Phonetic variation and the recognition of words with pronunciation variants

Meghan Sumner (sumner@stanford.edu), Chigusa Kurumada (kurumada@stanford.edu), Roey J. Gafter (gafter@stanford.edu), Marisa Casillas (middyp@stanford.edu)
Department of Linguistics, Margaret Jacks Hall, Bldg. 460
Stanford, CA 94305-2150 USA

Abstract

Studies on the effects of pronunciation variants on spoken word recognition have seemingly contradictory results – some find support for a lexical representation that contains a frequent variant, others, an infrequent (but idealized) variant. We argue that this paradox is resolved by appealing to the phonetics of the overall word. In two phoneme categorization studies, we examined the categorization of the initial sounds of words that contain either tap or [t]. Listeners identified the initial sound of items along a voiced-voiceless continuum (e,g, bottom–pottom, produced with word-medial [t] or tap). No preference for word-forming responses for either variant was found. But, a bias toward voiced responses for words with [t] was found. We suggest this reflects a categorization bias dependent on speaking style, and claim that the difference in responses to words with different variants is best attributed to the phonetic composition of the word, not to a particular pronunciation variant.

Keywords: phonetic variation, pronunciation variation, speech perception, phoneme categorization, lexical representation

Introduction

As listeners, we face a speech signal that is riddled with variation, with countless acoustic realizations of any given word. Words stream by listeners at a rate of about 5–7 syllables per second, further complicating the listener's task. How listeners understand spoken words despite this variation is an issue central to linguistic theory.

The finding that lexical representations are rich with phonetic detail along with associated theories of representation and lexical access have greatly advanced our understanding of this process (e.g., Goldinger, 1998; Johnson, 2006). Incorporating variation into theory was a major step toward a full explanation of spoken language understanding.[1] But, claims made by lexical-representation-based accounts are becoming increasingly difficult to validate or falsify.

Studies that examine the effects of pronunciation variants on spoken word recognition highlight this point. Two different realizations of a sound are considered pronunciation variants. For example, one can produce the word *baiting* with a [t], sounding like *bay-ting* or with a tap [ɾ], sounding more like *bay-ding*. Or, one can produce the word *center* with a [t], sounding like *sen-ter*, or without [n_] (though some acoustic residue is likely to remain), sounding like *sen-ner*. Studies that examine the recognition of words with pronunciation variants typically compare a frequent (commonly produced) variant (e.g., [ɾ] or [n_]) to an canonical, but infrequent variant (e.g., [t] or [nt]). Interestingly, in this area of research, two conceptually-identical studies have found evidence for lexical representations that are specified for a particular pronunciation variant. In one case, though, the data suggest that the frequent variant is stored (Connine, 2004). In the other case, the data suggest that the canonical variant is stored (Pitt, 2009). We call this the *representation paradox*. Specifically, these studies found:

(1) **Frequency bias**: A **cost** for words produced **with [t],** like *baiting* produced like *bay-ting,* (Connine, 2004) compared to those produced with the more common tap ([ɾ]) variant, and

(2) **Canonical bias**: A **benefit** for words **with [t]**, like *center* produced sounding like *sen-ter* (Pitt, 2009) compared those produced with the more common post-nasal deletion variant ([n_]) (sounding like *sen-ner*).

In this paper, we suggest that this paradox has resulted for two reasons. First, pronunciation variants are typically examined independent of the phonetic composition of the entire word (see also Andruski et al., 1994). While it is true that we may produce [t] or [ɾ] in a word like *baiting*, it is also true that each variant co-varies with a different set of acoustic correlates across the word. Second, in the examples in (1) and (2), it is not clear that listener responses are driven by stored lexical forms in this task, and not by these co-present acoustic cues.

It is undoubtedly the case that detailed representations exist. But, it is also the case that (1) listeners are highly sensitive to acoustic fluctuations in speech (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Green, Tomiak, & Kuhl, 1997; McMurray & Aslin, 2005; McMurray, Tanenhaus, & Aslin, 2009), (2) low-level acoustic mismatches result in major perceptual costs either from manipulations resulting in incongruent cues (Gaskell & Marslen-Wilson, 1996) or from intentionally mispronounced sounds (Gow, 2001, 2003; Sumner & Samuel, 2005), and (3) acoustic cues inform a listener not only about linguistic units, but provide expectations about the style of a speech event (Labov, 1966; among many others**)**

In this paper, we ground ourselves broadly in a phonetic perspective and make two suggestions. First, we suggest that *different pronunciation variants are processed equally*

[1] This is a move often discussed, but still largely absent from theories of spoken word recognition; see McLennan & Luce (2005) for related discussion.

well when presented in a congruent phonetic word frame. Note that this is not inconsistent with an exemplar account, but suggests only that these representations are not at play here, and the canonical bias in (2) results from an artificial bias toward the canonical variant. Second, we suggest that once we accept that all variants are processed equally well by listeners, we need to reconsider the exaplanatory burden placed on exemplars for theories accommodating variation during spoken word recognition.

Categorization and pronunciation variants

As mentioned, current studies diverge on how listeners respond to words with different pronunciation variants in spoken word recognition. The *frequent* variant is the one uttered by speakers with the highest frequency, and is often regarded as a reduced form, (i.e., [ɾ] and [n_]; see Patterson & Connine, 2001). The *canonical* variant is less common in casual conversation, but is more likely to be produced in careful speech, and may be more faithful to orthography (e.g., [t] and [nt]).

Through a series of phoneme categorization studies, Connine (2004) examined the perception of the initial sounds of words that contain either tap or [t]. Creating voiced-voiceless continua for words like *baiting* (*baiting–paiting*, produced either with word-medial [t] or tap), listeners were asked to identify the initial sound of items along the continuum. Listeners made more word-forming responses (in this case, "B" responses) to items with the frequent tap than to items with canonical [t]. She argued that in words like *baiting*, the tap is stored in the lexical form. Consistent with exemplar accounts of lexical access (Goldinger, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002), the **cost associated with [t]** is argued to result from access to a frequency-based lexical representation (see also LoCasto & Connine, 2002).

Through his own series of phoneme categorization studies, Pitt examined post-nasal [t]-deletion in words like *center*. Comparing responses to items along a *center–shenter* continuum ([nt]) to those along a *cenner–shenner* ([n_]) continuum, he found that listeners made more word-forming responses (in this case, "S" responses) to items with the canonical [nt] than to items with the frequent [n_]. In this case, the **benefit associated with [t]** is argued to result from access to a canonical representation.

Maybe differences encoding words or word forms exist, and there is no paradox at all. One would need to argue that *baiting* and *center* are treated differently, and that experience with one yields a surface-based representation and experience with another yields a canonical representation. In this case, tapping and post-nasal [t] deletion are different processes which affect representations differently (e.g. the former could be viewed as an altered form, and the latter a phonological deletion (though a nasal tap may be a residual phonetic cue to the t-deletion process)). Here we offer another alternative explanation: The apparent paradoxical dissolves when we consider the phonetic composition of the word frame that houses a particular pronunciation variant.

The phonetic perspective

Consider again the example of *baiting.* The use of the [t]-variant is constrained by speech style, occurring (though rarely) in extra *careful* speech (Shockey, 2003). When [t] is used, the entire word (word-level phonetic variation) is hyperarticulated, so that [t] co-occurs with other predictable acoustic values (longer stop closure, longer duration of the previous vowels, de Jong, 1998, p. 293). In contrast, the tap, when produced in casual speech, co-occurs with cues common to *casual* speech (short, centralized vowels, shorter overall duration, reduced amplitude).

Interestingly, the usage patterns of each pronunciation variant pair ([t]–[ɾ]; [nt]–[n_]) differ greatly. The production of tap is nearly categorical in American English (AE), produced nearly 97% of the time in running speech (Patterson & Connine, 2001), typically uttered in a casual frame with approximant-like characteristics, but is so often pronounced that it can occur in a careful phonetic frame (Tucker, 2011). The [t] variant is virtually never uttered, but when uttered, it is paired with a careful phonetic frame. In contrast, the same is not true for post-nasal t-deletion. While rampant in AE, it is less likely as the onset of a prominent syllable (Raymond et al., 2006), so as one shifts to a careful speaking style, post-nasal t-deletion becomes less likely. Critically, the stimuli used in both studies involved different pronunciation variants uttered in controlled, careful phonetic frames, biasing a listener against the frequent-variant in the Pitt study.[2]

Consider Figure 1, which illustrates 6 different productions of the word *beating*. Along a hyper-to-hypoarticulation continuum (Lindblom, 1991), half of these productions include the phonological variant [t], the other half include the phonological variant [ɾ].[3]

[2] Pitt (2009, page 903) mentions that the with-[t] production and deleted-[t] production differ by 55 msec (605 vs. 550), which, when carefully-articulated, is the approximate time needed to produce a voiceless alveolar stop, including release. Connine (2004) mentions that the two were phonetically-controlled, as she spliced the variants into a single token that served as the base form.

[3] These examples were created by asking a naïve speaker to produce the word *beating (*extremely carefully; carefully; casually; extremely casually). We created spectrograms for the longest and shortest productions that contained [t] (1, 3; left column) and for those that contained [ɾ] (3, 5; right column). We chose one of many productions in between the endpoints to represent the gradient productions along the continuum.

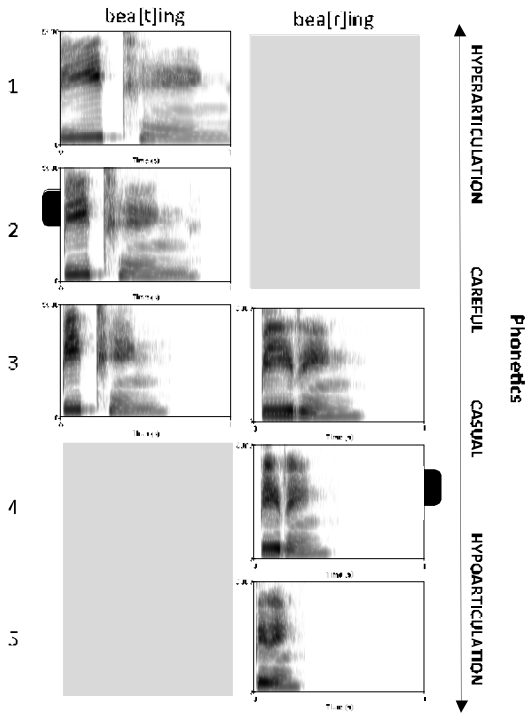

Figure 1. Sample productions of the word *beating* produced with [t] (left column) or [ɾ] (right column). The schema represents a categorical view of the pronunciation variants, with co-varying phonetic patterns, but also shows that a variant may be natural or forced in a particular phonetic frame.

Figure 1 illustrates the wide-range of variation that appears not only in different productions of each pronunciation variant, but also in the variation across word utterances. The spectrograms in row 3 likely illustrate the stimuli used in the studies discussed, as they are phonetically similar independent of the variant examined.

We are interested in comparing variants, like [t] or [ɾ], in *different phonetic frames*. In Figure 1, the spectrograms in the middle of each variant-specific continuum reflect the types of phonetic patterns we find in words uttered with one variant or another. We investigate listener responses to words with [t] that have a more carefully-articulated phonetic word frame to words with [ɾ] that have a more casually-articulated word frame. The stimuli that exemplify this comparison are marked with black tabs.

Across two experiments, we examine the perception of words with pronunciation variants dependent on the phonetic composition of critical words and fillers. Replicating the methods of prior studies, we examine the perception of word-initial stops of words with either [t] or [ɾ], but show that the effects can be attributed more to the phonetic composition of words and fillers (and the expectations and information those provide) than to the pronunciation variants themselves.

Experiment 1

In Exp. 1, we investigate responses to words with [t] and tap when the variants are embedded in phonetic frames that typically co-vary with each variant. We do this for two reasons: First, while this task is based on work by Ganong (1980) showing that lexical status drives categorization responses (listeners make more "T" responses on a task-dask continuum than a tesk-desk continuum), it is not immune to low-level perceptual responses that may also drive categorization (McMurray, et al., 2009; Sumner, 2011). Second, replicating within task as a first step enables us to better interpret past work.

Methods

Participants Thirty-five native monolingual speakers of AE participated in this experiment for credit. All were Stanford University undergraduate students. No participants reported any hearing-related issues.

Materials Eight critical words were used in this study. Four words were b-initial (e.g., *bottom*) and four were p-initial (e.g., *pattern*). In addition to the critical words, we included seven b-initial fillers, and seven p-initial (three for each onset without /t/, *believe, police*; four for each onset with final /t/, *bait, put*). Critically, the voiced/voiceless counterpart of all words (critical and filler) resulted in a pseudoword (e.g., bottom/*pottom, believe/*pelieve). The inclusion of fillers served two purposes (1) to control for response bias (Exp. 1) and to include word-external phonetic support for a casual or careful speaking style (Exp. 2). Each word was recorded, along with its voiced/voiceless pseudoword counterpart in two articulation types: Casual speech and Careful speech. This resulted in eight Careful/[t] and eight 1 Casual/[ɾ] critical words. A continuum was created for each word, as described below.

Stimuli Creation From our recordings, we created b-p continua, resulting in a word-pseudoword continuum for each item. To avoid naturalness differences across onsets, all items were manipulated from the nonword base. Using PSOLA in Praat (Boersma & Weenink, 2008), we then created a 10-step continuum for each item (from 0 to 45 msec in five msec steps) by increasing or decreasing the amount of aspiration in each word. We should note here that we expect an overall bias toward "P" responses, as on this continuum, there are more responses that typically fall within the English voiceless category.

Design This experiment was designed to examine the proportion of word-forming responses (e.g., "B" for *bottom-pottom* continua; "P" for *pattern-battern* continua) resulting from listening to Careful/[t] words and Casual/[ɾ] words. The design was a 2x2 within-subjects design, where the main factors were onset (p, b) and variant/articulation type ([t]/Careful, [ɾ]/Casual).

Procedure Participants completed the task individually or in groups of two or three in a sound-attenuated booth. All 160 critical items (8 critical words X 2 articulation types X 10 continuum steps) were randomized with 140 filler items and presented to participants one at a time in isolation over Sennheiser 390 Pro headphones at a comfortable listening level using E-Prime experimental presentation software. Participants were instructed to listen carefully to each word

presented, to decide whether the token they heard began with a P or B, and then press the corresponding button. Response categories were held constant for each participant, but randomized across participants, so the "B" button appeared equally on the right and in the left across participants. A new trial began one second after a response was recorded, and three seconds if no response was made.

Predictions Evidence for the *frequency bias* should result in more word-forming responses to words with tap than to words with [t]. Evidence for the *canonical bias* should result in more word-forming responses for words with [t] than words with tap. Evidence that this task better reflects pre-lexical responses independent of specified lexical representations should yield some pattern that reflects an influence of the phonetic frame of the words.

Results and Discussion

A mixed logit regression analysis was employed to predict the participants' word-forming responses. We report the results for the model with maximum random effect structure justified by the data based on model comparison (Jaeger, 2008), which contained random by-subject and by-item intercepts. Initial analyses were based on the proportion of word-forming responses, following past work, and show no main effect of articulation type (β = .11 p >.47). A closer look revealed that responses to b-initial words differed dramatically from those to p-initial words. Specifically, b-initial words resulted in a higher proportion of word-forming responses for Careful/[t] words than for Casual/[ɾ] words (β = 1.1 p <.002). Mean proportions of word-forming responses are provided in Figure 2. The onset-based differences suggest that when collapsing across onset, the effects cancel each other out.

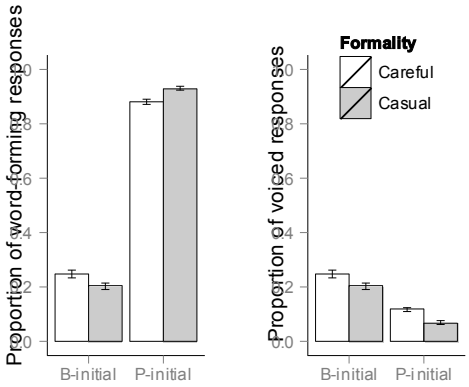

Figure 2. Two plots for resulting data depending on response label: Proportion word-forming responses (left); Proportion voiced responses (right).

The data pattern is unexpected if responses are driven by the pronunciation variant. Were these effects due to the activation of a lexical representation with one pronunciation variant or the other, we would expect the Careful/[t] and Casual/[ɾ] items to behave similarly for each onset with respect to word-forming responses. This is not the case. Here, a "B" response to b-initial words is consistent with both a "word" response and a "voiced" response (e.g., I heard a [b] not a [p]). For p-initial words, a "P" response corresponds with a word-forming response, but not a "voiced" response.

The right panel of Figure 2 plots the data by proportion *voiced* responses. Any influence of co-varying phonetic cues present in the congruent word frames is likely to surface independent of the lexicon—as a phonetic bias. Analyzing the data in terms of proportion voiced responses reveals that Careful [t] items result in a higher proportion of *voiced* responses ("B" regardless of lexical status) than Casual/[ɾ] items (β = -.5 p <.003). This suggests that listener responses depend on the phonetics, not on access to a stored lexical representation. A higher proportion of "B" responses reflects a different categorization boundary between the two articulation types, with more aspiration (longer VOT) required to prompt a "P" response in careful speech, resulting in a higher proportion of "B" responses.

One implication is that the pronunciation variants have little to do with the response patterns in this paradigm. The high number of "P" responses for p-initial words likely reflects combined influences of the asymmetrical breakup of the VOT continuum in English, and lexical status. In order for the variant effects to be attributed to lexical representations, the patterns of responses to words with [t] and words with tap must behave similarly across onsets, predicted by a view where the most accessible lexical representation is the best match to the incoming signal (Johnson, 2006). We would expect results analogous with the lexical effect; which we do not find.

If this paradigm is capturing phonetic responses rather than lexical responses, then we should reconsider claims made about the nature and activation of variant-dependent lexical representations more broadly. Certainly, it is now fact that listener memory for auditory events is detailed. But, this does not imply that all accommodation of variation is handled at the level of the lexicon. One prediction a phonetic account makes is that as the articulation type becomes more predictable, the phonetic categorization bias should be more robust. For example, if item presentation were to be blocked by articulation type, listeners would have information about the speech style well before each critical item. We cast the effect as a category boundary difference mediated by word-level phonetic variation. Therefore, the effects are not due to lexical activation. We predict, then, an increase in evidence of a speech style will reinforce the different VOT thresholds, and will result in a greater difference between the two articulation types.

Experiment 2

Our goal in Exp. 2 was to increase the predictability of a particular articulation type. One prediction of our claim that the basic effects are driven by the phonetic composition of the words and not by the pronunciation variant is that effects should fluctuate as evidence of a particular speech style increases. Blocking the stimuli by articulation type enabled us to test this prediction.

Methods

Participants Thirty-four native monolingual speakers of AE participated in this experiment for credit. All were Stanford University undergraduate students. No participants reported any hearing-related issues.

Materials The stimuli from Exp. 1 were used.

Design The design was identical to Exp. 1 with one exception: Stimuli were blocked by articulation type (careful vs. casual). Block order was randomized, as was the presentation order of items within a block.

Procedure The procedure was identical to Exp. 1.

Results and Discussion

Using the same statistical approach as in Experiment 1, we analyzed the data to predict proportions of voiced responses. Proportions of voiced responses by condition are provided in Table 1.

Table 1. Proportion of voiced responses from Exp. 2.

Type	Condition	N	Proportion Voiced Responses	Standard Deviation	Standard Error	Confidence Interval
B-initial	*Casual /[_]*	1119	0.235	0.424	0.012	0.024
	Careful/[t]	1117	0.286	0.452	0.013	0.026
P-initial	*Casual /[_]*	1141	0.105	0.306	0.009	0.017
	Careful/[t]	1141	0.175	0.380	0.011	0.022

We find that participants are more likely to respond "B" in the Careful/[t] condition than in the Casual/[ɾ] condition (β = -.76 p <.002). To investigate the phonetic effects across experiments, we conducted an additional analysis on the first 100 items for all conditions across Exp. 1 and Exp. 2. The first 100 trials were examined to minimize the influence of learning throughout the experiment. The first 100 trials give us the best picture of participant responses dependent on the nature of the filler items. The proportion of voiced responses for the first 100 items across experiments and conditions are provided in Figure 3

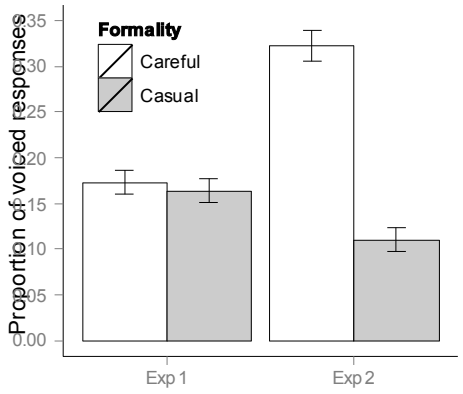

Figure 3. Proportion voiced responses to Careful/[t] items and Casual/[ɾ] items collapsed across onset type for Exp. 1 (left) and Exp. 2 (right) for first 100 trials.

The difference suggests that the word-external information available serves to stabilize different categorization criteria, resulting in a higher rate of voiced responses in the Careful/[t] condition than that found in Exp. 1.

General Discussion

We began this study based on our observations that (1) effects of pronunciation variants are typically examined independently from the phonetic composition of the word-frame in which they are uttered and (2) accounts of opposite pronunciation variant effects that intuitively seem incompatible with each other are both viable under an exemplar-theoretic interpretation. While it is accepted and verified that lexical representations are rich with phonetic detail, we sought to investigate phonetic effects in speech perception independent of the lexicon.

To do this, we investigated pronunciation variants that are embedded in congruent phonetic frames. We then examined the responses made to voiced- and voiceless-initial words when presented in a single block with careful and casual speech styles mixed (Exp.1). Finally, we strengthened the expectations based on speech style by blocking the stimuli by articulation type (Exp.2).

In Exp. 1, considering responses made by listeners as *word-forming* caused some difficulty. The data are more easily accounted for by considering the responses as voiced-voiceless, not as word-forming or pseudoword-forming. In a careful word frame, listeners require a longer VOT before they will switch to a "P" categorization than in a casual frame. Alternatively, this could be driven by an increased likelihood to press "P" at the slightest hint of aspiration in casual word frame.[4] In Exp. 2, we found that increasing the likelihood of a carefully-articulated word (via critical items with phonetically-congruent fillers) increased voiced responses compared to Exp. 1.

One implication of this work is that the canonical bias is, in part, artificially bolstered by our comparisons. And, reconsidering past work, there is support for this notion. A number of studies that have found a canonical effect examine a frequent variant embedded in an incongruent phonetic frame (Andruski, et al., 1994; Gaskell & Marslen-Wilson, 1996). Our data show that the effects here, and likely in some number of previous studies, are due more to congruence between a phonetic frame and a pronunciation variant and to expectation-based categorization than to the activation of a particular lexical representation (or of a more available lexical representation, if we assume there are within-word phonetic clouds). The next step is to consider how frequency-based accounts of lexical access are separate from and integrated with the pre-lexical processes listeners use to navigate a variable speech stream.

[4] While we cannot distinguish the two here, both are compatible with a phonetic explanation of the data rather than one dependent on the pronunciation variants.

Conclusion

We have discussed one limit of exemplar-accounts of variation effects, and have tailored our investigation to examine an apparent paradox in the literature in which two representative studies account for opposing data with the same broad representation-based interpretation. We highlighted both the ways in which phonetic variation might interact with pronunciation variants in speech production, and presented two experiments aimed at understanding the effects of this interaction. As listeners exhibited a strong bias toward voiced responses for Careful/[t] tokens, amplified by within-speech style blocking, we suggest that the difference between the conditions is entirely due to the phonetic composition of the word, absent the influence of detailed lexical representations.

Acknowledgments

We are grateful to Benjamin Lokshin for his assistance with data collection, and to Jason Grafmiller and Kyuwon Moon for valuable comments and discussion. This material is based upon work supported by the National Science Foundation Grant BCS - 0720054 to Meghan Sumner. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Andruski J. E., Blumstein S., & Burton M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163-187.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.0.43, retrieved from http://www.praat.org/.

Clayards, M., Tanenhaus, M.K., Aslin, R.N., and Jacobs, R.A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 804-809.

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. *Psychological Bulletin and Review*, 11, 1084-1089.

de Jong, K. J. 1998. Stress-related Variation in the Articulation of Coda Alveolar Stops: Flapping Revisited. *Journal of Phonetics*, 26, 283 –310.

Gaskell, M. G. & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144-158.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review,* 105, 251-79.

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

Gow, D. W. Jr., (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65, 575–590.

Green, K. P., Tomiak, G. R., Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception and Psychophysics*, 59, 675-692.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language 59*, 434-446.

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

Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In: Hardcastle, W & Marchal, A (Eds.), *Speech production and speech modeling,* (pp. 403-439). Dordrecht: Kluwer.

LoCasto, R. C., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysic*s, 64, 208–219.

Luce, P. A., & McLennan, C. (2005). Spoken word recognition: The challenge of variation. In D. B., Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception*, pp 591-609.

McMurray, B., and Aslin, R.N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, 95, B15-B26.

McMurray, B., Tanenhaus, M.K., and Aslin, R.N. (2009). Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition.

Patterson, D. & Connine, C.M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica*, 58, 254-275.

Pierrehumbert, J.B. (2002). Word-specific phonetics. *Laboratory Phonology*, 7, 101-139.

Pierrehumbert, J.B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

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.

Shockey, L. (2003*). Sound patterns of spoken English*. Malden, MA: Blackwell.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, 119, 131-36.

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

Tucker, B.V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, 39, 312-318.