Perceptual Discrimination of Proper Names and Homophonous Common Nouns

Chelsea Sanker
Stanford University

Various characteristics of homophone mates can result in acoustic differences in how they are produced, e.g., frequency, part of speech, and morphological breakdown. Some of these acoustic cues influence perceptual decisions, though resulting accuracy is usually quite low. In this study, I examine homophone mate pairs in which one of the words is a proper name, analyzing both the acoustic characteristics that differ between proper names and homophonous common nouns as well as how these characteristics influence listeners' decisions in identifying these items. Proper names are produced with longer duration, higher F0, and higher intensity than common nouns. All of these characteristics have a corresponding influence on how listeners identify a stimulus; longer duration, higher F0, and higher intensity increase the likelihood that a stimulus will be identified as being a proper name. I explain these acoustic differences as resulting from proper names receiving sentential stress more often than other words and listeners having corresponding expectations that proper names are more likely to be stressed.

Keywords: homophones, perception, proper names, sentential stress

1 Introduction

Homophone mates often exhibit acoustic differences when produced in natural speech. There are multiple ways that these patterns in production might be explained. On the one hand, they might reflect phonetic details that are an inherent part of the representation of each word, consistent with an exemplar model (cf. e.g., Goldinger 1998, Pierrehumbert 2002). On the other hand, they might be explained by effects of the environment and the process of lexical access, e.g., higher predictability in context facilitating lexical retrieval and subsequently characteristics like duration (Gahl et al. 2012, Kahn & Arnold 2012), and part of speech aligning with prosodic differences based on the position that a word typically occurs in (Sorensen et al. 1978, Conwell 2017).

In contrast to some of the relatively robust differences found in production, perceptual identification of homophone mates generally has low accuracy (e.g., Bond 1973), even under facilitating conditions, e.g., after recent exposure to the same words, with all items produced by the same speaker (Sanker 2022). While low accuracy in perceptual identifications might support the analysis that production patterns are merely effects of the environment, the fact that accuracy is ever above chance might be interpreted as favoring the analysis that phonetic details are part of the representation of individual words.

In this paper, I examine perceptual identifications of proper names and homophonous common nouns. Based on listeners' overall accuracy and the acoustic cues that predict their decisions, I argue that there are predictable prosodic tendencies of proper names. They are more likely to receive sentential stress than other words are, and listeners make use of these acoustic characteristics because they expect proper names to exhibit correlates of stress.

1.1 *Acoustic details in production* Homophone mates can differ in the acoustic characteristics that they are produced with. For example, lower frequency words have longer durations and larger vowel spaces

* I would like to thank the audience at WECOL 2022 for their thought-provoking questions and comments.

than higher frequency words (Gahl 2008, Guion 1995). Nouns have longer durations and larger vowel spaces than verbs (Lohmann 2018, Conwell 2017). Some studies find that segment duration is influenced by the morphological breakdown of a word, but results vary across studies and across different morphemes (Plag et al. 2017, Seyfarth et al. 2018).

There are two main possibilities for how these phonetic differences are represented. The first possibility is that they are an inherent part of the representation of specific words. The second possibility is that they are driven by syntactic and pragmatic context and cognitive processes in lexical access. In addition to the question of how these differences are represented, it is relevant to consider how these differences arise. Even if the acoustic patterns become part of the inherent part of specific words, effects of contextual factors are necessary in order to explain how the differences originate and why they have the observed patterns, e.g., why lower frequency words have longer durations rather than the opposite.

Many of the acoustic differences between homophone mates disappear when environmental factors are controlled for, suggesting that they are effects of the environment rather than being inherent characteristics in the representation of individual words. For example, Jurafsky et al. (2002) demonstrate that the frequency-based differences between homophone mates within a speech corpus are largely eliminated when factors like speech rate, predictability based on neighboring words, and surrounding segments, are included as factors. Guion (1995) finds that frequency-based differences are only present for words produced in meaningful sentences, and that they are eliminated when words are elicited in a frame sentence. Differences based on part of speech are also reduced when position in the sentence is controlled (Sorensen et al. 1978, Conwell 2017), though there are still some differences predicted by part of speech when position in the sentence is controlled (Conwell 2017, Lohmann & Conwell 2020).

Some work suggests that phonetic patterns caused by the context that a word frequently occurs in can become part of the representation of that word; some effects of word-specific informativity are significant even when the environmental factors are accounted for (e.g., Tang & Shaw 2021, Seyfarth 2014). Sóskuthy and Hay (2017) find that words which are often lengthened due to occurring utterance-finally are also longer than other words when they occur elsewhere. However, these results might reflect indirect effects, rather than the phonetic details becoming part of the representation of specific words. Informativity might influence ease of lexical retrieval, which in turn results in acoustic differences (Gahl et al. 2012, Kahn & Arnold 2012). The relationship between how frequently a word occurs utterance-finally and the typical duration of that word after accounting for position might also be an effect of word-specific informativity: A word might have longer duration due to low overall informativity, and words with lower informativity might also be more likely to occur in prominent positions, such as utterance-finally. Effects of informativity do not necessarily require word-specific phonetic targets, even when informativity are associated with the particular word rather than its context.

Contextual effects associated with part of speech are similar for real words and for nonce words (Conwell & Barta 2018). Given that nonce words do not have pre-existing representations, these patterns in nonce words must be attributed to the environment rather than being inherent to the (nonce) word's representation. Other environmental effects are also similar in real words and nonce words. For example, the first mention of a word has a longer duration than the second mention (Fowler & Housum 1987, Clopper & Turnbull 2018), which is also observed in nonce words (Keung 2013). These effects must be due to context rather than being inherent to particular words, since the variation is within the realization of an individual word. Goldinger (1998) demonstrates that the number of repetitions of nonce words can create patterns similar to the lexical frequency of real words, so the results for effects of repetition on acoustic characteristics may suggest that effects of lexical frequency could be explained in the same way. If these patterns can be predicted without word-specific phonetic details, using word-specific phonetic details to account for the same patterns in real words is unnecessary.

In convergence experiments, shifts are generalized to words that were not heard during exposure and also to sounds that were not heard during exposure but which have features shared with the exposure items, e.g., exposure to lengthened VOT in /p/ results in lengthened VOT in /p/ in novel words and also lengthened VOT in /k/ (Nielsen 2011). While generalization of a shift across words does not exclude the possibility that listeners also have word-specific phonetic targets, it raises the question of the weighting that each association would have, e.g., how strongly weighted word-specific phonetic details would need to be in order to outweigh category-level phonetic details, given that speakers encounter far more instances of each phonological category than of a specific word containing that phoneme.

1.2 *Acoustic details in perception* Patterns in perception provide a separate line of evidence that can help shed light on the status of phonetic details as they relate to individual words. There is evidence that listeners do have acoustically detailed memories, e.g., more accurately remembering that they have heard a word if it is presented again in the same voice (e.g., Hintzman et al. 1972) and more accurately identifying familiar tokens (Chiu 2000). However, these acoustically detailed memories are not limited to speech characteristics; listeners identify a word more accurately when it is presented with the same background noise (e.g., a phone ringing) the second time (Pufahl & Samuel 2014). This sensitivity to non-linguistic acoustic details might suggest that these studies are capturing something about the broad range of details retained in short-term memory, rather than indicating that phonetic details of particular speech events are integrated into the representation of each word.

Some of the tendencies that are present in production seem to influence listeners' expectations. For example, speakers make more accurate identifications for stimuli that exhibit reduction patterns that align with the reduction that they are typically produced with, e.g., deletion of /ə/ (Connine et al. 2008) and realization of underlying /t/ (Pitt et al. 2011). The existence of these perceptual effects might provide evidence in favor of phonetic details being part of word-specific representations. However, these expectations do not need to be based on word-specific phonetic representations; it is possible that listeners' decisions are based on expectations about broader patterns. For example, listeners might expect high frequency words to be reduced, rather than having separate expectations about the reduction of each particular word. A potential parallel comes from reduction with repetition; listeners have above-chance accuracy in deciding whether a stimulus was the speaker's first or second time saying that word (Fowler & Housum 1987), which cannot be due to phonetic details inherent in the representation of each word.

Listeners are also sensitive to the acoustic correlates of part of speech. Conwell (2015) demonstrates that noun vs. verb uses of polysemous real words like *hug* elicit different neural responses, but finds no significant effect for nonce words produced in the same sentences. While this could be interpreted as suggesting that acoustic differences are encoded in the representation of words, the results for real words and nonce words in this study might differ because nonce words do not activate a stage of processing in which they would be linked with part of speech. Infants habituated to noun forms of phonologically ambiguous items (e.g., *dance*) preferentially look towards stimuli of verbs, suggesting that they are sensitive to the acoustic patterns that are shared within each category (Conwell & Morgan 2012). Notably, this learning is at the level of the part of speech category, rather than being associated with individual words, because the infants were being habituated to the broader part-of-speech categories rather than one part of speech just for a particular word.

Listeners are also sensitive to correlates of emotional valence, e.g., duration, F0 mean, and F0 range. They can identify the emotion being conveyed by a speaker (Nygaard & Lunders 2002) and also will use these cues to emotion to identify the meaning of nonce words (Nygaard et al. 2009) and homophones (Nygaard & Lunders 2002). The use of these cues in nonce words indicates that listeners have associations between emotional valence and acoustic characteristics that are independent of the representation of specific words.

Although accuracy for distinguishing between most homophone mates is low, some homophone mates seem to have higher discriminability based on having strong tendencies in their prosody. Martinuzzi and Schertz (2022) demonstrate that listeners can distinguish between the attention-seeking vs. apology functions of "sorry" with high accuracy (64.7%). Several of the prosodic cues that distinguish these two functions in production are predictive of listeners' identifications: Duration, mean F0, intensity, and F0 contour. Accuracy in these decisions doesn't necessarily require word-specific memories (cf. intonational patterns for questions vs. statements); the prosodic differences between each function of "sorry" may be a by-product of syntax and pragmatics, rather than being inherent parts of the representation of the word. In part, listeners' accuracy for these items may be influenced the fact that both functions of "sorry" often occur in isolation, which could help listeners map the prosodic patterns from production onto the stimuli being heard in isolation.

1.3 *Proper names* Homophone mate pairs with proper names (e.g., *Phoenix, phoenix*) may be a useful group to examine, because proper names differ from other words in a range of ways. Little previous work has examined whether proper names and common nouns exhibit systematic phonetic differences, though there is some evidence for such differences, e.g., in duration (Whalen & Wenk 1994). There are several

reasons why differences might be expected.

Proper names can differ from other nouns syntactically; for example, determiners cannot combine with proper names in English (Longobardi 1994). Such patterns might suggest that proper names have different syntactic structure than other nouns, e.g., forming DPs on their own. Different phrasal structure could result in different prosodic patterns for proper names than for other nouns. If those prosodic differences set expectations that listeners use in identifying words, homophone mate pairs including proper names might produce perception results similar to what was found for the attention-seeking vs. apology functions of "sorry" (Martinuzzi & Schertz 2022). Use of prosodic cues might interact with the type of context expected for different types of proper name; personal names might be easier to identify than other words because they appear in isolation as vocatives, while most nouns usually do not appear in isolation.

Proper names might be processed differently than other words are. They differ from other words semantically because they have no clear meaning as such; they just function as reference to particular entities (e.g., Yasuda et al. 2000). This relative lack of meaning may underlie why people tend to remember names less accurately than other words, and why names with no associated meaning are remembered less accurately than words which do have an associated meaning, e.g., the name *Baker* is associated with the noun *baker* (Cohen 1990). The differences in processing of proper names are also reflected in different neural activity (Yasuda et al. 2000, Desai et al. 2023) and can produce differences in aphasia, with some patients exhibiting greater impairment for proper names than common nouns and others exhibiting greater impairment for common nouns than proper names (Semenza 2006). Differences in lexical access might result in phonetic differences, as discussed above (Gahl et al. 2012, Kahn & Arnold 2012), and the lack of semantic connections might impact predictability and subsequently the acoustic correlates of predictability.

Personal names are likely to be less predictable in context than other words; when the referent is predictable from the discourse context, a name is likely to be replaced by a pronoun. Other words may have high predictability in context based on their semantic connections, but names lack this semantic network (Yasuda et al. 2000). Less predictable words are more likely to be stressed (Pan & Hirschberg 2000), so proper names might be more likely to receive sentential stress than other words are. If proper names are more likely to be stressed, they should exhibit the characteristics of sentential stress, including longer duration, higher F0, and greater intensity (Breen et al. 2010).

Lexical frequency might also contribute to phonetic differences between proper names and other words. Proper names usually have lower frequency than other words: Among words in SUBTLEX that are listed as just having a noun usage and a proper name usage, the proper names have a median log frequency of 1.1, while the nouns have a median log frequency of 2.5. As discussed above, lexical frequency is correlated with duration and other reduction patterns (Gahl 2008, Guion 1995, Clopper & Turnbull 2018).

1.4 *This study* This study examines the perceptual identification of homophone mate pairs in which one is a proper name, using several categories of proper nouns with a range of lexical frequencies. What acoustic cues do listeners use to try to distinguish between these homophone mates, and are these the same acoustic differences that are present in production? The results can shed some light on the status of phonetic details in the representation of proper names and other words.

2 Methods

2.1 *Participants* The participants were 22 native speakers of American English, who were members of the Brown University community. Data was also collected from 2 additional participants, but they were excluded based on providing the same response for almost all of the trials for homophone mate pairs; that uniformity created convergence issues in models testing predictors of how a stimulus was identified, in addition to suggesting that those participants were not completing the task as intended.

2.2 *Stimuli* Stimuli were produced by two native speakers of American English, one male and one female, in meaningful sentences. The sentences were constructed so that homophone mates occurred in environments that were syntactically and phonologically as similar as possible (e.g., "John likes the Phoenix painting", "John likes the phoenix painting"). These sentences were elicited orthographically in randomized order using the software PsychoPy (Peirce 2007) in a sound-attenuated room with a stand-mounted Blue Yeti microphone using the Audacity software program, and digitized at a 44.1 kHz sampling

rate. The target items were extracted from these sentences and presented in pairs.

There were 5 categories of proper names: Brands (e.g., *Bobcat*), cities (e.g., *Buffalo*), human names (e.g., *Holly*), possessives (e.g., *Poppy's*), and teams/bands (e.g., *Dolphins*). There was also a category of definite phrases (e.g., *The Creature*).

2.3 *Procedure* The study was run in-person in a quiet room using PsychoPy (Peirce 2007); the acoustic stimuli were presented to participants over headphones. Participants were instructed that they would hear pairs of words, and that in each pair one item would be a word that occurs with a capitalized first letter, and one would be a word that occurs in lowercase. For half of the participants, the framing of the instructions was that they were deciding whether the "capitalized word" was the first word or the second word (e.g., *Phoenix, phoenix* or *phoenix, Phoenix*). For the other half of the participants, the framing of the instructions was that they were deciding whether the "plain/lowercase word" was the first word or the second word. Responses were given with the left and right arrow keys on the keyboard; instructions remained on the screen indicating which arrow corresponded to the "capitalized" or "lowercase" word being first and second.

There were 33 pairs of target items and 29 pairs of filler items. Filler trials contained unambiguous pairs (e.g., *Seattle, unicorn*); filler trials are not included in the analysis. Each pair appeared in both orders, for a total of 124 stimuli in each block. Each participant heard one block of items from each of the two speakers, with the same pairs in both blocks; the order of the speakers was balanced across participants.

Results come from mixed-effects regression models calculated with the lme4 package in R (Bates et al. 2015); p-values were calculated with the lmerTest package (Kuznetsova et al. 2015).

3 Results

3.1 *Accuracy* Table 1 presents the summary of an intercept-only mixed effects logistic regression model for accuracy in identifications of homophone mates. There were random intercepts for participant and for word pair.

Table 1 Regression model for accuracy.

	Estimate	SE	z-value	p-value
(Intercept)	0.24	0.060	4.0	< 0.0001

Overall accuracy was significantly higher than chance (56%); listeners could distinguish between proper names and homophonous common nouns, though accuracy was low as compared to decisions about the phonologically distinct filler items (86%).

Including the category of proper name did not significantly improve the model ($\chi^2 = 1.7$, df = 4, p = 0.78), so it was not included as a factor. However, Figure 1 presents the accuracy of identifications for homophone mate decisions divided by category and by word. The accuracy is for both words in each homophone mate pair (e.g., both *apple* and *Apple*), even though the words are organized based on the category of the proper name. Some words potentially could fall into multiple categories (e.g., *Harmony* was categorized as a city, but can also be a human name).

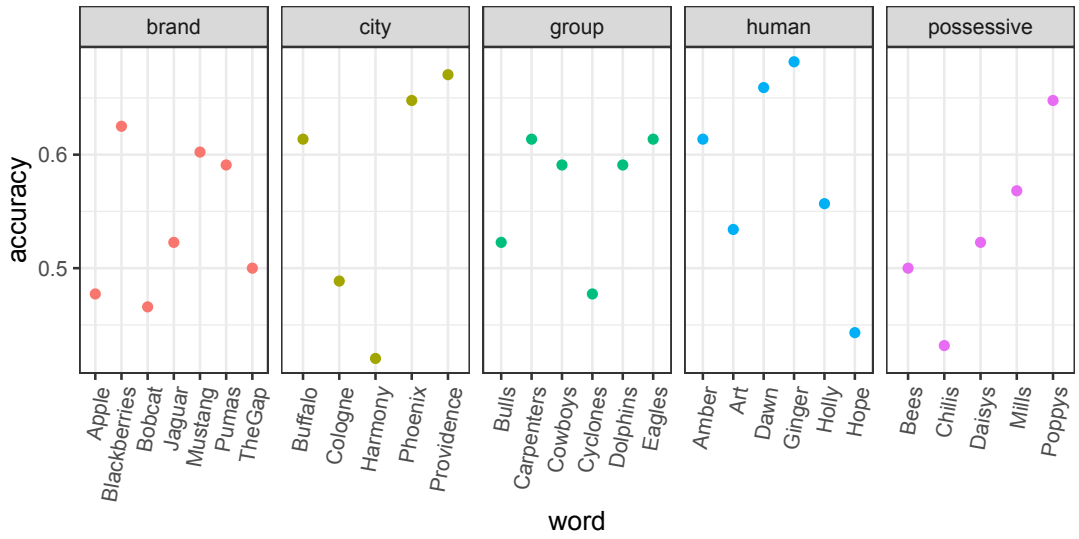
Adding trial number and block number marginally improved the model, suggesting a slight improvement with subsequent trials within a block ($\beta = 0.0029$, SE = 0.0017, z = 1.8, p = 0.077) but a decrease in accuracy in the second block, perhaps due to the switch to a different speaker ($\beta = -0.21$, SE = 0.12, z = -1.8, p = 0.074).

Lexical frequency was considered as a possible predictor, using the frequency of each word within the SUBTLEX corpus for US English (Brysbaert & New 2009). Analyses used log(1+Frequency), to handle words with a frequency of 0. Accuracy was not predicted by lexical frequency; adding log lexical frequency to the model did not significantly improve the fit ($\chi^2 = 0.97$, df = 1, p = 0.33). For a model restricted to accuracy just for identifications of the proper names, log lexical frequency also did not provide a better fit than a model without it ($\chi^2 = 0.26$, df = 1, p = 0.61). Given the low frequency of many of the proper names, it is likely that some of the proper names were not familiar to all of the listeners, particularly the names of bands and sports teams. The lack of effect of lexical frequency suggests that listeners'

identifications are not based on memories of word-specific acoustic details.

Some of the possessives could not be familiar to participants because they were not based on real businesses. Chili's is a well-known chain and Bee's is a restaurant in Providence (where the study was run), while the other three possessives were invented as plausible business names; some participants might have been familiar with actual businesses by these names, but most of them probably were not. Notably, these three were the possessives with the highest accuracy.

Figure 1 Accuracy of identifications by category and by word.



3.2 *Acoustic correlates of proper names* What are the acoustic correlates of proper names vs. common nouns that might influence decisions? Several acoustic differences might be predicted, based on differences in typical lexical frequency and also differences in how often proper names and common nouns receive sentential stress. Table 2 presents the mean values for five acoustic characteristics of the homophone mates used as stimuli, divided by whether they were proper names or common nouns: Proper names had longer duration, higher mean F0, larger F0 range, higher intensity, and lower spectral tilt.

Table 2 Acoustic characteristics of homophone mates based on whether they were proper names or common nouns.

	Word Duration	F0 mean	F0 range	Intensity	Spectral Tilt
Proper name	457 ms	153 Hz	73.6 Hz	56.4 dB	-4.1
Common noun	441 ms	142 Hz	64.6 Hz	55.9 dB	-3.0

Table 3 presents the summary of a mixed effects logistic regression model for "capitalized" identifications (vs. "lowercase") as predicted by acoustic characteristics of the stimulus relative to the paired item. The fixed effects were word duration ratio, F0 mean ratio, intensity ratio, and spectral tilt ratio; all were centered. There were random intercepts for participant and for word pair.

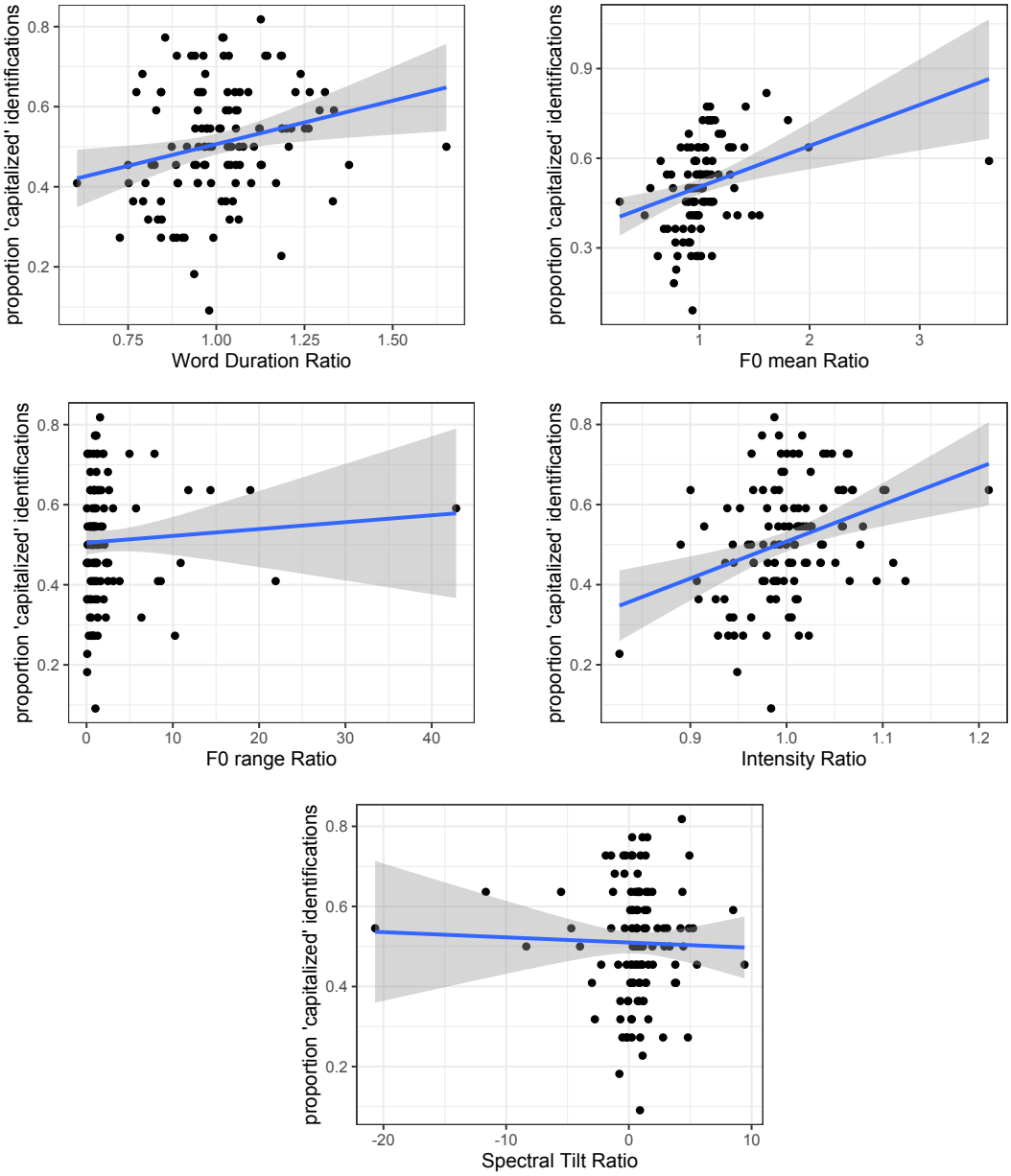
Table 3 Regression model for "capitalized" identifications.

	Estimate	SE	z-value	p-value
(Intercept)	0.071	0.096	0.73	0.46
Word Duration Ratio	1.1	0.27	4.2	< 0.0001
F0 Mean Ratio	0.48	0.14	3.5	0.00039
Intensity Ratio	4.1	0.85	4.8	< 0.0001
Spectral Tilt Ratio	0.013	0.013	1.0	0.31

Several acoustic characteristics of the stimuli were significant predictors of how a stimulus was identified. Listeners were more likely to identify a stimulus as being a proper name if it had longer duration, higher mean F0, or higher intensity. Figures 2a-e illustrate the relationship between acoustic characteristics and how listeners identified each stimulus.

A model including F0 range ratio was tested, but the strong correlation between F0 mean ratio and F0 range ratio within the stimuli ($r(56) = 0.78$, $p < 0.0001$) makes a model including both factors unreliable. In that model, a larger F0 range predicted significantly fewer "capitalized" identifications, which is the opposite of the relationship observed in production. In a model including F0 range and excluding F0 mean, there is no evidence for an effect of F0 range.

Figure 2 The proportion of "capitalized" identifications as predicted by the acoustic characteristic of each stimulus relative to its paired homophone mate: (a) Word duration ratio, (b) F0 mean ratio, (c) F0 range ratio, (d) Intensity ratio, (e) Spectral tilt ratio.



There is no evidence for response time influencing use of acoustic cues. No interactions between the acoustic predictors and log response time (as measured from the beginning of the first item of the pair) produced a model with a significantly better fit than a model without interactions.

The proper names were from five different categories: Brands (e.g., *Bobcat*), cities (e.g., *Buffalo*), human names (e.g., *Holly*), possessives (e.g., *Poppy's*), and teams/bands (e.g., *Dolphins*). Cue usage could potentially differ based on the type of proper name. Adding an interaction between construction and F0 mean ratio provides a significantly better fit than a model that includes both factors but no interaction (χ^2 = 11.4, df = 4, p = 0.022). The effect of F0 mean is strongest for city names and human names, and weakest for team names and possessives; Figure 3 illustrates. Adding an interaction between construction and word duration ratio also provides a significantly better fit than a model without the interaction (χ^2 = 9.9, df = 4, p = 0.042). The effect of word duration is strongest for city names and weakest for human names; Figure 4 illustrates. Note that the figures are based on the raw data, not the model output.

Figure 3 The proportion of "capitalized" identifications as predicted by F0 mean ratio, by type of pair.

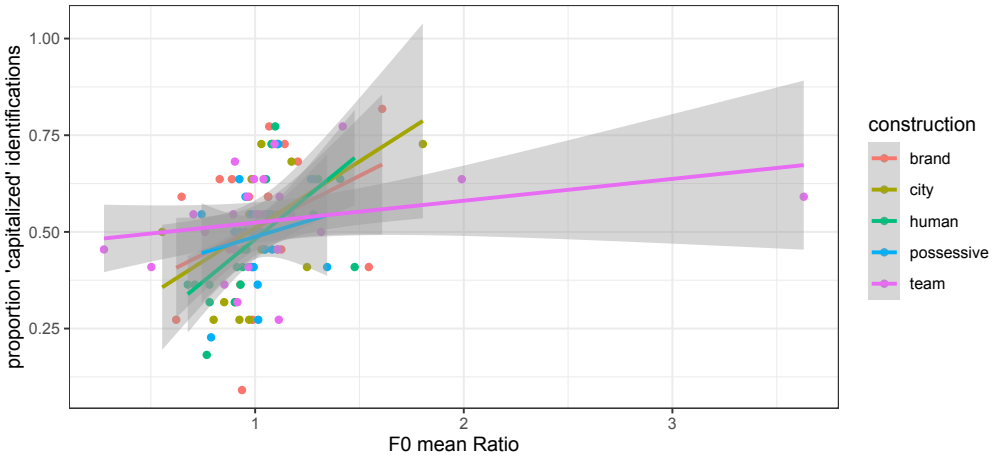
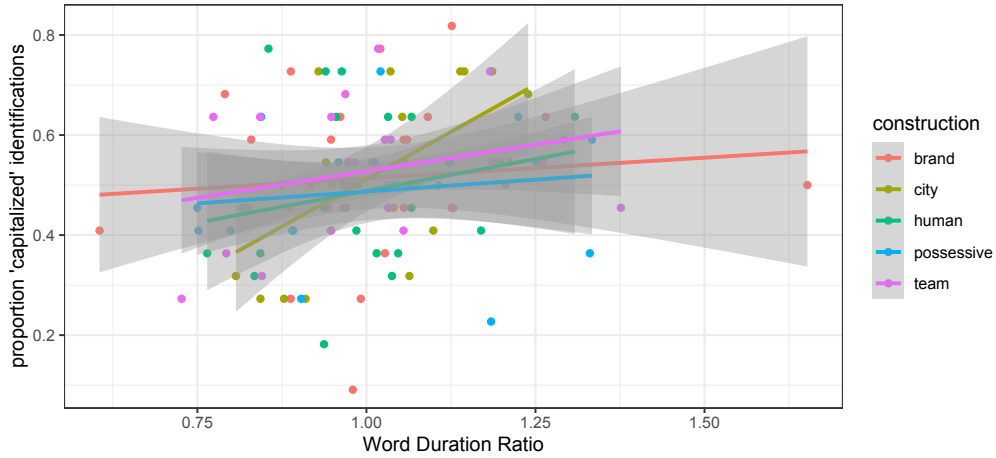


Figure 4 The proportion of "capitalized" identifications as predicted by word duration ratio, by type of pair.



3.3 *Lexical frequency* One of the potential factors driving the acoustic characteristics of proper names is lexical frequency. Previous work has demonstrated that lexical frequency is a predictor of word duration (Guion 1995, Gahl 2008), though it isn't as clear that it predicts the other acoustic characteristics that differ between proper names and common nouns.

Within the stimuli used in this experiment, log lexical frequency is a predictor of some acoustic characteristics: Word duration is negatively correlated with frequency (r(114) = -0.4, p < 0.0001), and F0

range is also negatively correlated with frequency ($r(114) = -0.18$, $p = 0.058$). The correlations between lexical frequency and F0 mean, intensity, and spectral tilt did not approach significance.

There is a confound between lexical frequency and being a proper name in this dataset; proper names have a much lower mean frequency than other words do (the mean log counts from SUBTLEX are 2.6 vs. 5.6). A similar difference is found in the lexicon more generally, as described above. Being a proper name seems to be a clearer predictor of the acoustic characteristics than lexical frequency.

Lexical frequency only predicts listeners' identifications to the extent that it is a predictor of acoustic characteristics. While lexical frequency does significantly predict identification decisions in a model with no acoustic predictors ($\beta = -0.047$ SE = 0.017, $z = -2.8$, $p = 0.0057$), adding lexical frequency to the model in Table 3 does not significantly improve the model ($\chi^2 = 1.4$, $df = 1$, $p = 0.24$).

3.4 *Sentential stress* One of the potential factors driving acoustic differences between proper names and homophonous common nouns is sentential stress. To examine this, the stimuli included four items that were capitalized vs. non-capitalized definite phrases (e.g., *the creature* vs. *The Creature*).

Table 4 presents the mean values for five acoustic characteristics of capitalized vs. non-capitalized definite phrases. While several of the effects are the same as what is observed in proper names vs. common nouns, F0 mean and F0 range are lower for the capitalized phrases than non-capitalized phrases, while they were both higher for proper names than for common nouns, suggesting a different form of emphasis.

Table 4 Acoustic characteristics of capitalized and not capitalized definite phrases.

	NP Duration	F0 mean	F0 range	Intensity	Spectral Tilt
Capitalized	509 ms	153 Hz	84.6 Hz	54.1 dB	-1.4
Not capitalized	500 ms	161 Hz	102.6 Hz	53.3 dB	1.7

Accuracy of identification of these items was higher than chance (60%). Table 5 presents the summary of a mixed effects logistic regression model for accuracy. The only fixed effect was type of pair (Definite Phrases, Bare Nouns). There were random intercepts for participant and for word pair.

Table 5 Regression model for accuracy among all homophone mates, comparing pair types. *Reference levels: Type = Definite Phrases.*

	Estimate	SE	z-value	p-value
(Intercept)	0.43	0.18	2.4	0.017
Type Bare Nouns	-0.19	0.19	-1.0	0.31

Accuracy for identifications of capitalized vs. non-capitalized definite phrases was significantly above chance. Accuracy for bare nouns (e.g., *Phoenix, phoenix*, as discussed in the previous sections) was slightly but not significantly lower than accuracy of identification of the definite phrases.

Table 6 presents the summary of a mixed effects logistic regression model for "capitalized" identifications (vs. "lowercase") for capitalized vs. non-capitalized definite phrases, as predicted by acoustic characteristics of the stimulus relative to the paired item. The fixed effects were word duration ratio, F0 mean ratio, intensity ratio, and spectral tilt ratio; all were centered. There were random intercepts for participant and for word pair.

Table 6 Regression model for "capitalized" identifications in capitalized vs. non-capitalized phrases.

	Estimate	SE	z-value	p-value
(Intercept)	0.068	0.13	0.54	0.59
Word Duration Ratio	3.1	1.0	3.0	0.0026
F0 Mean Ratio	-1.2	0.61	-2.0	0.048
Intensity Ratio	9.0	2.7	3.3	0.00092
Spectral Tilt Ratio	0.063	0.049	1.3	0.2

Listeners were more likely to identify a stimulus as being the capitalized phrase if it had longer duration, lower F0, or higher intensity.

4 Discussion

Listeners had above chance accuracy at distinguishing between proper names and homophonous nouns and at distinguishing between capitalized and non-capitalized definite phrases. Responses were strongly predicted by several acoustic cues: Word duration, mean F0, and intensity, which are also all correlates of these different categories in production. These results can be explained by listeners using prosodic cues based on expectations set by syntactic, semantic, and pragmatic characteristics.

The main factor that seems to drive the prosodic differences between proper names and common nouns is sentential stress. Proper names are probably more likely to be stressed than other words are, because they are more likely to be unpredictable in context and are often key elements of the utterances containing them. Their low lexical frequency may also contribute to how likely they are to receive stress, though the results suggest that lexical frequency is not directly driving the effects in this study. The primary differences observed between proper names and common nouns are characteristics of sentential stress: Longer duration, higher F0, greater intensity (Breen et al 2010).

The capitalized vs. non-capitalized definite phrases used in this study (e.g., *the creature* vs. *The Creature*) seem to exhibit a different type of intonational difference than proper names vs. common nouns, based on their acoustic characteristics. Adding capitalization to definite phrases does not seem to make them into proper names or create sentential stress, at least for the phrases used in this study. Using word-initial capitalization in phrases that are not proper names has been analyzed as indicating that the phrase refers to a well-established or prominent meaning (Linden 2020), which may be the function that has been captured in this study.

Listeners make use of the acoustic correlates of proper names and capitalized phrases, which is apparent both in overall above-chance accuracy and also the relationship between acoustic characteristics of a stimulus and how it was identified. Speakers and listeners use prosodic cues in a range of ways, such as indicating phrase boundaries and other syntactic contrasts (Shattuck-Hufnagel & Turk 1996, Cho et al. 2007). Thus, prosodic structure can set some expectations about syntactic, semantic, and pragmatic information; in this study, those expectations may be directly about proper names, or may be about sentential stress, which in turn is linked to the categories of words that are more likely to be stressed. Direct association of phonetic details with particular words is not necessary for listeners to make use of prosodic information. This study was set up to encourage decisions based on the broader group rather than considering each pair separately, as every trial asked listeners to identify whether the "capitalized" word (or "lowercase" word) was first or second in the pair, rather than asking listeners to identify the ordering of the particular words.

The results suggest that use of the prosodic cues associated with proper names is based around proper names as a broad category. Accuracy is above chance not just for names that are likely to be familiar but also for invented business names and for low-frequency names that are likely to be unfamiliar to many of the participants; there is no evidence that the use of acoustic cues is stronger for real names than for invented names or stronger for higher frequency names than for lower frequency names. This generalized cue usage supports the analysis that listeners' expectations about acoustic characteristics that distinguish between proper names and homophonous common nouns are set by syntactic and pragmatic factors rather than by phonetic details that are associated with the representations of particular words.

There was some evidence for differences in cue usage for different types of proper names (e.g., city names vs. team names), though the limited number of items in each category makes it somewhat unclear what might drive these differences, e.g., variation in word length, lexical frequency, or different typical prosodic environments (e.g., the fact that human names can appear in vocatives, while the other categories of proper names generally will not). In this study, the only pairs that differed morphologically were the possessive vs. plural pairs (e.g., *poppies* vs. *Poppy's*). If listeners are sensitive to differences in duration between different morphemes (cf. Plag et al. 2017), the effect of duration might be a stronger predictor for these items. However, this category did not exhibit a stronger effect of word duration than other categories; perhaps duration of the word is not sufficient to capture an effect of the duration of particular segments.

Use of prosodic cues associated with capitalized phrases is necessarily capturing a contextually-driven pattern based on factors outside of the individual word, because the capitalized and non-capitalized phrases contain exactly the same words. Expectations about prosody seem to be set by a particular pragmatic usage of word-initial capitalization. One way that this form of capitalization is used informally in written English

is to draw attention to the intended meaning being the most prominent or well-established meaning rather than something that is more limited to the particular discourse context (Linden 2020).

5 Conclusions

Proper names and homophonous common nouns exhibit several systematic phonetic differences, which seem to align with correlates of phrasal stress: Word duration, mean F0, and intensity. Listeners make use of these acoustic characteristics when identifying these items, demonstrating expectations for the acoustic characteristics of proper names vs. common nouns. The results can be explained based on connections between prosodic structure and semantic categories like proper names; similar results for real and invented names suggests that the results are not driven by phonetic details associated with individual words.

References

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1-48.

Bond, Zinny S. 1973. The perception of sub-phonemic phonetic differences. *Language and Speech* 16. 351-355.

Breen, Mara, Evelina Fedorenko, Michael Wagner & Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25. 1044-1098.

Brysbaert, Mad & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977-990.

Chiu, Ching-Yiu Peter. 2000. Specificity of auditory input and explicit memory: Is perceptual priming for environmental sounds exemplar specific? *Memory & Cognition* 28(7). 1126-1139.

Cho, Taehong, James M. McQueen & Ethan A. Cox. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35. 210-243.

Clopper, Cynthia G. and Rory Turnbull. 2018. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Francesco Cangemi, Meghan Clayards, Oliver Niebuhr, Barbara Schuppler & Margaret Zellers (eds), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, 25-72. Berlin: Mouton de Gruyter.

Cohen, Gillian. 1990. Why is it difficult to put names to faces? *British Journal of Psychology* 81. 287-297.

Connine, Cynthia M., Larissa J. Ranbom & David J. Patterson. 2008. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics* 70(3). 403-411.

Conwell, Erin. 2015. Neural responses to category ambiguous words. *Neuropsychologia* 69. 85-92.

Conwell, Erin. 2017. Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Child Language* 44(3). 734-751.

Conwell, Erin & James L. Morgan. 2012. Is it a noun or is it a verb? Resolving the ambicategoricality problem. *Language Learning and Development* 8. 87-112.

Conwell, Erin & Kellam Barta. 2018. Phrase position, but not lexical status, affects the prosody of noun/verb homophones. *Frontiers in Psychology* 9. Article 1785.

Desai, Rutvik H., Usha Tadimeti & Nocholas Riccardi. 2023. Proper and common names in the semantic system. *Brain Structure and Function* 228. 239-254.

Fowler, Carol & Jonathan Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26. 489-504.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 83(4). 474-498.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66. 789-806.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251-279.

Guion, Susan G. 1995. Word frequency effects among homonyms. In *Texas linguistic forum,* vol. 35, 103-116.

Hintzman, Douglas, Richard A. Block, & Norman R. Inskeep.1972. Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior* 11(6). 741-749.

Jurafsky, Daniel, Alan Bell & Cynthia Girand 2002. The role of the lemma in form variation. In Carlos Gussenhoven &

Natasha Warner (eds.), *Laboratory phonology VII*, 3-34. Berlin: Mouton de Gruyter.
Kahn, Jason M. & Jennifer E. Arnold. 2012. A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language* 67. 311-325.
Keung, Lap-Ching. 2013. *Effects of discourse status and planning difficulty on acoustic variation*. Chapel Hill, NC: University of North Carolina at Chapel Hill Bachelor of Science thesis.
Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2015. *lmerTest: Tests in linear mixed effects models.* https://CRAN.R-project.ord/package=lmerTest. R package version 2.0-29.
Linden, Josh. 2020. Contrastive Focus Capitalization: Nonstandard usages of capital letters in web-based English and their capital-I Implications. *Studies in the Linguistic Sciences: Illinois Working Papers 2020*. 116-138.
Lohmann, Arne. 2018. *Cut* (N) and *cut* (V) are not homophones: Lemma frequency affects the duration of noun-verb conversion pairs. *Journal of Linguistics* 54(4). 753-777.
Lohmann, Arne & Erin Conwell. 2020. Phonetic effects of grammatical category: How category-specific prosodic phrasing and lexical frequency impact the duration of nouns and verbs. *Journal of Phonetics* 78. Article 100939.
Longobardi, G. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry*, 25(4), 609-665.
Martinuzzi, C. & Schertz, J. 2022. Sorry, not sorry: The independent role of multiple phonetic cues in signaling the difference between two word meanings. *Language and Speech*, 65(1), 143-172.
Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39. 132-142.
Nygaard, Lynne C., Debora S. Herold & Laura L. Namy. 2009. The semantics of prosody: Acoustic and perceptual evidence of the prosodic correlates to word meaning. *Cognitive Science* 33. 127-146.
Nygaard, Lynne C. & Erin R. Lunders. 2002. Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition* 30(4). 583-593.
Pan, Shimei & Julia Hirschberg. 2000. Modeling local context for pitch accent prediction. In *Proceedings of the 38th annual meeting of the Association for Computational Linguistics*, 233-240.
Peirce, Jonathan W. 2007. PsychoPy–Psychophysics software in Python. *Journal of Neuroscience Methods* 162(1-2). 8-13.
Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (Eds.), *Laboratory Phonology VII*, 101-139. Berlin: Mouton de Gruyter.
Pitt, Mark A., Laura Dilley & Michael Tat. 2022. Exploring the role of exposure frequency in recognizing pronunciation variations. *Journal of Phonetics* 39(3). 304-311.
Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53(1). 181-216.
Pufahl, April & Arthur G. Samuel. 2014. How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology* 70. 1-30.
Sanker, Chelsea. 2022. Homophone discrimination based on prior exposure. *Journal of Phonetics* 95. Article 101182.
Semenza, Carlo. 2006. Retrieval pathways for common and proper names. *Cortex* 42. 884-891.
Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140-155.
Seyfarth, Scott, Marc Garellek, Gwendolyn Gillingham, Farrell Ackerman & Robert Malouf. 2018. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33(1). 32-49.
Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosodic tutorial for investigators of auditory sentence processing. *Journal of Psycholoinguistics Research* 25(2). 193-247.
Sorensen, John M., William E. Cooper & Jeanne M. Paccia. 1978. Speech timing of grammatical categories. *Cognition* 6(2). 135-153.
Sóskuthy, Márton & Jennifer Hay. 2017. Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166. 298-313.
Tang, Kevin & Jason A. Shaw. 2021. Prosody leaks into the memories of words. *Cognition* 210. Article 104601.
Whalen, Douglas H. & Heidi E. Wenk. 1994. Durational characteristics of proper names common words. *Journal of the Acoustical Society of America* 95. 2924.
Yasuda, Kiyoshi, Tetsuo Nakamura & Bobbie Beckman. 2000. Brain processing of proper names. *Aphasiology*,14(11). 1067-1089.