Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

________________________________ Date:_______________

Kelly Ruth McCormick

Cross-Linguistic Sound to Meaning Mappings in Relational Terms: The Role of Acoustic Form in Judgments of Word Meaning

By

Kelly Ruth McCormick

Master of Arts

Psychology

____

Lynne Nygaard

Advisor

____

Harold Gouzoules

Committee Member

____

Laura Namy

Committee Member

Accepted:

____ Lisa A. Tedesco, Ph.D. Dean of the James T. Laney School of Graduate Studies

____________________ Date

Cross-Linguistic Sound to Meaning Mappings in Relational Terms: The Role of Acoustic Form in Judgments of Word Meaning

By

Kelly Ruth McCormick

B.Sc. University of California San Diego, 2006

Advisor: Lynne Nygaard, Ph.D.

An abstract of a thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Arts

in Psychology

2012

Abstract

Cross-Linguistic Sound to Meaning Mappings in Relational Terms: The Role of Acoustic Form in Judgments of Word Meaning

By Kelly Ruth McCormick

Language is a hallmark of the human mind, and pervades human experience. Since ancient Hindu and Greek philosophers, humans have pondered how it is that words carry meaning. A central question is whether words can have inherent meaning (Plato's 'natural meaning'), or whether words assume meaning only by language-specific convention. The classic and predominant view of human language is that the relationship between sound and meaning is arbitrary (de Saussure, 1959; Hockett 1960; Pinker, 1994). The sounds that comprise words bear no inherent relationship to the things in the world that they represent. Over the past century, however, research has provided evidence for non-arbitrary mappings in language (Sapir, 1929; Nuckolls, 1999). To date, a majority of empirical work has focused on sound to meaning correspondences for concrete, sensory domains of meaning. I examine whether similar mappings exist for relational terms, which take their meaning largely from context, and in relation to one another. Acoustic analyses of proximal and distal relational terms across multiple languages were conducted to determine whether characteristics of the sound structure of these terms reliably map to meaning. An experiment was then conducted to determine whether listeners could judge word meaning in foreign relational terms from sound structure alone.

Cross-Linguistic Sound to Meaning Mappings in Relational Terms: The Role of Acoustic Form in Judgments of Word Meaning

By

Kelly Ruth McCormick

B.Sc. University of California San Diego, 2006

Advisor: Lynne Nygaard, Ph.D.

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Arts

in Psychology

2012

Acknowledgements

My sincerest thanks to my advisor, Dr. Lynne Nygaard, and FAC committee members, Dr. Harold Gouzoules, Dr. Laura Namy, and Dr. Larry Barsalou for their intellectual support of this project. I wish to acknowledge the undergraduate research assistants who helped with experiment design and data collection. Valentin Lazar, Michelle Linch, and Arielle Walzer have contributed a great deal of time and thought toward the project. I also with to thank my parents and my sister for their love, support and encouragement to pursue the things that interest me most. Without these people, the ideas that spawned this project would have remained chicken scratch in a notebook. My deepest gratitude to you all.

Table of Contents

Cross-Linguistic Sound to Meaning Mappings in Relational Terms: The Role of Acoustic Form in Judgments of Word Meaning

Introduction

Language is a hallmark of the human mind, and pervades human experience. Since ancient Hindu and Greek philosophers, humans have pondered how it is that words carry meaning. A central question is whether words can have inherent meaning (Plato's 'natural meaning'), or whether words assume meaning only by language-specific convention. The classic and predominant view of human language that has emerged since these original ponderings is that the relationship between sound and meaning is arbitrary (de Saussure, 1959; Hockett 1960; Pinker, 1994). The sounds that comprise words bear no inherent relationship to the things in the world that they represent. By this view, words come to represent meanings by learned associations, and in principle, any meaning could be represented by any combination of the finite inventory of sounds in a given language. Arbitrary reference is considered powerful because symbolic forms (e.g. a word) are unconstrained in relation to meaning (Monaghan & Christiansen, 2006; Gasser, 2004) and therefore in some respects highly efficient. Although referential arbitrariness has the advantage of indexical power, it poses a burden on the language-learner in that sound to meaning correspondences must be learned, stored and recalled word by word. In a truly arbitrary system the language-learner could not use structure of familiar words to posit meanings of novel words (Gasser, 2004).

In contrast to the view that sound to meaning mappings are exclusively arbitrary is the idea that sounds in language are systematically related, in some way, to the meanings they represent. Scholars since Plato have suggested that the words could be intrinsically or sound symbolically related to their referents (Sedley, 2003). There are numerous examples of nonarbitrary sound to meaning mappings throughout the world's languages. For example, onomatopoeic words are characterized by a sound-to-sound mapping in which the sound structure of a word emulates the sound being referenced (e.g. *meow, trickle, whoosh*). Although such words appear throughout the world's languages, they appear to be relatively rare, and the phonemic content differs across languages.

Sound to sound mappings, as in onomatopoeia, are not the only non-arbitrary mappings we find in language. Researchers have documented many sound to meaning correspondences that bridge sensory modalities (e.g., sound for shape, sound for brightness; Nuckolls, 1999). Many have argued that sound symbolism may be the result of cross-modal associations that are intrinsically biased by functional interconnectedness of different sensory-motor cortical areas (Ramachandran & Hubbard, 2001; Maurer, Pathman & Mondloch, 2006; Spector & Maurer, 2008; Kovic, Plunkett & Westermann, 2010). By this view, synesthetically corresponding percepts such as the sounds of language and particular meanings would be associated because they have a functionally interconnected basis of processing. Ramachandran and Hubbard (2001) used a paradigm in which participants were shown two shapes accompanied by two nonsense names, and asked to match the labels to the objects. As in the original study by Köhler (1947), subjects showed a robust preference for labeling the jagged object *kiki* and the rounded object *bouba*. Ramachandran and Hubbard hypothesize that this reliable sound-shape mapping is based in a synesthetic correspondence between shapes of the referents, and the motor-based articulation of these sounds (for *bouba,* mouth is rounded and sound is unobstructed, for *kiki*, stop consonants disrupt airflow repeatedly).

To date, most empirical research examining sound symbolism has been confined to sensory domains, focused primarily on terms describing concrete properties of objects or events (Kita, 1997, 2001; Ikegami & Zlatev, 2007, Ramachandran & Hubbard, 2001; Maurer, Pathman & Mondloch, 2006; Spector & Maurer, 2008). These are typically attributes that are directly perceptually experienced, for which we can point to singular exemplars in the physical world. When we refer to particular objects or events, we can draw on our knowledge of these experiences, and may use the sounds of language to represent objects and events with a kind of verbal caricature to simulate the referent (Kita, 1997; Ramachandran and Hubbard, 2001). The sounds in words like *click*, *waddle*, and *slurp* are sensorily evocative of the meanings they represent. But do we use sounds to represent concepts that are more context-dependent and abstract and if so, how would those relationships be established? For what classes of words, or domains of meaning do we find systematic sound-meaning correspondences?

In recent years, researchers have amassed evidence suggesting systematic sound-to-meaning correspondences exist for relational terms that are used contrastively along a given semantic dimension, e.g. spatial proximity (Woodworth, 1991), size (Sapir, 1929; Nuckolls, 1999; Thompson & Estes, 2010), brightness (Newman, 1933; Mondloch et al, 2004), and vertical height (Ben-Artzi & Marks, 1999). These mappings are distinct in that they are inherently *relational*, and convey information about some attribute along some dimension of meaning. Rather than simply representing the presence of an attribute, sounds could be symbolic of magnitude or intensity of an attribute. Although symbolic reference is often considered all-or-nothing (something either *is* or *is not* referenced by a symbol), recent work by Thompson and Estes (2010) suggests that some representations can be analog, with more instances of a specific sound in a word corresponding to relatively more size along a graded semantic dimension. In an experiment by Thompson and Estes (2010), participants were given nonsense words, all with the same number of syllables, but varying in how many high and low vowels they contained. Participants nominated the nonsense words with more low vowels as names for relatively larger objects, and words with more high vowels as names for relatively smaller objects. Based on these findings, there does seem to be sound-symbolic marking of object size- and this representation looks more analog. These findings could support a cross-modal basis for sound-meaning mappings, with different analog positions along some dimension of sound, corresponding to a similarly analog dimension of meaning. Here sounds and corresponding meanings are graded, not all-or-nothing. But the stimuli used by Thompson and Estes were nonsense words. Can we find similar analog sound to meaning mappings in natural language? Examining continuous acoustic measures could reveal analog sound to meaning mappings in relational terms.

Spatial terms are a particularly rich domain for study of sound to meaning mappings because they are not only relational, taking their meaning in contrast to one another, but they can be contrastive along scalar *or* polar dimensions. For example, terms *above* and *below* or *left* and *right* rely on one another to be meaningful, and are semantically contrastive in a polar sense, their meaning arising out of an opposition to one another. Other spatial terms are contrastive in a scalar sense- these terms describe meanings that vary in *degree*. For example *near* and *far* both represent distance in space relative to some reference point, with *far* representing a relatively greater distance. A distance that is termed *far* in one context could be termed *near* in another. As a consequence, there is no absolute perceptual value that corresponds to the concepts of proximity or distance. This raises the interesting question as to whether sound symbolic mappings might exist in this domain and if so, how we might ever come to associate a particular sound with a concept like distance.

Because space perception is highly multi-modal, affording many possible cross-sensory associations from which spatial terms may draw structure, many different mechanisms of association could give rise to different patterns of sound to meaning mappings. Perhaps the study of sound to meaning mappings for space could serve as a theoretical stepping stone towards understanding if and how more abstract meanings are reflected in the sound structure of spoken language.

In the present study, I examine whether sound symbolic mappings exist for spatial relational terms. Spatial terms are basic, relatively high frequency terms, which appear throughout the world's languages, affording the opportunity for cross-linguistic comparison. Although languages differ in how they categorize and delineate spatial relationships (Bowerman & Choi, 2003, Kemmerer 1999, Plauché and Bergen 1999), they are, perhaps surprisingly, consistent in certain distinctions they make. In addition, there is some evidence for non-arbitrary mappings between sound and meaning for this semantic domain. In a cross-linguistic survey of spatial terms in 26 languages, Woodworth (1991) analyzed phonemic features of the vowels in words, based on prototypical IPA transcriptions. Woodworth found a trend for the terms with more proximal meanings to have higher vowels than the terms with distal meanings. This finding suggests that sound may be used symbolically to mark deictic space. But while these featural analyses are compelling, and interesting in their own right, they cannot directly inform us about the true acoustic structure of the words, as uttered by native speakers. Indeed, where sound is used symbolically, the acoustic structure of a word could encode meaning in a less all-or-nothing manner than is encoded by particular phonemic features. It may, therefore, be appropriate to examine possible symbolic use of continuous acoustic parameters such as vowel formants, fundamental frequency, intensity and so on.

If it is the case that there is some naturally motivated association between the sounds in language and their meanings, and listeners are sensitive to these relationships, such sounds could evoke representations of meaning in a listener, even if he or she had no prior experience with the word (or even the language). Further, a listener could utilize these cues to meaning when tasked with learning new words in a language. One line of empirical work, notably Köhler (1947), Kunihira (1971), Nygaard et. al (2009), demonstrates precisely this phenomenon, with people guessing the meanings of unfamiliar words from foreign languages at rates better than predicted by chance. These findings suggest that listeners are able to infer meaning from the sound structure of words, and support the idea that sound-to-meaning mappings in language are not entirely arbitrary. An experiment conducted by Clepper, Nygaard, and Namy (2012 also suggests that there may be nonarbitrary correspondences between sound and meaning in spatial relationship terms across languages. In their study, native English-speaking participants were able to reliably match foreign synonyms for the meanings *near* and *far* drawn from ten different foreign languages with their correct meaning. These findings suggest that mappings for these terms were not based on language-specific conventions, but rather on properties of the sound to meaning mappings that were consistent cross-linguistically.

In order to identify sound to meaning correspondences for spatial terms and to determine whether people are sensitive to these mappings, I employed two approaches. First, I collected contrastive proximal and distal terms in 16 languages. Words for translation were *here, there, this, that, near* and *far*, and were chosen because they are basic spatial terms that are semantically contrastive in a binary sense. These words were analyzed for acoustic characteristics related to structure of the vowels, as well as overall measures of the word to identify whether specific attributes of words vary systematically with meaning across languages and language families. If acoustic attributes of sound are correlated with specific aspects of meaning cross-linguistically, then it would suggest a motivated basis for sound to meaning mappings for spatial terms. Particular patterns of correspondence may also help to elucidate the mechanisms underlying such mappings. For example, we might expect to find that pitch differs for proximal and distal spatial terms, corresponding to our statistical experience of pitch and location of objects in the world. Because higher pitch sounds do not travel as far as lower pitch sounds we might come to associate higher pitch with more proximal spatial locations. Linguistic sound structure may reflect our sensory perception of space. Alternatively, I could find that sound is used systematically within but not across languages, suggesting that salient aspects of sound are varied to mark differences in meaning. In this case, sound is not intrinsically related to meaning, but is used symbolically to delineate a contrast. Based on findings from the Woodworth (1991) study, I predict that the second and third vowel formants will be higher for words with proximal meanings than for distal meanings. However, even if such a pattern does not emerge from the overall cross-linguistic analysis, reliable differences may still be present within languages and be specific to particular languages. Such a pattern could suggest that the identity of the sound may be less important than the relationship between sounds that are used contrastively to mark meaning.

In addition to the acoustic analysis, I implemented a forced-choice behavioral experiment in which participants heard the foreign words from our set, and nominated which of two (contrastive) English words (e.g. ‘near’ and ‘far’) were the correct translation for each word they heard. For the behavioral test, based on Clepper et al (2012), I predict that listeners will be able to judge the meanings of these spatial terms at above chance rates. In addition, I predict that words with higher second and third formants will be mapped to proximal meanings, and words with lower second and third formants to distal meanings. Pitch and duration of the words could also be related to listeners’ judgments. By combining cross-linguistic analysis with judgments of meaning in a behavioral forced-choice task, I aim to identify specific acoustic characteristics of words that co-vary with meaning.

Cross-linguistic Acoustic Analysis

Materials.

Nineteen native speakers of 16 different languages[1,2] were asked to translate basic spatial terms from English to their respective native languages. The spatial terms were *near/far, this/that,* and *here/there* and were intended to refer to proximal (*near, this,* and *here*) or distal (*far, that,* and *there*) space relative to the speaker.

Translations were elicited from two groups of speakers. The first group included native speakers of ten languages who listed translations for the English words *near* and *far*. Languages were Albanian, Dutch, Gujarati, Indonesian, Korean, Mandarin, Romanian, Tamil, Turkish, and Yoruba. These utterances were drawn from a larger database of foreign equivalents for dimensional adjectives constructed by Clepper, et al (2012). Speakers were asked to nominate as many words as possible for each meaning. This process resulted in 61 words across languages for these particular terms (see Appendix for translations generated in each language). Once the list of words was established for each language, each speaker came to the laboratory and recorded the translations for their native language. Speakers were asked to read the appropriate word list using neutral, list-like prosody and were recorded using Audacity software. Utterances were digitized at a 44.1 kHz sampling rate.

The second group included native speakers of nine languages who provided translations of the English words *here, there, this, that, near, far*. Languages were Arabic, Bulgarian, Burmese, Farsi, Georgian, Korean, Mandarin, Portuguese, and Romanian. Speakers in this group were asked to provide a single best translation, but were encouraged to provide more than one if they felt other translations were equally fitting. This process resulted in 57 words across languages (see Appendix for translations generated in each language). After informants indicated that they were satisfied with their list of translations, they were immediately recorded using a Zoom H2 audio recorder with a built-in cardioid microphone. During the recording session, the experimenter read individual English words from the list, and the native speaker spoke the translation for the word. Utterances were digitized at a 44.1 kHz sampling rate.

Procedure.

All words were edited into separate files using Sound Studio sound processing software, downsampled to 22.050 Hz, then amplitude normalized. Acoustic analysis was conducted in Praat (Boersma & Weenink, 2005). For the analysis, each word was partitioned into individual phonemes and acoustic measures were collected for both the whole word and for the vowels within the words. Whole word measures were intended to capture the overall sound structure of the word and reflect segmental, prosodic, and speaker-specific properties. Vowel measures were intended to capture primarily segmental characteristics, and based on previous work, were hypothesized to correspond to spatial aspects of word meaning. Unlike analysis of phonemic features, the continuous acoustic measures collected for vowels permit direct quantitative comparison of the sounds in our items of interest. For the whole word analysis, fundamental frequency (f_0) and word duration were measured. For the vowel-level analysis, f_0, duration, intensity, and the first three vowel formants (f_1, f_2, and f_3) were measured. These first three formants are the frequency bands that define and differentiate vowels in spoken language. The first formant is inversely related to vowel height (how high in the mouth the vowel would be articulated). The second formant is correlated with vowel backness (how far back in the mouth the vowel is formed). A more complex relationship between the second and third formant corresponds to lip rounding when pronouncing the vowel. Vowels were partitioned based on several characteristics of the waveform and spectrogram. In order for an interval to be marked as a vowel, formant structure and voice pulses needed to be present. Interval boundaries of vowels were positioned where the acoustic waveform crossed the zero axis. Vowel midpoints were computed based on the entire interval of the vowel, and then shifted to the nearest point at which the waveform crossed the zero axis. Vowel formants were measured both at vowel midpoint and averaged across the entire vowel interval. These two techniques yielded similar values in preliminary analyses. As such, the analyses we report are based on values at vowel midpoints, which is a more conventional measurement (Ladefoged, 1996).

Results and Discussion.

In order to determine if the sound structure of the foreign words varied systematically as a function of meaning, one set of analyses was conducted for each whole word measure (e.g., f_0, duration) and separate sets of analyses were conducted for vowel composite measurements (e.g., duration, intensity, f_0, *f1, f2, f3*). Note that because speaker and language varied together, it is difficult to determine whether acoustic properties varied as a function of language or speaker. Thus, comparisons were made as repeated measures within speaker/language. As proximal (*this, here, near*) and distal (*that, there, far*) terms were produced by each speaker in each language, comparisons between types of terms could be made within speaker, reducing any influence of idiosyncratic speaker effects. Thus, for each acoustic measure, pairwise, by-speaker comparisons were conducted using two-tailed paired sample t-tests. Table 1 reports mean values for each acoustic measure as a function of word meaning (proximal and distal) for each English word pair and overall.

Word level analysis

Fundamental frequency (f_0). A paired samples t test comparing proximal terms to distal terms pairwise by native speaker, revealed that proximal terms had significantly higher fundamental frequency (f_0) or pitch at the word level than the corresponding distal terms ($t(18) = 3.008$, $p = .008$; see; Table 1). Figure 1 shows mean fundamental frequency for proximal and distal terms for each language. Twelve out of the sixteen languages had proximal terms with higher fundamental frequencies than the distal terms in the respective languages.

Duration. Comparison of word duration in proximal and distal terms was not significant ($t(18) = -.541$), $p = .595$; see Table 1). Proximal and distal terms did not appear to differ in overall word duration. Figure 2 shows mean word duration for proximal and distal terms for each language. In ten out of the sixteen languages, distal words were longer than proximal. Word duration in proximal and distal terms varies substantially across the different languages, being more or less even in some languages, and quite different in others (e.g. proximal and distal terms had very different durations in Burmese).

Vowel level analysis

Measures on individual vowels were combined to produce a word-level composite score for each of the vowel measurements. For each nominated word, the acoustic values for each measure were averaged across all the vowels in that word resulting in the composite score for each term. Vowel-level measures were not significant in the pairwise, by-speaker analysis; however, individual analyses and any trends that emerged are reported below.

Overall intensity. Overall vowel intensity did not differ between proximal and distal terms ($t(18) = -.592$, $p = .561$; see Table 1). Figure 3 shows mean vowel intensity for proximal and distal terms for each language. Eight out of sixteen languages had proximal terms with greater intensity than distal terms.

Fundamental frequency (f_0). The fundamental frequency of the vowels in our proximal and distal terms approached, but did not attain significance ($t(18) = 1.937$, $p = .069$; see Table 1). Figure 4 shows mean vowel fundamental frequency for proximal and distal terms for each language. Distal terms had vowels with higher fundamental frequencies in seven out of sixteen languages.

Duration. Although vowel duration in proximal and distal terms was not significantly different, there was a trend for proximal terms to have vowels with shorter durations ($t(18) = 1.759$, p = .095). Figure 5 shows mean vowel duration for proximal and distal terms for each language. Eleven out of sixteen languages use shorter vowel durations for proximal meanings over distal. Burmese patterns strongly in the opposite direction, from these eleven languages, with longer vowels in proximal words, which may obscure an effect.

Spectral Characteristics. No significant differences were found in first ($t(18) = .415$, $p = .683$), second ($t(18) = 1.288$, $p = .214$), or third ($t(18) = .284$), $p = .779$) formant values between proximal and distal terms. As for the other measures, Table 1 shows mean formant frequency values for proximal and distal terms for each vowel formant.

Although differences did not approach significance, for each measure, formant frequencies were higher for proximal than for distal terms.

In order to explore the possibility that individual languages might contrast proximal and distal terms using different acoustic formant distinctions, as predicted from previous work, I examined the direction of contrast (e.g., f1 higher or lower for proximal versus distal terms) for each language. Figures 6-8 show first, second, and third formant frequency values for proximal and distal terms in each language. As the figures illustrate, although proximal and distal terms differed in formant frequency values, the direction of the difference varied across languages. For first formants, frequency values were higher for distal than proximal terms for seven of the sixteen languages. For second formants, frequency values were higher for proximal than for distal terms for nine of the sixteen languages. For third formants, frequency values were higher for proximal than for distal terms for ten of the sixteen languages. Although there are not enough observations for a formal statistical comparison, these qualitative observations suggest that languages differed in their use of formant values, with second formants being the most consistently related to spatial contrasts.

Contrastive Word Pairs. To determine if acoustic properties differed across individual word pairs with contrastive meaning, three one-way ANOVAs were conducted. ANOVA tests were conducted comparing acoustic measures in proximal and distal values for each of three pairs (*here/there, near/far, this/that*). Acoustic measures tested were: word-level duration and fundamental frequency, vowel-level duration, intensity, f0, f1, f2 and f3. None of these measures differed significantly for individual proximal and distal paired word meanings (all *p*'s > .05; see Table 2). It is likely that the sample was too small to allow underlying patterns to attain significance.

Summary

The cross-linguistic acoustic analysis revealed reliable differences in whole word measures of fundamental frequency between proximal and distal words. Differences between terms in fundamental frequency suggest that aspects of the sound structure of spatial terms across languages are non-arbitrarily related to word meaning. The finding that word level measures reliably corresponded to meaning suggests that there may be aspects of spatial meaning encoded in word pitch. This variation in pitch could originate in one of two ways. It could be that segmental content of the words differs in fundamental frequency in a manner that is not captured in analysis of the individual vowels of the words. It could also be that speakers modulate prosodic or suprasegmental cues when they produce words with distal and proximal meanings, in a sense providing an extra layer of information for the listener.

Vowel-level analyses showed non-significant trends in two acoustic measures comparing proximal and distal terms. Vowel duration tended to be shorter and consistent with the word level measures, vowel pitch was relatively higher in proximal words relative to distal words produced by the same speaker. Either or both of these attributes may be systematically and sound symbolically related to meaning of these words.

The non-significant differences in the majority of acoustic measures, and in particular in vowel formant values, may either reflect a lack of distinctiveness in acoustic properties across these relational semantic domains, heterogeneity in how languages cue these kinds of contrasts, or alternatively, that the acoustic measures conducted for this analysis or the small sample size did not adequately capture the relevant distinctions.

Meaning Judgments

The acoustic analyses revealed that pitch varied across instances of proximal and distal words and these differences were reliable across languages and language groups. And, although it did not reach significance in the cross-linguistic acoustic analysis, vowel duration may also be a reliable cue to meaning. Do language users use these differences in fundamental frequency and/or other acoustic properties not captured in the analyses as cues to word meaning? Are language users sensitive to sound symbolic correspondences in spatial terms? If so, even unfamiliar or foreign words may be identified as appropriate labels for a particular concept. In order to examine these questions, I designed a behavioral experiment to determine whether native English speakers could correctly judge the meanings of foreign spatial terms, and to determine whether their judgments were related to any specific acoustic features.

Method

Participants

24 students (19 female, mean age 18.6 years) participated in the experiment for either course credit or pay. Participants were native speakers of English, who reported having normal hearing, no history of speech disorders, and no familiarity with any of the languages used in the stimulus set. Data from an additional 32 participants were

excluded because they indicated familiarity with one or more of the languages in the stimulus set. Two participants were excluded due to self-reported speech disorders.

Stimulus Materials

The stimuli were spatial terms as described above from 16 languages, which were generated by native speakers as translations for English words *near, far, here, there, this,* or *that*. The stimuli came from the same set of 118 words that were used for the acoustic analysis (see appendix for complete list).

Procedure

Participants were seated at a desktop computer in a sound-attenuated room and auditory stimuli were presented over headphones. Written instructions for the task were displayed on the monitor, "…*For each foreign word you hear, you will be presented with two English words on the computer screen. Your task is to choose which of the two English words is the correct meaning for each word you hear…".* Participants initiated the experiment when ready. Participants heard an unfamiliar foreign word, and were presented, in written text, with two possible English translations for the word. Participants selected the English word they thought was the correct translation of the foreign word they had heard. Words were presented in random order, and all stimuli were presented within a single block. For each trial, one of the two response options was a correct translation of the foreign word, and the other option was a semantically contrastive term. The response options were spatial terms that were contrastive in meaning. Possible response pairs were: *near/far, here/there, this/that*. The two response options were presented side by side on the monitor, with (left-right) positioning counterbalanced within the block. Across participants, response mappings were counterbalanced such that for half of the trials, response options (possible English translations for the word) were listed with the proximal term on the left side of the screen and distal term on the right side of the screen (e.g. *here there*), and response options were listed in the reverse order (e.g. *there here*) for the other half of the stimuli. Responses were made using the left and right buttons on a button box, and no time limit was imposed.

Results and Discussion.

Judgments with response times greater than 2 standard deviations above each participant's mean response time and response times less than 200ms were excluded from the analyses in order to eliminate responses that were not stimulus-dependent.

Figure 9 shows the proportion of correct judgments of word meaning by word type (proximal or distal meanings). Foreign proximal terms (those corresponding to *here, near,* or *this*) and distal terms (those corresponding to *there, far, that*) were matched to the appropriate translations in English at rates significantly greater than chance. Comparisons to chance revealed that participants matched both foreign words with proximal (M = .55, $t(23) = 2.70$, $p = .013$) and with distal meanings (M = .56, $t(23) = 3.17$, $p = .004$), to the English translations significantly above chance.

Although across spatial terms, participants were able to guess word meaning at rates significantly greater than chance, performance varied across individual items. Participants were reliably able to select the correct English translations for some foreign spatial terms but not others. Figure 10 shows the proportion of correct judgments for each individual word meaning. Comparisons to chance revealed that foreign words meaning *near* (M = .53, $t(23) = 1.84$, $p = .04$), *far* (M= .62, $t(23) = 6.63$, $p < .001$), *this* (M = .60, $t(23) = 3.61$, $p < .001$) and *here* (M= .56, $t(23) = 1.72$, $p = .05$), were matched to English translations significantly above chance. Performance for foreign words meaning *there* (M = .46, $t(23) = -1.14$, $p = .13$) and *that* (M = .49, $t(23) = -.409$, $p = .34$) did not differ from chance.

Correlations between acoustic properties and judgments of meaning.

In order to determine which properties of the spatial terms listeners might have used to judge word meaning, the word- (duration and pitch) and vowel-level acoustic measurements (duration, intensity, pitch, formants) were linked to word judgment performance from our participants by examining acoustic characteristics and meaning judgments across stimulus items. Correlations were conducted comparing the values of each acoustic measure for each word with the proportion of proximal judgments for that item. Figures 11-18 show the correlations on each of the measures described above.

<u>Word level measures</u>

Fundamental Frequency (f0). Proximal and distal judgments were not significantly correlated with fundamental frequency of items (r = -.037, $p = .345$; see Figure 11).

Duration. Duration was significantly negatively correlated with judgments of proximal meaning (r = -.167, $p = .035$; see Figure 12). Words with shorter durations were more likely to be judged to have proximal meanings than words with longer durations.

Vowel level measures

Intensity. Mean intensity of the vowels was negatively correlated with judgments of proximal meaning (r = -.192, p = .019; see Figure 13). Words judged to have proximal meanings were composed of vowels with lower intensity than words judged to be distal.

Fundamental Frequency (F0). Fundamental frequency of the vowels did not co-vary significantly with judgments meaning. (r = .022, p = .406; see Figure 14).

Duration. Words judged to be proximal and distal did not differ in vowel duration (r = -.071, p = .221; see Figure 15).

Formants (f_1, f_2, f_3): Values of the first (r = -.10, p = .141) and second (r = .112, p = .114) vowel formants were not significantly correlated with judgments of word meaning (see Figures 16 and 17). However, words with higher third formant values were more likely to be mapped to proximal than to distal meanings (r = .177, p = .027; see Figure 18).

Summary. Several acoustic characteristics co-varied with judgments of meaning. Words judged to have proximal meanings had significantly shorter durations, lower intensity vowels, and higher values of the third formant. Whereas the cross-linguistic analysis found a trend for vowel-level duration to differ in proximal and distal terms, participants in the behavioral task appeared sensitive to word-level (but not vowel-level) duration, and were more likely to judge a word as having a proximal meaning when word duration was shorter. Although it is not clear why duration per se would be associated with proximal and distal meanings, the relationship is consistent with findings from Shintel, Nusbaum, and Okrent (2006) in which listeners associated the way in which items were produced with properties of objects in a visual display.

In addition to duration, listeners were more likely to judge words as having proximal meanings when they had higher third formants. Because *f*3 is related to vowel height and roundedness, this finding is consistent with Woodworth (1991), who reported that proximal terms contained higher vowels than distal terms across a sample of 27 languages.

Finally, words with lower vowel intensity were more often judged to have proximal meanings. Although not as well documented in the sound symbolism literature, intensity could certainly be another acoustic property that might be used to infer word meaning. Nygaard, Herold, and Namy (2009) found that overall word intensity was one acoustic cue that reliably differentiated dimensional adjective meanings. Certainly, the relationship between intensity of a sound and spatial proximity is coupled in experience, and could be the basis for sound to meaning mapping in language.

General Discussion

This study examined the relationship between sound structure and meaning in relational spatial terms. An acoustic analysis of proximal and distal terms across 16 languages was conducted in order to examine if distinct acoustic profiles differentiate proximal and distal terms cross-linguistically. In general, spatial terms were reliably distinguished by word-level acoustic properties, with less consistent evidence for differences in the segmental properties of contrasting terms.

A behavioral test was also conducted to determine whether listeners attend to aspects of the sound structure of spoken words in attributing meaning to unfamiliar words. In the forced choice translation task, participants accurately judged the meanings of the foreign spatial terms significantly above chance. This study replicates previous work by Clepper, Nygaard, and Namy (2012) in which participants were able to reliably match foreign synonyms for the meanings *near* and *far* drawn from ten different foreign languages with their correct meaning. These findings suggest that some aspect of the acoustic and/or phonological structure of these words across languages is non-arbitrary, and that listeners are able to use this structure as a cue to meaning. Analysis of the relationship between acoustic characteristics and judged meaning revealed that words with higher third formants were more often mapped to proximal meanings, which, although not significant in our own cross-linguistic analysis, is consistent with cross-linguistic patterns observed by Woodworth (1991). In addition, vowel-level intensity and duration measures were correlated with meaning judgments, suggesting that these features were used by listeners to infer meaning.

The finding that characteristics such as word-level pitch was associated with word meaning in the cross-linguistic acoustic analysis and that vowel-level duration and intensity were associated with listeners' judgments of meaning suggest that phonological features are not the only components of spoken utterances that appear to be used systematically to index meaning. The finding is consistent with research that has found that the manner in which linguistic forms are articulated can carry meaning. For example, Nygaard, Herold and Namy (2009) found that when producing infant-directed speech, talkers systematically changed specific acoustic features depending on word meaning. Nygaard et al (2009) found that listeners associated nonsense words that were spoken more loudly, and spoken for a longer duration with meanings *big* and *tall* rather than *small* and *short*. Similarly, Shintel, Nusbaum, and Okrent (2006) found that language users systematically varied the duration of their pronunciation of an utterance depending on the direction and speed of the stimulus they were being asked to describe and that listeners were sensitive to these cues to meaning. These previous studies, in conjunction with the present study, suggest that changes to acoustic features of speech seem to encode semantic information, and that listeners use these as cues to meaning in the relational domain of spatial terms.

Contrary to predictions, the cross-linguistic study did not reveal consistent cross-language differences in the majority of acoustic properties between proximal and distal terms. Although patterns were not consistent across languages, it did appear that individual languages might use different formant values for proximal and distal terms. Unlike more concrete, sensory sound symbolic words, which are thought, by many to rely on perceptual simulation (Kita, 1997; 2001), synesthetic connections (Ramachandran & Hubbard, 2001; Spector & Maurer, 2008), or some direct cross sensory mapping (Marks, 1975; Melara and Marks, 1990), sound symbolism for relational terms such as the contrastive spatial terms in this study may be more related, not due to any iconicity or resemblance, but simply because languages make use of perceptual distinctiveness to mark semantic differences. This opens the possibility of systematically representing concepts that do not have natural perceptual correlates (as may be the case for concrete, perceptually-based concepts). Languages may map specific dimensions of sound attributes to specific dimensions of meaning because differences along the dimension are salient. In this case, we apply a kind of sensory-based metaphor by which we associate specific dimensions of perceptual and conceptual experience. Languages map conceptual dimensions onto acoustic dimensions simply because they are perceptually salient and differentiable. For example, we might associate qualities of sounds (e.g. high and low pitch) with some other perceptual domain (e.g. high and low space). For two domains with similar dimensional structures (scalar gradation or polar opponency) associations between the perceptual and conceptual domains could be formed on the basis of their alignable differences (Gentner & Markman, 1997, Lakens et al, 2011). For example, languages might choose to modulate f_2 to mark differences in meaning, simply because it is acoustically salient, and readily manipulable. If languages do map sound attributes to conceptual dimensions in this way, we would expect languages to differ in their mappings of meaning to sound- a proximal meaning could just as well correspond to a high f_2 as a low f_2, and languages should be just as likely to have one mapping as the other. Such mappings would, most likely, be devised and used by cultural convention to help people think and communicate about concepts. Further, the same acoustic characteristics would likely be modulated for various dimensions of meaning, not just one semantic domain (e.g. a language may use f_2 to differentiate contrastive concepts).

Recent studies by Lakens and colleagues (2011) examined the role of anchoring and polarity in percept-concept mappings, and suggest a possible hybrid explanation for cross-domain association of polar correspondences. Lakens and colleagues suggest that where people consistently align perceptual and conceptual poles, the basis for mappings between the two poles may differ. For instance, while correspondence between one perceptual and conceptual pole may be motivated, the opposite poles may come to be associated with one another solely by virtue of their opponency to their polar counterparts. In this case dimensions are aligned at one pole due to intrinsic association, and at the other pole by their coactivation as opposites to their coupled counterparts. Therefore, when we find participants' responses demonstrate a preference for one mapping over another, this does not necessarily indicate stable associations at both poles of the dimension, but rather, can result from structural factors in experimental designs (Lakens, 2011).

Systematic mapping of sound to meaning, or "sound symbolism" has been suggested to exist on many levels, from segments (e.g. phonesthemes) to words (e.g., onomatopoeia, mimetics) and researchers have attempted to describe these different classes of sound symbolism (Firth, 1935; Jesperson, 1922; Nuckolls, 1999). Although multiple types of sound symbolic correspondences appear to exist in natural language, the mechanisms by which meaning is encoded in symbolic form are poorly understood.

Research in cognition and representation offers several accounts of how such perceptual-conceptual associations might arise. These theories of mechanism underlying sound symbolism are not mutually exclusive, and it is quite possible that several such forces are at play when it comes to language processing and representation of meaning. These theories offer several different models for how sound to meaning correspondences may pattern in language, and predict different patterning in the cross-linguistic analyses and response patterns in our forced-choice translation task. If the association between sound and meaning arises by language-specific conventions, we should find patterns in sound-to-meaning mappings within languages, but no coherent patterns emerging across languages. If these associations reflect intrinsic connectivity within the human nervous system, we might expect to find similar sound-meaning mappings across languages. Similarly, if such associations originate in experience-based functional connectivity across modal systems, we might expect to find similar associations across languages and cultures. With respect to the current study, on the one hand, the results of the cross-linguistic acoustic analyses were mixed, suggesting considerable variation across languages and thus, the possibility of within-language cues to meaning. On the other hand, the results of the test of perceptual sensitivity suggest cross-linguistic agreement among language users on the relation between the acoustic form of a word and its meaning, suggesting the possibility of consistent cross-modal correspondences between sound and meaning in this domain. Regardless of which account can ultimately explain the nature of sound to meaning correspondence for spatial terms, it would almost certainly be the case that these sound symbolic forms would occur probabilistically in language.

Conclusion

In the existing sound symbolism literature, we find some cases of sound symbolic mappings that are conventionalized within a language, and other mappings that appear across languages. The fact that people can reliably nominate correct meanings of words from unfamiliar languages (in a number of domains) strongly suggests that these sound-to-meaning correspondences are not due to language-specific phonological conventions, and are likely much more pervasive throughout languages than we currently recognize. Although few acoustic patterns reached significance in the cross-linguistic analysis, fundamental frequency reliably distinguished proximal from distal meanings. In our behavioral study, English speakers could reliably judge the meanings of unfamiliar spatial words and that judgments of meaning appeared to be based on third formant structure and vowel intensity and duration.

Where we find consistent sound-meaning correspondences across languages, this could be compelling support for a common perceptual-cognitive basis for such mappings. However, finding cross-linguistic patterns in sound to meaning mappings does not, in itself, allow us to distinguish between naturally and experientially based biases since it is likely that humans have similar statistical coupling across modal experiences. Cross-modal associations might result out of common cognitive experience or natural biases to which humans are predisposed.

References

Ben-Artzi, E., & Marks, L. E. (1999). Processing linguistic and perceptual dimensions of speech: interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 579-595.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. (Version 5.3.13).

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. (D. Gentner & S. Goldin-Meadow, Eds.)*Language in mind Advances in the study of language and thought*. MIT Press.

Clepper, L., Namy, L.L. & Nygaard, L.C.,(in preparation). *Cross-linguistic consistency and within-language variability of sound symbolism in natural languages.*

Firth, John Rupert (1935), The Use and Distribution of Certain English Sounds. *English Studies*, 17: 8-18.

Gasser, M. (2004). The Origins of Arbitrariness in Language. *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 26, p. 4–7).

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45-56. American Psychological Association.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(5662), 88-96

Ikegami, T., Zlatev, J. (2007). From pre-representational cognition to language. in Ziemke, T., Zlatev, J., & Frank, R. (Eds.), *Body, language, and mind. Vol. 1, Embodiment* (241-283). (Vol. 1). Berlin, NY: Mouton de Gruyter.

Jesperson, O. (1922). Language: its Nature, Development, and Origin. London: Allen and Unwin.

Kemmerer, D. (1999). “Near” and “far” in Language and Perception. *Cognition*, *73*(1), 35-63.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics, 35,* 379-415.

Kita, S. (2001). Semantic schism and interpretive integration in Japanese sentences with a mimetic: A reply to Tsujimura. *Linguistics*, 39, 419-436.

Köhler, W. (1947). Gestalt psychology (2nd ed.). New York: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. Cognition, 114(1), 19-28.

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 10*, 427-429.

Ladefoged, P. (1996). Elements of Acoustic Phonetics. University of Chicago Press.

Lakens, D., Semin, G. R., & Foroni, F. (2011). But for the bad, there would not be good: Grounding valence in brightness through structural similarity. *Journal of Experimental Psychology: General*.

Marks, L. E. (1987). On cross-modal similarity: auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384-394.

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

Melara, R. D., & Marks, L. E. (1990). Processes underlying dimensional interactions: correspondences between linguistic and nonlinguistic dimensions. *Memory cognition*, *18*(5), 477-495.

Monaghan, P., & Christiansen, M.H. (2006). Why form-meaning mappings are not entirely arbitrary in language. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1838-1843). Mahwah, NJ: Lawrence Erlbaum.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136.

Nuckolls, J. B. (1999). The Case for Sound Symbolism. Annual review of anthropology, 28(1), 225-252. Annual Reviews.

Newman, S.S. (1933). Further Experiments in phonetic symbolism. *Am. J. Psychol*. 45: 53-75.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181-186.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The Semantics of Prosody: Acoustic and Perceptual Evidence of Prosodic Correlates to Word Meaning. *Cognitive Science*, *33*(1), 127-146.

Pinker, S. (1994). *The Language Instinct*. *Cognitive and Behavioral Neurology* (Vol. 1, p. 494). Harper Perennial.

Plauché, M. C., & Bergen, B. K. (1999). Markedness and the evolution of binary spatial deictics :French voilà and voici. In Steve Chang et al. (Eds.), *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society.

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. *J. Experimental Psychol*. 12:225.39

Saussure F. de (1959). Course in General Linguistics, EC Bally, A Sechehaye, A Riedlinger. Transl. W. Baskin. New York: McGraw-Hill.

Sedley, D. (2003). Plato's Cratylus. New York: Cambridge University Press.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, 55(2), 167-177.

Spector, F., & Maurer, D. (2008). The colour of Os: naturally biased associations between shape and colour. *Perception*, *37*(6), 841-847.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology (2006)*, (November), 37-41.

Thompson, P., & Estes, Z. (2010). Sound Symbolic Marking for Size and Shape. *Proceedings of* the *thirty-second Annual Conference of the Cognitive Science Society.*

Woodworth, N. L. (1991). Sound Symbolism in Proximal and Distal Forms. Linguistics, 29(2), 273-299.

Footnotes

[1] Mandarin, Romanian, and Korean were included in both samples.

[2] Terms in German and Georgian were also collected. However, six German words and two Georgian words were omitted from the behavioral experiment due to experimenter error. As a consequence, these were subsequently removed from the cross-linguistic analysis.

Table 1. Acoustic measures for all proximal and distal terms generated by native speakers of 16 languages.

	proximal		distal	
	n=60		n=58	
	mean	s.d.	mean	s.d.
Word Duration (seconds)	0.49	0.14	0.51	0.20
Word Fundamental Frequency (Hz)	153.11	55.76	144.13	52.32
Mean vowel duration (seconds)	0.11	0.05	0.12	0.05
Mean Vowel Intensity (dB)	68.8	10.4	69.6	10.1
Mean Vowel Pitch (Hz)	155.1	58.2	146.9	54.9
Mean First Vowel Formant (Hz)	559.7	214.1	549.8	124.4
Mean Second Vowel formant (Hz)	1656.7	367.7	1575.0	354.9
Mean Third Vowel Formant (Hz)	2730.1	254.4	2711.8	255.4

Table 2. Acoustic measures for foreign words translated from each of the six English spatial meanings (*here, there, near, far, this, that*)

	near		*far*		*here*		*there*		*this*		*that*	
	n=39		*n=35*		*n=12*		*n=14*		*n=9*		*n=9*	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Word Duration (seconds)	0.52	0.12	0.51	0.14	0.45	0.14	0.55	0.33	0.44	0.22	0.49	0.17
Word Fundamental Frequency (Hz)	156.48	60.89	143.54	55.15	154.98	51.34	159.53	51.82	136.04	36.11	122.43	35.56
Mean vowel duration (seconds)	0.10	0.04	0.12	0.05	0.11	0.07	0.11	0.03	0.13	0.09	0.14	0.06
Mean Vowel Intensity (dB)	71.8	8.7	73.0	8.8	64.2	12.3	66.5	10.9	61.6	9.9	61.3	7.5
Mean Vowel Pitch (Hz)	156.9	62.8	147.4	58.9	161.4	57.9	154.5	55.9	138.9	35.2	133.0	35.9
Mean First Vowel Formant (Hz)	574.1	188.1	549.1	110.1	448.0	101.4	533.1	138.4	646.0	357.3	578.6	162.1
Mean Second Vowel formant (Hz)	1591.6	364.6	1613.7	322.2	1801.6	428.6	1522.3	380.1	1745.1	228.6	1506.3	452.6
Mean Third Vowel Formant (Hz)	2700.8	271.4	2707.2	250.5	2795.9	210.2	2683.7	254.1	2769.5	233.3	2773.7	295.4

Figure Captions

Figure 1. Mean fundamental frequency for all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 2. Mean duration for all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 3. Mean intensity of vowels in all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 4. Mean fundamental frequency of vowels in all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 5. Mean duration of vowels in all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 6. Mean frequency of first formant of vowels in all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 7. Mean frequency of the second formant of vowels in all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 8. Mean frequency of the third formant of vowels in all proximal and distal terms generated by individual native speakers. Speakers on the left-hand side generated words for *here, there, near, far, this, that.* Speakers on the right-hand side were asked to generate as many synonyms as possible for near and far.

Figure 9. Percent accuracy in meaning judgments is plotted for proximal and distal meanings. Error bars represent Standard Error.

Figure 10. Percent accuracy in meaning judgments of meaning in force-choice task is plotted for individual word meaning; *here, there, near, far, this, that*. Error bars represent Standard Error

Figure 11. Correlation of word fundamental frequency with proportion of proximal judgments across items (r=-.037, p=.345).

Figure 12. Correlation of word duration with proportion proximal judgments in forced-choice task (r=-.071, p=.221).

Figure 13. Correlation of vowel intensity correlated with proportion proximal responses on the forced-choice task. Intensity is significantly correlated with proximal judgments (r=-.192, p=.019).

Figure 14. Correlation of vowel fundamental frequency with proportion of proximal judgments across items.

Figure 15. Correlation of vowel duration with proportion proximal judgments in forced-choice task.

Figure 16. Correlation of first formant frequency with proportion proximal judgments in forced-choice task.

Figure 17. Correlation of second formant frequency with proportion proximal judgments in forced-choice task.

Figure 18. Correlation of third formant frequency with proportion proximal judgments in forced-choice task.

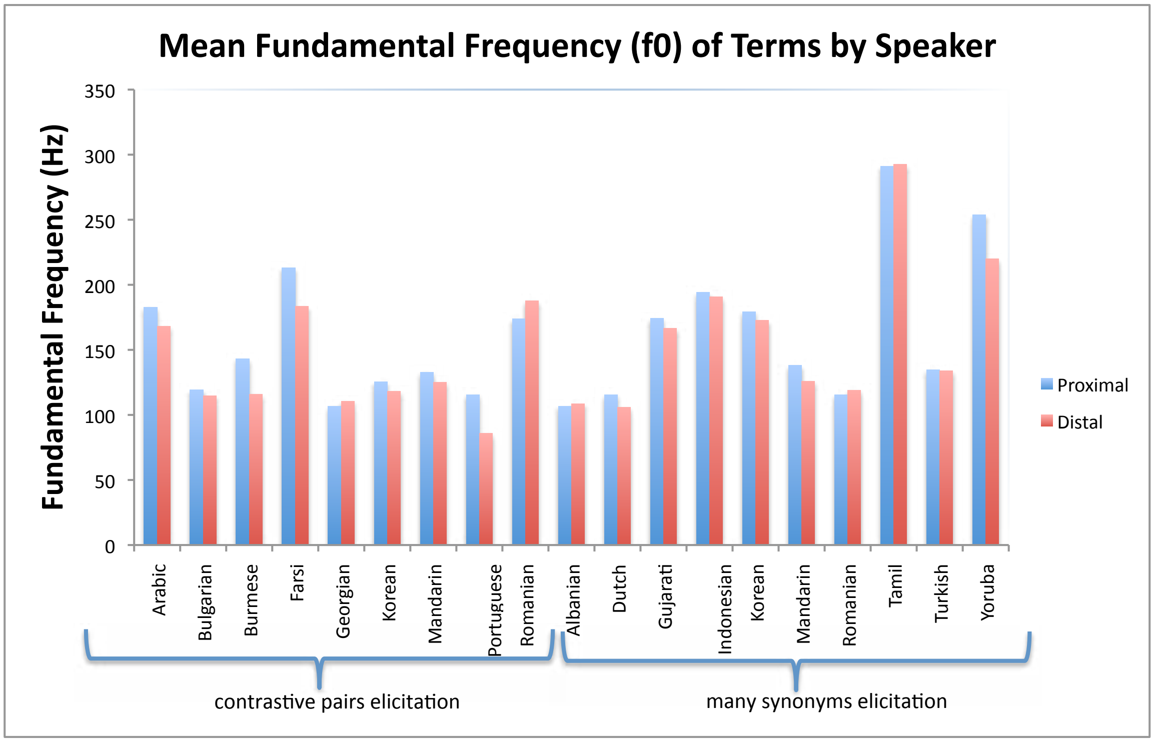

Mean Fundamental Frequency (f0) of Terms by Speaker
Fundamental Frequency (Hz)
350
300
250
200
150
100
50
0
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
contrastive pairs elicitation
many synonyms elicitation
Proximal
Distal

Figure 1.

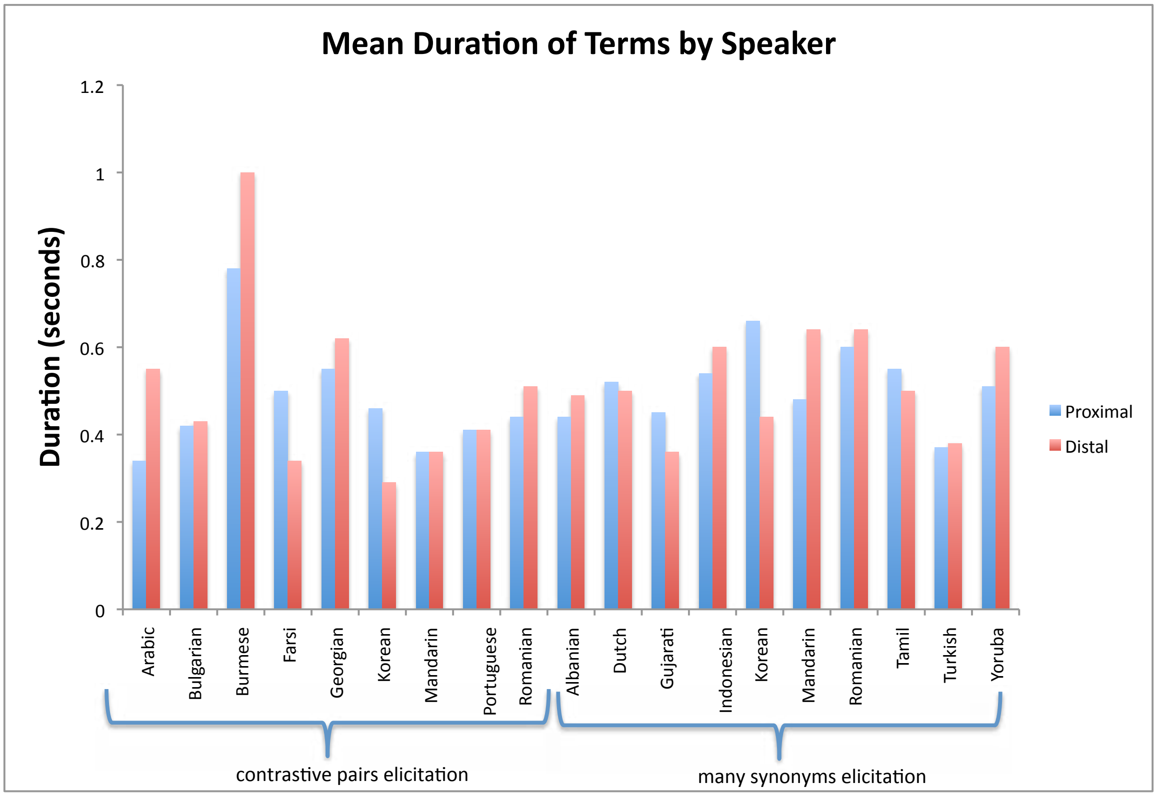

Mean Duration of Terms by Speaker
Duration (seconds)
1.2
1
0.8
0.6
0.4
0.2
0
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
contrastive pairs elicitation
many synonyms elicitation
Proximal
Distal

Figure 2.

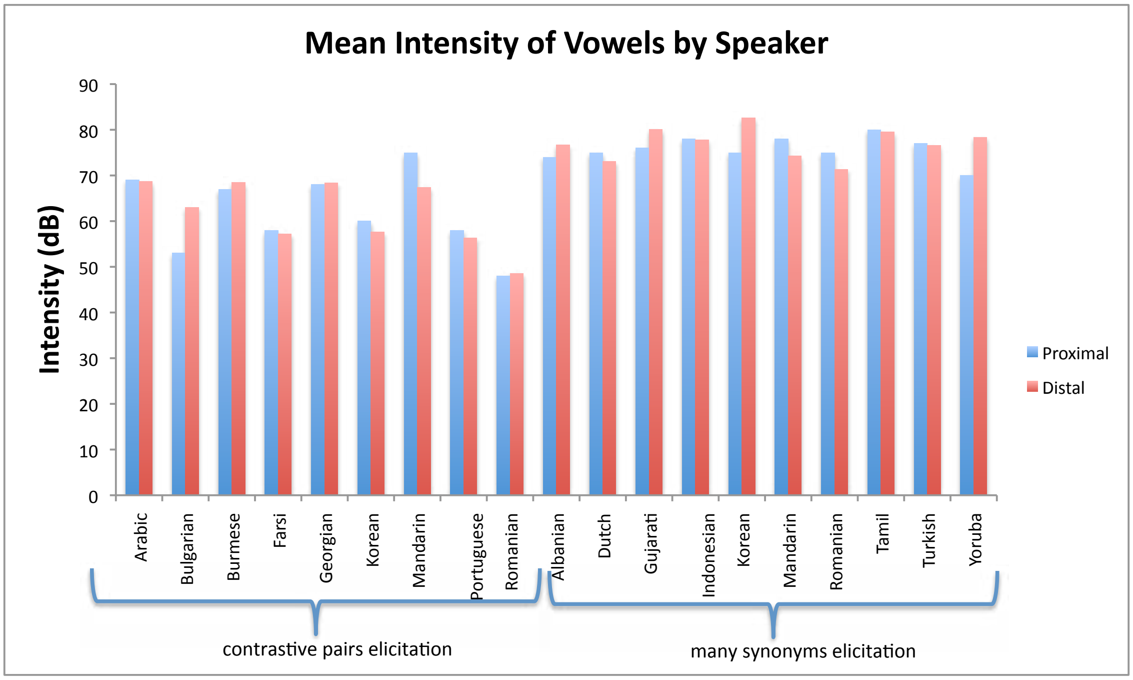

Mean Intensity of Vowels by Speaker
Intensity (dB)
90
80
70
60
50
40
30
20
10
0
Proximal
Distal
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
contrastive pairs elicitation
many synonyms elicitation

Figure 3.

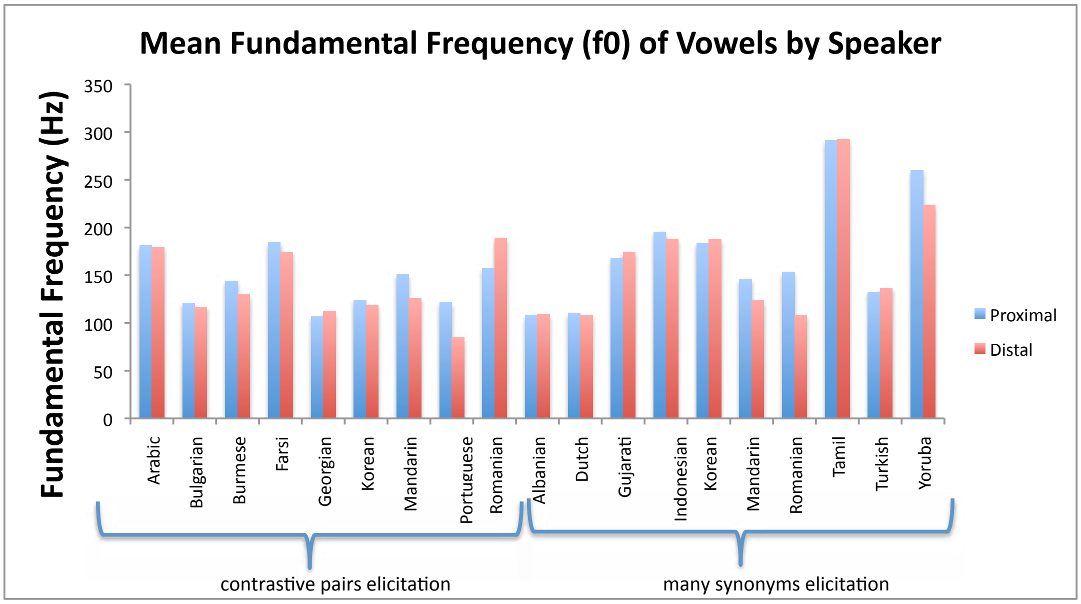

Mean Fundamental Frequency (f0) of Vowels by Speaker
Fundamental Frequency (Hz)
350
300
250
200
150
100
50
0
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
Proximal
Distal
contrastive pairs elicitation
many synonyms elicitation

Figure 4.

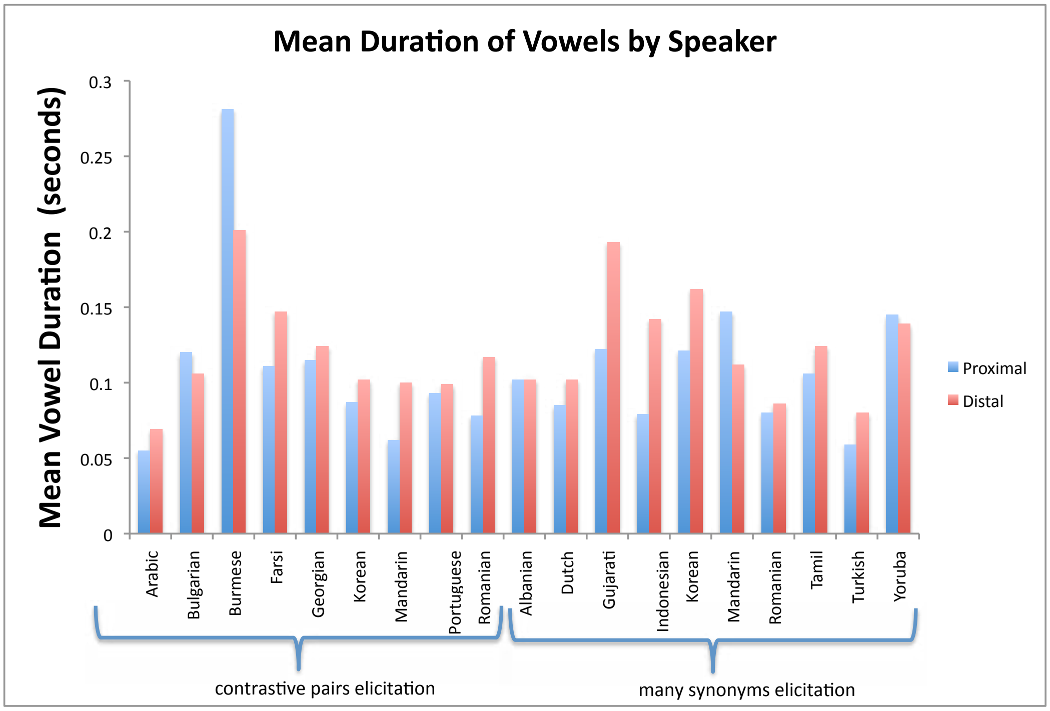

Mean Duration of Vowels by Speaker
Mean Vowel Duration (seconds)
0.3
0.25
0.2
0.15
0.1
0.05
0
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
Proximal
Distal
contrastive pairs elicitation
many synonyms elicitation

Figure 5.

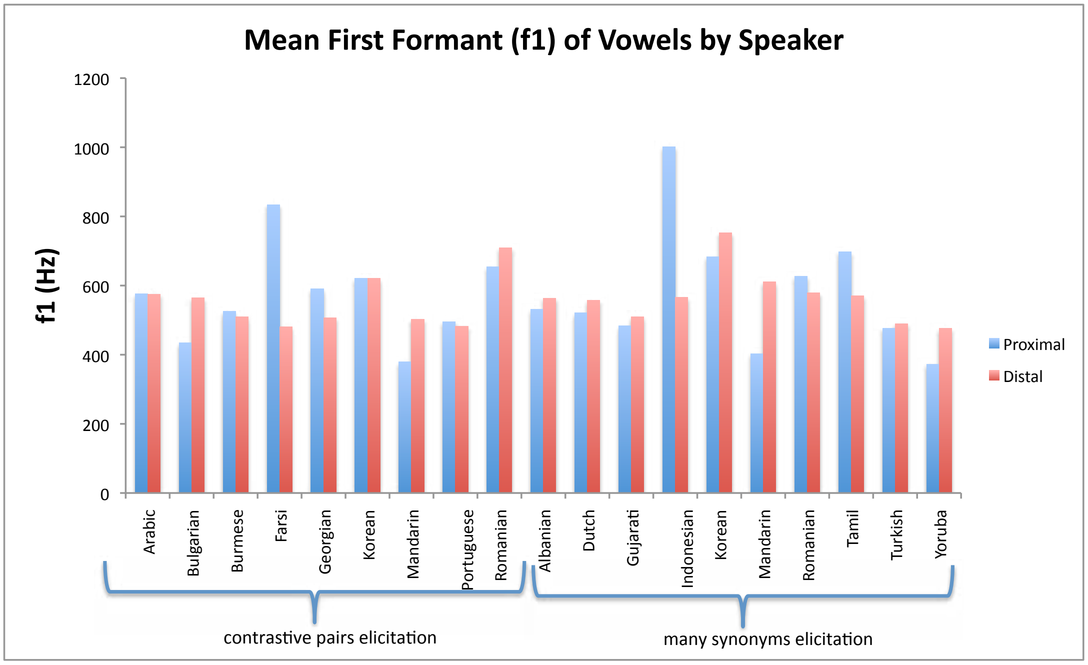

Mean First Formant (f1) of Vowels by Speaker
f1 (Hz)
1200
1000
800
600
400
200
0
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
contrastive pairs elicitation
many synonyms elicitation
Proximal
Distal

Figure 6.

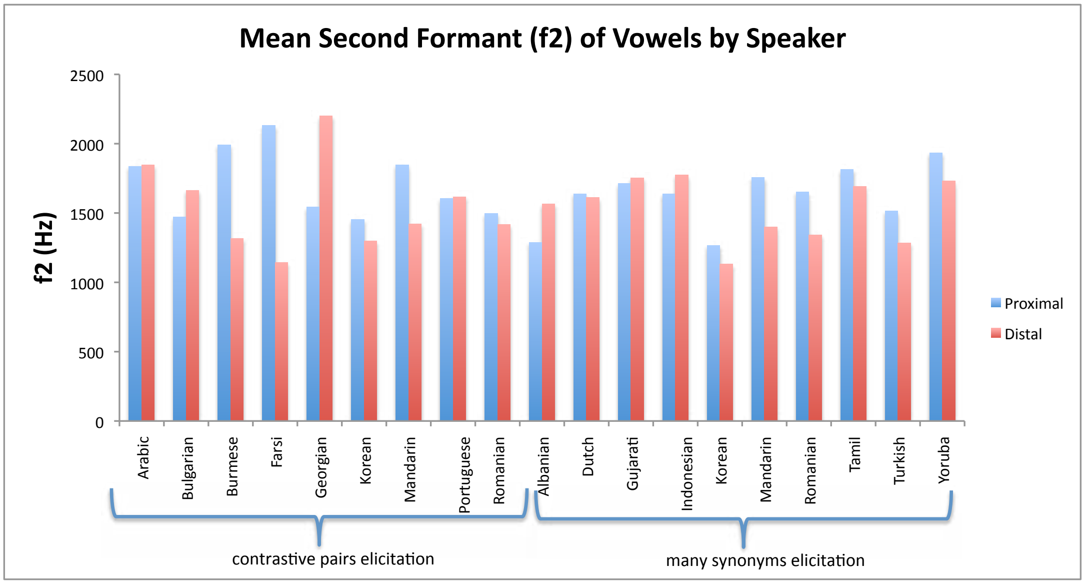

Mean Second Formant (f2) of Vowels by Speaker
f2 (Hz)
2500
2000
1500
1000
500
0
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
contrastive pairs elicitation
many synonyms elicitation
Proximal
Distal

Figure 7.

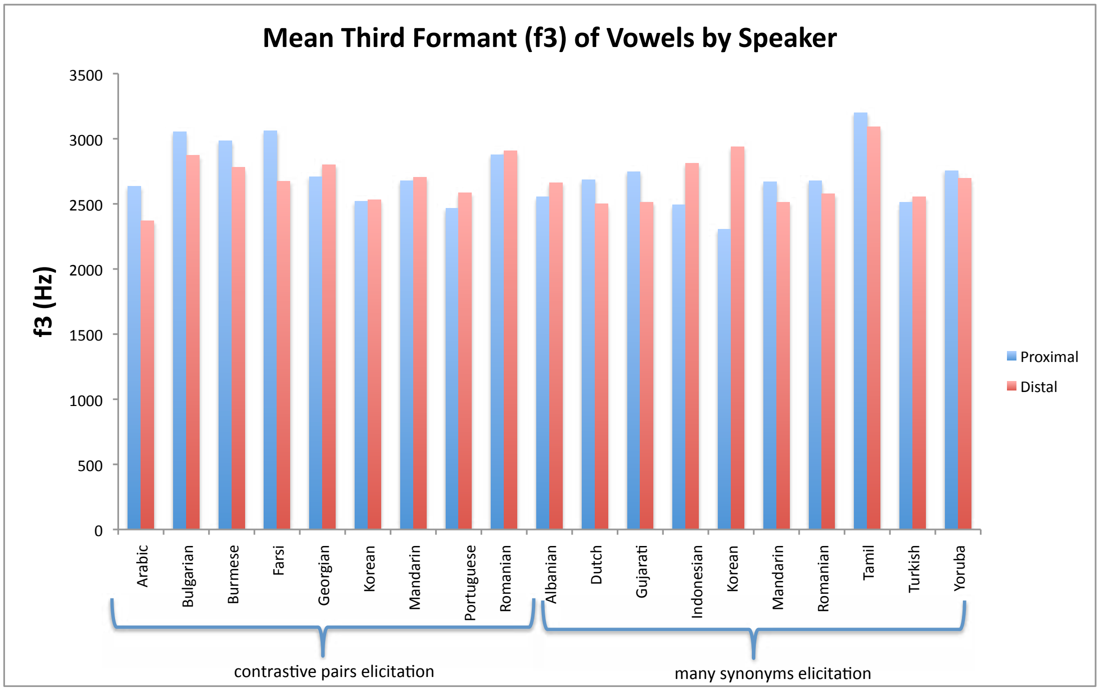

Mean Third Formant (f3) of Vowels by Speaker
f3 (Hz)
3500
3000
2500
2000
1500
1000
500
0
Proximal
Distal
Arabic
Bulgarian
Burmese
Farsi
Georgian
Korean
Mandarin
Portuguese
Romanian
Albanian
Dutch
Gujarati
Indonesian
Korean
Mandarin
Romanian
Tamil
Turkish
Yoruba
contrastive pairs elicitation
many synonyms elicitation

Figure 8.

Behavioral Results

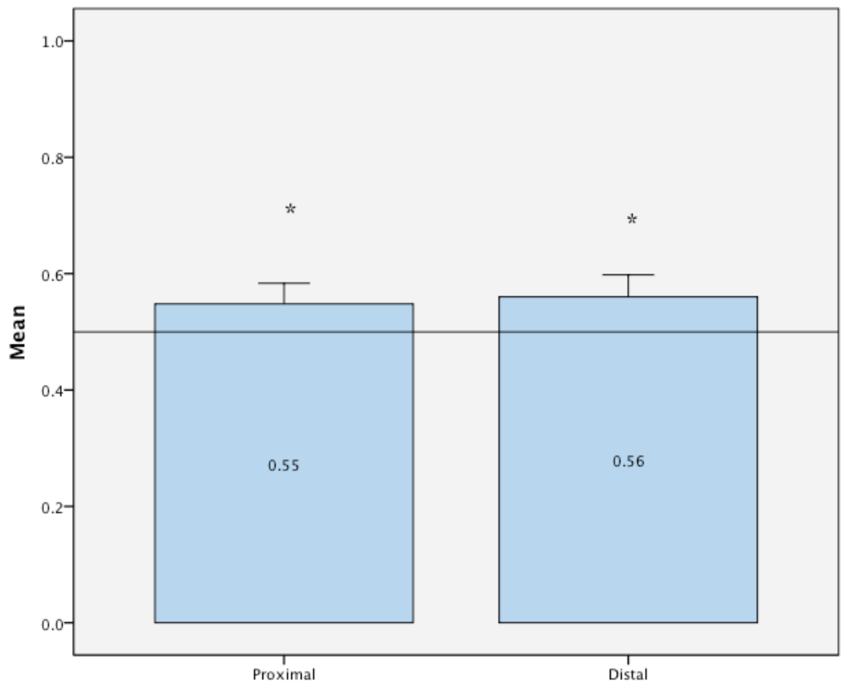


Figure 9.

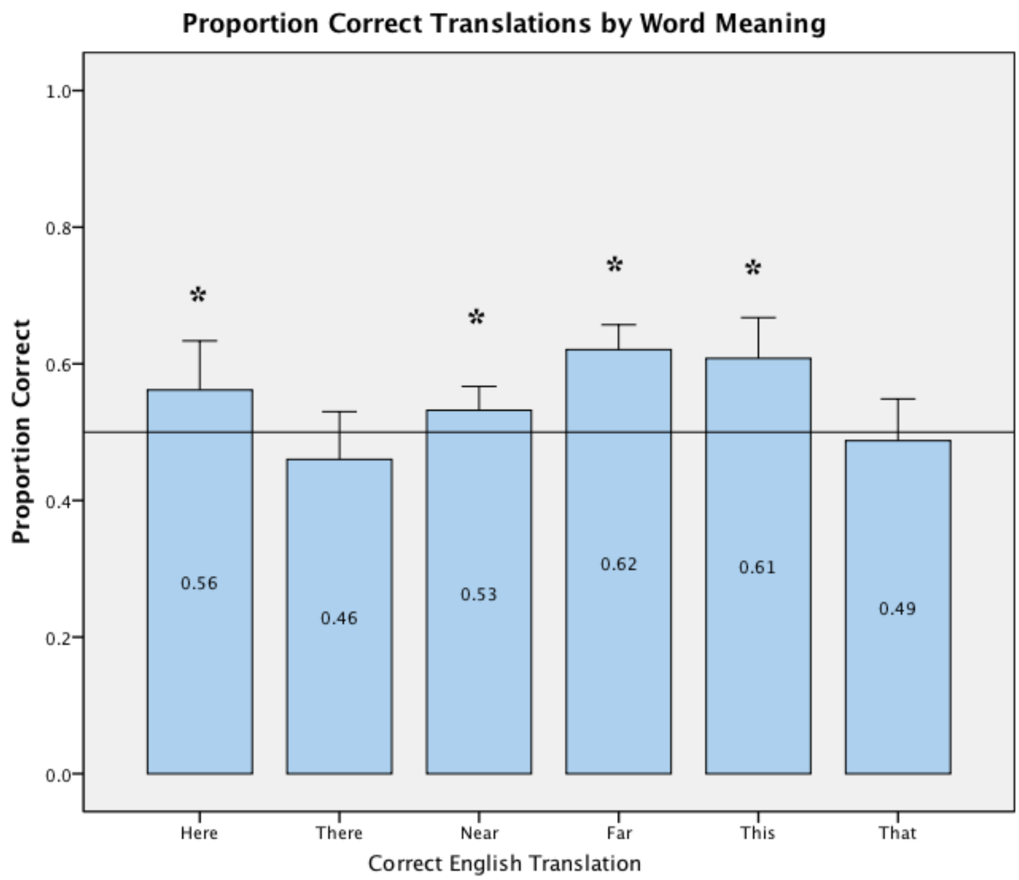
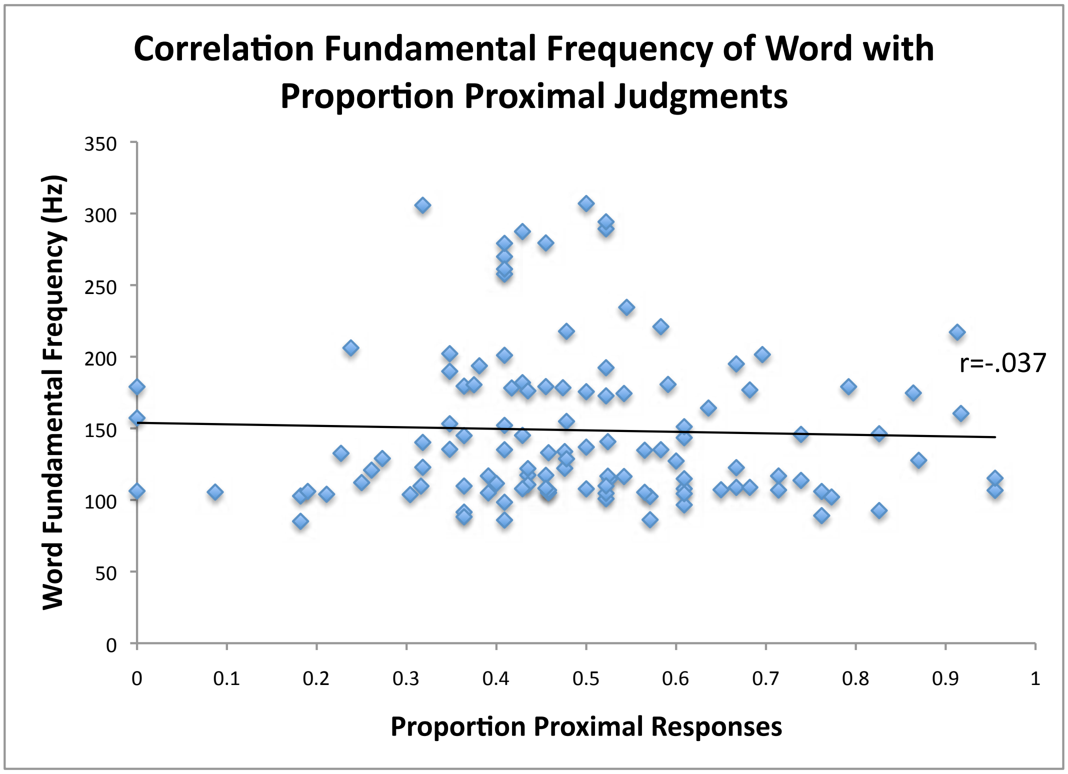
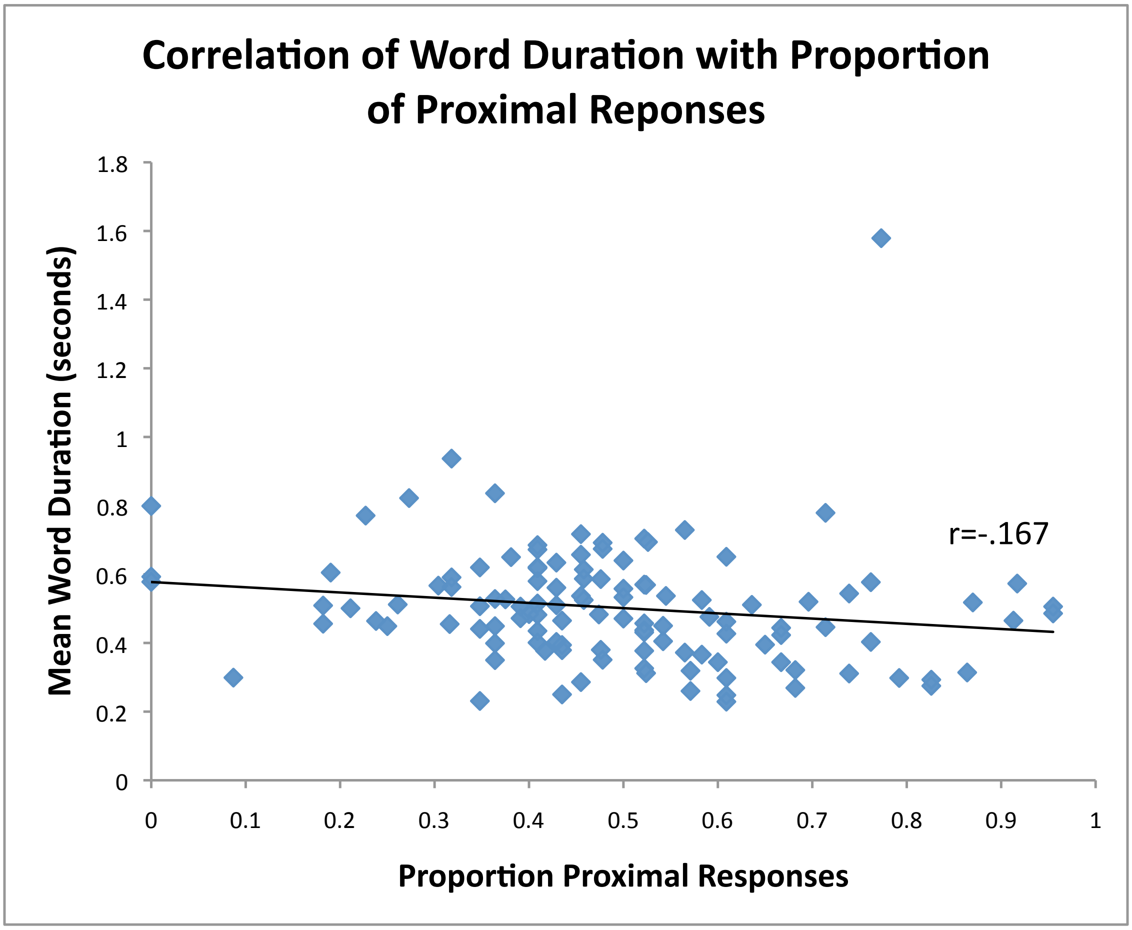

Proportion Correct Translations by Word Meaning
Proportion Correct
1.0
0.8
0.6
0.4
0.2
0.0
*
*
*
*
0.56
0.46
0.53
0.62
0.61
0.49
Here
There
Near
Far
This
That
Correct English Translation

Figure 10.



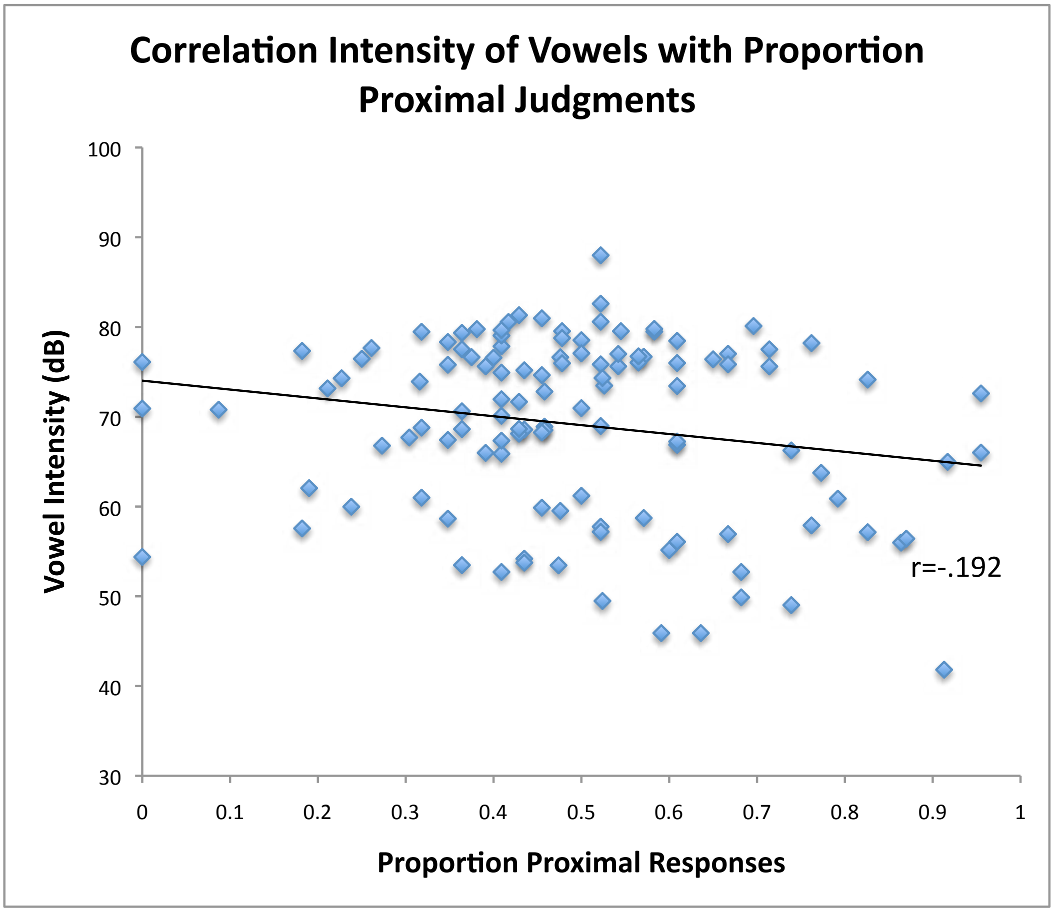
Correlation Fundamental Frequency of Word with Proportion Proximal Judgments
Word Fundamental Frequency (Hz)
350
300
250
200
150
100
50
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
r=-.037

Figure 11.



Correlation of Word Duration with Proportion of Proximal Reponses
Mean Word Duration (seconds)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
r=-.167

Figure 12.



Correlation Intensity of Vowels with Proportion Proximal Judgments
100
90
80
70
60
50
40
30
Vowel Intensity (dB)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
r=-.192

Figure 13.

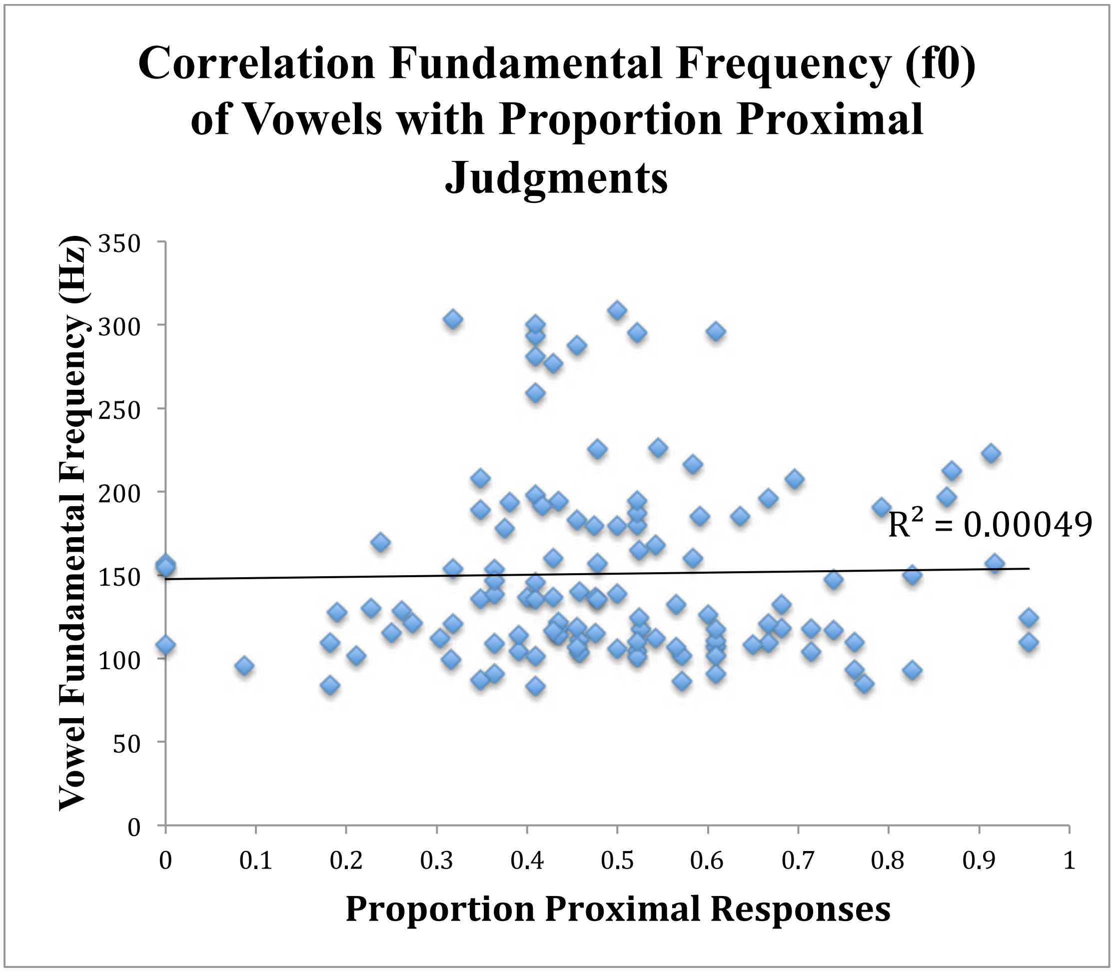

Correlation Fundamental Frequency (f0) of Vowels with Proportion Proximal Judgments
Vowel Fundamental Frequency (Hz)
350
300
250
200
150
100
50
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
R² = 0.00049

Figure 14.

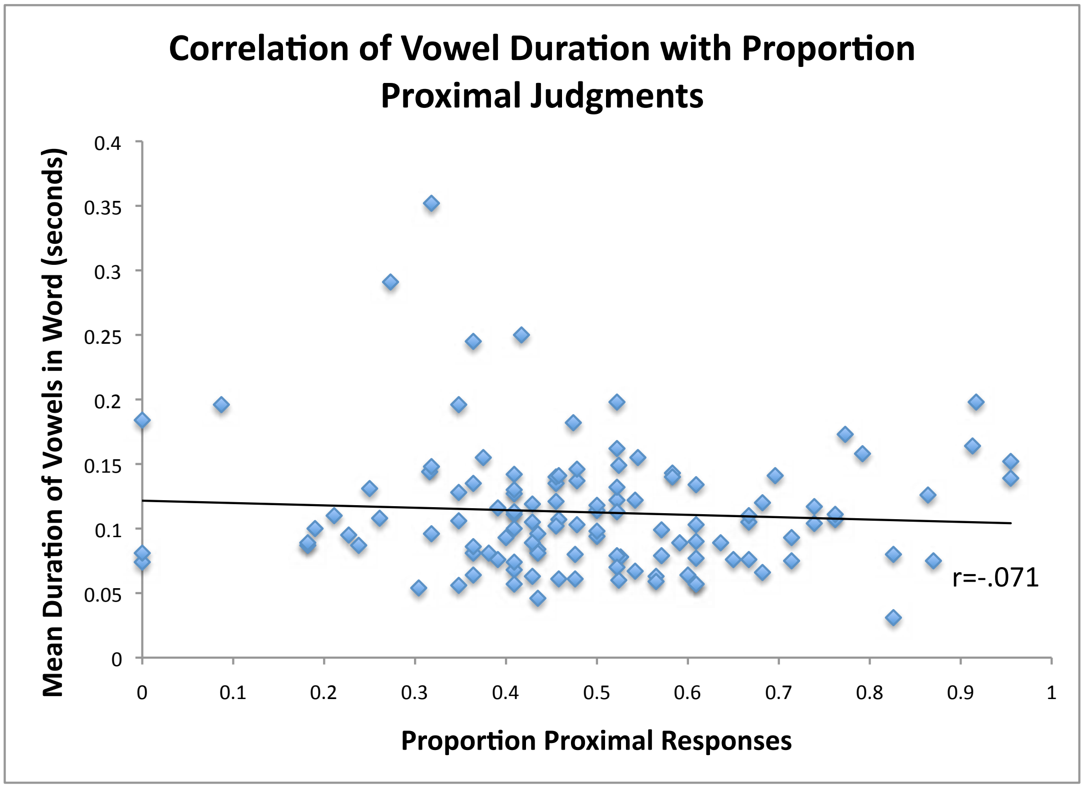

Correlation of Vowel Duration with Proportion Proximal Judgments
Mean Duration of Vowels in Word (seconds)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
r=-.071

Figure 15.

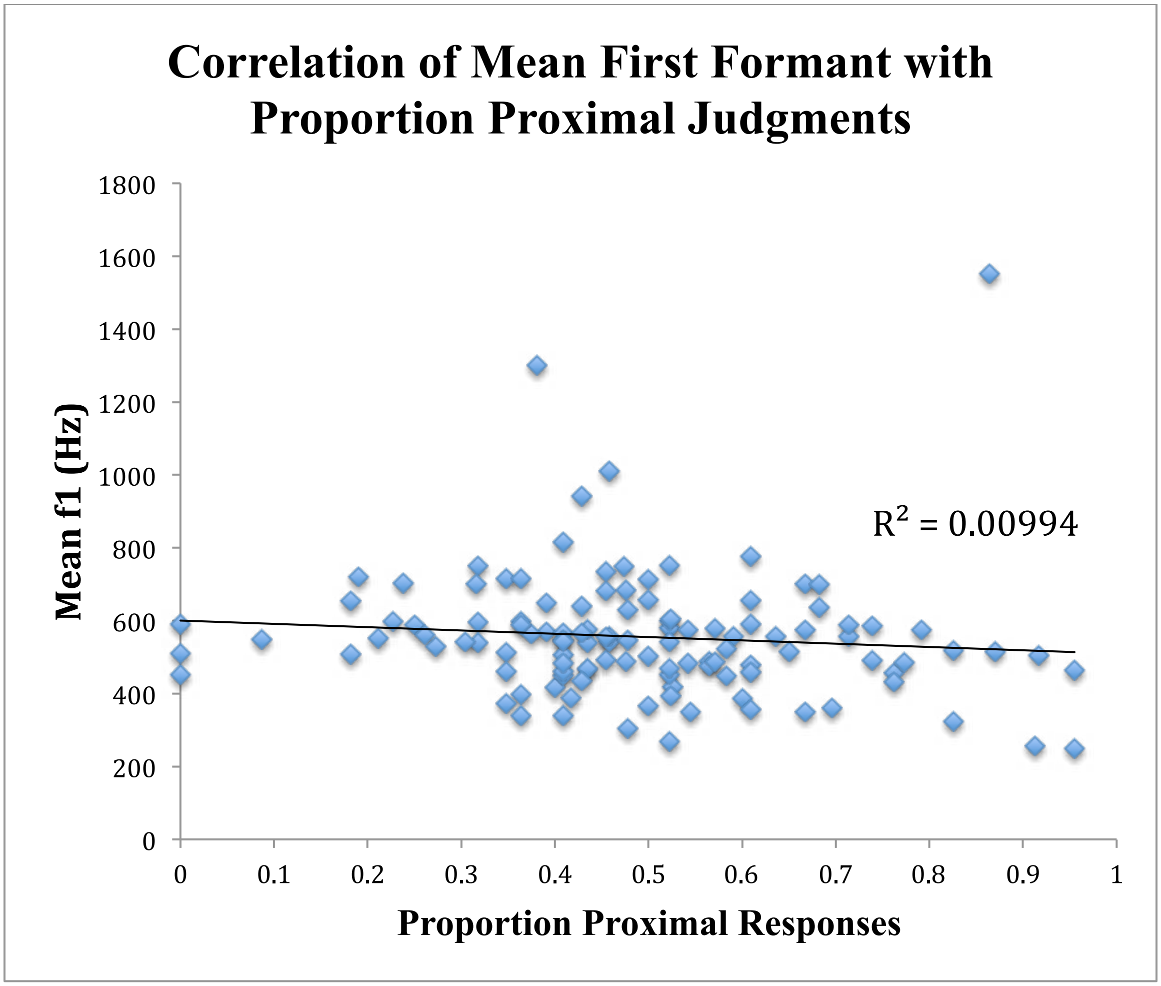

Correlation of Mean First Formant with Proportion Proximal Judgments
Mean f1 (Hz)
1800
1600
1400
1200
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
R² = 0.00994

Figure 16.

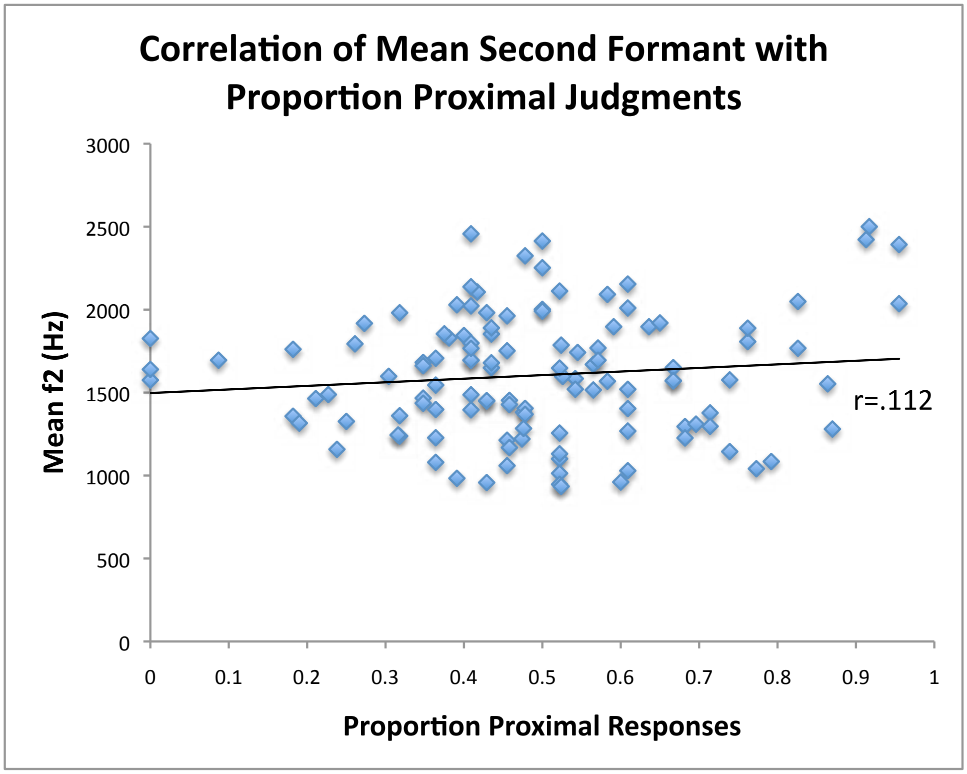

Correlation of Mean Second Formant with
Proportion Proximal Judgments
Mean f2 (Hz)
3000
2500
2000
1500
1000
500
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
r=.112

Figure 17.

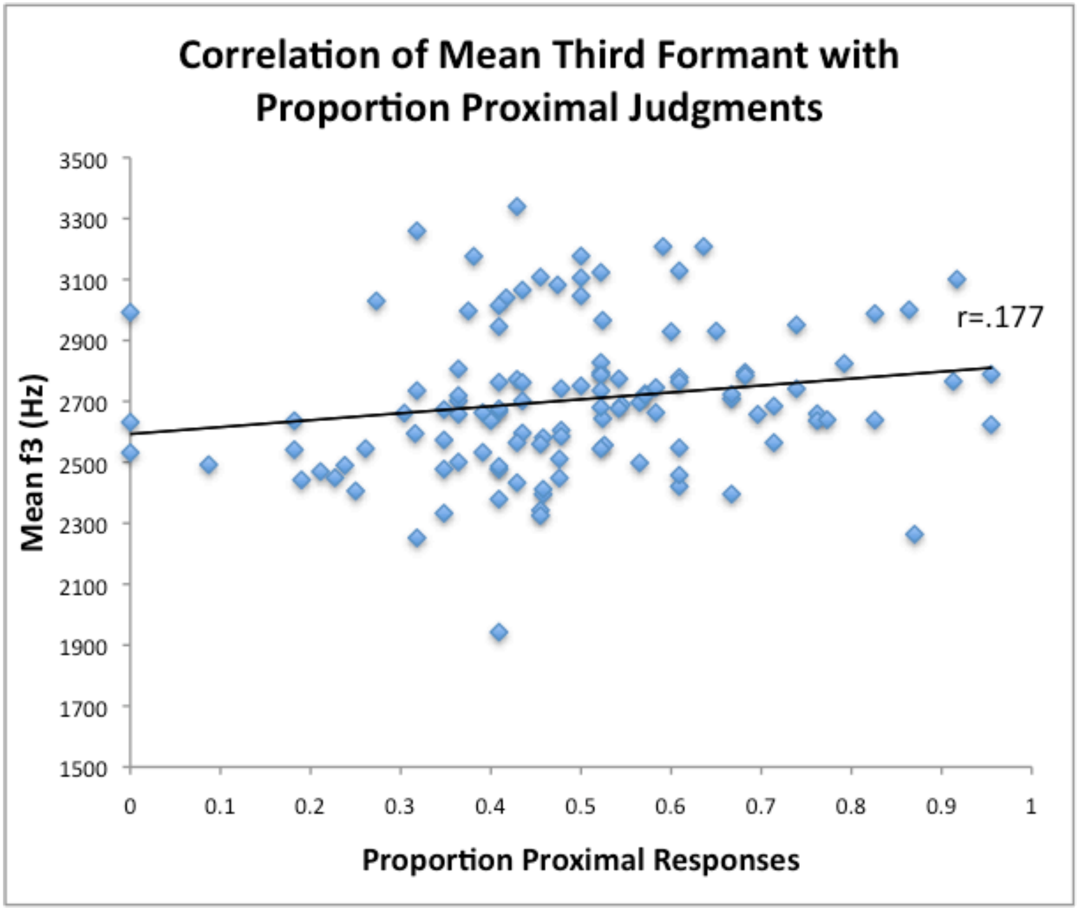

Correlation of Mean Third Formant with Proportion Proximal Judgments
Mean f3 (Hz)
3500
3300
3100
2900
2700
2500
2300
2100
1900
1700
1500
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Proportion Proximal Responses
r=.177

Figure 18.

Appendix: Word counts for stimuli used in acoustic analysis and in behavioral experiment.

Number of Words Translated from English by each Foreign Language Speaker

Elicitation Method	Language	*near*	*far*	*here*	*there*	*this*	*that*	Total words generated
Generate many words for near/far	Albanian	4	4	1	1	~	~	10
	Dutch	6	6	~	~	~	~	12
	Gujarati	1	2	~	~	~	~	3
	Indonesian	2	1	~	1	~	~	4
	Korean*	1	1	~	~	~	~	2
	Mandarin*	4	3	~	~	~	~	7
	Romanian*	3	2	1	1	~	~	7
	Tamil	3	2	~	~	~	~	5
	Turkish	1	1	~	~	~	~	2
	Yoruba	2	2	2	3	~	~	9
Provide best translation for near/far here/there this/that	Arabic	1	1	1	1	1	1	6
	Bulgarian	1	2	1	1	1	1	7
	Burmese	1	1	1	1	1	1	6
	Farsi	1	1	1	1	1	1	6
	Georgian	2	2	~	~	1	1	6
	Korean*	2	1	1	1	1	1	7
	Mandarin*	1	1	1	1	1	1	6
	Portuguese	2	1	1	1	1	1	7
	Romanian*	1	1	1	1	1	1	6
	total	39	35	12	14	9	9	118

Note: for Korean, Mandarin and Romanian, two different speakers contributed to the sample. One in each method of elicitation.