d	o
g	i	b	i		g	ʊ	d	u
g	I	b	i		g	o	g	u
g	e	d	e		g	u	g	o
g	e	d	i		g	ʊ	g	o
g	ɛ	d	e		g	ʊ	g	u
g	ɛ	d	i					
g	i	d	e					
g	I	d	e					
g	i	d	i					
g	e	g	i					
g	ɛ	g	e					
g	ɛ	g	i					
g	i	g	e					
g	I	g	e					

Stops	**Voiced**	**Unrounded**	**Front**
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**
g	I	g	i

Stops	**Unvoiced**	**Unrounded**	**Front**		**Stops**	**Unvoiced**	**Rounded**	**Back**
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
k	e	k	i		k	o	k	u
k	ɛ	k	e		k	u	k	o
k	ɛ	k	i		k	ʊ	k	o
k	i	k	e		k	ʊ	k	u
k	I	k	e		k	o	p	o
k	I	k	i		k	o	p	u
k	e	p	e		k	u	p	o
k	e	p	i		k	u	p	u
k	ɛ	p	e		k	ʊ	p	o
k	ɛ	p	i		k	ʊ	p	u
k	i	p	e		k	o	t	o
k	I	p	e		k	o	t	u
k	i	p	i		k	u	t	o
k	I	p	i		k	u	t	u
k	e	t	e		k	ʊ	t	o
k	ɛ	t	e		k	ʊ	t	u
k	ɛ	t	i		p	o	k	o
k	i	t	e		p	o	k	u
k	I	t	e		p	u	k	o
k	i	t	i		p	u	k	u
p	e	k	e		p	ʊ	k	o
p	e	k	i		p	ʊ	k	u
p	ɛ	k	e		p	o	p	u
p	ɛ	k	i		p	u	p	o
p	i	k	e		p	ʊ	p	o
p	I	k	e		p	ʊ	p	u
p	i	k	i		p	o	t	o
p	e	p	i		p	o	t	u
p	i	p	e		p	u	t	o
p	I	p	e		p	u	t	u
p	e	t	e		p	ʊ	t	o
p	e	t	i		p	ʊ	t	u
p	ɛ	t	e		t	o	k	o
p	i	t	e		t	o	k	u
p	I	t	e		t	u	k	o
p	i	t	i		t	u	k	u

Stops	Unvoiced	Unrounded	Front		Stops	Unvoiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
t	e	k	e		t	ʊ	k	o
t	e	k	i		t	ʊ	k	u
t	ɛ	k	e		t	o	p	o
t	i	k	e		t	o	p	u
t	I	k	e		t	u	p	o
t	I	k	i		t	u	p	u
t	e	p	e		t	ʊ	p	o
t	e	p	i		t	ʊ	p	u
t	ɛ	p	e		t	o	t	u
t	ɛ	p	i		t	u	t	o
t	i	p	e		t	ʊ	t	o
t	I	p	e		t	ʊ	t	u
t	I	p	i					
t	e	t	i					
t	ɛ	t	e					
t	ɛ	t	i					
t	i	t	e					
t	I	t	e					
k	e	k	i					
k	ɛ	k	e					
k	ɛ	k	i					
k	i	k	e					
k	I	k	e					
k	I	k	i					
k	e	p	e					
k	e	p	i					
k	ɛ	p	e					
k	ɛ	p	i					
k	i	p	e					

Affricate/ Fricatives	Voiced	Unrounded	Front		Affricate/ Fricatives	Voiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
dʒ	e	dʒ	i		dʒ	o	dʒ	u
dʒ	ɛ	dʒ	e		dʒ	u	dʒ	o
dʒ	ɛ	dʒ	i		dʒ	ʊ	dʒ	o
dʒ	i	dʒ	e		dʒ	ʊ	dʒ	u
dʒ	I	dʒ	e		dʒ	o	v	o
dʒ	I	dʒ	i		dʒ	o	v	u
dʒ	e	v	e		dʒ	u	v	o
dʒ	e	v	i		dʒ	u	v	u

Affricate/ Fricatives	Voiced	Unrounded	Front		Affricate/ Fricatives	Voiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
dʒ	ɛ	v	e		dʒ	ʊ	v	o
dʒ	ɛ	v	i		dʒ	ʊ	v	u
dʒ	i	v	e		dʒ	o	z	o
dʒ	I	v	e		dʒ	o	z	u
dʒ	i	v	i		dʒ	u	z	o
dʒ	I	v	i		dʒ	u	z	u
dʒ	e	z	e		dʒ	ʊ	z	o
dʒ	ɛ	z	e		dʒ	ʊ	z	u
dʒ	ɛ	z	i		v	o	dʒ	o
dʒ	i	z	e		v	o	dʒ	u
dʒ	I	z	e		v	u	dʒ	o
dʒ	I	z	i		v	u	dʒ	u
v	e	dʒ	e		v	ʊ	dʒ	o
v	e	dʒ	i		v	ʊ	dʒ	u
v	ɛ	dʒ	e		v	o	v	u
v	i	dʒ	e		v	u	v	o
v	I	dʒ	e		v	ʊ	v	o
v	i	dʒ	i		v	ʊ	v	u
v	I	dʒ	i		v	o	z	o
v	e	v	i		v	o	z	u
v	ɛ	v	e		v	u	z	o
v	ɛ	v	i		v	u	z	u
v	i	v	e		v	ʊ	z	o
v	I	v	e		v	ʊ	z	u
v	I	v	i		z	o	dʒ	o
v	e	z	e		z	o	dʒ	u
v	e	z	i		z	u	dʒ	o
v	ɛ	z	e		z	u	dʒ	u
v	ɛ	z	i		z	ʊ	dʒ	o
v	i	z	e		z	ʊ	dʒ	u
v	I	z	e		z	o	v	o
v	i	z	i		z	o	v	u
v	I	z	i		z	u	v	o
z	e	dʒ	e		z	u	v	u
z	e	dʒ	i		z	ʊ	v	o
z	ɛ	dʒ	e		z	ʊ	v	u
z	ɛ	dʒ	i		z	o	z	u
z	i	dʒ	e		z	u	z	o
z	I	dʒ	e		z	ʊ	z	o
z	i	dʒ	i		z	ʊ	z	u

Affricate/ Fricatives	**Voiced**	**Unrounded**	**Front**
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**
z	I	dʒ	i
z	e	v	e
z	e	v	i
z	ɛ	v	e
z	ɛ	v	i
z	i	v	e
z	I	v	e
z	i	v	i
z	I	v	i
z	e	z	i
z	ɛ	z	e
z	ɛ	z	i
z	i	z	e
z	I	z	e
z	I	z	i

Affricate/ Fricatives	**Unvoiced**	**Unrounded**	**Front**		**Affricate/ Fricatives**	**Unvoiced**	**Rounded**	**Back**
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
f	e	f	i		f	o	f	u
f	ɛ	f	e		f	u	f	o
f	ɛ	f	i		f	ʊ	f	o
f	i	f	e		f	ʊ	f	u
f	I	f	e		f	o	s	o
f	I	f	i		f	o	s	u
f	e	s	e		f	u	s	o
f	e	s	i		f	u	s	u
f	ɛ	s	e		f	ʊ	s	o
f	ɛ	s	i		f	ʊ	s	u
f	i	s	e		f	o	tʃ	o
f	I	s	e		f	o	tʃ	u
f	i	s	i		f	u	tʃ	o
f	I	s	i		f	u	tʃ	u
f	e	tʃ	e		f	ʊ	tʃ	o
f	e	tʃ	i		f	ʊ	tʃ	u
f	ɛ	tʃ	e		s	o	f	o
f	ɛ	tʃ	i		s	o	f	u
f	i	tʃ	e		s	u	f	o
f	I	tʃ	e		s	u	f	u

Affricate/ Fricatives	Unvoiced	Unrounded	Front		Affricate/ Fricatives	Unvoiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**		**Consonant 1**	**Vowel 1**	**Consonant 1**	**Vowel 1**
f	i	tʃ	i		s	ʊ	f	o
f	I	tʃ	i		s	ʊ	f	u
s	e	f	e		s	o	s	u
s	e	f	i		s	u	s	o
s	ɛ	f	e		s	ʊ	s	o
s	ɛ	f	i		s	ʊ	s	u
s	i	f	e		s	o	tʃ	o
s	I	f	e		s	o	tʃ	u
s	i	f	i		s	u	tʃ	o
s	I	f	i		s	u	tʃ	u
s	e	s	i		s	ʊ	tʃ	o
s	ɛ	s	e		s	ʊ	tʃ	u
s	ɛ	s	i		tʃ	o	f	o
s	i	s	e		tʃ	o	f	u
s	I	s	e		tʃ	u	f	o
s	e	tʃ	e		tʃ	u	f	u
s	e	tʃ	i		tʃ	ʊ	f	o
s	ɛ	tʃ	e		tʃ	ʊ	f	u
s	ɛ	tʃ	i		tʃ	o	s	o
s	i	tʃ	e		tʃ	o	s	u
s	I	tʃ	e		tʃ	u	s	o
s	i	tʃ	i		tʃ	u	s	u
s	I	tʃ	i		tʃ	ʊ	s	o
tʃ	e	f	e		tʃ	ʊ	s	u
tʃ	e	f	i		tʃ	o	tʃ	u
tʃ	ɛ	f	e		tʃ	u	tʃ	o
tʃ	ɛ	f	i		tʃ	ʊ	tʃ	o
tʃ	i	f	e		tʃ	ʊ	tʃ	u
tʃ	I	f	e					
tʃ	i	f	i					
tʃ	I	f	i					
tʃ	e	s	e					
tʃ	e	s	i					
tʃ	ɛ	s	e					
tʃ	ɛ	s	i					
tʃ	i	s	e					
tʃ	I	s	e					
tʃ	i	s	i					
tʃ	I	s	i					
tʃ	e	tʃ	i					

Affricate/ Fricatives	**Unvoiced**	**Unrounded**	**Front**
Consonant 1	**Vowel 1**	**Consonant 1**	**Vowel 1**
tʃ	ɛ	tʃ	e
tʃ	ɛ	tʃ	i
tʃ	i	tʃ	e
tʃ	I	tʃ	e
tʃ	I	tʃ	i

Appendix 2: Auditory Stimuli for Experiment 2

Consonant Manner	**Consonant Manner**	**Vowel Voicing**	**Vowel Roundedness**	**C 1**	**V 1**	**C 2**	**V 2**	**Roundedness**
sonorants	sonorants	voiced	round/back	m	o	n	o	Very Rounded
sonorants	sonorants	voiced	round/back	n	u	m	o	Very Rounded
sonorants	sonorants	voiced	round/back	n	ʊ	l	o	Very Rounded
sonorants	sonorants	voiced	round/back	l	ʊ	n	o	Very Rounded
stops	obstruents	voiced	round/back	b	ʊ	b	u	Very Rounded
sonorants	sonorants	voiced	unround/front	n	ɛ	m	e	Less Rounded
sonorants	sonorants	voiced	unround/front	l	ɛ	m	e	Less Rounded
sonorants	sonorants	voiced	unround/front	n	i	m	i	Less Rounded
af/frics	obstruents	unvoiced	round/back	f	u	f	o	Less Rounded
af/frics	obstruents	unvoiced	round/back	f	ʊ	s	o	Less Rounded
af/frics	obstruents	unvoiced	unround/front	s	e	f	i	Neutral
af/frics	obstruents	unvoiced	unround/front	tʃ	ɛ	s	e	Neutral
sonorants	sonorants	voiced	unround/front	l	I	n	i	Neutral
af/frics	obstruents	unvoiced	round/back	tʃ	ʊ	f	u	Neutral
stops	obstruents	voiced	unround/front	g	ɛ	g	i	Neutral
stops	obstruents	unvoiced	unround/front	p	e	t	e	Less Pointed
af/frics	obstruents	voiced	unround/front	dʒ	e	v	i	Less Pointed
stops	obstruents	voiced	unround/front	g	e	d	e	Less Pointed
stops	obstruents	voiced	unround/front	d	I	d	e	Less Pointed
stops	obstruents	voiced	unround/front	d	i	b	i	Less Pointed
stops	obstruents	unvoiced	unround/front	t	i	t	e	Very Pointed
stops	obstruents	unvoiced	unround/front	p	I	k	e	Very Pointed
stops	obstruents	unvoiced	unround/front	k	i	t	i	Very Pointed
stops	obstruents	unvoiced	unround/front	k	I	p	i	Very Pointed
stops	obstruents	unvoiced	unround/front	t	ɛ	p	i	Very Pointed