Title
Voices and Variants: Effects of Voice on the Form-Based Processing of Words with Different Phonological Variants

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
King, Sharese
Sumner, Meghan

Publication Date
2014

Peer reviewed

Voices and Variants: Effects of Voice on the Form-Based Processing of Words with Different Phonological Variants

Sharese King (sharese@stanford.edu)
Department of Linguistics, Margaret Jacks Hall, Bldg. 460
Stanford, CA 94301-2150 USA

Meghan Sumner (sumner@stanford.edu)
Department of Linguistics, Margaret Jacks Hall, Bldg. 460
Stanford, CA 94301-2150 USA

Abstract

Spoken words have robust acoustic variation. How listeners understand spoken words despite this variation remains an issue central to theories of speech perception. Current models predict listener behavior based on the frequency of a variant in production. A phonological variant, though, is often investigated independent of phonetic variation that provides listeners with information about talkers. In this study, we investigate whether standard variants in words produced by a talker with a standard voice are recognized more quickly than standard variants in words produced by a talker with a non-standard voice. Conversely, we investigate whether non-standard variants in words produced by a talker with a standard voice are recognized more slowly than standard variants in words produced by a talker with a non-standard voice. These comparisons enable us to assess limitations of current theory, illuminating the understudied influence of talker voice in the understanding of spoken words with different phonological variants.

Keywords: spoken-word recognition; speech perception; variation; dialect; African American Vernacular English

Introduction

Speech varies across speakers based on a variety of social and linguistic factors. While variation was, viewed as problematic noise (e.g., Verbrugge, Strange, Shankweiler, & Edman 1976), researchers have turned to investigating the potential contribution phonetic variation has in the quick and adept ability of listeners to understand spoken words. For example, listeners are highly sensitive to variation in speech (Bradlow & Bent, 2008; Bradlow & Pisoni, 1999; Clopper & Pisoni, 2004; Johnson, 2006; Sumner & Samuel, 2009), use this information to process upcoming words (e.g., Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Salverda, Kleinschmidt, & Tanenhaus, 2014), store detailed talker-based acoustic detail in memory (Goldinger, 1998; Nygaard, Sommers, & Pisoni, 1994), and depend on acoustic patterns in speech to activate acoustically-similar representations (Johnson, 2006).

Many contemporary theories oriented toward accommodating phonetic variation in speech perception are episodic in nature. Such theories posit that a listener's ability to access a lexical item is contingent upon the encoding of detailed episodes of spoken words (Goldinger, 1998; Johnson, 2006). Incoming speech is perceived against the clusters arising from the storage of phonetically-rich lexical representations. This leads to an activation benefit of more frequently experienced acoustic patterns, as a common structure benefits from the shared activation of a rich, dense cluster of stored word forms, making up the form component of a form-meaning lexical representation. In the most simplistic and extreme interpretation of such a theory, listeners understand and recognize frequent word forms faster than and/or more accurately than less frequent word forms.

The bulk of studies that have supported this view have investigated talker-specific variation and its effect on the recall and recognition of spoken words. For example, Johnson (2006) found that words produced by women with more typical female voices are recognized more quickly than words produced by women with less typical female voices. Nygaard and colleagues (Nygaard et al., 1994; Nygaard and Pisoni, 1998) have shown that words are recognized better and recalled more accurately upon second presentation when the first presentation matched in talker voice, and speech rate.

While these studies have provided evidence for specificity in form at the lexical level and in a benefit for more typical or frequent forms, words forms vary phonologically, too. For example, speakers of General American English (GA) may produce the word *center* with a medial [nt] sequence, or with a medial [n_] sequence, stemming from a post-nasal t-deletion process common in GA, and across regional and ethnic varieties of American English more broadly. Recent work has investigated the composition of lexical form-based representations of words with different pronunciation variants, as well. Typically, these studies compare the effects of words produced with one variant to those of words produced with a different variant. In other words, the comparisons are typically purely phonological and categorical. The similar thread tying this work to those with episodic-based approaches to variation has been the link to frequency. From a representational standpoint, researchers have wondered whether one variant is dominant compared to another. Additionally, they wonder if evidence exists as to whether representations are tied to production frequency or tied to a canonical, or idealized, form of a word.

These studies have produced mixed results. Some work has found that listeners are more likely to recognize a word when it contains the variant that is most often produced (e.g., *beetle* with a medial tap, which occurs 96% of the time in American English, is recognized as a word ***more often*** than the same word with a medial [t]; Connine, 2004). Others have found the opposite; listeners are more likely to recognize a word when it contains the variant that is uncommon, but socially idealized (e.g., *center* with a medial [n_], which occurs almost categorically in GA, is recognized as a word ***less often*** than the same word with medial [nt]**;** Pitt, 2009).

Recently, Sumner (2013) has argued that these mixed results stem from using an approach that is highly sensitive to acoustic patterns in speech, without necessarily considering how those acoustic patterns interact with the phonological variants, leading to less than optimal comparisons. For example, while it is true that the [n_] variant in a word like *center* is overwhelmingly the frequent variant, that variant also occurs overwhelmingly in a casually-articulated, phonetically-reduced word frame. Using the semantic priming paradigm to investigate the dependence of a phonological variant on the phonetic word frame, Sumner found that words with both variants are equally able to facilitate recognition to a semantically related target as long as each variant is produced in its typically occurring phonetic word frame (see also Gow, 2001; McLennan, Luce, & Charles-Luce, 2003; Sumner & Samuel, 2005).

Having now established that the recognition of words with phonological variants depends greatly on the phonetic word frame, we might now consider how these two interact in more nuanced uses of variants across regional and ethnic varieties of American English. While understudied, some work has been conducted. For example, Sumner and Kataoka (2013) investigated the semantic priming of targets preceded by primes ending in either rhotic (-er) or non-rhotic (-uh) vowels (e.g., slend-er/-uh – THIN) across voices with different accents (GA-er; New York City (NYC)-uh; Southern Standard British English (BE)-uh). They found that the presence or absence of priming cannot be tied to a particular variant. Specifically, for a population of GA listeners, strong semantic priming was found for the GA/-er pairing *and* for the BE/-uh pairing. No priming was found for the NYC/-uh pairing. This asymmetry indicates that the voice carrier of a word impacts spoken word recognition greatly. Recent work by Sumner and colleagues (Sumner, Kim, King, & McGowan, 2014) suggests that form-based representations and access to those representations depend on both the linguistic and social information conveyed through acoustic patterns in speech. And, that listeners integrate both in speech recognition in order to build representations and understand spoken words.

This study builds on this work and examines the effects of voices and variant on the form-based processing of spoken words. Specifically, we shift to the underexplored area of ethnic variation. Ethnic variation provides us with an opportunity to explore the effects of standard and nonstandard voices and variants, increasing our understanding of how phonological and phonetic variation interact. This study investigates whether standard variants (e.g., [nd] in *friendly*) are recognized more quickly when produced in a standard voice (e.g., GA) or in a nonstandard voice (e.g., African American Vernacular English (AAVE)). Additionally, we investigate whether non-standard variants (e.g., [n_] in *friendly*) are recognized more quickly when produced in a nonstandard voice or in a standard voice.

To do this, we investigate two types of variation: Dialect-independent variation and dialect dependent variation. Dialect-independent variation refers to a production pattern across American English that spans across regions and ethnicities and cannot be tied to any one particular speaker population. Dialect-dependent variation indicates a production pattern that is well-documented to be highly common in and specific to one particular speaking population. Consonant cluster deletion (CCD) is an example of dialect-independent variation, and TH-fronting is an example of dialect-dependent variation. We describe each of these processes in the following section. By comparing variants that occur generally across voices *and* variants that are tied to particular voices, we can investigate both variant-specific frequency-based predictions and voice-dependent context-based predictions stemming from past work, providing insight into the complex process of spoken language understanding.

Dialect-Independent vs. Dialect-Dependent Variation

Certain types of variation can occur commonly across dialects or specifically within dialects. Particularly, CCD, the reduction of a consonant cluster to a single sound (CC → C_), is a dialect-independent pattern that occurs more generally across all dialects of English including GA and AAVE. TH-fronting, the production of a syllable-final interdental fricative as a labiodental fricative (e.g., [θ] → [f]; *booth* → *boof*) is a dialect-dependent process, occurring more restrictively in dialects like AAVE (Thomas, 2007), but not in GA. These types of variation patterns make it possible for us to explore ethnic dialectal variation and the effects of this variation on speech perception when produced in the context of different voices.

Production Patterns

The process of CCD results in the production of a reduced cluster, as seen in the pronunciation of *friendly* as *frienly*, having the variants [nd] and [n_], respectively. For ease of explication, the symbols [nd] will be used in this paper to represent an unreduced cluster and [n_] will be used to represent the reduced variant. Speakers across dialects tend

to delete the final stop of a cluster when that stop is followed by a consonant (Thomas, 2007). Although CCD occurs more generally across dialects, this pattern is still attributed to AAVE. This may result from the fact that an AAVE speaker may also reduce the cluster when the following sound is a vowel.

The process of TH-fronting results in the production of the dental fricative as a labiodental fricative as seen in the pronunciation of *booth* as *boof* with the variants [θ] and [f], respectively. We refer to the standard variant, for simplicity, as [θ] and the nonstandard variant as [f], though both voiced and voiceless pairs were included in the study.

The acoustic similarity of these fricatives may make this distinction in consonants less perceptible. At the fricative-vowel boundary, Jongman, Wayland and Wong (2000) observed a significantly higher F2 onset in dental fricatives than labiodental fricatives. It is possible that this could be a cue to distinguishing these consonants. Previous research has shown that in careful versus casual speech, the acoustic distances in minimally different places of articulation are enhanced in clear speech (Maniwa, Jongman, & Wade, 2009). Additionally, Maniwa et al. (2009) show that differences between /f/ and /θ/ are more discriminable in clear speech compared to conversational speech. Listeners are therefore likely to perceive this contrast when produced in isolated words in a carefully articulated speech style.

Predictions

This study investigates whether GA listeners recognize words with standard, infrequent, and dialect-independent variants produced by a GA voice faster than when produced by an AAVE voice. And, whether GA listeners recognize words with nonstandard, dialect-dependent variants produced by an AAVE voice faster than when produced by a GA voice.

To do this, we employed a cross modal form-priming paradigm. Participants are presented with an auditory stimulus followed by a visual target. Previous research suggests that the phonological similarity between the prime and target influences the rate at which a participant responds, with slower reaction times for unrelated primes than identical ones (Radeau, Morais, & Segui, 1995; Sumner & Samuel 2009). Additionally, primes that mismatched by a single feature, word finally, [flut] versus [flus], showed slower response times if that particular variant ([flus]) was not part of the participant's dialect (Sumner & Samuel 2005; 2009).

The predictions made here are nuanced and depend greatly on perspective and theoretical framing. From a purely variant frequency-based approach (e.g., the more often I hear a variant, independent of the phonetic context, the more easily I recognize a word with that variant), we predict the that: (1) For CCD, the more frequent variant [n_] should show greater priming effects in comparison to the less frequent variant [nd] for both the GA and AAVE primes, as this pattern is common across both varieties, and (2) For TH-fronting, we expect an asymmetry to emerge as [f] is infrequent in GA and more frequent in AAVE. Words with [f] should induce greater priming of form-related targets when that variant is uttered by an AAVE speaker than when it is uttered a GA speaker. But, words with [θ] should induce greater priming of form-related targets when uttered by a GA speaker than an AAVE speaker. In a theory in which the recognition of words with different variants depends on the phonetic context in which those variants are experienced, we would expect to find evidence of an interaction between voices and variants. Specifically, the processing of words with these different variants depends on the voice that houses each variant. In a perspective in which the voice conveys social meaning, and we infer social or talker-based properties from a voice, we may expect the emergence of social differences where the standard variants are recognized better when housed in a GA frame than in a AAVE frame and nonstandard variants are recognized better when housed in a AAVE frame than in a GA frame.

The Experiment

Methods

Participants Fifty-two participants participated in this study for pay. Participants included local residents and students in Palo Alto, CA. All participants were monolingual American English speakers, and none were AAVE speakers.

Stimuli The two authors of this paper recorded stimuli. Both are from Rochester, NY and share traits of the Inland North dialect. One is an AAVE speaker and the other, a GA speaker. All stimuli were recorded at a comfortable speaking rate in a sound-attenuated booth.

For the CCD stimuli, we included words with final consonant clusters (e.g., friend) followed by suffix or another word, such as *kindly (kind +ly)* or *handbag (hand +bag)* We chose this word structure because in production, deletion most naturally occurs between a consonant cluster and a following consonant.

For TH-fronting stimuli, they consisted of words with interdental fricatives in coda position (e.g., *booth*, *athlete*). Both words with the [θ]-[f] alternation and words with the [ð]-[v] alternation were included. Words where the [θ] to [f] change results in ambiguity (i.e. *Ruth* to *Ruf/roof*) were not included.

Four types of stimuli were produced by each talker: (1) CCD words with the standard variant [nd]; (2) CCD words with the nonstandard variant [n_]; (3) TH words with the standard variant [θ]; and (4) TH words with the nonstandard [f].

Design We collected 44 mono or bisyllabic words per type of variation (CCD and TH-fronting) for a total of 88 critical target words. For each variation, half of the targets (=22)

were paired with a related audio prime. A related prime is a prime that is identical to the target, or a prime that is mismatched by only a single sound (*birth; birf*). The other half of the targets were paired with unrelated audio primes like *car*. The unrelated prime and target pairs were created by pairing each target with a different prime from the list. In addition to the 88 target words, there were 264 fillers for a total of 352 words in a single list. The critical items represented 25% of the trials in the experiment. Half of the filler targets (=132) were real words with no interdental fricatives. The other half (=132) were pseudo words. There were 176 real word fillers and 88 pseudo word fillers. The design was between subject with half of the subjects receiving words produced by the AAVE speaker with nonstandard variants ([n_] and [f]) and the GA speaker with the standard variants ([nd] and [θ]), and the other half receiving words produced by the AAVE speaker with the standard variants and the GA speaker with the nonstandard variants. Figure 1 displays example stimuli across conditions. We varied relatedness (unrelated vs. related), voice (AAVE vs. GA), and variant (CCD vs. TH-fronting). No subject heard both variant types (dialect-independent vs. dialect-dependent) in the same voice.

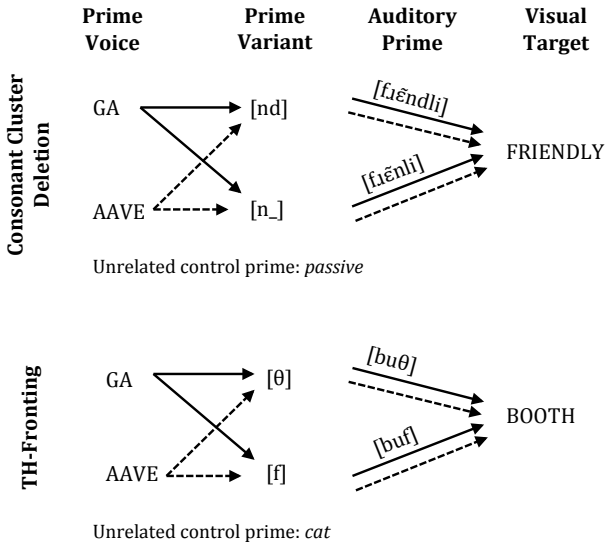

Figure 1: Sample stimuli across conditions

Procedure Participants were presented with an audio prime-visual target pair where the primes are related or unrelated to the target. We are comparing the time it takes a listener to respond correctly to a visual target (e.g., FRIENDLY) when preceded by a related word (e.g., friendly GA/[nd]; GA/[n_]; AAVE/[nd]; AAVE/[n_]) compared to when that word is preceded by a prime unrelated in form.

Listeners were tested in a sound-attenuated booth. They were asked to perform a lexical decision to the visual target as quickly and accurately as possible. Participant accuracy and response latencies were recorded.

Results Analyses focused on response times (RT) to correctly identified targets. RTs more than 4 standard deviations from the mean were removed (<6%). Due to oversight in the development of the stimuli, responses to six words from the CCD list were discarded, as these words contained clusters followed by a vowel, eliminating the appropriate context licensing deletion. Mean reaction times across voices and variants for related and unrelated conditions are provided in Table 1.

Table 1: Mean latencies for correct responses across speaker

	Voice:	AAVE	GA
Variant	**Prime Type**		
[nd]	Related	518	529
	Unrelated	582	623
[n_]	Related	535	535
	Unrelated	609	582
[θ]	Related	493	500
	Unrelated	559	589
[f]	Related	527	515
	Unrelated	589	556
Mean		552	554

We conducted a three factor Omnibus ANOVA (voice (GA/AAVE) x variant ([θ], [f], [nd], [n_]) x relatedness (RELATED/UNRELATED) on the log RTs to investigate the effects of voice, variant and relatedness on the recognition of form-related targets. No main effect of voice was found, suggesting that listeners were equally fast at identifying a visual target whether followed by a GA voice or an AAVE voice ($F(1, 39) < 1$; $p = 0.712$). Establishing an overall priming effect, a main effect of relatedness was found, where targets preceded by form-related primes were recognized faster than targets preceded by unrelated control primes ($F(1, 39) = 215.959$, $p < .001$). We also found a main effect of variant ($F(3, 39) = 8.745$, $p < .001$). Critically, a voice by variant interaction was found, suggesting that the processing of words with different variants depends greatly on the voice in which that variant was produced ($F(3, 77) = 4.884$ $p < .05$). We also found a voice by relatedness interaction ($F3, 77) = 2.626$, $p < .01$), suggesting that priming for related words differed by voice. A marginal interaction of variant by relatedness was also found ($F3, 77) = 2.262$, $p = 0.0792$).

Given the voice by variant interaction, we conducted a set of planned comparisons on the difference scores from related-unrelated pairs in order to assess the dependencies between voices and variants in the priming paradigm. Figure 2 plots the differences between the unrelated and related means across voices and variants by variation type (CCD vs. TH). As shown, the priming patterns for CCD and those for TH-fronting are strikingly similar despite the

facts that CCD is a dialect-independent process and TH-fronting is a dialect-dependent process.

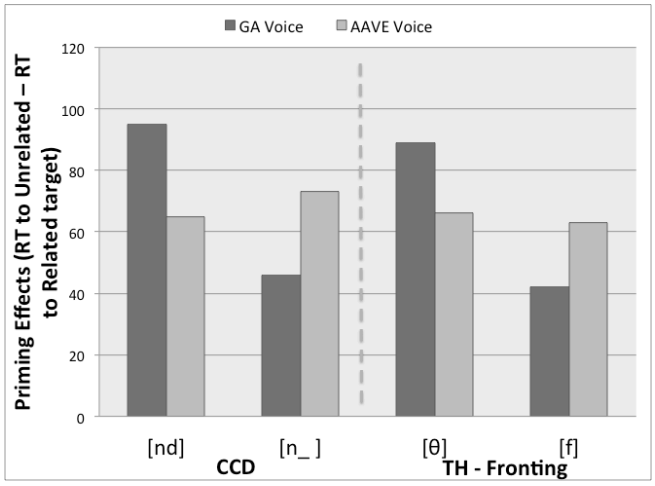

Figure 2: Differences between unrelated and related mean response times across voices and variants.

To calculate difference scores, each participant's unrelated mean was used as a baseline RT and the reaction time for each related observation was subtracted from this baseline score. The standard variant [nd] induced greater priming when produced with the GA voice than when produced with the AAVE voice ($t(331.826) = 1.948$, $p = .052$), though marginal. The standard variant [θ] induced greater priming when produced with the GA voice than when produced by the AAVE voice ($t(398.963) = 2.443$, $p < .05$). These two analyses support the idea that words with standard variants facilitate recognition to form-related targets when the standard variants co-occur with a subjectively perceived standard voice. Highlighting the important role of a phonetic frame, this offers at least speculative support that social information inferred by a talker voice influences word recognition.

Moving to our second hypothesis, that words with nonstandard variants are recognized more quickly when produced in a nonstandard voice than a standard voice, we have some support. Specifically, the non-standard variant [n_] induced greater priming when produced with the AAVE voice than when produced with the GA voice ($t(318.37) = -2.959$, $p < .01$). And, the non-standard variant [f] induced marginally greater priming when produced with the AAVE voice than when produced with the GA voice ($t(393.929) = -1.784$, $p = .07$).

Finally, to address the predictions of an account based purely on variant frequency, it is important to note that in addition to GA [nd] facilitating recognition to related targets more than AAVE [nd], it also facilitates recognition to related targets more than the most common variant ([n_]). In the GA voice, the less common variant ([nd]) showed greater priming than the GA ([n_]) ($t(331.949) = -3.595$, $p < .001$). Interestingly, though, the GA variant ([nd]) did not show greater priming than the frequent [n_] variant when housed in an AAVE voice frame ($t(334.75) = -.501$ $p > .05$).

Discussion

The purpose of this study was to examine the interaction of voices and variants across dialects in immediate processing tasks. Specifically, we investigated whether GA listeners recognize words with standard, infrequent, and dialect-independent variants produced by a GA voice faster than when produced by an AAVE voice. And, whether GA listeners recognize words with nonstandard, dialect-dependent variants produced by an AAVE voice faster than when produced by a GA voice.

Support for both hypotheses ensued. Specifically, we found across variants that words with standard variants facilitate recognition to form-related targets *more* when produced in a GA voice frame than when produced in an AAVE voice frame. And, we found that the nonstandard variant [n_] facilitates recognition to form-related targets *more* when produced in an AAVE voice than when produced in a GA voice, and the same, though marginal for the [f] variant. What is interesting about the patterning of [n_] is that this is a dialect-independent variant. In other words, GA listeners are regularly (and mostly) exposed to words that are produced with the [n_] variant. And, it is not the case that words with [nd] simply facilitate recognition to form-related words better than words with [n_] because statistically there is no priming difference between GA/[nd] and AAVE/[n_]. Rather, the results are highly nuanced and depend less on frequency of a particular variant, or even frequency of a particular variant housed in a particular voice. If the latter were supported, we should have found that the strongest priming was induced by [n_] across voices. But, that is clearly not the case.

From these data, we have clear evidence that the phonetic frame that carries a particular variant has an effect on the priming of a form-related target. But, it is also clear that neither a frequency-based nor a canonical-variant approach alone is sufficient to account for the data. Speculatively, what appears to be happening is that listeners are attributing non-standard variants to stigmatized speech varieties, despite the fact that one of the non-standard variants in the study is used prevalently by GA speakers. We found reduced priming in the AAVE voice for the standard variants and increased priming in the AAVE voice for the nonstandard variants, despite investigating variants that differ in terms of dialect-dependency in production.

One limitation of the present study is that we do not operationalize experience as in previous work on regional variation (Sumner & Samuel 2009). It is difficult to do so for an ethnic variety given the various sources of exposure from which one can draw. To date, few studies have focused on assessing experience with an ethnic dialect (Staum Casasanto, 2008) or foreign-accented English (McGowan,

2011). Future work will seek a method for doing so in order to test listeners from a population of AAVE speakers.

To summarize, we have investigated the interaction of voices and variants. Recognizing words with different phonological variants depends greatly on the perceived talker properties conveyed through a voice. While more research is needed to understand the role of experience in processing, at a minimum, variant frequency-based accounts of speech perception do not support the divergence of behavior from production patterns observed in this study. The interplay of voices and variants might be better explained by exploring the linguistic and social information conveyed by acoustic variation in the phonetic frame of spoken words.

Acknowledgments

The authors would like to acknowledge Kevin McGowan, Seung Kyung Kim, Ed King, John R. Rickford, and Paul Kiparsky for their feedback on this paper. This material is based upon work supported by the National Science Foundation Grant No. 0720054 to Meghan Sumner and Grant No. DGE-114747 to Sharese King. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, 133(4), 2350-2366.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, 106, 2074.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, 47(3), 207-238.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, 11(6), 1084-1089.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, 105(2), 251.

Gow, D. W. Jr., (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133–159.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485-499.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108, 1252.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America,* 125, 3962.

McGowan, K. (2011). *The Role of Socioindexical Expectation in Speech Perception*. Dissertation. University of Michigan Department of Linguistics.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4), 539.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42-46.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355-376.

Pitt, M.A. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance,* 35, 896-910.

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6), 1297.

Salverda, A.P., Kleinschmidt, D., & Tanenhaus, M.K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition, *Journal of Memory and Language*, 71(1), 145-163.

Staum Casasanto, L. (2008). *Experimental Investigations of Sociolinguistic Knowledge*. Dissertation, Stanford University Department of Linguistics.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final/t. *Journal of Memory and Language*, 52(3), 322-338.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4), 487-501.

Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America*, 134(6), EL485-EL491.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1015.

Thomas, E. R. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass*, 1(5), 450-475.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space?. *The Journal of the Acoustical Society of America*, 60, 198.