Distribution Agreement

Daphne Weiss April 2, 2021

Speaking Big and Small:

Integrating Prosodic Size and Meaning in Spoken Language

by

Daphne Weiss

Dr. Lynne Nygaard
Adviser

Neuroscience and Behavioral Biology

Dr. Lynne Nygaard

Adviser

Dr. Jack Hardy

Committee Member

Dr. Hillary Rodman

Committee Member

2021

Speaking Big and Small:

Integrating Prosodic Size and Meaning in Spoken Language

by

Daphne Weiss

Dr. Lynne Nygaard
Adviser

Neuroscience and Behavioral Biology

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

2021

Abstract

Speaking Big and Small: Integrating Prosodic Size and Meaning in Spoken Language

By Daphne Weiss

Spoken language is a rich and multidimensional signal, combining linguistic units with prosodic features like pitch, tone, and speech rate. Although these two information streams are sometimes thought to be separate, a growing body of evidence suggests that prosodic cues may instead interact with lexical processing and carry referential meaning. Such a phenomenon could be grounded in cross-modal sound-to-meaning correspondences, which link acoustic features to specific perceptual characteristics of a word's referent. One relevant feature may be an object's size, which is known to be a key element of visual representations, and which is associated with consistent prosodic characteristics. Size has also been implicated in lexical processing of written language: S. C. Sereno et al. (2009) found a processing advantage for written words with big meaning, though the role of semantic size in spoken language processing has not been investigated. This study investigates the broader role of prosody in spoken language processing by asking whether semantic size, size-specific prosody, or an interaction of the two factors influence lexical access. We performed a lexical decision task in which 45 participants judged whether an utterance was a word. Stimuli were concrete nouns associated with big and small meanings, as well as matched pseudowords, spoken in large, small, or neutral prosody. Acoustic analysis confirmed that the stimuli differed in pitch and duration by condition. We predicted that words with big meanings would be processed faster, consistent with S. C. Sereno et al. (2009). We further predicted that, if prosody interacts with lexical processing, size-congruent prosody would speed word recognition. We found that participants responded significantly faster to words with big meaning, with no decrease to accuracy. We also found a main effect of prosody in which participants responded more quickly to words produced with small prosody. However, we did not observe the expected interaction between prosody and semantic size. These results suggest that semantic size is automatically accessed during spoken word recognition, a relationship that could support a grounded cognition account of spoken word processing. More research is necessary to disambiguate the role of prosody.

Speaking Big and Small:

Integrating Prosodic Size and Meaning in Spoken Language

by

Daphne Weiss

Dr. Lynne Nygaard
Adviser

Neuroscience and Behavioral Biology

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

2021

Acknowledgements

My sincerest thanks to Dr. Nygaard for her mentorship—and for recording 1,500 stimuli right before COVID-19 forced us out of the lab. Thanks also to the members of the Speech and Language Perception Laboratory, especially Kaitlyn Matthews for her gracious and frequent technical assistance, and to Dr. Jack Hardy and Dr. Hillary Rodman for their guidance as committee members. Finally, I am deeply grateful to the Center for Mind, Brain, and Culture for supporting this project, and to my cohort of CMBC undergraduate fellows, who have provided excellent feedback and support.

Contents

List of Figures

List of Tables

Introduction

Non-Arbitrariness in Language

Arbitrariness is widely considered a fundamental property of language. Many assume that, with limited exceptions, there is no fundamental relationship between the form of a lexical sign (such as the sound of a word) and its meaning (de Saussure, 1956; Pinker & Bloom, 1990). Most words, by this understanding, have meaning only by convention. There is no "correct" sign for a given meaning: the Spanish *perro* is no more nor less correct than the English *dog* or the French *chien*. Along the same lines, there is no "natural" meaning to most strings of sounds.

Without some degree of arbitrariness, the complexity of human language would likely be impossible (Gasser, 2004). Consider, for instance, if every word had to sound like its meaning. To build an entire lexicon with this limitation would be infeasible. However, this understanding is now sometimes taken to its logical extreme, with some believing that convention *and only convention* determines the meaning of a word (Monaghan, Shillcock, Christiansen, & Kirby, 2014; Perniss, Thompson, & Vigliocco, 2010). By this account, the mental lexicon acts essentially as a dictionary—Levelt (1993) conceptualizes phonetic representations as "access keys" mapped to meanings (p.8). The sounds themselves are irrelevant, and spoken sounds are processed with no feedback from other lexical or higher-order perceptual processes (Levelt, 1993). All that matters, by this model, are learned associations between sound and meaning.

The interpretation of form as inconsequential ignores converging evidence for iconicity in spoken language. Iconicity refers to a relationship between linguistic sign (which may be a spoken signal, a written word, or a physical movement) and its meaning (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Perniss et al., 2010). Considering only acoustic signs, systematic sound-to-meaning mappings seem to exist between properties such as hue and pitch or vowel shape and physical shape, among many others (Spence, 2011). These instances of iconicity are generally thought to be grounded in connections between a concept's perceptuomotor properties and its lexical sign (Dingemanse, Schuerman, Reinisch, Tufvesson, & Mitterer, 2016; Perniss et al., 2010).

The most frequently cited and studied example of a systematic sound-to-meaning mapping is the bouba-kiki effect, originally described by Köhler (1929) and extended by Ramachandran and Hubbard (2001). The bouba-kiki effect describes participants' tendency, when asked to name a new shape with a pseudoword, to associate names like "kiki" or "takete" with pointed shapes, and "bouba" or "malouma" with rounded ones (Köhler, 1929).

Although some (Cuskley, Simner, & Kirby, 2015) argue that orthography may confound claims of iconicity in the bouba-kiki effect, since *bouba* looks rounder than *kiki*, even infants show sensitivity to the bouba-kiki effect (Ozturk, Krehm, & Vouloumanos, 2013). Additionally, the association has been documented across cultures and language backgrounds, including among a remote Namibian population that has little exposure to Western languages and that does not use written language (Bremner et al., 2013). The pervasiveness of this naming effect, as well as its emergence early in development, suggest that an innate cross-modal correspondence could underlie the phenomenon.

McCormick, Kim, List, and Nygaard (2015) and others have attempted to explain the bouba-kiki effect in terms of sound symbolism, which refers to a specific kind of iconicity in which the phonemes, or sound segments, of a word are thought to represent some aspect of that word's meaning. McCormick et al. (2015) asked participants to rate pseudowords for their roundedness or pointiness. They found that, across both forced-choice and Likert rating tasks, participants consistently associated certain consonants like /t/ and unrounded vowels with pointiness. Conversely, they associated rounded vowels with round shapes (McCormick et al., 2015). These associations could explain why people tend to ascribe names like "kiki" to pointy shapes and names like “bouba” to rounder ones.

It’s important to note that “bouba” and “kiki” are pseudowords chosen intentionally to *sound* round and pointy. Thus, while this consistent naming effect indicates that certain sounds may be linked to certain meanings, it does not on its own speak to iconicity in natural language.

To address this limitation, Blasi, Wichmann, Hammarström, Stadler, and Christiansen (2016) reported an extensive analysis of 6,400 word lists representing 62% of the world’s languages and 85% of its lineage. They identified 74 distinct sound-meaning associations associated with meanings related to body parts, shapes, and concepts such as *big* and *small* (Blasi et al., 2016). This analysis shows that sound-symbolic mappings are present in natural language, which could suggest that there is some benefit to sound symbolic language that favors its development during language evolution.

Sources and Benefits of Iconicity

Some degree of arbitrariness is integral to human language (Gasser, 2004). Aribitrariness extends the set of possible words whereas iconicity constrains it, so without arbitrary form-meaning mappings, a system would lack the communicative flexibility that characterizes human language (Dingemanse et al., 2015). At the same time, Lockwood and Dingemanse (2015) argue that iconicity presents its own advantages, with arbitrariness making communication more *efficient* but iconicity making it more *effective* by grounding abstract language in familiar perceptual space.

There is evidence, for instance, that iconicity facilitates word learning and recognition (Lockwood, Dingemanse, & Hagoort, 2016; Tzeng, Nygaard, & Namy, 2017). Perry, Perlman, Winter, Massaro, and Lupyan (2017) also reported that early vocabulary tends to contain a relatively higher proportion of sound symbolic words, as does speech that adults direct at young children, implying that sound symbolism could facilitate language acquisition. These patterns suggest that iconic mappings might develop and persist in natural language because words associated with such mappings are easier to process.

The Role of Prosody

Although sound-to-meaning mappings are often studied on small segments of speech, such as individual syllables, it's possible that prosodic countours could also convey referential information in spoken language. Prosody refers to features such as stress, pitch, tone, and speech rate—the properties of spoken language that arise when combining phonemes into longer utterances.

There is evidence that prosody conveys meaning in certain contexts, a function that could draw on cross-modal correspondences between prosodic properties and meaning. For instance, Nygaard and Queen (2008) found that words with emotional meanings are processed faster when produced in a tone of voice corresponding to that meaning. For instance, a happy word like *cheer* is processed more quickly when it is produced in a happy emotional tone than when it is said in a sad tone (Nygaard & Queen, 2008).

Considering non-emotional prosodic features, Shintel and Nusbaum (2008) reported a processing advantage when speech rate was congruent with perceived motion. Participants listened to sentences spoken at a “fast” or “slow” speech rate. They were then shown images and asked whether the subject of the image matched the subject of the sentence. In the congruent motion condition, fast speech was matched with an image implying rapid motion. For instance, given the sentence “The horse is brown” spoken quickly, the congruent condition would show a brown horse in mid-stride; the incongruent condition would show a brown horse standing still. Participants were significantly faster to identify the subject for congruent fast-speech/moving-object trials compared to fast-speech/static-object trials. This trend could mean that the faster speech rate activated the concept of “speed” in the same way that the moving-object images did. If so, then the congruency effect would suggest that prosody can influence the way that listeners automatically activate mental representations in spoken language processing (Shintel & Nusbaum, 2008). That speech rate can evoke a sense of physical speed would be highly suggestive of a cross-modal correspondence and, perhaps, an iconic mapping unique to spoken language.

In another study of prosodic correlates to meaning, Tzeng, Duan, Namy, and Nygaard (2018) found that, when asked to produce novel names for colors varying in brightness, speakers used higher pitch and amplitude for brighter color swatches. Listeners performed above chance when asked to associate these recordings with the corresponding color (Tzeng et al., 2018). This finding adds to a broader body of literature suggesting that speakers draw on cross-modal associations in producing prosodic cues, and that listeners are sensitive to these differences.

These studies already suggest that prosody can serve a referential function and may be implicated in iconicity. Some have gone even further to argue that prosody is inseparable from sound-symbolic cross-modal mappings. For instance, Dingemanse et al. (2016) reported that listeners performed above chance when guessing the meaning of iconic words, but only when exposed to the utterances that contained both segmental and prosodic cues to meaning. When participants heard spoken utterances resynthesized to preserve either only segmental or only prosodic information, they were unable to use iconicity as a cue to meaning (Dingemanse et al., 2016). This study offers evidence that prosody informs word meaning through non-arbitrary sound-to-meaning mappings, and that such prosodic cues may be necessary to derive a processing advantage from iconic words (Dingemanse et al., 2016).

All of these findings speak to a larger debate about the role of prosody in speech processing. To understand speech, the brain must rapidly process both linguistic cues, or the words and grammatical structures, alongside non-linguistic cues like gestures, speaking style, and emotion tone of voice. Under this framework, prosody has traditionally been considered non-linguistic and therefore processed separately from linguistic content. If emotional tone, speech rate, and similar markers are parsed separately from meaning, there's no mechanism by which they should influence the processing of word meaning. That we observe congruency effects based on prosodic features suggests not only that prosody may exhibit iconicity, but also that prosody may serve as a carrier of meaning in spoken language.

Size as a Semantic Variable of Interest

Iconic prosody, to the extent that it exists, links acoustic features to salient perceptual characteristics of a word's referent. Given the importance of size in cognitive representations, it's likely that size is one such perceptual feature.

Konkle and Oliva (2012) showed that, when presented with images of common objects and asked to indicate which image is larger, participants respond more quickly when the relative image sizes were congruent with their real-world sizes. This finding suggests that size is automatically activated in object recognition. Similarly, Sensoy, Culham, and Schwarzer (2020)'s study of infant looking times found that children as young as 12 months can successfully identify the appropriate size for familiar objects. The fact that children so young are sensitive to real-world object size highlights the importance of size in perceptual processing.

The brain clearly stores object size as a component of mental representations, and there is evidence that this property is accessed during visual processing. Although the evidence is less conclusive, size may also be activated during lexical processing. S. C. Sereno, O'Donnell, and Sereno (2009) proposed semantic size, or the real-world size of a written word's meaning, as potentially being activated in written word recognition. Participants were asked to identify whether a written stimulus was a word or a pseudoword. Half of the actual words were bigger than a human body, such as *river*, and half were smaller, such as *robin*. They found that words referring to a big object were recognized faster than words with smaller referents, or meanings. This suggests that semantic size, or the real-world size of a word's referent, may impact lexical processing speed (S. C. Sereno et al., 2009).

Yao et al. (2013) replicated and extended these findings with a separate set of stimuli representing both concrete and abstract words. Their larger stimulus set comprised words rated by participants on scales of both largeness and abstractness. In a standard lexical decision task following the same protocol as S. C. Sereno et al. (2009)'s, participants responded more quickly to words with a big meaning across both concrete and abstract conditions. Participants also responded more accurately to big words than to small ones (Yao et al., 2013).

Complicating the picture is Kang, Yap, Tse, and Kurby (2011)'s inability to replicate S. C. Sereno et al. (2009) despite using the same stimuli and a larger sample. Although they reported a numerical advantage for reaction time to "big" words, the difference was not statistically significant (Kang et al., 2011).

In a more applied setting, Wei and Cook (2016) found that, while semantic size did not affect early word processing in a reading task, there was a congruency effect based on semantic size. They found that total re-reading time was greater when the semantic size was incongruent with the sentence's context (for instance, "The *bag* contained the bear she had just seen.") than when it was congruent ("The *cage* contained the bear she had just seen.") Their results, again, suggest that real-world size is implicated in word recognition.

Despite somewhat contradictory evidence, S. C. Sereno et al. (2009), Wei and Cook (2016), and Yao et al. (2013) present compelling evidence that the brain may activate semantic size during lexical processing. Proponents of grounded cognition as described by Barsalou (1999), would argue that such a lexical processing advantage is intrinsically linked to sensory processing. Grounded cognition theory posits that hearing or reading a word activates sensory-motor areas associated with the word's referent. By this account, which S. C. Sereno et al. (2009) cite as a possible explanation for their results, hearing a big word simulates seeing whatever that big word refers to (Barsalou, 1999). If the brain automatically activates relevant sensory areas when processing language, then it holds that a processing advantage for a certain feature in one domain (such as visual processing) might impart a corresponding advantage to language processing, since the sensory simulation occurs faster.

Although visual processing is highly complex, some evidence points to such an advantage for big objects that could carry over to lexical processing: in humans, visual gamma band response shows sensitivity to stimulus size (Busch, Debener, Kranczioch, Engel, & Herrmann, 2004), and in primates, the magnocellular pathway processes low spatial-frequency information more quickly; since large objects convey more visual information in the same timeframe relative to smaller objects, representations of large objects may be accessed more quickly (M. Sereno, 1993; S. C. Sereno et al., 2009). In a grounded cognition framework, such findings, even if inconclusive, point to the possibility that semantic size facilitates processing of lexical stimuli.

Prosodic Cues to Size

Whether or not grounded cognition plays a role in word recognition, there is other evidence that could connect semantic size to spoken language.

Nygaard, Herold, and Namy (2009) reported that speakers naturally produce characteristic prosodic features corresponding to object size, and that participants may use these cues to disambiguate the meaning of novel pseudowords. Nygaard et al. (2009)

asked participants to produce pseudowords as though they were referring to a large or small object using infant-directed speech. Speakers' utterances differed significantly in pitch, amplitude, and duration as a function of size. Specifically, large objects were conveyed with longer spoken utterances with a lower fundamental frequency, or pitch (Nygaard et al., 2009). Not only did the utterances differ on these metrics, but listeners responded above chance when asked to guess the size of a fake object based on these pseudoword utterances. If participants can recognize prosodic cues to size without prompting, then there may be natural sound-to-meaning mappings between acoustic features and semantic size. Moreover, listeners may be sensitive to such associations and use them to disambiguate meaning.

The use of prosody to convey size may also extend to naturalistic communicative contexts. In a story reading task, Perlman, Clark, and Falck (2014) showed that participants read texts in a lower average pitch if the text contained phrases with big meaning vs. otherwise-identical texts with phrases of small meaning. Similarly, in a corpus of social media posts, posters more frequently duplicated vowels in the larger word of an adjective pair than the smaller one (ie., posters commonly wrote “biiig” but rarely wrote “smaaall”) (Fuchs, Savin, Solt, Ebert, & Krifka, 2019). Although the latter example derives from written texts, it seems to illustrate an attempt to convey size using a simulation of “prosody”—in this case, with reduplicated vowels suggesting increased utterance duration.

Taken together, these studies clearly point toward a link between prosody and semantic size, although the exact nature of the relationship is unclear. Given that listeners associate certain acoustic features with larger size, it's possible that speakers draw on these perceptual sensitivities to produce prosody that matches semantic size. By the same token, if speakers use prosody to convey referential meaning, then listeners may integrate these prosodic cues with word meaning to facilitate spoken language processing.

The Present Study

When and how prosodic cues to meaning influence lexical processing remains an open question with implications for how language is represented in the brain. This study attempts to contribute to this broader question by determining whether an iconic mapping between prosody and semantic size is automatically activated in spoken language processing.

First, we ask whether semantic size imparts a lexical processing advantage during the recognition of spoken words. Previous research suggests that semantic size is one feature of lexical items that influences how quickly a written word is recognized (S. C. Sereno et al., 2009; Wei & Cook, 2016; Yao et al., 2013). We extend this examination to spoken language.

Second, we are interested in the broader effect of prosody on lexical processing and the role of cross-modal cues to size as a specific kind of iconic mapping in spoken language. We specifically consider whether and how prosodic cues to size interact with word meaning (or semantic size) during lexical processing.

To evaluate our research questions, we recorded spoken words referring to big and small objects in stereotypical large and small prosody, as described in Nygaard and Queen (2008). We then conducted a lexical decision task in which participants indicated, as quickly as possible, whether each utterance was a real world or a pseudoword.

A lexical decision task is a standard tool in psycholinguistics research thought to measure lexical processing speed. In a lexical decision task, participants hear or read a stimulus, which is either a word or a pseudoword. They are asked to respond as quickly as possible as to whether the stimulus is a valid word. Since participants respond so quickly, it’s thought that reaction times represent the influence of automatic processes rather than conscious ones. Thus, manipulating variables in the lexical decision can indicate whether the manipulated variable impacts automatic lexical processing (Chumbley & Balota, 1984).

In this case, we presented stimuli that were either size-congruent (big meaning/large prosody, small meaning/small prosody) or size-incongruent (big meaning/small prosody, small meaning/big prosody). This set of stimuli allowed us to assess whether semantic size, prosodic size, or their interaction influence lexical processing.

If prosody influences spoken word recognition in this task, then it would suggest that prosodic information is integrated relatively early in language processing. If prosodic features are only processed after a word is recognized, as Levelt (1993) and others suggest, then prosody should should not influence lexical decision time.

Aim 1: Does Semantic Size Influence Spoken Word Recognition?

If semantic size influences lexical processing, as S. C. Sereno et al. (2009) and Yao et al. (2013) suggest, the effect may extend to spoken language.

> **H1:** If semantic size is activated during processing of spoken utterances, then there will be an advantage for spoken words with large meanings compared to small meanings during lexical access.

If humans naturally access sensory perceptual characteristics of a word’s meaning during word recognition, then there may be an advantage for spoken words that refer to bigger, easier-to-image referents.

If there is a lexical processing advantage for words with big meanings, then we expect that participants will more quickly recognize words that refer to big items than small ones. Average reaction times in a lexical decision task should show faster reaction times to words with big meanings.

Aim 2: How Does Prosody Interact With Word Recognition?

Although we expect to replicate S. C. Sereno et al. (2009)'s finding that large words carry a processing advantage, we also consider whether prosody might affect this relationship.

If there is also an iconic or cross-modal mapping between prosody and semantic size, then it's possible that these mappings could impact lexical processing speed. Thus, if there is an cross-modal mapping for size and prosody, and if that mapping influences lexical processing, then we expect a congruency effect in which participants will respond faster if a stimulus' prosodic condition matches its semantic size.

> **H2:** If prosody conveys information about the size of the reference and is integrated with the semantic size of a word or its meaning, then prosody that is congruent with word meaning should facilitate lexical processing.

If humans naturally produce (Fuchs et al., 2019; Nygaard et al., 2009; Perlman et al., 2014) and attend to (Nygaard et al., 2009) prosody rooted in semantic size, then it's plausible to expect that a sound-to-meaning mapping links certain prosodic characteristics with semantic size.

In the context of this study, that would mean that a big word said in large prosody (low pitch, longer duration) should be recognized faster than a small word said in large prosody. Likewise, a small word said in small prosody (high pitch, short duration) should be recognized faster than a big word said in small prosody.

Critically, unlike preceding studies of prosody in lexical processing, we forced our participants to make their judgements purely on the basis of auditory information; we did not elicit congruency judgements using visual stimuli, nor did we present written words. This novel study design means that any observed relationship between prosody and meaning must be based purely on auditory information. An interaction effect, therefore, could be taken as stronger evidence of a cross-modal correspondence rooted in iconic prosody.

Predictions

In sum, if there is a main effect of semantic size on spoken word recognition, then big words should be recognized faster across prosodic conditions. If there is an iconic mapping between prosody and semantic size, then there should be an interaction effect between prosodic size and semantic size such that congruent prosody imparts a processing advantage. The net effect is that, if both hypotheses hold, participants should respond the fastest to congruent large prosody, big size recordings.

Stimulus Generation & Acoustic Analysis

Before conducting our primary experiment, we first collected and analyzed recordings of words with big referents (Big words) and words with small referents (Small words) with both large and small prosodic features. We conducted acoustic analysis to verify that the stimuli differed on the expected acoustic measures.

Materials & Methods

Stimuli

A total of 278 words were drawn from S. C. Sereno et al. (2009) and Yao et al. (2013)'s sets of stimuli (see Appendix). The Big and Small words taken from S. C. Sereno et al. (2009) were matched for length, number of syllables, and frequency. The words from Yao et al. (2013) were also matched for age of acquisition.

Matching pseudowords were generated using the program Wuggy, which creates pseudowords that resemble valid English words (Keuleers & Brysbaert, 2010).The resulting pseudowords had the same number of syllables as their corresponding words.

Participants

One female, native speaker of American English recorded the set of words and pseudowords. The speaker was recruited from the Emory University Speech and Language Perception Laboratory and was not a naive participant with respect to the aims of the experiment.

Procedure

The speaker was presented with the list of words and pseudowords in random order and was instructed to pronounce each word in three ways: as they normally would, as if to convey that the word referred to something small, and as if to convey the word referred to something big.

The participant recorded stimuli in random order in a sound-attenuated room, using a Zoom 2 Cardioid microphone. The utterances were recorded and digitized at a 44.1 kHz sampling rate using Audacity software. Each utterance was then down- sampled at 22.05 kHz, which is standard for speech, and amplitude-normalized using Audacity.

Acoustic properties for each recording were obtained using Jadoul, Thompson, and De Boer (2018)'s Parselmouth Python library, which interfaces with the Praat phonetics software package (Boersma & Weenink, 1996). We wrote a Python script to query from Praat, for each utterance, measures of fundamental frequency (F_0), F_0 variation (as measured by F_0 SD), and utterance duration.

F_0 measures the number of oscillations per second in a periodic sound, which listeners interpret as pitch. The standard deviation of F_0 indicates the degree of variability in pitch across a single utterance. F_0 and F_0 SD are reported in Hertz (Hz). Duration is the length of the utterance in ms and indicates speaking rate or pronunciation speed.

Results & Discussion

Based on Nygaard et al. (2009)'s acoustic analyses of prosodic correlates to object size, we expected our speaker to produce large prosody utterances with a lower F_0 and a longer duration than small prosody utterances. Nygaard et al. (2009) did not report a difference in F_0 SD between large and small prosody.

To determine whether large and small prosody productions differed with respect to these acoustic measures, we conducted separate two-way analyses of variance with prosody as a within-item factor and word meaning as a between-item factor for each acoustic measure, (F_0 , F_0 SD, and duration). In all analyses, each item represented a distinct utterance. We used SPSS software to conduct the ANOVA tests at a significance level of 0.05.

Fundamental frequency F_0 . As expected, F_0 differed significantly as a function of prosody type, $F(2,308) = 936.825$, $p < .001$, $\eta_p^2 = .859$ (Table 1). Mean F_0 was significantly higher for small prosody than for neutral ($p < .001$) or large prosody ($p < .001$). Mean F_0 was significantly lower for words produced with large prosody than for words produced with neutral prosody ($p < .001$).

Duration. Duration differed significantly as a function of prosody, consistent with the findings of Nygaard et al. (2009), $F(2, 308) = 15.446$, $p < .001$, $\eta_p^2 = .091$ (see Table 1). The mean duration for words spoken with large prosody was significantly longer than for words produced with both neutral ($p < .001$) and small prosody ($p < .001$). Mean duration did not differ between words spoken with neutral or small prosody ($p > .294$).

F_0 Variability. We observed a significant main effect of prosodic condition on F_0 variability, $F(2,308) = 350.301$, $p < .001$, $\eta_p^2 = .695$ (Table 1). Small prosody utterances had the most variation in pitch ($M = 131.83$ Hz). F_0 SD for words produced with small prosody differed significantly from F_0 SD for words produced with both neutral ($M =$

49.65 Hz; *p*< .001) and large (*M*= 46.37 Hz; *p*< .001). F_0 variation did not differ for words produced with large or neutral prosody (*p*> .330). Nygaard et al. (2009) did not report a difference in F_0 variability as a function of size when averaged across three speakers, so it is possible that individual speakers instantiate prosodic cues to size with different sets of acoustic properties. Nevertheless, the differences in F_0 and duration as a function of prosodic size is consistent with previous work.

Interpretation. There were no interaction effects between word meaning (or semantic size) and prosody size for any of the acoustic measures described. That no significant interaction between semantic size and prosody condition was found suggests that the speaker did not vary their productions of prosodic cues to size as a function of the word's size.

Overall, we observed significant differences in acoustic characteristics between small, neutral, and large prosody utterances, consistent with the differences in pitch and duration observed by Nygaard et al. (2009).

Lexical Decision

Having confirmed that our stimuli differed by prosodic condition, we used a subset of the recordings to perform a lexical decision task. The task assessed the effects of semantic size and prosodic size on lexical processing speed and accuracy.

Materials & Methods

Participants

A total of 45 participants were recruited on the online Prolific platform (www.prolific.co). Prolific is a UK-based participant recruitment platform catering specifically to academic research (Palan & Schitter, 2018). We used Prolific's filters to pre-screen participants so that all respondents were right-handed speakers with English as a first language. Participants were compensated $7.25 for their participation. One participant was excluded because they reported having a diagnosed hearing problem.

Stimuli

A subset of 150 words and their matched pseudowords was selected from the original list of 278 words (see Appendix). The total set of stimuli comprised 900 recordings of the original 1668: 75 Big words and 75 Small words, each recorded with small, neutral, and large prosody, along with their matched pseudowords, also recorded in small, neutral, and large prosody. Each participant heard every word and its matched pseudoword one time in either small, neutral, or large prosody for a total of 300 trials per participant. Prosody conditions were counterbalanced across participants such that, across participants, all words were evenly represented in all conditions.

Task

Participants completed a standard lexical decision task in which they were asked to judge whether or not a given stimulus represented an existing English word. Participants completed the task, created and hosted on the Gorilla online platform (www.gorilla.sc), on their home computers (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). To help minimize errors in timing or stimulus presentation from varying internet speeds, we configured Gorilla's platform to reject participants whose estimated internet speed was below 12 Mbps. Gorilla also rejected participants who attempted to access the experiment from tablets or mobile devices.

Participants were randomly assigned to one of the three counterbalanced groups. Every participant heard each utterance only once, but across groups, all stimuli were presented in all prosodic conditions.

Participants were shown written instructions telling them to answer as quickly and accurately as possible by pressing "j" to indicate that the stimulus was a word and "f" to indicate that it was not a word. In each trial, participants were presented a fixation cross for 250 ms, after which was a 100 ms pause before a stimulus played. As soon as the participant made a keystroke, indicating a judgement on the stimulus, there was a 100 ms pause, then the fixation cross appeared again. Each participant completed 3 blocks of 100 randomized trials. In between blocks, participants could take a break for two minutes so long as they did not exit the browser tab.

Data Analysis

Data trimming. Data were trimmed twice. First, within participants, trials with reaction time outliers (defined as 2.5 standard deviations from the participant's average reaction time) were eliminated. Then, item analysis was performed to identify stimuli with average accuracy below 70%. Data from these stimuli were eliminated for all participants before calculating averages. These trimming steps eliminated 10.0% of observations. Data trimming and item analysis were performed using Python with the NumPy statistical package (Harris et al., 2020).

Quality checks. Two participants were excluded from analysis based on their performance. One failed to meet our accuracy threshold of 70%, and one exhibited significant reaction time variability even after removing outliers. The latter participant was excluded, specifically, because their average reaction time and the variability of their reaction times far exceeded participant averages.

Accuracy. Participant accuracy was defined as the percentage of correct responses to words only, after the data were trimmed.

Reaction times. Reaction times in ms were measured from the moment that the fixation cross disappeared to the moment that the participant responded with a keystroke.

Reaction time adjustment. Because word duration varied as a function of type of prosody, with large prosody having longer word durations, each reaction time was adjusted by subtracting the duration of the corresponding stimulus (Table 1). This is an accepted adjustment method consistent with Nygaard and Queen (2008) and imposes no assumptions about lexical processing dynamics. To ensure that the systemic difference in duration did not affect our analyses, all statistics were run using these adjusted reaction times rather than the raw data.

Analysis strategy. The reaction times and participant accuracy were analyzed using SPSS software. We performed two sets of analyses of variance for each dependent variable (adjusted reaction time and lexical decision accuracy.) We first used a 3 (large, small, neutral prosody) x 2 (Big, Small meaning) within-subject repeated measure ANOVA to examine the effects of prosody and word meaning on reaction time and accuracy. We further used a 2 (large prosody, small prosody) x 2 (congruent meaning, incongruent meaning) ANOVA to identify a potential congruency effect. Neutral recordings were excluded in considering congruence since these recordings did not feature characteristic prosody that would correspond with Big or Small meaning. We also performed follow-up T-tests to assess differences between individual conditions when ANOVA indicated a significant interaction.

Results

Lexical Decision Accuracy

Figure 1 shows average accuracy for words in each stimuli condition, excluding pseudowords, which do not have associated semantic sizes. Note that in both Figures 1 and 2, the y-axis is truncated and begins at 70%; this is because our data trimming threshold for both individual items and participants was 70%.

Accuracy ANOVA. There was a significant main effect of prosody on lexical decision accuracy, $F(2,40) = 5.102$, $p= .008$, $\eta_p^2= .111$, but not of semantic size. There was also a significant interaction effect between prosody and size, $F(2,40) = 18.71$, $p < .001$, $\eta_p^2 = .313$.

Accuracy T-Tests. Accuracy differed as a function of prosody. A follow-up analysis using paired-sample T-Tests showed that words produced in large prosody ($M = 94.0\%$) were responded to significantly more accurately than neutral ($M = 92.1\%$, $p = 0.018$) or small ($M = 92.3\%$, $p = 0.009$) prosody stimuli. The difference in accuracy between small and neutral prosody conditions was not significant.

Big words were responded to more accurately in the large prosody ($M = 95.0\%$, $p < 0.001$) and small prosody ($M = 94.0\%$, $p = 0.001$) conditions compared to neutral prosody ($M = 89.5\%$). There was no significant difference in accuracy for Big items between large and small prosody conditions.

Accuracy congruency analysis. There was no main effect of congruence on accuracy ($p = .218$), but there was a significant interaction effect between prosodic condition and congruence, $F(1,41) = 11.99$, $p = .001$, $\eta_p^2 = .226$. In the large prosody condition, there was a slight numerical advantage for congruent prosody, $M = 95.0\%$ for congruent vs. $M = 93.0\%$ for incongruent. However, in the small prosody condition, there was a slight numerical advantage for incongruent prosody over congruent ($M = 90.1\%$ for congruent vs. $M = 94.0\%$ for incongruent) (Figure 2).

Reaction Time Results

Reaction times ANOVA. There was a significant main effect of semantic size, $F(1,41) = 6.778$, $p = .013$, $\eta_p^2 = .142$. Across prosody conditions, participants responded more quickly to Big words ($M = 119.55$ ms) compared to Small words ($M = 137.19$ ms). A significant main effect of prosody condition was also found, $F(2,40) = 104.504$, $p < .001$, $\eta_p^2 = .839$ (Figure 3). No significant interaction was found between prosody and semantic size, $F(2,40) = 1.373$, $p = .265$.

Reaction times T-Tests. Lexical decision times differed as a function of prosody. Participants responded more quickly to words produced with small prosody (M= 59.96 ms) than to words produced with neutral (M= 149.27 ms; p< .001) or with large prosody (M= 177.02 ms; p< .001). Likewise, words produced with neutral prosody were responded to more quickly than words produced with large prosody (p< .002).

Reaction times congruency analysis. Assessing congruent vs. incongruent prosody, there was a main effect of prosody on reaction time (F(1,41) = 201.157, p< .001, η_p^2= .831), but there was no significant effect of congruence (p= .895 (Figure 4). However, there was a significant interaction effect between prosody and congruence, F(1,41) = 13.350, p = .001, η_p^2= .246.

On average, for recordings in the large prosody condition, participants responded faster to congruent trials, which were Big words. The average reaction time for congruent large prosody trials was M= 164.44 ms, compared to M= 190.27 ms for incongruent trials. For small prosody trials, participants responded more quickly when word meaning was incongruent (ie. to Big words), reflecting that participants responded more quickly to Big meaning even when produced with small prosody (Figure 4).

Discussion

In this study, participants completed a lexical decision task incorporating Big and Small words expressed with large and small prosody. We assessed whether participants respond more quickly to Big words, and whether prosody interacted with the effect of semantic size.

Processing Advantage For Big Meaning

Our primary finding indicates that, as predicted, there is a processing advantage for spoken words with Big semantic size.

When asked to identify, as quickly as possible, whether utterances were real English words, participants responded significantly faster to words with Big meanings than Small meanings in both large and small prosodic conditions (Figure 3). Accuracy was also higher for Big words than Small words within large and small prosody conditions, so the increased response rate did not limit recognition accuracy (Figure 1).

This finding extends S. C. Sereno et al. (2009)'s observation that Big meaning carries a lexical processing advantage in written word recognition, suggesting that this advantage extends to spoken word recognition as well. Other work has shown that meaning can affect lexical decision time, indicating that referential properties of words may be activated early in processing, but we are the first to describe this pattern for semantic size within spoken language(Chumbley & Balota, 1984).

Barsalou (1999) proposed that lexical processing is grounded in sensory-motor associations, and that during lexical processing, the brain partially reactivates sensory-motor areas, effectively simulating perception. S. C. Sereno et al. (2009) argue that their processing advantage is rooted in this simulation. The brain activates visual areas during mental imagery (Besle et al., 2008; Sekiyama, Kanno, Miura, & Sugita, 2003), and embodied cognition accounts would suggest that the same occurs during lexical access. If this is the case, then Big words might carry an advantage because visual imagery is available more quickly during simulation (S. C. Sereno et al., 2009). Such a relationship could be explained by the fact that the visual system processes perceptual information about big objects more quickly than small ones (S. C. Sereno et al., 2009).

Our study could similarly be taken to provide evidence of grounded cognition in spoken language processing. If the brain automatically accesses visuo-spatial representations when processing spoken language, and if big representations are accessed faster as M. Sereno (1993) suggests, then it makes sense that Big concrete words would be processed faster than Small words.

Our participants completed questionnaires designed to measure the degree to which they engage in mental imagery. In the future, we plan to analyze those results alongside our reaction time findings to determine whether participants who engage in more visual imagery exhibit a more pronounced processing advantage. If so, then there would be additional evidence that language users are accessing visuo-spatial representations when processing word meaning.

However, grounded cognition is not the only possible explanation for our results. Yao et al. (2013) present a slightly different interpretation of the processing advantage for Big vs. Small, invoking arousal as a potential mediator. Since Yao et al. (2013) found that the processing advantage for Big meaning extended even to abstract words rated as Big, rather than just concrete ones, sensory grounding could be insufficient to explain the effect unless we understand abstract concepts to also be grounded in sensory-motor perceptions. Yao et al. (2013) argue, instead, that listeners cannot visualize an abstract concept in the same way that they could a concrete object, nor can they easily situate an abstract concept in spatial dimensions, so there could be another variable implicated in the processing advantage for abstract words. Specifically, they suggest arousal, based on participant ratings that showed Big words were rated as significantly more arousing than Small ones (Yao et al., 2013).

Other research has linked arousal to lexical processing, including Aryani, Isbilen, and Christiansen (2020), who showed that “kiki-like” words elicited higher arousal ratings than “bouba-like” words. They suggest that cross-modal correspondences could, therefore, be rooted in differences in arousal. For instance, spiky shapes could be linked to harmfulness, inducing a heightened state of arousal (Aryani et al., 2020). By this explanation, emotional arousal links otherwise unconnected sensory modalities, resulting in sound-to-meaning correspondences associated with extreme, abnormal, or potentially dangerous referents (Aryani et al., 2020; Yao et al., 2013).

If Big words are consistently considered more arousing, then reading or hearing big words might increase alertness and thus speed reaction times, which could explain the effect without invoking an embodied cognition account (Yao et al., 2013).

Despite all this, the effect of arousal does not seem sufficient to explain the lexical processing advantage we observed. Yao et al. (2013) found that arousal had a stronger impact on abstract word processing specifically, perhaps due to their emotional valence. Considering only concrete words, arousal was not sufficient to explain the effect of Bigness on processing speeds. Since our stimuli comprised only concrete nouns, including concrete nouns drawn from Yao et al. (2013)'s list, it seems unlikely that arousal could have driven our observation. The processing advantage for Big words seems, instead, to point toward sensory grounding.

To disambiguate the effect of arousal, we could obtain arousal ratings for our specific stimuli, as Yao et al. (2013) did, and evaluate whether arousal mediates word recognition. We could also eliminate words that refer to clearly dangerous referents, such as *tiger* and *lion*, which would likely be linked to greater affective arousal.

The Influence of Prosody

In addition to semantic size, we were also interested in the potential effect of prosody, particularly since Nygaard et al. (2009) previously described the existence of consistent prosodic cues to size.

We predicted a congruency effect in which participants would respond more quickly to Big and Small words spoken with size-congruent prosody. Our results were not consistent with this hypothesis, as we did not see a facilitatory effect of size-specific prosody (Figure 4). While there was a main effect of prosody on reaction times, with participants responding more quickly to words in the small prosody condition regardless of semantic meaning, there was no significant interaction between prosody and meaning for response times (Figure 3). Thus, prosody affected lexical processing, but not in the manner expected.

Interestingly, there *was* a significant interaction between prosody and meaning for lexical decision accuracy. This effect seemed to be driven by anomalous findings in the neutral prosody condition; in the small and large prosody conditions, there was an accuracy advantage for Big meaning, just as there was for reaction times. However, accuracy for Big items in the neutral condition was significantly lower than Small items, and Big items in neutral prosody were responded to significantly less accurately than Big items in large or small prosody (Figure 1).

Although speculative, arousal could explain our observed differences between prosodic conditions. The acoustic analysis revealed significantly higher pitch and greater pitch variability for small prosody recordings (Table 1). It could be that these acoustic features increased arousal and sped response to words produced with small prosody (Aryani et al., 2020; Nygaard & Queen, 2008).

Similarly, both large and small prosody differed from neutral prosody on pitch and duration, although in opposite directions 1. Since the large and small prosody conditions used exaggerated prosody relative to normal speech, *both* large and small prosody may have been more arousing than neutral prosody by virtue of being more expressive. Such an effect, if present, could have been heightened by the fact that the exaggerated large and small prosody stimuli were presented alongside the neutral stimuli, which were meant to lack any abnormal prosodic cues. This explanation would align with the accuracy advantage for Big items in both the small and large prosodic conditions, but not in neutral prosody.

Additional work is necessary to determine if arousal functions in this manner. To test this explanation, a future study could account more explicitly for arousal, perhaps by obtaining arousal ratings for each type of prosodic instantiation. To explain the pattern of results, participants would need to rate both large and small prosody, as our speaker instantiated them, to be more arousing than neutral prosody.

Prosody's Interaction with Meaning

Although this study does indicate that prosody had some effect on lexical processing, we did not observe prosody being integrated with meaning in a size-congruent manner. Therefore, it remains unclear whether and when prosodic cues to size are integrated with word meaning. Had we seen a congruency effect, as predicted, it might have indicated that prosody interacts during lexical processing, speeding word recognition by constraining the set of possible referents for an utterance. That we did not see a congruency effect, however, does not exclude the possibility that prosody influences lexical processing.

One possibility is that the listener draws on semantic size before recruiting prosodic cues to meaning, which could explain why we saw a strong effect of semantic size and an effect of prosody, but no congruency effect between the two factors.

Another possibility is that prosodic cues to size may constrain or influence word recognition when the communicative or referential context is ambiguous, but not in other contexts. For example, Tzeng, Namy, and Nygaard (2019) showed that speakers employed iconic prosody only when their utterance was otherwise ambiguous. Participants were asked to tell a listener which of two color swatches to select. One swatch was light and the other was dark. In the ambiguous condition, both swatches were red, while in the unambiguous condition, one swatch was purple. In the ambiguous condition only, speakers produced significantly higher pitched utterances for bright swatches, and listeners reliably chose the correct red swatch based on those utterances, suggesting that iconic prosody may be employed when necessary to resolve lexical ambiguity (Tzeng et al., 2019). Although our task was difficult, requiring participants to identify real words from other word-like pseudowords, there was little ambiguity that would require participants to draw on prosodic cues; the recording was either a word or a pseudoword. Thus, the specific task may have obscured an effect of prosody if it is contingent on communicative ambiguity.

To test this possibility, we could use a similar paradigm to Tzeng et al. (2019), but ask participants to choose the shape that matches the target word on each trial. An unambiguous trial could use two different shapes while an ambiguous one could use the same shape but at different sizes.

A third possibility is that prosody conveys referential information, but that the effect is post-lexical. That is, we did not see an effect on lexical processing because there is no effect on lexical processing, but there is still an iconic association linking semantic size to prosody. One way to assess this would be to obtain size ratings for the pseudowords in different prosody types. If large prosody pseudowords receive reliably Big ratings, we could have more confidence that our stimuli not only differed acoustically, but that listeners infer meaning based on those acoustic differences. Alternatively, we could ask participants to judge the size of the words produced in large and small prosody. The benefit of this alternative is that it invokes an explicit size judgement rather than an implicit one. If prosody influences the speed or accuracy of size judgements, we could infer that participants rely on prosodic cues in making size judgements.

Summary

This study was motivated by two aims: to ascertain whether semantic size is activated during spoken word processing, and if so, to determine whether prosody affects this relationship.

Aim 1

Our evidence supports Hypothesis 1, which stated that words with Big meaning should be processed more quickly. This prediction derived its theoretical basis from embodied cognition, which implies that the reactivation of visual processing areas during lexical access could influence word recognition (Barsalou, 1999; Besle et al., 2008). Although other variables, such as arousal, could explain the observed effect, the data do fit the pattern predicted by a grounded cognition account. Based on the data we present, there is some evidence that, rather than being divorced from lexical access, perceptual features of a word's referent are automatically activated and inform word recognition during spoken language processing. It seems that, based on S. C. Sereno et al. (2009)'s work and the present study, semantic size is such perceptual feature that is activated, although it remains unclear whether that activation is rooted in simulation of visual processing.

Future analysis should consider imageability ratings for our specific set of stimuli, as well as the extent to which participants report relying on visual imagery. If either measure correlates with greater accuracy or faster reaction times, there would be additional evidence for the embodied cognition account.

Aim 2

Hypothesis 2 was rooted in two primary lines of evidence: that prosody can convey referential information (Tzeng et al., 2019), and that there are consistent prosodic correlates to Big and Small meaning, perhaps rooted in an iconic cross-modal sound-to-meaning mapping (Nygaard et al., 2009). If prosody interacts with lexical access, rather than being accessed post-lexically or simply in parallel with word recognition, then prosody could impact lexical processing speed. We specifically predicted that there would be an interaction between prosody and meaning in the form of a congruency effect. Instead, we observed a main effect of prosody with no interaction.

Our findings with reference to Aim 2 are more difficult to interpret, and more research is necessary to elucidate the exact role of size-based prosody in spoken word recognition. We do not see evidence that prosodic cues to semantic size influence lexical access, but arousal may have been an important confound, there may have been task-related effects that obscured the relationship.

Additional work is also necessary to determine whether prosodic cues to size can be considered a form of iconic mapping. Perniss et al. (2010) argue that cross-modal correspondences involving prosodic contours should be considered alongside sound symbolism as manifestations of iconicity. If size-specific prosody conveys information about the referent, as Nygaard et al. (2009) suggests, then there could still be an iconic cross-modal mapping, just not one that impacts lexical access. A size judgement task using the same stimuli could give more concrete evidence about this association.

Conclusion

This study asked whether a sound-to-meaning mapping between prosody and semantic size is activated during, and influences, lexical processing. We predicted that, in a lexical decision task, participants would respond more quickly to large referents and to words recorded with size-congruent prosody.

We found a processing advantage for words with Big meanings in large and small prosody conditions, demonstrating that the processing advantage for semantically big words described by S. C. Sereno et al. (2009) extends to spoken words. We also saw an advantage for small prosody, but no interaction effect.

This study lends support to Barsalou (1999)'s account of embodied cognition, suggesting that language perception might activate visual processing areas to speed word recognition.

Further research should seek to isolate the effect of prosody on spoken word recognition within this and other referential contexts.

Table 1

Acoustic properties of large, neutral, and small prosody stimuli.

Condition	Mean F_0 (Hz)	Mean SD F_0 (Hz)	Mean Duration (ms)
Large Prosody	179.53 (*SD*= 41.89)	46.37 (*SD*= 37.58)	1133.79 (*SD*= 215.27)
Big Meaning	179.25 (42.63)	44.60 (36.04)	1166.26 (234.05)
Small meaning	179.81 (41.44)	48.10 (39.18)	1102.15 (191.46)
Neutral Prosody	194.04 (25.45)	49.65 (32.20)	1047.14 (163.72)
Big Meaning	193.32 (22.45)	49.62 (28.50)	1049.72 (167.78)
Small meaning	194.04 (28.19)	49.68 (35.62)	1044.61 (160.69)
Small Prosody	339.94 (54.64)	131.83 (29.75)	1062.71 (176.51)
Big Meaning	337.04 (55.14)	133.89 (30.25)	1070.76 (182.30)
Small meaning	342.76 (54.35)	129.82 (29.31)	1054.86 (171.49)

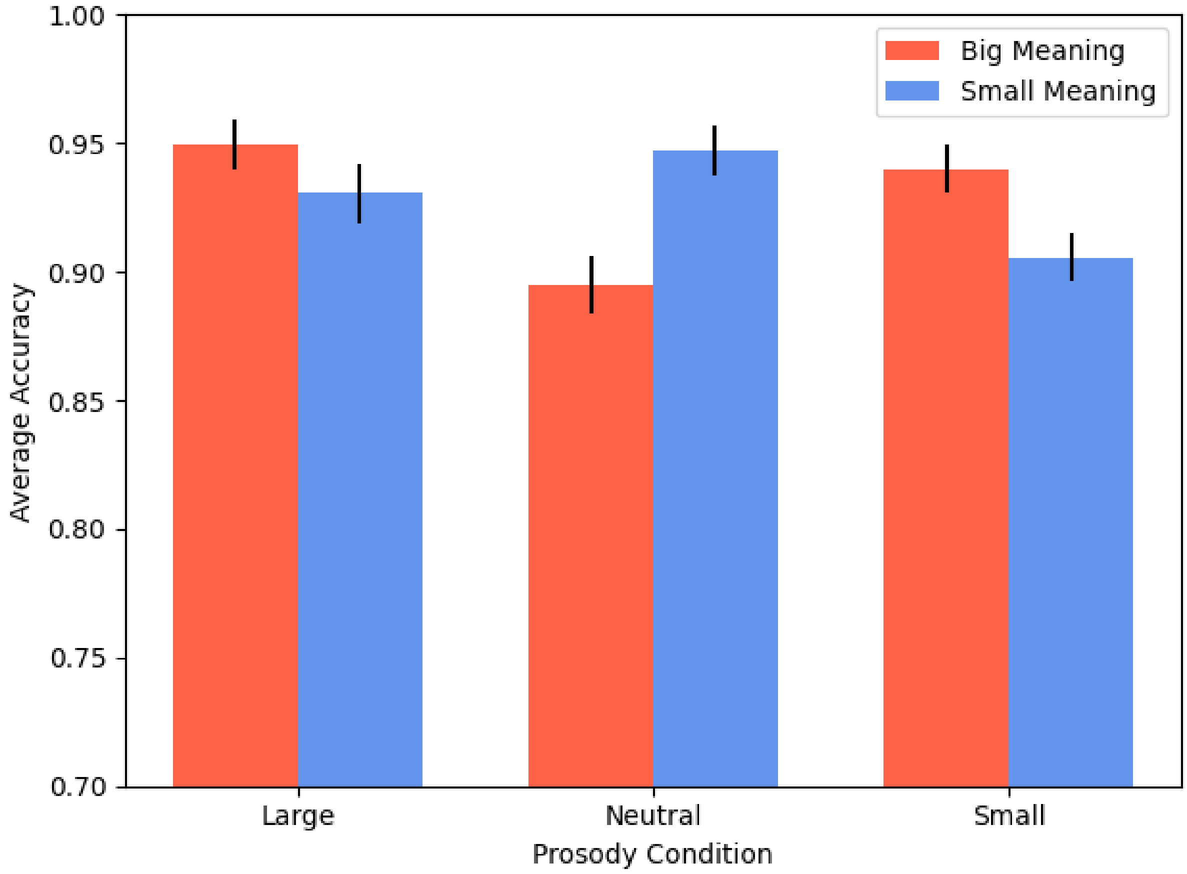

Figure 1

Participant accuracy by prosody and semantic size.

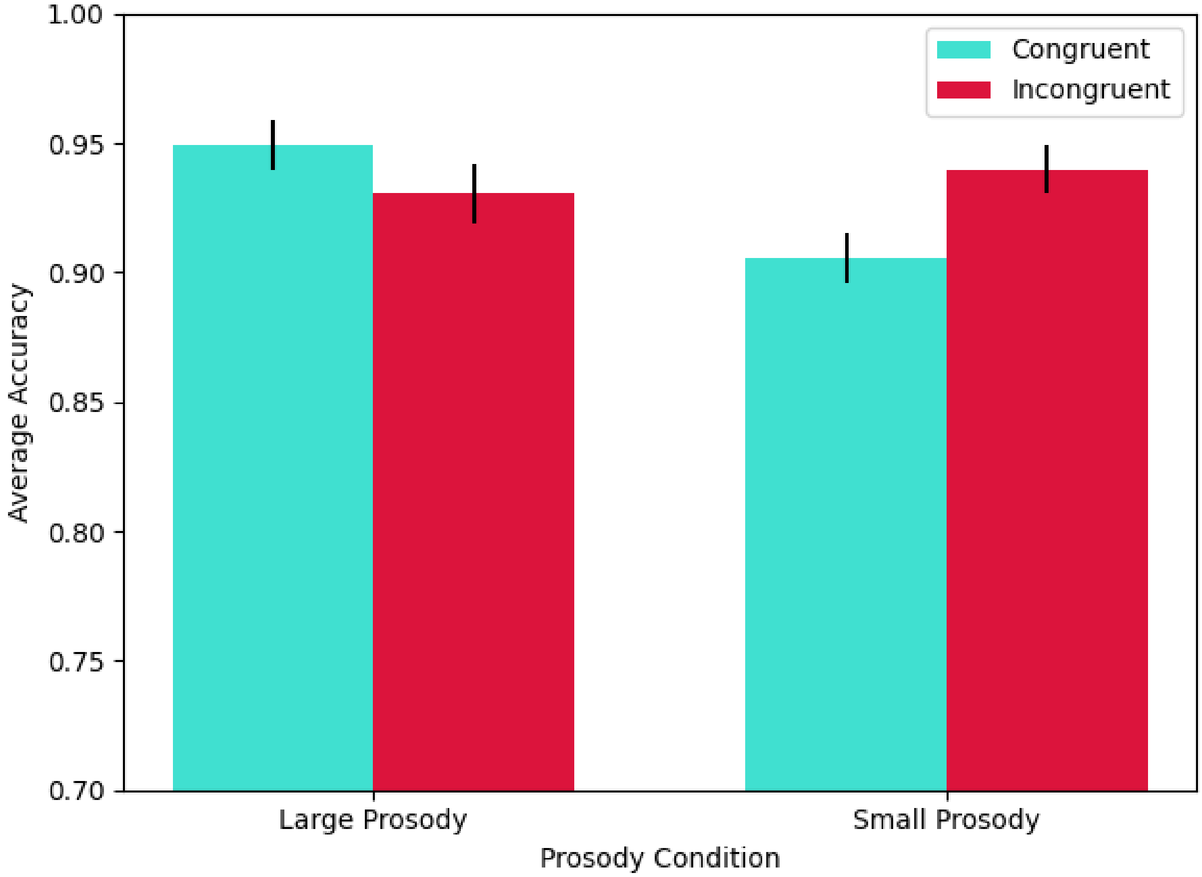

Figure 2

Participant accuracy by prosodic condition and congruence with semantic size.

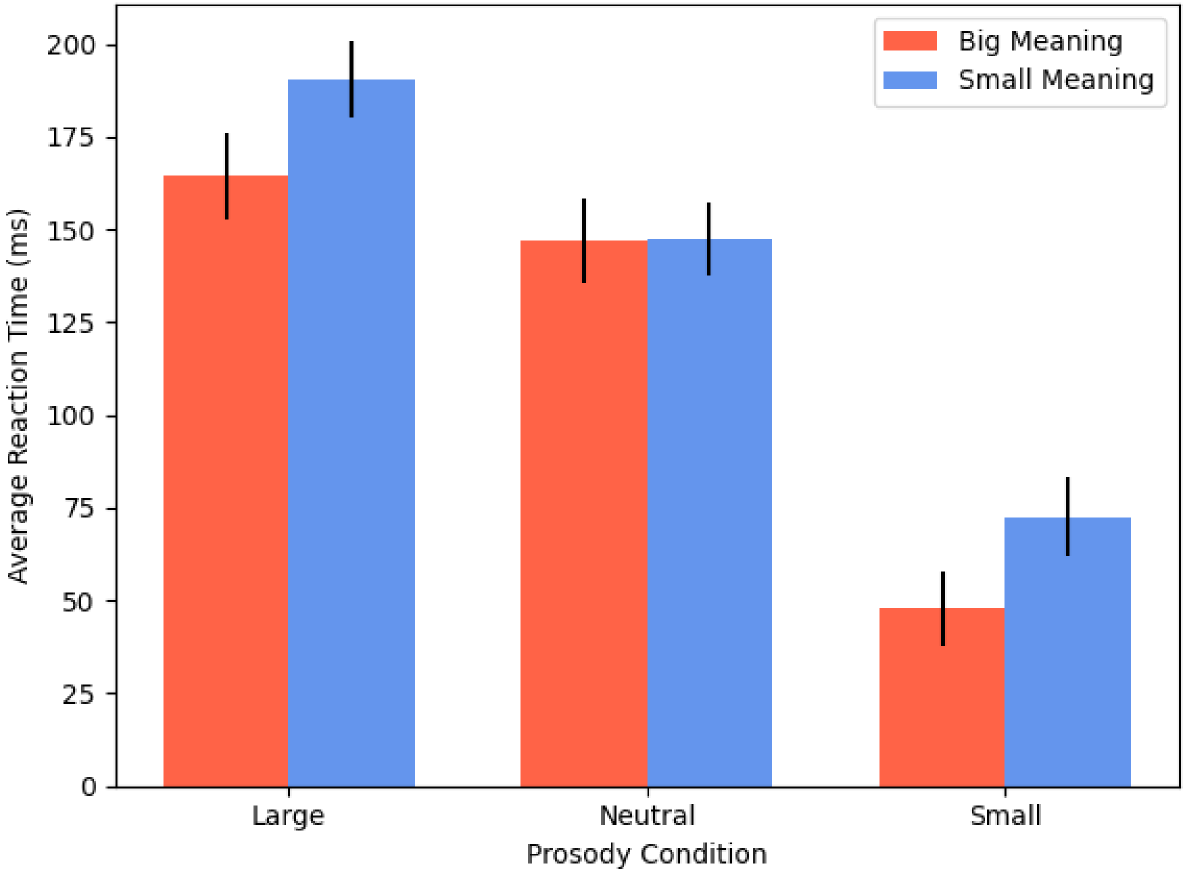

Figure 3

Average adjusted reaction times by prosody and semantic size.

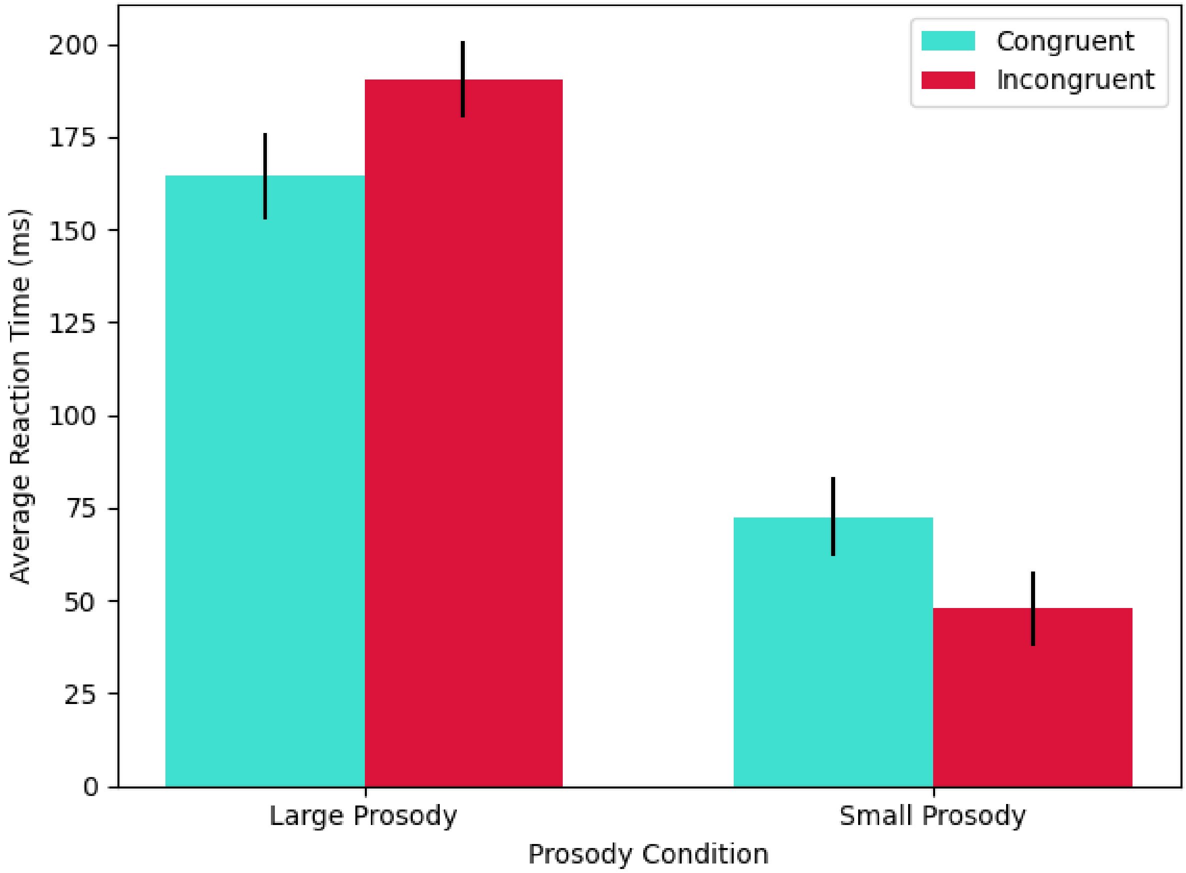

Figure 4

Average adjusted reaction times by prosodic condition and congruence with semantic size.

Appendix

Word Lists

Table A1

Words and matched pseudowords used in lexical decision task.

Small Words		Big Words	
plum	flum	jet	jat
coin	coge	galaxy	valewy
bird	bith	windmill	windfime
tulip	tumap	cathedral	cawredret
sparrow	sparrap	comet	codel
leaf	peaf	ladder	fidder
crown	crowl	museum	rupeup
glass	glath	lake	lape
glove	glive	stadium	stoniup
ladybird	hadyband	horse	hofts
bracelet	briftlet	downtown	dongtorm
grape	flape	dragon	slayon
vitamin	witason	storm	stoge
battery	fentery	cow	fow
thumb	thues	bull	burs
olive	osike	rocket	ducket
pearl	pearn	cannon	candin
ticket	hocket	castle	camble
lemon	feson	skyscraper	glystriper
berry	dorry	giant	wiase
bacteria	hortenia	rainbow	railpow

nose	nole	hippopotamus	hynmopotalas
apple	adfle	harbour	hanvour
fly	fla	gorilla	vavilla
pebble	ribble	giraffe	gisatch
insect	ingews	tank	tane
parsley	martley	bed	bes
butterfly	bostersty	valley	challop
magazine	supazine	tiger	fider
cup	hup	audience	aycients
drink	drigs	wardrobe	warerobs
bullet	dillet	lion	hoon
biscuit	bispeat	tractor	traccar
sandwich	sandfach	truck	sluck
molecule	polecome	planet	sparet
pill	pids	tower	tiper
thorn	snorn	mountain	meantail
ring	rire	ship	shap
robin	rigin	tree	blee
rose	mose	factory	furlory
sausage	savnage	hotel	lotet
pin	hin	monument	rarument
hummingbird	lummingbale	elephant	eletrite
lip	lig	parade	maraps
handkerchief	landkerproof	cattle	bastle
coffee	codlee	typhoon	tycreen
mosquito	potshito	tornado	fervado

fingernail	finnercads	whale	whame
emerald	iteract	stag	stap
cherry	shurry	camel	balel
walnut	watput	circus	cirdis
neck	nens	shore	shogs
peanut	deaput	jungle	vengle
parasite	penamite	volcano	gurcano
aspirin	atciren	river	sover
smile	smale	walrus	walkis
tape	tark	sofa	moda
pocket	sicket	flood	spood
needle	nooble	moose	mools
cigarette	cegaroled	submarine	sultarits
button	bullon	motorway	motorcop
letter	tutter	supermarket	supernarnue
spot	prot	shark	shage
page	parl	mansion	mangeen
phone	phove	buffalo	beggalo
diamond	riemond	ocean	omoan
tomato	fotalo	bear	beal
paper	raner	palace	pasaws
matches	rangees	dinosaur	dinomolt
snail	snage	piano	roano
peach	peart	farm	darm
mouse	mouls	train	trame
apricot	awrirat	beach	beals

radio	sonio	forest	terest
pencil	piscil	monster	munsher

Table A2

All words originally recorded and acoustically analyzed.

Small Words		**Big Words**	
Word	**Pseudoword**	**Word**	**Pseudoword**
acorn	avern	admire	adhore
amuse	anose	agony	etory
apple	adfle	ambition	etvition
apricot	awrirat	anger	arker
aspect	aldect	audience	aycients
aspirin	atciren	bang	baws
assist	assats	bay	baw
bacteria	hortenia	beach	beals
battery	fentery	bear	beal
berry	dorry	bed	bes
bird	bith	boast	boams
biscuit	bispeat	bookcase	booklare
bland	blate	brave	brams
book	beek	brutal	flucal
bother	busher	buffalo	beggalo
bracelet	briftlet	bull	burs
brief	crief	camel	balel
bullet	dillet	cannon	candin
butterfly	bostersty	career	catier

button	bullon	castle	camble
cherry	shurry	cathedral	cawredret
chestnut	chessmut	cattle	bastle
cigarette	cegaroled	circus	cirdis
coffee	codlee	comet	codel
coin	coge	courage	coatage
cowardly	cofently	cow	fow
crown	crowl	crisis	grinis
cup	hup	demolish	decadish
diamond	riemond	devotion	peletion
dimple	sumple	dinosaur	dinomolt
drink	drigs	disaster	denarter
emerald	iteract	downtown	dongtorm
episode	epimise	dragon	slayon
excuse	excend	dynasty	dyminty
fingernail	finnercads	eager	euber
flicker	clacker	elephant	eletrite
flutter	spetter	empire	embere
fly	fla	epidemic	edigacic
frail	frage	eternal	everpac
freckle	druckle	expanse	elpaste
glass	glath	factory	furlory
glimpse	glilths	farm	darm
glove	glive	flood	spood
grape	flape	forest	terest
handkerchief	landkerproof	galaxy	valewy

haste	haids	gallant	gallord
hint	hing	genius	weniln
humble	hupple	giant	wiase
hummingbird	lummingbale	giraffe	gisatch
hush	hule	gloom	scoom
impulse	impyons	glory	thoty
incident	incivate	gorilla	vavilla
insect	ingews	greed	cheed
instance	instoles	harbour	hanvour
intimate	intirent	heavy	heity
ladybird	hadyband	hippopotamus	hynmopotalas
leaf	peaf	honor	fonur
least	leals	horse	hofts
lemon	feson	hotel	lotet
letter	tutter	huge	hulp
lip	lig	infinity	onmifaty
literal	faveral	jet	jat
magazine	supazine	jungle	vengle
matches	rangees	ladder	fidder
meek	heek	lake	lape
minor	misur	liberty	fixerry
minute	manuce	lion	hoon
mischief	missteaf	loud	poud
molecule	polecome	mansion	mangeen
mosquito	potshito	marvel	mancel
mouse	mouls	miracle	suradle

narrow	narrap	misery	sanery
near	nean	monster	munsher
neck	nens	monument	rarument
needle	nooble	moose	mools
nose	nole	motorway	motorcop
olive	osike	mountain	meantail
ornament	unrament	museum	rupeup
page	parl	ocean	omoan
paper	raner	palace	pasaws
parasite	penamite	panic	minic
parsley	martley	parade	maraps
pause	pancs	paradise	parafods
peach	peart	park	pade
peanut	deaput	passion	mastion
pearl	pearn	piano	roano
pebble	ribble	planet	sparet
pencil	piscil	pride	prite
phone	phove	prosper	stotper
phrase	phrals	quest	thest
piece	piefs	rainbow	railpow
pill	pids	rich	riss
pin	hin	river	sover
pity	pury	roar	roak
plum	flum	rocket	ducket
pocket	sicket	shark	shage
polite	potise	ship	shap

portion	mortian	shock	shick
prefix	pregax	shore	shogs
pronoun	prosian	skyscraper	glystriper
quiet	swiel	smash	glash
radio	sonio	sofa	moda
rare	rast	stadium	stoniup
remark	remept	stag	stap
reminder	recurder	storm	stoge
ring	rire	submarine	sultarits
robin	rigin	supermarket	supernarnue
rose	mose	sure	sugs
rumour	rusier	tank	tane
sandwich	sandfach	terror	tessor
sausage	savnage	thrill	thrist
scarce	scarbs	tiger	fider
seed	sood	torment	turdent
shallow	challop	tornado	fervado
shell	shers	tower	tiper
short	shork	tractor	traccar
shrug	shrum	train	trame
smile	smale	tree	blee
snail	snage	truck	sluck
soon	hoon	trust	trums
sparrow	sparrap	truth	trupe
spot	prot	typhoon	tycreen
syllable	sottable	valley	jallep

tape	tark	virtue	vectue
teaspoon	teatheen	volcano	gurcano
thimble	phamble	walrus	walkis
thorn	snorn	wardrobe	warerobs
thrifty	squincy	wealth	dealth
thumb	thues	whale	whame
ticket	hocket	wide	wice
tidy	liny	windmill	windfime
tight	tilks	wisdom	wishop
tomato	fotalo	zeal	beal
trace	trass		
trick	trins		
tulip	tumap		
tweezers	tweagers		
unique	unanks		
video	waveo		
vitamin	witason		
walnut	watput		
weird	weich		

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019, apr). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. doi: 10.3758/s13428-019-01237-x

Aryani, A., Isbilen, E. S., & Christiansen, M. H. (2020, jul). Affective arousal links sound to meaning. *Psychological Science*, *31*(8), 978–986. doi: 10.1177/0956797620927967

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008, dec). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *Journal of Neuroscience*, *28*(52), 14301–14310. doi: 10.1523/jneurosci.2875-08.2008

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016, sep). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823. doi: 10.1073/pnas.1605782113

Boersma, P. P. G., & Weenink, D. J. M. (1996). *Praat: doing phonetics by computer: version 3.4.* Instituut voor Fonetische Wetenschappen.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013, feb). “bouba” and “kiki” in namibia? a remote culture make similar shape–sound matches, but different shape–taste matches to westerners. *Cognition*, *126*(2), 165–172. doi: 10.1016/j.cognition.2012.09.007

Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., & Herrmann, C. S. (2004, aug). Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clinical Neurophysiology*, *115*(8), 1810–1820. doi: 10.1016/j.clinph.2004.03.015

Chumbley, J. I., & Balota, D. A. (1984). A word’s meaning affects the decision in lexical decision. *Memory and Cognition*, *12*(6), 590-606.

Cuskley, C., Simner, J., & Kirby, S. (2015, sep). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, *81*(1), 119–130. doi: 10.1007/s00426-015-0709-2

de Saussure, F. (1956). *Course in general linguistics*. Philosophical Library.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615. doi: 10.1016/j.tics.2015.07.013

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), e117–e133. doi: 10.1353/lan.2016.0034

Fuchs, S., Savin, E., Solt, S., Ebert, C., & Krifka, M. (2019, jan). Antonym adjective pairs and prosodic iconicity: evidence from letter replications in an english blogger corpus. *Linguistics Vanguard*, *5*(1). doi: 10.1515/lingvan-2018-0017

Gasser, M. (2004). The origins of arbitrariness in language. *Science*.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020, sep). Array programming with NumPy. *Nature*, *585*(7825), 357–362. doi: 10.1038/s41586-020-2649-2

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing parselmouth: A python interface to praat. *Journal of Phonetics*, *71*, 1–15.

Kang, S. H., Yap, M. J., Tse, C.-S., & Kurby, C. A. (2011). Rapid communication: Semantic size does not matter:“Bigger” words are not recognized faster. *Quarterly Journal of Experimental Psychology*, *64*(6), 1041–1047.

Keuleers, E., & Brysbaert, M. (2010, aug). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. doi: 10.3758/brm.42.3.627

Konkle, T., & Oliva, A. (2012). A familiar-size stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology:*

Human Perception and Performance, *38*(3), 561–569. doi: 10.1037/a0028294

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

Levelt, W. J. (1993). The architecture of normal spoken language use. In *Linguistic disorders and pathologies: An international handbook* (pp. 1–15). Walter de Gruyter.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*, 1246. Retrieved from `https://www.frontiersin.org/article/10.3389/fpsyg.2015.01246` doi: 10.3389/fpsyg.2015.01246

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1274–1281. doi: 10.1037/xlm0000235

McCormick, K., Kim, J., List, S., & Nygaard, L. C. (2015). *Sound to Meaning Mappings in the Bouba-Kiki Effect.*

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299. doi: 10.1098/rstb.2013.0299

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009, January). The Semantics of Prosody: Acoustic and Perceptual Evidence of Prosodic Correlates to Word Meaning. *Cognitive Science*, *33*(1), 127–146. Retrieved 2020-01-22, from `http://doi.wiley.com/10.1111/j.1551-6709.2008.01007.x` doi: 10.1111/j.1551-6709.2008.01007.x

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017–1030. Retrieved 2020-01-25, from `http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.34.4.1017` doi:

10.1037/0096-1523.34.4.1017

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013, feb). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. doi: 10.1016/j.jecp.2012.05.004

Palan, S., & Schitter, C. (2018, mar). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. doi: 10.1016/j.jbef.2017.12.004

Perlman, M., Clark, N., & Falck, M. J. (2014, oct). Iconic prosody in story reading. *Cognitive Science*, *39*(6), 1348–1368. doi: 10.1111/cogs.12190

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, *1*. Retrieved 2020-07-30, from `https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00227/full` (Publisher: Frontiers) doi: 10.3389/fpsyg.2010.00227

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017, may). Iconicity in the speech of children and adults. *Developmental Science*, *21*(3), e12572. doi: 10.1111/desc.12572

Pinker, S., & Bloom, P. (1990). Natural selection and natural language. *Behavioral and Brain Sciences*, *13*(4).

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of consciousness studies*, *8*(12), 3–34.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003, nov). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, *47*(3), 277–287. doi: 10.1016/s0168-0102(03)00214-1

Sensoy, O., Culham, J. C., & Schwarzer, G. (2020). Do infants show knowledge of the familiar size of everyday objects? *Journal of Experimental Child Psychology*, *195*.

Sereno, M. (1993). *Neural computation of pattern motion: Modeling stages of motion analysis in the primate visual cortex.* MIT Press.

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Short article: Size matters: Bigger is faster. *Quarterly Journal of Experimental Psychology*, *62*(6), 1115–1122. doi: 10.1080/17470210802618900

Shintel, H., & Nusbaum, H. C. (2008). Moving to the Speed of Sound: Context Modulation of the Effect of Acoustic Properties of Speech. *Cognitive Science*, *32*(6), 1063–1074. Retrieved 2019-11-22, from `http://onlinelibrary.wiley.com/doi/abs/10.1080/03640210801897831` doi: 10.1080/03640210801897831

Spence, C. (2011, jan). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. doi: 10.3758/s13414-010-0073-7

Tzeng, C. Y., Duan, J., Namy, L. L., & Nygaard, L. C. (2018, April). Prosody in speech as a source of referential information. *Language, Cognition and Neuroscience*, *33*(4), 512–526. Retrieved from `https://doi.org/10.1080/23273798.2017.1391400` doi: 10.1080/23273798.2017.1391400

Tzeng, C. Y., Namy, L. L., & Nygaard, L. C. (2019). Communicative Context Affects Use of Referential Prosody. *Cognitive Science*.

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017, aug). Developmental change in children's sensitivity to sound symbolism. *Journal of Experimental Child Psychology*, *160*, 107–118. doi: 10.1016/j.jecp.2017.03.004

Wei, W., & Cook, A. E. (2016). Semantic size and contextual congruency effects during reading: Evidence from eye movements. *Discourse Processes*, *53*(5-6), 415–429.

Yao, B., Vasiljevic, M., Weick, M., Sereno, M. E., O'Donnell, P. J., & Sereno, S. C. (2013). Semantic size of abstract concepts: It gets emotional when you can't see it. *PloS one*, *8*(9), e75000.