Sound symbolism, speech identity, and size

Andrew Shibata

May, 2017

PhonLab, UC Berkeley

Advisors: Keith Johnson, Alice Shen

Second Reader: Susan Lin

Abstract

Sound symbolism is the hypothesized property for sounds to convey semantic meaning. Shinohara and Kawahara (2010) proposed that features of vowels (frontness, height) and obstruents (voicing) cause listeners to perceive words as either larger or smaller. Study 1 firstly replicates the original experiment then repeats the experiment using a speech perception paradigm. The speech perception experiment assesses whether listeners perceive sizes differently between spoken language and visual reading. The results from Study 1 were consistent with Shinohara and Kawahara (2010) except that words with /u/ were perceived as smaller in our results. We hypothesized that this result may be due to u-fronting which is an iconic feature of Californian English so we repeated both the written word and speech perception experiments in Study 2 with non-Californian English speakers. Our results support Shinohara and Kawahara's claims and suggest that speakers perceive dialect-specific phonetic properties from written word.

1. Introduction

1.1 *Traditional perspective on form and meaning*

One long-held assumption in linguistics is that the physical form of a linguistic signal is independent of its meaning. This idea is attributed to Ferdinand de Saussure who stated in his *Course in General Linguistics* that "the bond between the signifier and the signified is arbitrary" (De Saussure & Baskin, 2011). Saussure argues that the idea conveyed by a word is not linked to the ordered set of sounds that make up the word. He states that the existence of different words for an idea in different languages proves that the signifier is not tied to signified concept.

Charles Hockett listed arbitrariness as one of his design features of language (1959). He defined arbitrariness using a similar line of reasoning to Saussure's and claimed that it is an essential feature in determining whether a communication system is a language or not. Hockett

argues that the fact that language speakers can productively use an arbitrary communicative system proves that the system is a language.

While both Saussure and Hockett claim that human language arbitrary in nature, they both briefly acknowledge that there are words that seem to have a non-arbitrary connection between form and meaning: onomatopoeias.

1.2 *Onomatopoeia*

Onomatopoeias are words that try to capture sounds in the world. For example, the English onomatopoeia for the sound a cat makes is "meow". Onomatopoeias are problematic under Saussure's and Hockett's arbitrary systems because the physical signal (the word) is necessarily linked to the idea of the sound in the world. Saussure (2011) attempts to reconcile the existence of non-arbitrary onomatopoeias by noting the limited quantity and suggesting that the sounds that languages choose to label with an onomatopoeia are chosen arbitrarily. Hockett (1959) simply makes a distinction between iconic communication and arbitrary communication and states that "Human language is almost wholly arbitrary" (p.34). This study does not contest the claim that human language is predominately arbitrary, but seeks to focus on the problematic set of iconic words.

Bredin (1996) contemplates what about onomatopoeia makes the phenomenon so problematic. Definitions for onomatopoeia seem to agree that it describes the relationship between the sound of a word and some concept, but diverge when discussing how the relationship functions. Bredin identifies three categories for onomatopoeias: direct onomatopoeia, associative onomatopoeia, and exemplary onomatopoeia. Direct onomatopoeias are words like "bang", "buzz", and "splash" for which "the sound of the word resembles the sound that it names" (p. 558). Associative onomatopoeias are words that have sounds that

resemble a sound associated with the referent that the word denotes. For example, the word "cuckoo" refers to a bird, but resembles the sound that the bird makes. Exemplary onomatopoeias are words for which the meaning of the word is reflected in the articulatory effort used to produce the word. Bredin claims that words like "nimble" and "dart" require less effort than "sluggish" and "slothful" and the iconicity for these words are found within the qualities of the production. This third category does not seem to fit with the conventional understanding of onomatopoeias, but it introduces the idea that the phonetic attributes of a word can be iconic for the referent.

1.3 *Sound symbolism*

While onomatopoeia describes word-level iconicity, sound symbolism explores semantic information conveyed at the individual sound level. Ohala (1997) defines Sound Symbolism as the "hypothesized systematic relationship between sound and meaning". Ohala traces the idea for non-arbitrary relationships between the physical form of speech and meaning back to Plato's *Cratylus* dialogue (Sedley, 2003) in which Socrates argues that individual letters give words attributes of meaning. Ohala further supports the notion of sound symbolism through cross-linguistic data and an evolutionary argument for how sound symbolism may have shaped social communication.

Ohala begins with the claim that there are "phonetically natural classes of speech sounds that are systematically associated with expressions of size" (p. 1). He provides words and morphemes expressing the idea of "smallness" and "largeness" in languages such as Ewe, Yoruba, Spanish, English, and French. Within these languages, words for "small" seem to use [i] and [e] sounds while words for "large" seem to have [a] or [o]. While the list is not even

remotely exhaustive of languages in the world, there seems to be a distinction between vowels used to describe small things and vowels used to describe large things.

Ohala next turns to the domain of animal communication by arguing that pitch is associated with displays of aggression or submission. Animals that want to present themselves as intimidating or powerful can produce sounds with a low F0. Alternatively, animals that wand to appear submissive or non-threatening can use higher F0 productions. For example, dogs aggressively "growl" with a low F0, but can submissively "yelp" or "whimper", using a high F0. Ohala states that these forms of F0 modulation extend to human vocalizations for argumentation or cooperation.

Human facial displays convey important social information, but Ohala argues that facial displays such as smiles and "o-face" have acoustic origins. When people smile, they are trying to communicate cooperation and non-threat. Smiling effectively shortens the vocal tract by drawing back the mouth, allowing for higher resonant frequencies. Ohala defines "o-face" as a facial display involving lip protrusion and mouth constriction, used to convey hostility or disagreement. This set of changes effectively lengthens the vocal tract and conversely allows for lower resonant frequencies. Ohala suggests that these aggressive "o-face" productions may be relevant to human evolutionary sexual dimorphism and sexual selection. Historically, the male social role involves more aggressive and dominant behavior. Perhaps anatomical features contributing to longer vocal tracts improved a male's ability to secure resources such as food and territory for further producing offspring. Ohala combines statistical, cross-linguistic evidence with an anthropological argument for social communication to demonstrate the existence of sound symbolism.

Many other researchers have explored the role of sound symbolism in other domains such as perceived texture, speed, and weight. The earliest experimental evidence comes from Sapir (1929) in which he found that the vowel [a] was perceived as larger than [i] and that [a] was perceived as larger than [ɛ].

Many studies have investigated a perceived sharp vs. smooth distinction. Köhler (1970) first looked at this by showing Spanish speakers a spiky shape and a curvy shape and asking them which one was a 'takete' and which one was a 'maluma'. These speakers robustly named the spiky shape 'takete' and the curvy shape 'maluma', suggesting that there something about these names contributed texture meaning. Ramachandran & Hubbard (2001) repeated this experiment except with the words 'kiki' and 'bouba' and found that both English and Tamil speakers had similarly robust 'kiki'-spiky and 'bouba'-curvy associations. They hypothesized that the [k] stop feature for 'kiki' may be symbolic of sharpness and the [b] lip rounding may be connected to perceived smoothness or curviness. Westbury (2005) looked at the time-course involved for this texture categorization. He presented stop and continuant consonants within either a spiky or curvy frame image and asked subjects whether the target symbol was a letter or a number. Subjects correctly responded more quickly when stops were presented in spiky image frames and continuants were presented in curvy frames. It appears that texture sound symbolism is consistent across the studied languages and influences the time-course for semantic retrieval.

One very large study that explored sound symbolism across many domains is Klink (2000). Klink tested for sound symbolism in brand name meaning. He looked at distinctions between words with front vs. back vowels, stops vs. fricatives, voiced vs. voiceless stops, and voiced vs. voiceless fricatives. He investigated many domains such as perceived size, weight, gender, speed, and temperature. He found evidence that brand names that contained front

vowels, fricatives, voiceless stops, or voiceless fricatives were perceived to be smaller, faster, colder, more feminine, lighter, and many other qualities compared to their counterparts.

Table 1: Summary of significant sound symbolic domains from Klink (2000)

Creating brand names with meaning (**Klink, 2000**)		
Words with	**are perceived as _____ than**	**Words with**
Front Vowels	smaller, faster, lighter$_1$, milder, thinner, softer, colder, more bitter, more feminine, friendlier, weaker, lighter$_2$, prettier	Back Vowels
Voiceless Obstruents	faster, lighter$_2$, more feminine	Voiced Obstruents

1: lighter as opposed to darker; 2: lighter as opposed to heavier

The directions for these sound symbolic domains such as size and texture appear to be consistent across the literature (shown in **Figure 1**). Voiceless sounds, high vowels, and front vowels seem to be associated with properties that align with Ohala's predictions for "non-threatening" displays while voiced sounds, low vowels, and back vowels are associated with more intense or "aggressive" properties. While many domains have been found to have sound symbolic properties, this study focuses on the size domain.

Figure 1: Sound Symbolism directions for vowels and obstruents
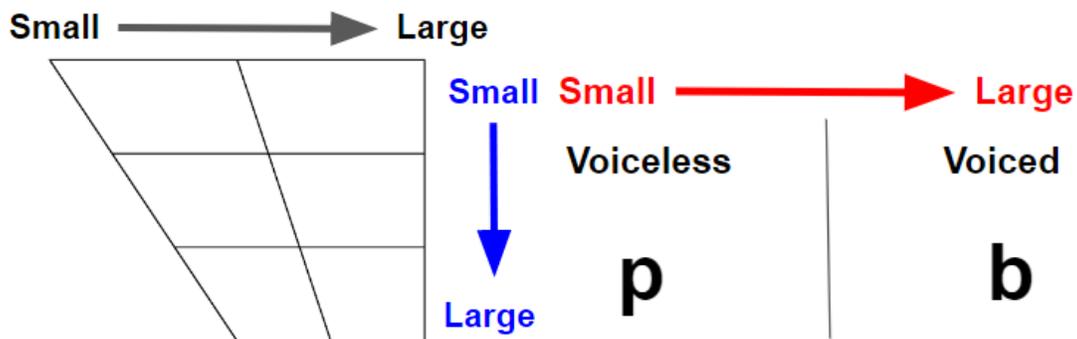
Objects with labels containing {front vowels, high vowels, voiceless obstruents} are perceived as smaller than objects with labels containing {back vowels, low vowels, voiced obstruents}

1.4 *Shinohara & Kawahara (2010)*

Shinohara & Kawahara (2010) investigated sound symbolism with respect to perceived size, specifically across the features of vowel height, vowel backness, and obstruent voicing. They tested Chinese, Japanese, Korean, and English speakers, but our current study will only concern the English speaker data. They asked subjects to read stimulus words and rate how large or small they perceived the object to be on a 1-4 Likert scale with 1 being 'very small' and 4 being 'very large'. The stimuli consisted of VCVC words with the consonants specifically being obstruents. Shinohara & Kawahara (henceforth referred to as "S&K") found that high vowels, front vowels, and voiceless obstruents are perceived as smaller than their counterparts. These findings are consistent with the directions in the literature (see **Figure 1**) and were observed for all the tested languages.

1.4.1 *Extending Shinohara & Kawahara (2010)*

While these results are consistent with the literature, there are potential confounds that our study seeks to address. It is unclear whether the stimulus words are intended to be IPA or English orthography since the symbols for the tested segments are identical for both systems. Subjects may be rating perceived sizes for non-standardized readings depending on whether they are familiar with IPA notation or not. For example, the letter 'a' could be read as [a], [æ], [ə], or [e]; each of which has very different phonological features.

Additionally, subjects' existing lexical inventory may bias them to associate nonce words with English words that look similar. S&K claimed that "All of [the stimuli] are nonce words in all the target languages" (p. 398), but that is not a true assumption. One notable example we encountered when replicating S&K's study was the word "isis". This was a particularly salient and meaningful word to our subjects due to its association with the violent extremist group.

Looking solely at the direction of size features indicated through the literature, [isis] is expected to be one of the smallest perceived words in the stimulus set because it contains [i], a [+high, +front] vowel, and [s], a voiceless obstruent. However, the word 'isis' is not associated with a small perceived size in conventional English usage. This potential bias can be explained through the TRACE model of speech perception (McClelland & Elman 1986) which describes how known lexical units influence how individual phonemes are perceived.

In summary, the stimuli used in the S&K study have ambiguous readings for English speakers and our current study seeks to build upon S&K's work by disambiguating the stimulus target words.

2. Study 1

Study 1 involves two components: a replication of the Shinohara & Kawahara (2010) study (Experiment 1), and a speech recognition experiment (Experiment 2) that addresses the concerns discussed above.

Experiment 1 was a replication of the S&K study. Subjects received a paper questionnaire formatted after the sample formatting and instructions outlined in the S&K paper.

Experiment 2 used the same instructions as Experiment 1, but asked subjects to listen to words before rating the perceived size. By presenting subjects with standard audio stimuli, it is unambiguous which word they are expected to rate.

2.1 *Hypotheses*

We expect to find results consistent with S&K, namely that vowel height, vowel frontness, and obstruent voicing are factors that influence perceived size. We expect to find a difference in perceived size between written word stimuli and audio stimuli.

2.2 *Methods*

2.2.1 *Participants*

There was a total of 40 included subjects (20 – Exp 1, 20 – Exp 2) and 2 excluded subjects who were not native English speakers[1]. All participants were UC Berkeley affiliated volunteers.

2.2.2 *Materials*

The Shinohara & Kawahara (2010) replication used a paper questionnaire. The speech perception experiment involved a Google Form questionnaire with the stimuli audio embedded as YouTube videos. Subjects were asked to use earphones so that the audio could be heard clearly.

2.2.3 *Stimuli*

The replication experiment used the instructions provided by S&K. The three changes for the computer questionnaire were changing the pronoun "she" to "he" to match the gender of the speaker, changing "read" to "listen to" to match the task at hand, and adding the instruction "Please use earphones so that you can hear the words properly." The instructions from S&K are:

> Imagine an exotic language that you don't know. The language has a rich lexical
>
> inventory of adjectives that express a variety of "largeness" or "smallness". Now, a
>
> speaker of this language looks inside a box and finds a jewel. She verbally expresses how
>
> large or small it looks using one of these adjectives. Your task is to read each of the
>
> following words and guess its meaning — i.e., how large or small it is.

[1] Native English speaker is defined here as having begun the process of learning English by age 3.

Both experiments used the same stimuli listed by S&K. They were VCVC nonce words with a vowel from the set {i, e, a, o, u} and a consonant from the set {p, b, t, d, s, z, k, g}. Crossing these two sets results in a total of 40 stimulus words shown below in **Table 2**.

Table 2: List of stimulus words

Vowels		Voiceless Consonants				Voiced Consonants			
		p	**t**	**s**	**k**	**b**	**d**	**z**	**g**
	i	ipip	itit	isis	ikik	ibib	idid	iziz	igig
	e	epep	etet	eses	ekek	ebeb	eded	ezez	egeg
	a	apap	atat	asas	akak	abab	adad	azaz	agag
	o	opop	otot	osos	okok	obob	odod	ozoz	ogog
	u	upup	utut	usus	ukuk	ubub	udud	uzuz	ugug

The speech perception experiment stimuli involve audio recordings of a volunteer who was blind to the research assumptions. S&K only specify that the stimuli are disyllabic, but do not discuss how the VCVC words were broken into 2 syllables. We chose to use V.CVC words rather than VC.VC words because they qualitatively sounded more natural. The audio recording took place in the UC Berkeley PhonLab using a sound-proof booth. We selected sound clips with productions that were similar in duration and intonation. We used Adobe Premiere to make video files with the text and audio and uploaded the video files to YouTube for embedding within the Google Form questionnaire.

At the end of both questionnaires, subjects were asked for current age, gender, age that they began learning English, and other languages spoken (both fluently or non-fluently). This demographics section primarily assessed whether the subject was a native English speaker or not.

2.2.4 *Procedure and design*

Subjects completed either a paper (Exp 1) or computer (Exp 2) questionnaire titled "Large or Small?". For each trial, subjects were asked to either read or listen to the stimulus word and rate the perceived size on a Likert scale of 1-4 from very small to very large.

We produced 2 randomized order lists and randomly assigned a list to each subject in Experiment 1 to address potential order effects between subsequent trials.

2.3 *Results and analysis*

2.3.1 *Comparing Experiment 1 and Experiment 2*

To evaluate the potential factor effects, we used a linear mixed effects regression model of the form:

Size (normalized) ~ Frontness * Height * Voicing + List + Stimulus Type

We chose to check the interactions for the acoustic properties that were crossed between the vowel and consonant stimulus features. We did not expect acoustic properties and stimulus list order to have a significant interaction and the null hypothesis for this study assumes that there is no interaction between stimulus type (written vs audio) and acoustic properties.

From the linear model summary, the factors that were found to be significant were Vowel Frontness ($p < 0.001$), Vowel Height ($p < 0.001$), and Obstruent Voicing ($p < 0.001$). No interactions were found to be significant (all $p > 0.1$). The List ordering was not found to be significant ($p = 0.847$), indicating that there is no significant effect of stimulus order on perceived size. Contrary to our expectations, there was not a significant effect due to Stimulus Type ($p = 0.89$).

Figure 2 below visually presents the mean perceived sizes by vowel between the Written Word experiment (Exp 1) and the Speech Perception experiment (Exp 2).

Figure 2: Comparing perceived size between written word stimuli and audio stimuli
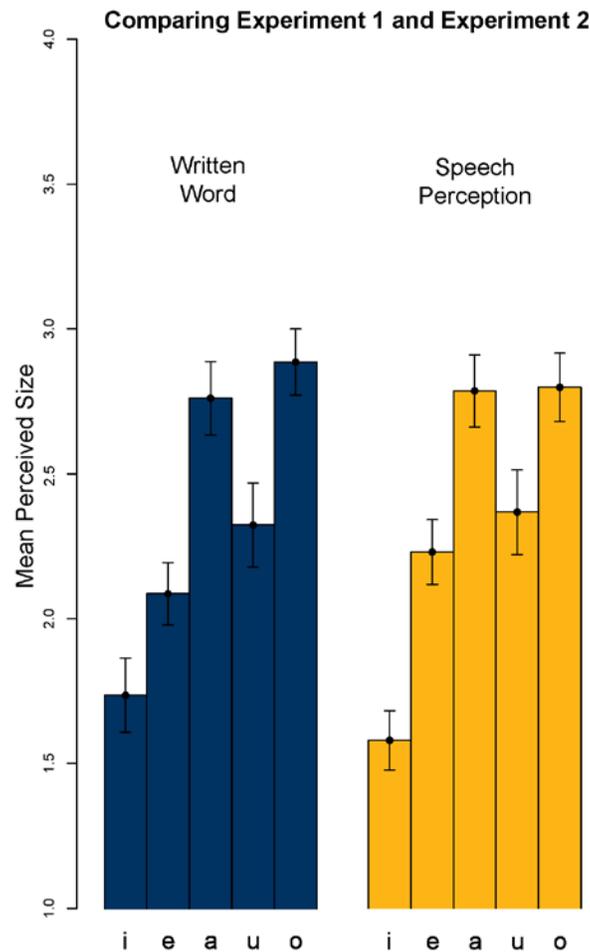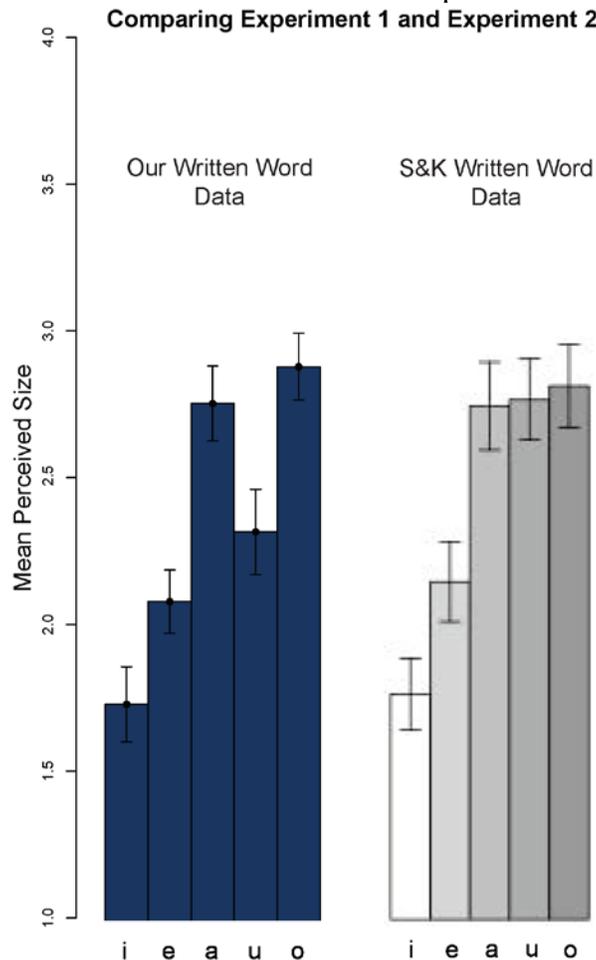

No observed difference in perceived size between written word and speech perception experiments

There is no difference in perceived size based on written versus audio stimuli. This is surprising because we expected that controlling for the ambiguous stimuli readings for written word would result in differences in perceived size. We next decided to compare our results with Shinohara & Kawahara's results.

2.3.2 *Comparing to Shinohara & Kawahara (2010)*

Figure 3 below compares S&K's results to the results from our written word replication. Since there was no difference between written word and speech perception and since S&K did not use a speech perception paradigm, comparing those results are not appropriate.

Figure 3: Comparing perceived size for written word in Experiment 1 and S&K results

Mean perceived sizes are comparable between our study and S&K's study except for /u/

While we did not have Shinohara & Kawahara's values from their experiment, there is a clear difference in perceived size for the vowel /u/. Assuming that sound symbolic features are binary, /u/ is interesting because it is both a [+ front] vowel (large perceived size) and a [+ high] vowel (small perceived size). Our results are consistent with the explanation that these features neutralize each other, causing /u/ to have a neutral perceived size. Shinohara & Kawahara account for a large perceived /u/ size by claiming that [- front] vowels involve more lip rounding (p. 403), effectively reducing vocal tract length and lowering pitch (similar to Ohala's 'o-face' description). We decided to carry out Study 2 to investigate the difference in perceived size for /u/.

2.4 *Discussion – Investigating smaller /u/ perceived size*

Since our results indicate that front vowels, high vowels, and voiceless obstruents correspond to small perceived sizes and the opposite features correspond to large perceived sizes, our results are consistent in direction with the literature (directions indicated in **Figure 1**). While there is no difference in results between Experiment 1 and Experiment 2, there is a striking difference between our results and S&K's results for /u/ perceived size.

In addition to the feature neutralization hypothesis outlined above, the smaller /u/ perceived size may attributable to u-fronting in Californian English. Since data collection took place at the University of California, Berkeley, most of the subjects in Study 1 were native Californian English speakers. One feature of Californian English that is highly salient is u-fronting (Hinton et al., 1987). Since vowel frontness is correlated with a smaller perceived size, perhaps the process of fronting /u/ to produce [ʉ] causes Californian English speakers to perceive [ʉ] as smaller compared to a fully backed [u]. We designed a second study investigating the potential influence of Californian u-fronting by repeating Study 1, except with non-Californian English speakers.

3. Study 2

The stimuli and procedure were identical to those in Study 1. To identify subjects as being non-Californian English speakers, we asked the question "Did you attend high school in California?" with the assumption that high school speech community is a major influence in shaping a speaker's dialect. While u-fronting is also present in Southeastern and Mid-Atlantic English (Labov et al., 2005), we designed Study 2 with the assumptions that u-fronting is a highly salient feature of Californian English and that non-Californian subjects were not consciously aware that u-fronting was the feature of interest when they participated.

3.1 *Hypothesis*

We expect to find a difference in perceived /u/ size between Californian English speakers (Study 1) and non-Californian English speakers (Study 2). More specifically, we expect that Californian English speakers perceive /u/ as smaller than non-Californian English speakers.

3.2 *Methods*

3.2.1 *Participants*

There was a total of 42 included subjects (22 – Exp 3a, 20 – Exp 3b) and 5 excluded subjects who were not native English speakers[2]. All participants were UC Berkeley affiliated volunteers. This study asked participants whether they attended high school in California or not as a determiner of speech identity. Subjects in Experiment 3a responded to written word stimuli and subjects in Experiment 3b responded to audio stimuli.

3.2.2 *Materials*

For ease of distribution, Experiment 3a was conducted using a Google Form rather than a paper questionnaire as in the analogous Experiment 1. This is admissible because stimulus list was not found to be a significant factor in Experiment 1. Experiment 3b used a Google Form identical to that of Experiment 2 except asking whether the subject attended high school in California or not in the demographics section.

3.3 *Results and analysis*

For Study 2 we focused on perceived size for /u/ so we compared the mean perceived sizes for /u/ between Californian English speakers (Experiment 1) and non-Californian English speakers (Experiment 3a) using a two-tailed t-test. This test yields a p-value < 0.001, indicating

---

[2] Native English speaker is defined here as having begun the process of learning English by age 3.

that Californian English speakers have smaller /u/ perceived size compared to non-Californian English speakers when responding to written stimuli.

Repeating the t-test for the Speech Perception Experiments (Experiment 2 and Experiment 3b) yields a p-value = 0.2421, indicating that Californian English speakers do not have a smaller /u/ perceived size compared to non-Californian English speakers when responding to audio stimuli. The speaker who recorded the audio stimuli is from California, so his speech included u-fronting. The result that CA and non-CA English speakers had comparable /u/ perceived sizes indicates that both CA and non-CA English speakers are sensitive to u-fronting and that they perceive similar object sizes from hearing [ʉ].

Figure 4: Comparing /u/ perceived size from written word between CA and Non-CA Speakers
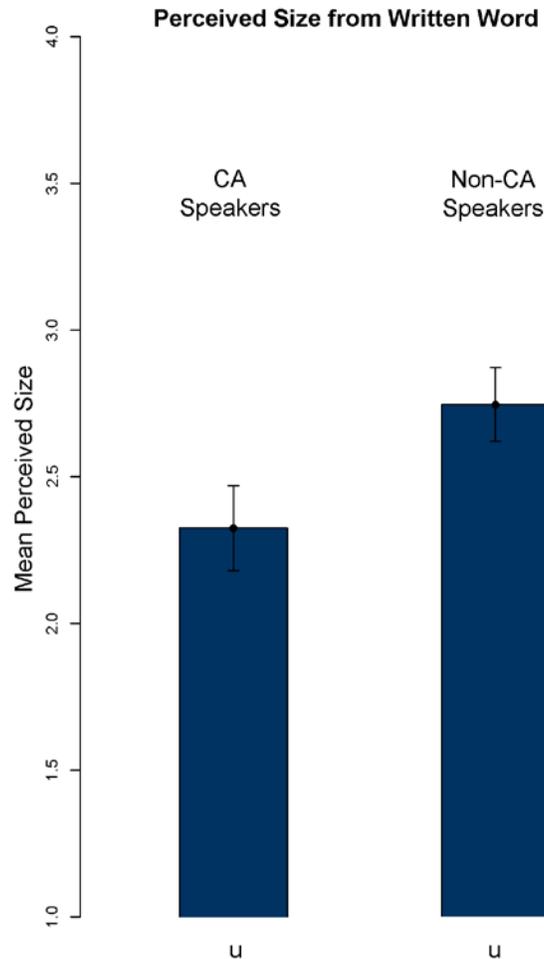

Californian English speakers perceive /u/ as smaller than non-Californian English speakers from written word

4. Conclusions and General Discussion

Our results suggest that u-fronting causes Californian English speakers to perceive objects as small when their labels contain an /u/. From this we conclude that English speakers perceive phonetic features that are characteristic of their dialect when reading written word and that these dialect-dependent features contribute to perceived referent size. While this current study focused on the relationship between sound symbolism and size, dialect features may influence other semantic domains such as perceived gender, speed, and other domains examined by Klink (2000).

The fact that Californian and non-Californian English speakers perceive /u/ size from different allophones provides insight into the order of sound symbolic processing. **Figure 5** compares expectations for two potential orderings that English speakers may process Fine Phonetic Detail, Sound Symbolism, and Phonlogical Abstraction. Our results suggest that Sound Symbolism is processed before Phonological Abstraction since /u/ perceived size is dependent on u-fronting and speakers did not perceive size solely from the phonological /u/.

Figure 5: Theoretical orderings for symbolic processing
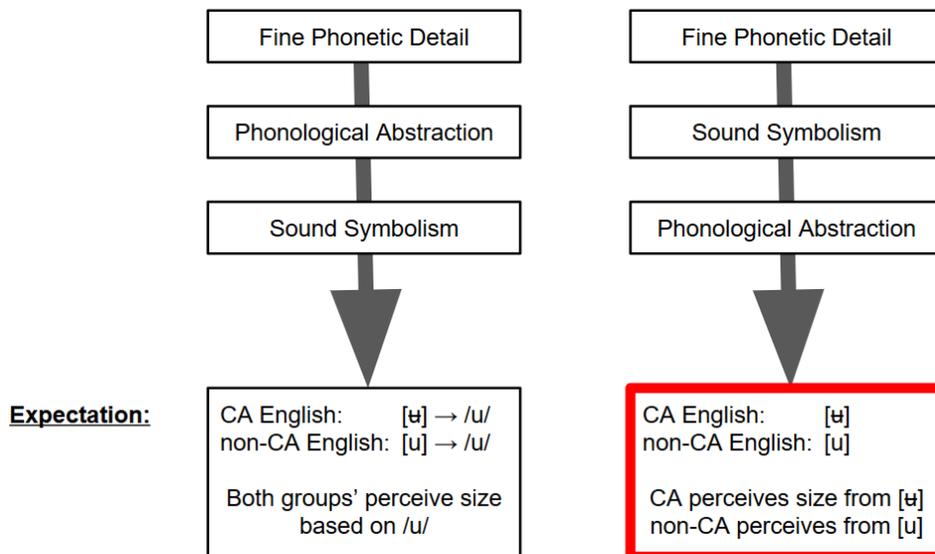
Sound Symbolism applies before Phonological Abstraction

It is particularly interesting that we observed this /u/ perceived size difference in written word since we could not control how subjects read the stimulus words. Many Californian English speakers are not overtly aware of their u-fronting, but our results indicate that Californian subjects consistently read the stimulus words with u-fronting. These results may extend to Vygotsky's notion of inner speech (1987, p. 256) or theories on subvocalization (Cleland & Davies, 1963). When Californian English speakers read words with /u/, their inner voices produce readings with u-fronting. In this sense, their inner voices are also Californian English speakers. From a subvocalization perspective, Californian English speakers may produce subvocal articulations for /u/ that are fronted, which then contribute to a smaller perceived size. Further research could explore the link between production and symbolic perception by explicitly asking subjects to vocalize their readings before rating perceived size.

The insignificant differences between perceived sizes in Study 1 make sense following the results from Study 2. Experiment 3a demonstrated that Californian English speakers perceive written /u/ with u-fronting and Experiment 3b demonstrated that Californian English speakers perceived the u-fronting in the audio stimuli. Since most participants in Study 1 were Californian English speakers, their /u/ perceived sizes should be comparable between written stimuli and audio stimuli. Perhaps there would be a difference within Californian English speakers' /u/ perceived size if the audio stimuli were from a non-Californian English speaker who did not have u-fronting in their speech.

While our study focused on the u-fronting feature characteristic of Californian English, more research is necessary to assess to what degree dialect-dependent size perception applies for features of other English dialects. Further Sound Symbolism research can investigate just how differently speakers of various English dialects perceive objects in the world.

6. Acknowledgements

7. References

Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. *New Literary History*, *27*(3), 555-569.

Cleland, D. L., & Davies, W. C. (1963). Silent Speech: History and Current Status [with Comment]. *The Reading Teacher*, *16*(4), 224-228.

De Saussure, F., & Baskin, W. (2011). *Course in general linguistics [1916]*. New York, NY: Columbia University Press.

Hinton, L., Moonwomon, B., Bremner, S., Luthin, H., Van Clay, M., Lerner, J., & Corcoran, H. (1987, September). It's not just the Valley Girls: A study of California English. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 13, pp. 117-128).

Hockett, C. F. (1959). Animal "languages" and human language. *Human Biology*, *31*(1), 32-39.

Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, *11*(1), 5-20.

Köhler, W. (1970). *Gestalt psychology: An introduction to new concepts in modern psychology*. WW Norton & Company.

Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

Ohala, J. J. (1997). Sound symbolism. *Proc. 4th Seoul International Conference on Linguistics [SICOL]* 11-15 Aug 1997. 98-103.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--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*(3), 225.

Sedley, D. (2003). *Plato's Cratylus*. Cambridge University Press.

Shinohara, K., & Kawahara, S. (2010, August). A cross-linguistic study of sound symbolism: The images of size. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 36, No. 1, pp. 396-410).

Vygotsky, L. S., & Rieber, R. W. (1987). *The collected works of LS Vygotsky: Volume 1: Problems of general psychology, including the volume Thinking and Speech* (Vol. 1). Springer Science & Business Media.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and language*, *93*(1), 10-19.