Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech

Susanne Gahl[1], Yao Yao[2], and Keith Johnson[1]

[1]University of California at Berkeley

[2]Hong Kong Polytechnic University

Address for correspondence:

Susanne Gahl

gahl@berkeley.edu

Department of Linguistics

University of California at Berkeley

1203 Dwinelle Hall

Berkeley, CA 94720-2650

Phone: (510) 643-7621

Fax: (510) 643-5688

Abstract

Frequent or contextually predictable words are often phonetically reduced, i.e. shortened and produced with articulatory undershoot. Explanations for phonetic reduction of predictable forms tend to take one of two approaches: Intelligibility-based accounts hold that talkers maximize intelligibility of words that might otherwise be difficult to recognize; production-based accounts hold that variation reflects the speed of lexical access and retrieval in the language production system. Here we examine phonetic variation as a function of phonological neighborhood density, capitalizing on the fact that words from dense phonological neighborhoods tend to be relatively difficult to recognize, yet easy to produce. We show that words with many phonological neighbors tend to be phonetically reduced (shortened in duration and produced with more centralized vowels) in connected speech, when other predictors of phonetic variation are brought under statistical control. We argue that our findings are consistent with the predictions of production-based accounts of pronunciation variation.

Keywords: Phonological neighborhood density; Word production; Lexical access; Audience design; Pronunciation variation; Phonetic reduction; Word duration; Vowel dispersion; Vowel centralization; Spontaneous speech; Corpora.

Introduction

Many studies have noted a relationship between pronunciation and predictability of utterances. For example, Liebermann (1963) observed that tokens of the word "nine" were shorter and less intelligible when excised from the context "A stitch in time saves ___" than from "The next word will be ___". Similar observations have been made for words that are frequent, repeated within a discourse, or contextually predictable based on semantic, syntactic, or phonological criteria, creating wide-spread consensus that highly predictable items tend to be phonetically reduced. Phonetic reduction is usually understood to mean not only durational shortening, but also articulatory undershoot resulting in consonant lenition, increased coarticulation, and vowel centralization (Aylett & Turk, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Bybee, 2001; Fowler & Housum, 1987; Gahl, 2008, 2009; Gahl & Garnsey, 2004; Hunnicutt, 1985; Jurafsky, 2003; Quené, 2008; Tily, et al., 2009).

Despite this broad consensus, it remains unclear <u>why</u> highly-predictable items reduce – or why, conversely, items of low predictability tend to be lengthened and hyperarticulated. Broadly speaking, explanations of phonetic variation – and variation at other levels of linguistic structure - tend to take one of two approaches, which may be termed "<u>intelligibility-based</u>" and "<u>production-based</u>", respectively. Intelligibility-based accounts (sometimes termed "listener-oriented" or stated with reference to Audience Design (Clark, Brennan, Resnick, Levine, & Teasley, 1991; Galati & Brennan, 2010)) note that speakers may adjust their speech so as to ensure intelligibility of words that might otherwise be difficult to understand (Aylett & Turk, 2004; Flemming, 2010; Fox Tree & Clark, 1997; Lindblom, 1990; van Son & Pols, 2003 for pronunciation variation;

and Lockridge & Brennan, 2002, Levy & Jaeger, 2007, Galati & Brennan, 2010, and Jaeger, 2010 for variation at other levels of linguistic structure).[1] Production-based (or "speaker-internal") accounts, by contrast, attribute variation to production-internal mechanisms, such as variation in the speed of lexical access, retrieval, and encoding in language production. Reduced forms, on this view, occur because articulation reflects the time course of lexical access and retrieval (see for example Bell et al. 2009, Ferreira, 2008; Gahl, 2008 for pronunciation variation; Ferreira & Dell, 2000, Ferreira, 2008 for variation in syntactic realization and word choice) . Both of these two approaches, then, attribute variation to speed and ease of retrieval. They differ in that the relevant retrieval processes underlie either word recognition (in intelligibility-based accounts) or production (in production-based accounts).

---

[1] Several of these proposals (Van Son & Pols,. 2003, Aylett & Turk, 2004, Levy & Jaeger, 2007, Jaeger 2010) are based on information theory and relate the reduction of highly-predictable forms to the pacing of information density throughout utterances. Since estimates of information density are based on the probability of recognition, i.e. from the listener's perspective, these approaches have typically aligned themselves with intelligibility-based approaches to variation. Depending on how information density is modeled, information-theoretic approaches can in principle arrive at the same predictions as production-based approaches, a possibility that is explicitly mentioned in Jaeger (2010): "[w]hether speakers consider their interlocutors' perspective when estimating information density is an empirical question that remains for future research." (Jaeger, 2010: 51).

Comparing the merits of production-based and intelligibility-based approaches is complicated by the fact that these approaches often yield identical predictions: High frequency and high predictability generally makes words good candidates for shortening on the basis of ease of retrieval for production, and it also enables listeners to cope well with poor intelligibility. At the core of this ambiguity is the fact that, "[f]or the most part, the same things that make a word easy to understand make that word easy to say." (Dell & Gordon, 2003, p. 9).

To understand the relationship between pronunciation and predictability of utterances, then, one must ask which retrieval speed matters for the articulation of more vs. less predictable items: production retrieval speed or recognition retrieval speed? The goal of the present paper is to address this question.

With that goal in mind, we focus here on a property of words that affects production and recognition processes differently. As Dell and Gordon (2003) point out, a lexical variable that has this property is phonological neighborhood density. Phonological neighborhood density is a measure of the number of words in the lexicon that are phonologically similar to a given target word. By the most common metric (Luce, Pisoni, & Goldinger, 1990; Nusbaum, Pisoni, & Davis, 1984; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985), two words are considered neighbors if they differ by deletion, insertion, or substitution of one segment (but see Goldrick, Folk, & Rapp, 2010 for an evaluation of different neighborhood metrics as predictors of speech errors). Importantly for the current discussion, words with many neighbors are recognized more slowly and less accurately than words with few neighbors (Luce & Pisoni, 1998; McClelland & Elman, 1986; Vitevitch & Luce, 1998). The relationship between neighborhood density and confusability conforms to many people's intuitions: It is easy to imagine a listener mishearing, for example, *cat* as *hat* or *cap* or some other similar-sounding word. In recognition, then, high phonological neighborhood density creates competition between the target and its neighbors. Interestingly, the effects of phonological neighborhood density on production are quite different: Having many neighbors facilitates word production, as evidenced in speech error rates (Stemberger, 2004; Vitevitch, 1997, 2002; Vitevitch & Sommers, 2003) and naming latencies (Vitevitch, 2002; Vitevitch & Sommers, 2003) in neuro-typical speakers, and in speakers with acquired language disorders (Goldrick, et al., 2010; Gordon, 2002). Phonological neighborhood density thus appears to have inhibitory effects on recognition, but facilitative effects on production.

The inhibitory effect of high phonological neighborhood density has been modeled in several models of word recognition, such as the TRACE model (McClelland & Elman, 1986), the Shortlist model (Norris, 1994), and the Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998). The basic mechanism for modeling the competition between a target and its neighbors in all of these models is that presentation of a target word activates the target along with its neighbors. The activation of other words besides the target word causes a delay or possibly failure in recognizing the target.[2]

---

[2] It should be noted that the notion of activation in the current discussion represents a construct in models of lexical access and retrieval, and in the memory literature more broadly (Anderson, 1983). "Activation", in that literature, refers to a gradient property of nodes in a network that is used to predict interactions among nodes in the network and maps onto processing times for retrieving items from long-term memory. The modeling constructs of "activation", and of "accessibility", differs from the use of

The facilitative effect of high phonological neighborhood density on language production has been modeled more recently (Dell & Gordon, 2003) in the two-step interactive model of lexical access (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). The two-step interactive model of lexical access is a spreading-activation model containing a conceptual semantic level, a “lemma” level, which represents words as semantic/syntactic units, and a level of phonological segments. Importantly, the model assumes that activation may flow in both directions: from lemmas to phonological segments, and from phonological segments to lemmas. As a consequence, once activation has spread from a target lemma to the desired phonological segments, it spreads from those segments to the lemma representations of the target’s phonological neighbors, each of which is linked to all but one of the target’s phonological segments. The target’s neighbors, once activated, send activation to their phonological segments – and the segments, in turn, send activation back to all lemmas linked to them, including the target lemma.

Dell and Gordon’s account anchors the seemingly paradoxical effects of phonological neighborhood density in one of the most fundamental properties of talking and listening: For the most part, speakers start out with an intention to convey some meaning, and they select suitable forms. Listeners, by contrast, start out being confronted with some form whose meaning they must work out. In production, a target word’s main

---

those terms in discussions of salience in discourse, for example, where the words “activate” and “activation” are sometimes used in the sense of “make/be salient” or “bring to someone’s attention”.

competitors and the main source of speech output errors are semantically related words, not phonologically related words (Dell et al., 1997). High neighborhood density facilitates production because feedback from the neighbors' segments to the target lemma increases activation of the target lemma, without increasing the activation of the target's semantic competitors (unless the semantic competitors also happen to be phonologically similar to the target). Word recognition, by contrast, is driven by form. A recognition target's main competitors are phonologically related words: Listeners are far more likely to mistake *cat* for *hat* than for *dog*. Therefore, "production and comprehension differ in their response to neighborhood density in the model because production and comprehension tasks create different competitive environments. When the task dictates that phonological neighbors are serious competitors, a densely populated phonological neighborhood is detrimental to fast and accurate retrieval. When the task dictates that other words are the main competitors, neighborhood density promotes accurate retrieval of the target" (Dell & Gordon, 2003: 28).

The fact that high neighborhood density facilitates production, yet inhibits recognition, means that this variable allows us to tease apart the role of production-based vs. intelligibility-based factors in pronunciation variation. Intelligibility-based accounts would lead one to expect that words with many neighbors should be lengthened and strengthened, to compensate for their low intelligibility. Production-based accounts, on the other hand, would lead one to expect that words that are retrieved quickly tend to be phonetically reduced – provided that fast retrieval speed translates into fast production speed. Whether that is the case may depend on a number of other factors, which we discuss below.

Previous studies of neighborhood density effects on pronunciation variation

A number of studies have examined effects of neighborhood density on pronunciation. Most of these studies have focused on <u>vowel dispersion</u> as a measure of phonetic realization. Vowel dispersion (and its opposite, <u>vowel centralization</u>) refers to the distribution of vowel tokens in vowel formant space. It is commonly quantified by measuring vowel formants (F1 and F2) in word tokens produced by a talker and calculating the Euclidean distance of individual tokens from the center of the space. The more central vowels are in F1/F2 space, the more schwa-like and "reduced" they are. Figure 1 illustrates the F1/F2 space for a talker in the Buckeye corpus of conversational speech (Pitt, et al., 2007).

------------Insert Figure 1 about here---

Increased vowel dispersion is known to be associated with greater intelligibility (Bradlow, Torretta, & Pisoni, 1996). Furthermore, increased vowel dispersion is a feature of "clear speech", i.e. a speaking style speakers adopt, for example, when talking to, or when asked to imagine themselves talking to, a person with a hearing loss (Moon & Lindblom, 1994; Picheny & Durlach, 1985). This makes vowel dispersion a natural variable to focus on for determining whether speakers modify vowel dispersion in such a way as to counteract neighborhood density effects on intelligibility.

The first study to investigate whether neighborhood density affected vowel dispersion (1997, 2004) examined two groups of monosyllabic (CVC) words read in isolation, selected from a database of recordings from 10 speakers (Torretta, 1995). The two groups of words differed in neighborhood density and word frequency. The first group of words, termed the "easy" words, were from sparse neighborhoods and had relatively high frequencies compared to their neighbors. The second group (the "hard" words) were words from dense neighborhoods and had relatively low frequencies, relative to their neighbors. It was found that vowels were significantly more centralized in the high-frequency, low-density words than in the low-frequency, high-density words. This overall effect was carried by the "point" vowels /i,u,a/, i.e. those vowels maximally distant from the (articulatory and acoustic) center of vowel space. Since word frequency and neighborhood density covaried in the stimulus set, the results do not indicate which of these variables was responsible for the observed effect.

A subsequent study (Munson & Solomon, 2004) probed the effects of word frequency and neighborhood density by factorially manipulating these two variables in a single-word naming task: It was found that low frequency and high density were each associated with increased vowel dispersion relative to high frequency and low density. There was also a significant interaction between the two variables, such that words that were of low frequency <u>and</u> high density exhibited the greatest degree of dispersion. It should be noted that the two sets of high-frequency words did not differ in the number of neighbors, but rather in frequency-weighted neighborhood density, a measure combining neighbor count and neighbor frequency. If pronunciation reflects neighborhood size, i.e. the number of neighbors, rather than frequency-weighted density, then the observed interaction could have arisen due to the fact that neighborhood size was not manipulated in the high-frequency group.

Increased vowel duration is usually associated with increased vowel dispersion (Moon & Lindblom, 1994), raising the possibility that variation in vowel dispersion could reflect variation in vowel duration. The correlation between vowel dispersion and vowel duration in Munson and Solomon's study was weak, suggesting that the observed pattern of dispersion did not result from variation in duration. Watson and Munson (2007) confirmed the association of high neighborhood density and increased vowel dispersion in young adult and elderly adult speakers. A further follow-up study (Munson, 2007) likewise reported greater vowel dispersion for words with high neighborhood density than words with low neighborhood density, again using a single-word naming task. Frequency and density were manipulated factorially in that study and had different effects: While high frequency was associated with reduced vowel dispersion and shorter vowel durations, there was no effect of density on duration. The effects of high density were also found in a delayed naming task, where participants were asked to respond after a 1000 ms delay. No effects of frequency on vowel duration or dispersion were found in the delayed naming condition. Similar patterns of greater vowel dispersion for words in dense neighborhoods were reported in Scarborough (2010), in which participants produced a set of short sentences with the target word in final position (though as pointed out in Flemming, 2010, neighborhood density appears to have been confounded with segmental context in that study), and in Kilanski (2009), in which participants produced target words in a short carrier phrase ("Say __ again.").

Scarborough (2009) investigated the degree of nasal coarticulation (nasality in vowels adjacent to nasal stops) in monosyllabic words with nasals in syllable onsets (e.g. *snack, next*) or rimes (e.g. *dunk, home*), along with vowel duration and vowel dispersion.

Scarborough found greater degrees of nasality on the vowels in words from dense neighborhoods than in words from sparse neighborhoods. Scarborough further found greater vowel dispersion in words from dense neighborhoods than in words from sparse neighborhoods, consistent with the patterns reported in Wright (1997, 2004) and Munson and Solomon (2004). Vowel duration did not differ across conditions. It should be noted that neighborhood density in that study was estimated as the sum of the target word frequency and the neighbors' frequency. It is not entirely clear, then, whether the observed pattern was due to target word frequency or phonological neighborhood density, or both. Coarticulation was also investigated in an earlier, more extensive study (Scarborough, 2005). Here, the independent variable was the target word frequency relative to the summed frequency of the target and the frequency of its phonological neighbors, as a measure of confusability of the target with its neighbors. It was found that high confusability, based on target frequency relative to summed neighbor frequency, was associated with increased degrees of nasal coarticulation and vowel-to-vowel coarticulation.

A further acoustic measure in studies of neighborhood effects is voice onset time (VOT), i.e. the time between the release of a stop closure and the onset of subsequent vocal fold vibration. Goldinger and Summers (1989, cited in Wright, 1997) found that, when talkers read pairs of CVC words that differed only in the voicing of the initial consonant (like *bat* and *pat*), VOT differed more in pairs from sparse neighborhoods than in pairs from dense neighborhoods. A more recent study (Baese-Berk & Goldrick, 2009) found that VOT in monosyllabic (CVC or CVCC) words with minimal-pair neighbors differing only in voicing of an initial stop consonant, such as *pox* (vs. *box*), was longer than in words that did not have such neighbors, e.g. *posh* (vs. **bosh*). It was found that this effect was stronger when both words were presented simultaneously on a computer screen than when only the target word was presented, without its neighbor. A subsequent study (Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2010) confirmed that the effect was present even when the minimal pair neighbor was not presented in the stimulus set. It should be noted that the variation in VOT in these studies was not a function of neighborhood density generally, but specifically of the existence of a minimal pair differing in the initial stop consonant.

Few studies of neighborhood density so far have focused on durational measures, other than the duration of the target vowel in studies of vowel dispersion. To date, the most extensive study of effects of neighborhood density on word or segment duration is Kilanski (2009). As mentioned above, high neighborhood density was found in that study to be associated with greater vowel dispersion. The findings for the duration measures, however, indicated that high-frequency words had shorter durations than low-frequency words, consistent with many previous studies. Interestingly for the current context, high neighborhood density was also associated with significantly shorter word and segment durations. This pattern of shortening in words from dense neighborhoods appears to have been carried by the vowel and the word-final consonants (the words in the stimulus set were CVC words).

The studies mentioned so far used a variety of different measures of neighborhood density. As mentioned above, the stimuli examined in Wright (1997, 2004), were contained in a database (Torretta, 1995) classifying words as “hard” or “easy” based on a criterion taking into account target frequency relative to neighbor frequency along with neighborhood size. Another measure of neighborhood density is weighted by the frequency of the neighbors (this measure is used e.g. in Munson, 2007). Another criterion that has been used is the sum of the target frequency and the neighbor frequencies (Scarborough, 2009), or the log frequency of the target divided by the (log) sum of the target frequency and the log frequencies of the neighbors (Scarborough, 2005), as an index of the frequency of a target word relative to its neighbors.

Importantly, previous studies of effects of neighborhood density on pronunciation variation have without exception focused on words produced in isolation or in short carrier phrases, such as “Say __ to me again” or “The first word is __. The word after __ is ___ ” (Scarborough, 2005). This fact is relevant because the relationship between lexical retrieval and phonetic realization may very well be task-dependent. Speakers tend to read word lists at a regular pace (Kello & Plaut, 2000, 2003), in effect setting themselves a deadline for each item. If speakers hold speaking rate constant, then fast lexical retrieval leaves extra time for pronunciation. By contrast, claims about the effects of word frequency have for the most part been based on word duration in conversational speech. This difference is striking, given that word frequency is <u>not</u> reliably associated with shortening when words are produced in isolation or in short carrier phrases. For example, one study (Geffen & Luszcz, 1983) found that, while lists of high-frequency words were read aloud more quickly than lists of low-frequency words when words were blocked by frequency, the difference in speaking tempo was due to differences in pause duration, not articulation time (see also Damian, 2003; Guion, 1995; Whalen, 1991). Nevertheless, there is broad consensus that high word frequency is associated with reduction, based on connected speech data. Analogous evidence on effects of neighborhood density on pronunciation variation in connected speech has not been available so far. The current study fills that gap.

To preview our results: We find that words with many neighbors are shorter in duration and contain more centralized vowels than words with few neighbors, when other factors influencing word duration and vowel dispersion are controlled for.

Methods

We examined the effect of phonological neighborhood density on two aspects of phonetic realization: word duration and vowel dispersion. Mixed-effects regression models were used to bring other known or suspected determinants of word duration and vowel dispersion under statistical control. We constructed two sets of models with token duration (in the first set of models) and vowel dispersion (in the second set) as the outcome variable, Word type and Talker as random effects, and the variables described below as fixed effects. All analyses were carried out using the lme4 (Bates & Maechler, 2010; Bates, Maechler, & Dai, 2008) and languageR (Baayen, 2008b) packages in R (R Development Core Team, 2008).

All data came from the Buckeye Corpus of conversational speech (Pitt, et al., 2007; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005), which consists of ca. one hour of spontaneous speech from each of 40 talkers from Columbus, Ohio, segmented into utterances, words, and phonological segments. One half of the talkers were male. One half of the talkers were under 40 years of age, and half over 40 years of age.

The current study focused on CVC monomorphemic content words in the corpus. Information about several of the control variables, described below, was obtained from the MRC Psycholinguistics database (Wilson, 1988), the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993). Words which did not appear in these databases were excluded from the analysis. A total of 175 word types were excluded because they were frequently used as function words or as discourse markers (e.g. *right* or *like),* their orthographic forms corresponded to multiple phonological forms (e.g. *read, lead, live* and *route)*, or represented personal names (e.g. *Wayne*). The corpus contained 594 word types that met the inclusion criteria. The word types that were included in the analyses did not differ significantly in neighborhood density from the word types that were excluded (mean neighborhood density 21.6 vs. 21.1, t = -0.57). We divided each talker's interview into stretches of speech delimited by changes of turns, non-linguistic sounds such as laughter, and pauses longer than 0.5 seconds. Stretch-initial and stretch-final word tokens, as well as word tokens immediately following or immediately preceding a filled pauses such as *um* and *uh* were excluded from analysis, in order to control variation due to utterance-initial and utterance-final prosody. In addition, we excluded word types with bigram probabilities of 1. Since such words generally represent parts of fixed expressions and/or hapax legomena, their properties may not generalize. The final data set contained 534 word types, represented by 12,414 tokens. A detailed description of the treatment of the data can be found in Yao (2011).

The Buckeye corpus is not currently annotated for syntactic or prosodic structure, both of which affect word duration and possibly other aspects of pronunciation (Warren, 1996; D. Watson & Gibson, 2004). Our decision to limit our investigation to CVC content words, which are all stressable, and to exclude utterance-initial and utterance-final words, was in large part driven by the desire to control for effects of prosody. Also in an attempt to control for effects of prosody, we included syntactic and semantic lexical properties in the model. As we have argued elsewhere (Gahl, 2008, 2009), measures such as familiarity, imageability, and syntactic category capture differences between words belonging to different syntactic categories, information that in turn affects the likely position of a word within prosodic constituents, and hence, its duration.

The analysis of vowel dispersion further excluded words with central (schwa-like) vowels and the diphthongs. Central vowels such as schwa and /ɚ/ are by their nature close to the center of vowel space. Studies of vowel dispersion therefore ordinarily exclude these vowels, along with the diphthongs /aɪ, oy, aʊ/, whose degree of dispersion cannot straightforwardly be measured in the same way as for monophthongs. These exclusion criteria are the same as in previous studies of phonological neighborhood density and vowel dispersion (Munson & Solomon, 2004; Wright, 2004). The exclusion of central vowels and diphthongs meant that the set of words in the analysis of vowel dispersion was a subset of the words in the analysis of word durations. The two sets of words were analyzed in two separate models, which we present in turn.

Model 1: Word durations

The outcome variable of the model of word duration was the log-transformed token duration. Durations were log transformed to take into account the fact that a given absolute difference in duration will amount to a more minor difference in tokens of longer duration. The transformation was further motivated by inspection of the univariate distributions: The distribution of log-transformed token durations was more nearly normal than the distribution of the raw durations. Log-transforms were also applied to several of the predictor variables, as noted in the description of each variable. After all relevant transformations, numerical variables were centered, by subtracting the mean transformed value from each raw value, following the recommendations in Baayen (2008a).

The model of word durations included Word type and Talker as random effects, and the variables described below as fixed effects, presented here in alphabetical order. Treatment coding was used for categorical predictors. Summary statistics for the outcome variable and the control variables are shown in Tables 1 (for numerical predictors) and 2 (for the categorical predictors).

----------------------Insert Table 1 about here -----------------------------

----------------------Insert Table 2 about here -----------------------------

Age: The corpus annotations only indicate two age groups (below and above 40 years), so Age was included as a binary categorical variable in the model. The majority of the talkers mention their age in the course of the interviews, and the ones that do not reveal their approximate age to within a small number of years. Talker age ranged from late teens to late seventies, but was distributed unevenly across age groups. Preliminary versions of the model included more fine-grained information on age, with no change in the pattern of results (Yao, 2011).

Baseline word duration: Phonological segments differ in duration. For example, tense vowels tend to be longer in duration than lax vowels, and nasal stops tend to be longer than voiceless oral stops (Bent, Bradlow, & Smith, 2008; Crystal & House, 1988; Peterson & Lehiste, 1960; Smiljanić & Bradlow, 2008). Word durations can therefore be expected to vary in part as a function of their segmental content. We calculated the average duration of each segment type across the entire Buckeye corpus (Pitt, et al., 2007). We then summed the average durations of each segment in the citation form of each word type. That sum represented the word's baseline duration. The baseline durations were log-transformed and centered.

The purpose of the Baseline duration variable is to capture the fact that word durations can be expected to vary due to segment-level properties, in addition to lexical-level properties. It will be noted that the Baseline durations likely overestimate the duration of the word tokens in our corpus, for two reasons: The Baseline values were estimates of citation forms, but conversational speech is characterized by many segment deletions (Johnson, 2004). Also, the average segment durations were estimated from the whole corpus, including utterance-final words and segments, as well as material before and after speech disfluencies. Since words and segments lengthen in utterance-final positions and near disfluencies, and since we excluded utterance-final and disfluent tokens from the regression analyses, average segment durations in the sample we analyzed are likely to be shorter.

Bigram probability given the word preceding / following the target: The probability of a word, given the immediately preceding or following word in an utterance, has proven a strong predictor of word durations in connected speech (Bell, et al., 2003;

Fosler-Lussier & Morgan, 1999). Bigram probabilities were estimated based on the entire Buckeye corpus. As mentioned before, word types with average bigram probabilities of 1 were excluded from further analysis. The bigram probabilities were log-transformed and centered around their respective means.

Familiarity: Subjective familiarity ratings, like frequency estimates, tend to be significant predictors of the speed of lexical retrieval (Gernsbacher, 1984; Nusbaum, et al., 1984; Pisoni, et al., 1985). Familiarity ratings were those in the MRC Psycholinguistics database (Coltheart, 1981; Wilson, 1988).

Frequency: Frequent words tend to shorten and undergo other types of phonetic reduction (Bell, et al., 2009; Bybee, 2001; Gahl, 2008; Schuchardt, 1885).The frequency measure used in the current model was each word's American English SUBTLEX frequency (Brysbaert & New, 2009). We adopted this measure because it has been shown to predict lexical decision times and accuracy better than several more widely-used measures of word frequency, including CELEX (Baayen, et al., 1993; Kučera & Francis, 1967). For category-ambiguous items, such as *nap,* we used the cumulative frequencies, e.g. the summed frequencies of the noun *nap* and the verb *nap*. The word frequency variable was log transformed and centered.

Phonological neighborhood density: The number of phonological neighbors for each word type was obtained from the Hoosier mental lexicon (Nusbaum, Pisoni, & Davis, 1984).

Orthographic length: The length of each word, in letters. Previous work (Warner, Jongman, Sereno, & Kemps, 2004) has shown that orthographic length can affect word durations even when segmental content and syllable count are controlled for. Orthographic length was centered.

Phonotactic probability: Two separate phonotactic probability estimates for each word type were obtained through the web-based phonotactic probability calculator (Vitevitch & Luce, 2004). One was the average bi-phone positional probability, the other was the average single-phone positional probability. Since measures of phonotactic probability and neighborhood density tend to be highly correlated, and since phonotactic probability has been found to facilitate production when neighborhood density is controlled (Vitevitch, Armbrüster, & Chu, 2004), we examined the behavior of phonotactic probability and neighborhood density closely, in a separate set of models, as described below. The Phonotactic probability measures were log-transformed and centered.

Previous mention: Using the same word multiple times in a discourse tends to promote shortening and possibly other types of phonetic reduction (Bard, et al., 2000; Bell, et al., 2009; Fowler, 1988; Fowler & Housum, 1987; Gahl, 2009). This information was entered into the model as a binary variable coding whether the talker had used the target word previously in the course of the interview.

Speech rate: Two speech rate measures, both measured as syllables per second, were coded for each word token: one for the stretch of speech preceding the target within the utterance, and the other for the stretch of speech following the target. The speech rates, measured in syllables per second, were log-transformed and centered.

Sex: Talker sex was coded as a binary variable, based on the Buckeye corpus information.

Syntactic category (part of speech): Each word type was coded as noun, verb, adverb, or adjective, based on its syntactic category in the CELEX database. The corpus is not syntactically annotated, and hand-disambiguating each token was not feasible. For category-ambiguous items, we therefore used the category with the highest frequency for that item.

Modeling procedure

We used the following procedure to ascertain which of the predictor variables significantly predicted word duration and vowel dispersion: First, we fitted models using only the control predictors, i. e. without the critical variable Neighborhood Density, beginning with a model containing all control variables and retaining only those variables that showed a significant effect, using an alpha level of .15. Significance was estimated based on comparisons between pairs of models with and without each control variable. Then, we added Neighborhood Density to the "control" model and used backward elimination to make the final decisions as to which predictors to retain in the model, i.e. based on comparisons between successively less complex models. At each step, we removed one variable and refit the model. We then compared the Log-Likelihoods of the models with and without the variable in question. When the null hypothesis is true, the change in Log Likelihood (multiplied by 2) follows a chi-square distribution (for sufficiently large datasets) with the difference in the number of parameters between the two models as the degrees of freedom. Predictors that did not significantly lead to significant model improvement, based on this criterion, were eliminated from the model.

In the backward elimination procedure for the models of word duration, we removed variables in the following order: (1) Neighborhood Density; (2) Speaking rate preceding the target; (3) Speaking rate following the target; (4) Bigram probability of the target, given the preceding word; (5) Bigram probability of the target, given the following word; (6) Baseline duration; (7) Part of Speech; (8) Target word frequency. In the backward elimination procedure for the models of vowel dispersion, the order was as follows: (1) Neighborhood Density; (2) Vowel duration; (3) Speaking rate following the target word; (4) Consonant duration; (5) Bigram probability, given the preceding word; (6) Place of articulation of the consonant preceding the target vowel. The least complex models of word duration and vowel dispersion contained only the random effects (Talker and Word). The p-values associated with the beta coefficients in the final model were estimated using the procedure described in Baayen, Davidson and Bates (2008), based on the posterior distribution of model parameters generated by Markov Chain Monte Carlo (MCMC) sampling procedure (10,000 samples). We also conducted model comparisons comparing the full model to models omitting each of the predictors in turn. Each of the predictors in the final models that we arrived at using backward elimination yielded significant model improvement based on those comparisons, and the direction of predicted effects was the same for all predictors regardless of modeling strategy. In prior work (Yao, 2011) and in preliminary work for the current study, we explored the behavior of the control variables further. Since the order in which predictors are included affects the resulting models, we were interested to see whether the behavior of the Neighborhood Density variable remained stable under various different orders of entry. This was found to be the case.

Table 3 shows the bivariate correlations between pairs of variables in the final model.

----------------------Insert Table 3 about here -----------------------------

Word Duration Model: Results

Six predictors – Talker Age, Sex, Orthographic length, Familiarity, Imageability, and Previous mention - did not yield significant model improvement based on the change in log-likelihood and were eliminated. We also explored some non-linear relationships between predictors and word duration, by testing the ability of quadratic and cubic functions of the continuous predictor variables to improve the model. This was the case for the quadratic effect of Speaking rate in the region preceding the target word. We also examined the interaction between Neighborhood size and word frequency, and the three-way interaction between neighborhood size, frequency, and Sex. Neither of these produced significant model improvement, so they were eliminated from the final model. With random effects and fixed effects, the final model accounted for 41% of the observed variability in word duration. A model with only the random effects (Word and Talker) and without any fixed effects accounted for 38% of the variance. A comparison of the random-effects-only model vs. in the model with the fixed effects showed that including the fixed effects reduced the standard deviation of the random effect for Word by 42%.

Model comparisons also revealed that including random slopes for the neighborhood density variable did not yield significant model improvement. This is unsurprising, given that many words in our sample only occurred a very small number of times in the speech of a given talker. Given the large number of control variables, we were concerned about possible multicollinearity. We assessed the degree of collinearity following the procedure in Baayen (2008). The condition number for the model of word durations was 6.4, suggesting a level of multicollinearity that is unlikely to be problematic (Belsley, Kuh, & Welsch, 1980, cited in Baayen et al. 2007) . A summary of the final model is shown in Tables 4 and 5.

----------------------Insert Table 4 about here -----------------------------

----------------------Insert Table 5 about here -----------------------------

The relationship of the control variables to word duration was what one would expect, given previous studies: Longer baseline duration was associated with longer word durations. Increasing Frequency, Bigram probabilities, and Speaking rates were associated with shorter word durations. The proportion of variability accounted for is low compared to some previous models of word and segment duration in connected speech (Bell, et al., 2009; Gahl, 2008; Quené, 2008). This difference is likely to be due in part to the fact that the studies just cited included utterance-final and pre-pausal tokens. Phrase-final position and disfluencies produce large effects on word duration, making it possible to account for a substantial portion of variability in duration based on these two predictors alone.

Crucially for the point of the study, increased Neighborhood density was associated with shorter word durations. Comparison of models with and without this predictor indicates that including this variable resulted in a significant improvement in model fit ($\chi^2(1) = 25.42$, $p < .0001$). The contribution of neighborhood density to word duration, although subtle, approaches that of well established predictors of duration: The difference between the predicted word durations of words with the smallest vs. the largest number of neighbors was 40 ms (269 vs. 229 ms when other predictors are held at their median values). For comparison, the difference in predicted duration of words with the lowest vs. highest frequency in the dataset was 61 (300 ms. vs. 239 ms.).

Figure 2 shows the partial effects of all fixed effects in the final model of word durations.

-----------------------Insert Figure 2 about here ---------------------------

Given the high bivariate correlation between neighborhood density, i.e. the critical variable of interest, and phonotactic probability measures, we scrutinized the behavior of these variables in a separate set of modeling steps, as follows: We first fitted simple linear regression models, predicting neighborhood density from phonotactic probability and *vice versa.* The residuals of these models represent the portion of variability in one variable (e.g. Neighborhood density) not attributable to the other (e.g. Phonotactic probability). We then added the resulting residuals to our mixed-effects regression models of word durations. This allowed us to see the individual contribution of Phonotactic probability and Neighborhood density to variability in word duration.

We used two different measures of Phonotactic probability: The single-phone positional probability and the biphone positional probability (Vitevitch & Luce, 2004).

Since these two measures are highly correlated with neighborhood density and with one another ($r = .62$ for the correlation between biphone positional probability and neighborhood density, $r = .58$ for the correlation between single-phone positional probability and neighborhood density in our data), separate linear regression models were fitted, regressing neighborhood density on each phonotactic probability measure in turn. The simple regression models are summarized in Table 11 in the Appendix.

The effects of neighborhood density were stable, regardless of whether phonotactic probability or neighborhood density were given a chance to explain the variability that could be attributed to phonotactic probability or to neighborhood density: In all models, neighborhood density, or the residual neighborhood density measure representing density not attributable to Phonotactic probabiliy, neighborhood density was associated with shorter word durations (all $p_{MCMC} < .0001$).

The effects of phonotactic probability were more variable: When single phone positional probability was given a chance to explain all the variability attributable to neighborhood density or phonotactic probability, it did not yield a significant effect ($t = -1.46$, $p_{MCMC} = .17$), while residual neighborhood density remained significant ($t = -5.92$, $p_{MCMC} < .0001$). Likewise, when biphone positional probability was given a chance to explain all the variable attributable to neighborhood density or phonotactic probability, it also did not yield a significant effect ($t = -0.74$, $p_{MCMC} = .50$), while residual neighborhood density still remained significant ($t = -5.84$, $p_{MCMC} < .0001$). On the other hand, in models where neighborhood density was given a chance to explain all the variability ambiguously attributable to density or phonotactic probability, residual single-phone and residual biphone positional probability were each associated with lengthening to a significant or marginally significant degree ($t = 3.02$, $p_{MCMC} = .009$ for single-phone probability; $t = 1.74$, $p = .09$ for biphone positional probability); in both of these latter models, neighborhood density was associated with significant degrees of shortening ($t = -6.008$, $p_{MCMC} = .0001$ and $t = -5.25$, $p_{MCMC} = .0001$, respectively). We conclude that the observed effect of neighborhood density is unlikely to be due to phonotactic probability.

Whereas the model just described measures neighborhood density as the number of neighbors, some earlier studies (e.g. Munson, 2007) used a frequency-weighted measure of neighborhood density. To facilitate comparison of our results to those earlier studies, we repeated the analysis, this time using a frequency-weighted measure of phonological neighborhood density (the sum of the neighbors' log frequencies). The frequency-weighted measure of neighborhood density was associated with shorter word durations ($t = -5.2.91$, $p_{MCMC} = <.0001$), just like the unweighted measure of neighborhood size. The pattern of significance and the direction of the predicted effects also remained unchanged.

In summary, the models of word duration suggest that, other things being equal, words with many phonological neighbors are shorter than words with few neighbors. To examine the effect of phonological neighborhood density on phonetic reduction more closely, and to facilitate comparison of our data with earlier studies, we now turn to the analysis of vowel dispersion.

Model 2: Vowel dispersion

Methods

The data set for the analysis of vowel dispersion was smaller than the data set for word durations, in part due to the exclusion of central vowels and diphthongs. One speaker's data (speaker s35, 222 tokens) were removed due to errors in the transcript, which contained incorrect time labels for a sizable portion of the vowels. An additional 125 word tokens had to be excluded because extremely short durations or low intensity precluded reliable formant measurements. The final dataset for the analysis of vowels contained 414 word types, represented by 9,075 tokens from 39 talkers.

Vowel formant analyses were carried out using Praat (Boersma & Weenik, 2002-2005). The onset and offset of the vowels were those in the Buckeye segmentation. The duration of the analysis window was 25 ms, and the time steps were 2.5 ms. For each token, we extracted the mean F1 and F2 over the middle 50% of the vowel. Tokens with mean formant values at least 2.5 standard deviations away from the speaker- and vowel-specific means were manually checked: Where possible, formants for such tokens were measured by hand. Tokens for which estimates of the formant values were impossible to obtain, e.g. because of excessively short duration, were removed from the dataset. Fewer than 1% of the tokens in the database were removed for this reason. Further details about the treatment of the data and preliminary analyses can be found in Yao (2011).

The center of each talker's F1/F2 space was estimated by obtaining the average F1 and F2 values for the mid central vowel [ʌ] in all CVC monomorphemic content words (e.g. *hub*) produced by that talker (41 tokens on average). Figure 1 above shows the vowel space of one of the talkers (s26, female). The center of the talker's F1/F2 space is marked with a plus sign.

Following earlier work (Bradlow, et al., 1996), vowel dispersion was quantified as mean Euclidean distance between the F1 and F2 of each vowel token and the center of each talker's F1/F2 space. That distance measure was then normalized, to control for between-vowel differences in vowel dispersion: For example, tokens of the vowel [i] are further from the F1/F2 center, on average, than tokens of the vowel [ɑ]. We calculated the standardized distance of each token as a z-score, i.e. as the difference between the token's distance from the F1/F2 center and the mean distance from the center for all tokens of a given vowel type, divided by the standard deviation of the distance from the center for all tokens of a given vowel type. Increased distance from the F1/F2 center, compared to other tokens of a given vowel, increases standardized distance.

Normalizing the distance measurements in this way meant that the exact location of the designated center of each speaker's vowel space would not substantially affect the results: The standardized distance represented the distance of particular token from the center, relative to the typical distance from the center for tokens of that vowel type for a given speaker. For example, tokens of the vowel [i] have a certain average distance from whatever reference point one might choose. The standardized distance of a particular token is the difference between the token's F1/F2 coordinates and the coordinates of the average [i] values, normalized by the standard deviation of F1/F2 values of [i] (to take into account the spread of F1/F2 values for tokens of [i]). If the chosen reference point were at an extreme point of the vowel space, the estimates of standardized distance would be distorted. To check whether the choice of reference point unduly affected the outcome, we repeated our analyses using a different center, based on the average F1/F2 of two sets of four non-schwa vowels ([a, æ, i, o] and [a, æ, i, u], respectively). The pattern of results was unchanged.

The model included Word type and Talker as random effects. Most of the fixed-effect variables in the vowel dispersion model were the same as in the word duration model. The model of vowel dispersion additionally included several variables, described below, that pertain to the analysis of single segments. As in the model of word duration, continuous variables were centered and log-transformed where appropriate. Tables 6 and 7 present summary statistics of the numerical (Table 6) and categorical (Table 7) variables. Table 8 shows the pairwise correlations between the predictors.

The following variables were specific to the vowel dispersion model:

Vowel duration: Vowel dispersion is in part a function of vowel duration (Lindblom, 1964), both in that short vowels have a tendency to centralize, and in that the formants of short vowels tend to be similar to those of surrounding consonants. Therefore, reduced vowel dispersion could easily result from variation in vowel duration alone. We therefore entered vowel duration in the model. Vowel durations were log-transformed and centered.

Consonant duration: To control for effects of word duration outside of the target vowel itself, we also controlled for the duration of the consonants preceding and following the target vowel, i.e. the target word duration minus the duration of the vowel (recall that all target words were CVC words). Durations were log-transformed and centered.

Place and manner of articulation (before, after the target vowel): Neighboring consonants can affect vowel formants, due to coarticulation. For example, vowels near nasal consonants tend to have lower F2 values, whereas vowels near alveolar consonants tend to have higher F2 values. To control for the influence of the consonants in the target words, we added categorical variables coding place (front vs. back) and manner (glide vs. nasal vs. obstruent) of the consonants preceding and following the target vowel.

----------------------Insert Table 6 about here -----------------------------

----------------------Insert Table 7 about here -----------------------------

----------------------Insert Table 8 about here -----------------------------

Results: Vowel dispersion model

Several variables (Vowel type, Talker age, Sex, Frequency, Part of speech, Manner of articulation, Voicing of neighboring segments, Bigram probability given the following word, Speaking rate preceding the target word, and Previous mention) were not associated with significant model improvement and were removed from the model. Random slopes for neighborhood density also did not improve the model and were eliminated. Neither the Frequency * Density interaction, nor the three-way interaction of Frequency, Density and Sex, yielded significant effects. The control variables that did give rise to significant effects in the final model did so in the expected direction: Other things being equal, vowels were more centralized (less dispersed) following non-back consonants, and before stretches of speech with higher speaking rates. Vowels were more dispersed in tokens with greater vowel and consonant durations. The model accounted for 34 % of the observed variability in vowel dispersion. The final model is summarized in Tables 9 and 10. The partial effects are shown in Figure 3.

-----------------------Insert Figure 3 about here ---------------------------

Turning to the neighborhood density variable, we observed that high neighborhood density and squared neighborhood density were both associated with reduced vowel dispersion, to a significant degree (Neighborhood density: t.= -1.695, p_{MCMC} = .04; Squared neighborhood density: t = -2.687, p_{MCMC} = .0076).

----------------------Insert Table 8 about here -----------------------------

We examined the contribution of phonotactic probability, using the same residualization and model comparison techniques as with the model of word durations: We residualized neighborhood density on phonotactic probability and *vice versa* using simple linear regression. We then fitted mixed-effects models with the same random and fixed effects as in the final model of vowel dispersion, except that instead of the measure of neighborhood density, we entered fixed effects probing the contributions of neighborhood density and phonotactic probability. For example, in one model, single-phone positional probability was entered along with residual neighborhood density, i.e. the variability in neighborhood density that could not be predicted from single-phone positional probability.

The pattern of results was simple. Measures of phonotactic probability (single-phone positional probability and biphone probability) did not give rise to significant effects in any of these models, regardless of whether phonotactic probability was residualized on neighborhood density or the other way around (all $t < 1.8$, all $p_{MCMC} > .18$). Neighborhood density, by contrast, gave rise to a significant effect in all models and was consistently associated with decreased vowel dispersion. This was the case regardless of whether neighborhood density was regressed on a measure of phonotactic probability or *vice versa* (all $|t| > 2.25$, all $p_{MCMC} < .03$). We conclude that the observed effect of neighborhood density was unlikely to be due to phonotactic probability. We note that the inability of Phonotactic probability to account for variability in vowel dispersion may have to do with competition from the Place of articulation variable, which models some of the same segment-to-segment coarticulatory effects that would lead one to expect effects of phonotactic probability.

To facilitate comparison of our results to earlier studies, we also fitted a model with a frequency-weighted measure of neighborhood density, in place of the neighborhood size variable. The frequency-weighted density measure did not yield a significant effect (beta = -0.020, $t = -1.125$, $p_{MCMC} = .37$).

An anonymous reviewer points out that there is some evidence suggesting a tendency for talkers to produce novel dialectal variants more readily in contexts that are predictable semantically (Clopper & Pierrehumbert, 2008) or based on word frequency or frequency-weighted neighborhood density (P. J. Watson & Munson, 2007). The effect we observed was not restricted to particular vowel types, which one would expect if the pattern were driven by dialect variation. We therefore believe that dialect variation is unlikely to be the source of the effect.

There is some evidence in previous studies (Munson, 2007; Munson & Solomon, 2004) of an interaction between frequency and neighborhood density, such that the effect of neighborhood density was stronger, or possibly restricted to, low-frequency words. We did not observe such an interaction. Nevertheless, it is of course possible that some effects of neighborhood density are restricted to, or are strongest in, low-frequency words, which are underrepresented in spontaneous speech corpora.

In summary, neighborhood density – the number of a word's neighbors in the lexicon – was associated with reduced vowel dispersion.

Discussion

Our central finding was that, in conversational speech, words from dense phonological neighborhoods were shorter and contained more centralized (less dispersed) vowels than words from sparse phonological neighborhoods. These findings resemble a familiar pattern of phonetic reduction in words that are of high frequency or high contextual predictability (Aylett & Turk, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Bell, et al., 2003; Gahl, 2008).

Our aim in investigating the effects of neighborhood density on word durations and vowel dispersion was to understand the role of lexical retrieval and intelligibility in pronunciation variation of predictable forms. Neighborhood density provides a means to adjudicate between competing explanations of pronunciation variation, because it has been shown to yield facilitative effects on production (Stemberger, 2004; Vitevitch, 1997, 2002), yet detrimental ones on intelligibility (e.g. Vitevitch & Luce, 1998). Therefore, production-based accounts of pronunciation variation lead one to expect phonetic reduction of words in dense neighborhoods, whereas intelligibility-based accounts would lead one to expect the opposite. Our findings are consistent with the predictions of production-based accounts of pronunciation variation in spontaneous speech.

We begin our discussion by considering some limitations of the current study, before comparing our findings to those reported in earlier studies.

Limitations and alternative explanations for the observed pattern

Some limitations of the current study are inherent in data from spontaneous speech: Our findings may reflect uncontrolled variation in the corpus. Secondly, our measure of neighborhood density was based on citation forms. Conversational speech is characterized by many instances of omissions of segments or entire syllables (Johnson, 2004). In fact, this was one of the reasons for our decision to restrict our analysis to tokens in which all segments present in the citation form were actually produced. It remains as a topic for future research whether neighborhood density effects in conversational speech perhaps reflect neighborhood characteristics of forms as they are actually produced. Furthermore, like all previous studies of the effects of neighborhood density on pronunciation variation, we used a position-independent measure of neighborhood density, meaning that “cap” and “fat” were counted equally as neighbors of “cat”. As an estimate of lexical competition, that measure is problematic in a number of ways (see Goldrick, et al., 2010).

The uncontrolled nature of conversational speech data makes it especially important to consider alternative explanations of the observed patterns. One candidate for such an alternative might be word frequency: The measure of word frequency that we chose (Brysbaert & New, 2009) has been shown to be a good predictor of lexical decision and naming times. The decision to use a corpus-external frequency measure leaves open the possibility that our results might have been due to a positive correlation between phonological neighborhood density and corpus-specific word frequency. We therefore examined the role of frequency within the corpus in a set of follow-up analyses.

The Buckeye corpus consists of one-on-one interviews. As a result, many words, particularly content words, occur frequently in some interviews, and hence in the speech of some talkers, but not in others. Overall frequency in the corpus is a poor index of word frequency in any one talker's speech. To check if the observed effect was due to usage frequency within the corpus, we therefore examined the relationship between talker-specific word frequency and neighborhood density: If words used frequently by individual talkers tended to reside in dense neighborhoods, then the observed pattern of reduction of high-density words could have come about due to talker-specific frequency in our sample. To investigate this possibility, we determined, for each talker, the Spearman rank correlation between talker-specific word frequency and neighborhood density. These correlations turned out to be weak, ranging from -.10 to .02. A total of 33 out of these 40 correlations were negative, three of them significantly so at an alpha level of .05. None of the seven positive correlations were significant at an alpha level of .05 (all $p > .65$). In light of this, we consider it unlikely that the observed association of high neighborhood density with shortening and vowel reduction was due to talker-specific word frequency in the Buckeye corpus. If anything, there was a slight tendency for words in dense neighborhoods to occur less frequently in a given interview; therefore, effects of corpus-specific frequency should counteract the overall observed association of high neighborhood density and reduction.

The more general possibility remains, of the observed effect resulting from uncontrolled variation. For example, our model does not control for effects of upcoming material, except through the bigram probability of the target word given the word immediately following it. Future, more complete, models of spontaneous speech generally, and of the Buckeye corpus in particular, may provide alternative explanations for the observed pattern.

Comparison to previous results

Previous studies (Kilanski, 2009; Munson, 2007; Munson & Solomon, 2004; P. J. Watson & Munson, 2007; Wright, 1997, 2004) found increased vowel dispersion for words in dense neighborhoods compared to words in sparse neighborhoods, contrary to our findings. What might account for this apparent discrepancy? We see several methodological differences, including the different measures of neighborhood density and our use of a normalized measure of vowel dispersion. We discuss these differences next, before turning to what we believe is the main source of differences between our results and previous studies, which is the fact that our observations are based on conversational speech, as opposed to single-word production.

As mentioned above, using a frequency-weighted measure of neighborhood density in place of the measure of neighborhood size left the pattern of results unchanged in the model of word duration. When entered into the model of vowel dispersion, frequency-weighted neighborhood density did not give rise to a significant effect. It is thus possible that our use of an unweighted neighborhood density measure was responsible for the difference in findings concerning vowel dispersion.

Our use of a normalized measure of vowel dispersion constitutes another source of differences between the present findings and previous results. Whereas the greater dispersion of vowels in “hard” words in Wright (1997, 2004) was only observed in the “point” vowels /i,a,u/, we found an across-the-board effect of neighborhood density on vowel dispersion, for all vowel types. Presumably, our dispersion normalization procedure is responsible for part of this difference: Despite vowel-to-vowel differences in absolute dispersion, when dispersion is expressed as a z-score relative to the range of acoustic variation typically seen for a particular vowel, the degree of dispersion is seen to be constant across vowels. Normalization does not change the direction of the result, but the normalization procedure may explain why the observed effect did not depend on vowel type in our data.

We suspect that the main reason for the discrepancy between previous findings and ours is the fact that we examined conversational speech, as opposed to words presented in isolation or in short carrier phrases. It is clear that temporal characteristics of the material analyzed in previous studies differ from ours: Wright (1997, 2004), for example, presented words one at a time and instructed talkers to say each word “at a ‘medium’ rate” (Wright, 1997: 475). Even when speakers are not specifically instructed to keep their speaking rate constant, they tend to produce word lists at an even pace (Kello & Plaut, 2000, 2003). By contrast, the current study is based on word tokens excised from running conversational speech, which is highly variable and very fast, compared to words produced in isolation (Bard & Aylett, 2005). As importantly, attentional demands in elicited isolated utterances and conversational speech differ. We believe that these differences in temporal and attentional constraints may explain the apparent discrepancy between the current findings and previous studies.

Increased vowel dispersion is associated with greater intelligibility (Bradlow, et al., 1996). Given that neighborhood density inhibits word recognition, it is natural to attribute variation in vowel dispersion to speakers' attempts to maximize intelligibility, and several previous accounts have done so (e.g. Scarborough, 2005; Wright, 1997, 2004), building on Lindblom (1990). Previous authors have also noted other possible explanations for the increased vowel dispersion for words in dense neighborhoods, based articulatory target drift (Pierrehumbert, 2001) and perceptual factors unrelated to speakers' attempts to modify intelligibility (Baese-Berk & Goldrick, 2009; Munson, 2007; Munson & Solomon, 2004). For example, Baese-Berk and Goldrick (2009) attribute their observed pattern of longer VOTs for words with minimal-pair neighbors differing only in voicing of an initial stop consonant (*pox* vs. *box*), compared to words without such neighbors (*posh* vs. **bosh*) to "higher activation levels for words in dense neighborhoods" (Baese-Berk & Goldrick, 2009, p. 531). Activation, in the model that study is situated in, models lexical retrieval speed. If Baese-Berk and Goldrick's proposal is correct, then faster retrieval speed for production might be associated with maximally intelligible pronunciation more generally – or more accurately, with the more precise realization of articulatory targets. High word frequency has been argued to cause articulatory targets to "drift" towards more phonetically reduced productions (Pierrehumbert, 2001); high neighborhood density, by contrast, does not have this effect. Taken together with the current results, and with the observation that word lists tend to be produced at a regular pace (Kello & Plaut, 2003), Baese-Berk and Goldricks' and Pierrehumbert's proposals leads to a different understanding of the previously observed association of high neighborhood density and intelligibility: Given that people tend to read word lists at an even pace, fast retrieval leaves speakers time to realize extreme articulatory targets, which in turn tend to be highly intelligible.

Production speed aside, conversational speech may also create different attentional demands than word lists or short, scripted utterances. In single-word naming tasks, for example, speakers are only faced with the task of planning whatever word is required for the current trial. Conversational speech, on the other hand, requires the language production system to coordinate grammatical and phonological encoding of upcoming material during lexical retrieval, phonological encoding, and articulation of current targets. In single-word naming tasks, this is not the case, freeing speakers to realize more or less extreme articulatory targets as temporal and attentional demands allow, and as articulatory target selection may favor.

Conclusion

Neighborhood density effects in conversational speech yielded a pattern of shortening and vowel centralization in words that are generally found to be challenging targets for word recognition, yet easy production targets. Our findings are consistent with the generalization that pronunciation variation associated with lexical access and retrieval -- “early”, automatic processes in language production -- are speaker-centric (Bard & Aylett, 2005). In our view, these results are fully compatible with the notion that variation at some levels of linguistic structure, with different levels of planning and encoding, may reflect speakers’ models of their listeners and of their surrounds. Clearly, speakers do take their listeners’ needs into account, and this fact is reflected in referential form and other dimensions of linguistic structure (Arnold, 2008; Brennan & Clark, 1996). More generally, we see no reason to doubt, for example, the existence of foreigner talk, “clear speech”, or baby talk.

Previous research studying situations in which speakers’ and listeners’ needs are pitted against each other suggests limits of intelligibility-based behavior (Arnold, 2008; Bard & Aylett, 2005; Ferreira, 2008; Ferreira & Dell, 2000), partly as a function of demands on attention and working memory (Wardlow Lane & Ferreira, 2008; Wardlow Lane, Groisman, & Ferreira, 2006). Our findings suggests that conversational speech is a situation of just this kind. It is our hope that they current study will inspire further scrutiny of the mechanisms – be they production-based or otherwise – linking what is known about lexical access and retrieval to the study of the phonetic realization of conversational speech.

Appendix A:

Results of residualizing phonotactic probability on neighborhood density and *vice versa*

-------------------------Insert Table 11 about here -----------------------------

Appendix B:

Summary of word duration models using residualized measures of phonotactic probability or neighborhood density

-------------------------Insert Table 12 about here -----------------------------

References

Anderson, J. R. (1983). Spread of Activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 791-798.

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence and Duration in Spontaneous Speech. Language and Speech. *Language and Speech, 47*(1), 31-56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America, 119*(5), 3048-3058.

Baayen, H. (2008a). *Analyzing linguistic data: A practical introduction to Statistics using R*. Cambridge: Cambridge University Press.

Baayen, H. (2008b). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics" (Version 0.953).

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database (CD-ROM).

Baayen, H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency words: A regression study across tasks and modalities. *The Mental Lexicon, 2*(3), 419-463.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*, 527-554.

Bard, E. G., & Aylett, M. P. (2005). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use* (pp. 173-191). Cambridge, MA: MIT Press.

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-33. http://CRAN.R-project.org/package=lme4.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. http://lme4.r-forge.r-project.org.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*(1), 92-111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*(2), 1001-1024.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics. Identifying influential data and sources of collinearity*. New York: Wiley

Bent, T., Bradlow, A. R., & Smith, B. L. (2008). Production and perception of temporal patterns in native and non-native speech. *Phonetica, 65*, 131-147.

Boersma, P., & Weenik, D. (2002-2005). Praat: doing phonetics by computer (Version 4.3.04) (Version 4.0.26). Amsterdam.

Bradlow, A. R., Torretta, G., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

Brysbaert, M., & New, B. (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.

Clark, H. H., Brennan, S. E., Resnick, L. B., Levine, J. M., & Teasley, S. D. (1991). Grounding in communication *Perspectives on socially shared cognition.* (pp. 127-149). Washington, DC, US: American Psychological Association.

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America, 124*, 1682-1688.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33*(A), 497-505.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America, 83*(4), 1553-1573.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 416-431.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production* (pp. 9-47). New York: Mouton De Gruyter.

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 210-246). Amsterdam: Elsevier.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*(4), 296-340.

Flemming, E. (2010). Modeling listeners: Comments on Pluymaekers et al. and Scarborough. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 587-606). Berlin: Mouton De Gruyter.

Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word predictability on conversational pronunciations. *Speech Communication, 29*, 137-158.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*(2), 151-167.

Gahl, S. (2008). "Time" and "thyme" are not homophones: Word durations in spontaneous speech. *Language, 84*(3), 474-496.

Gahl, S. (2009). Homophone duration in spontaneous speech: A mixed-effects model. UC Berkeley.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35-51.

Geffen, G., & Luszcz, M. A. (1983). Are the spoken durations of rare words longer than those of common words? *Memory & Cognition, 11*(1), 13-15.

Goldinger, S. D., & Summers, W. V. (1989). *Lexical neighborhoods in speech production: A first report*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language, 62*(2), 113-134.

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language, 82*(2), 113-145.

Guion, S. (1995). Word frequency effects among homonyms. *Texas Linguistic Forum, 35*, 103-116.

Jaeger, F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23-62.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29-54). Tokyo: The National International Institute for Japanese Language.

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 719-750.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language, 48*(1), 207-232.

Kilanski, K. (2009). *The effects of token frequency and phonological neighborhood density on native and non-native speech production.* University of Washington, Seattle.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Levy, R., & Jaeger, F. (2007). *Speakers optimize information density through syntactic reduction*. Paper presented at the Twentieth Annual Conference on Neural Information Processing Systems.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172-187.

Lindblom, B. (1964). Articulatory activity in vowels. *Speech, Music and Hearing, Quarterly Progress Report, 1*, 1-15.

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

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic bulletin & review, 9*(3), 550-557.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear & Hearing, 19*(1), 1-36.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*(1), 40-55.

Munson, B. (2007). Lexical Access, Lexical Representation, and Vowel Production. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9: Phonology and Phonetics* (pp. 201-227). Berlin: Mouton de Gruyter.

Munson, B., & Solomon, N. P. (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *Speech, Language, and Hearing Research, 47*, 1048-1058.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words*. Bloomington, IN: Psychology Department, Indiana University.

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2010). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of cognitive neuroscience, 23*(3), 593-603.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32*, 693-703.

Picheny, M. A., & Durlach, N. I. (1985). Speaking Clearly for the Hard of Hearing I. *Journal of Speech, Language, and Hearing Research, 28*, 96-103.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. [References] *Bybee, Joan (Ed); Hopper, Paul. (2001). Frequency and the emergence of linguistic structure. Typological studies in language, vol. 45.* (pp. 137-157). Amsterdam, Netherlands: John Benjamins Publishing Company.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]

R Development Core Team (2008). R: A language and environment fo statistical computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Scarborough, R. A. (2005). Coarticulation and the Structure of the Lexicon (Doctoral dissertation, UCLA Department of Linguistics, 2004). *Dissertation Abstracts International, A 66*(02), 574.

Scarborough, R. A. (2009). *Lexical similarity and speech production: Neighborhoods for Nonwords*. Paper presented at the NELS Workshop on Phonological Similarity.

Scarborough, R. A. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In L. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Laboratory Phonology 10* (pp. 557-586). Berlin, New York: De Gruyter Mouton.

Smiljanić, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36*(1), 91-113.

Stemberger, J. P. (2004). Neighbourhood effects on error rates in speech production. *Brain and language, 90*(1-3), 413-422.

Torretta, G. (1995). The "easy-hard" word multi-talker speech database: An initial report *Research on Spoken Language Processing Progress Report* (Vol. 20, pp. 321-333). Bloomington, IN: Speech Research Laboratory, Indiana University.

van Son, R. J. J. H., & Pols, L. C. W. (2003). *Information structure and efficiency in speech production.* Paper presented at the 2003 Eurospeech conference, Geneva, Switzerland.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech, 40*(3), 211-228.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 735-747.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and Lexical Representations in Speech Production: Effects of Phonotactic Probability and Onset Density. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 514-529.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325-329.

Vitevitch, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments & Computers, 36*(3), 481-487.

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31*(4), 491-504.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1466-1481.

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science, 17*, 273-277.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*(2), 251-276.

Warren, P. (1996). Prosody and parsing: An introduction. *Language & Cognitive Processes, 11*(1-2), 1-16.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*(6), 713-755.

Watson, P. J., & Munson, B. (2007). *A comparison of vowel acoustics between older and younger adults*. Paper presented at the ICPhS.

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America, 90*, 2311.

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20*, 6-11.

Wright, R. (1997). *Lexical competition and reduction in speech: A preliminary report*: Indiana University.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden & R. Temple (Eds.), *Papers in Laboratory Phonology VI* (pp. 26-50). Cambridge: Cambridge University Press.

Yao, Y. (2011). *Effects of Neighborhood Density on Pronunciation Variation.* University of California at Berkeley, Berkeley.

Table 1. Summary statistics of the numerical variables in the model of word durations. See text for additional information about each variable.

	Median	Mean (SD)	Range
Token duration	241 ms	256 ms (89)	10 – 1043 ms
Baseline duration	250 ms	252 ms (34)	188 – 378 ms
Bigram probability, given the preceding word	.005	.027 (.070)	7.90e-5 - .75
Bigram probability, given the following word	.005	.030 (.078)	7.90e-5 - .83
Familiarity	7.0	6.95 (0.13)	2.4 – 7.0
Frequency	523.1	799.1 (763.3)	0.43 – 3141.0
Neighborhood density (number of neighbors)	21.0	20.65 (6.84)	3-40
Frequency-weighted neighborhood density	40.68	43.0 (15.47)	4.4-92.0
Orthographic length (in letters)	4.0	4.05 (0.70)	3-7
Phonotactic probability:			

Phoneme probability	.046	.048 (.016)	.012 – .098
Biphone probability	.002	.003 (.002)	.000 - .016
Speech rate (before) (in syllables/sec)	5.94	6.25 (2.28)	0.9 – 33.3
Speech rate (after)	5.25	5.32 (1.70)	0.42 – 41.0

Table 2. Summary statistics of the categorical control variables in the model of word durations. See text for additional information about each variable.

Age	Young (< 40): 5,450
	Old: (> 40): 6,964
Part of speech	Adjective: 2,399
	Noun: 4,530
	Verb: 4,981
	Adverb: 504
Previous mention	True: 8,811
	False: 3,603
Sex of talker	Female: 5,910
	Male: 6,504

Table 3. Pairwise (Spearman) correlations between variables in the model of word durations

	Dur	Age	Base	BigrA	BigrB	Fam	Freq	Len	ND	PoS	BiPh	SPh	Prev	RateA	RateB	Sex
Dur	1	0	0.17	-0.09	-0.04	0.05	-0.12	0.05	-0.01	-0.08	-0.03	-0.04	0.1	-0.12	-0.11	0.01
Age	0	1	-0.01	0	-0.01	0	-0.04	-0.01	0	-0.01	0	-0.02	0.06	-0.05	0.01	0.04
Base	0.17	-0.01	1	-0.05	-0.01	0.03	-0.09	0.09	0.17	-0.02	0.04	-0.02	0.09	0.01	-0.01	0
BigrA	-0.09	0	-0.05	1	-0.01	-0.11	0.07	-0.05	0.04	-0.12	0.09	0.07	-0.02	-0.06	0.02	0.01
BigrB	-0.04	-0.01	-0.01	-0.01	1	-0.01	0.15	0.04	0	-0.21	0.01	0.02	-0.09	0.02	-0.03	0
Fam	0.05	0	0.03	-0.11	-0.01	1	-0.1	0.16	-0.1	-0.18	-0.11	-0.15	0	0	-0.01	0.02
Freq	-0.12	-0.04	-0.09	0.07	0.15	-0.1	1	0.06	-0.04	-0.03	-0.06	-0.13	-0.26	0.05	0.04	-0.01
Len	0.05	-0.01	0.09	-0.05	0.04	0.16	0.06	1	-0.28	-0.02	-0.29	-0.38	0.02	0.01	-0.01	-0.03
ND	-0.01	0	0.17	0.04	0	-0.1	-0.04	-0.28	1	0.13	0.44	0.44	0.07	0	-0.01	0.02
PoS	-0.08	-0.01	-0.02	-0.12	-0.21	-0.18	-0.03	-0.02	0.13	1	-0.16	-0.08	0.07	0.03	0.02	-0.02
BiPhono	-0.03	0	0.04	0.09	0.01	-0.11	-0.06	-0.29	0.44	-0.16	1	0.73	0.02	0	0.02	0.01
SPhono	-0.04	-0.02	-0.02	0.07	0.02	-0.15	-0.13	-0.38	0.44	-0.08	0.73	1	0.03	0	0.02	0.02
PrevMen	0.1	0.06	0.09	-0.02	-0.09	0	-0.26	0.02	0.07	0.07	0.02	0.03	1	-0.02	-0.02	0.01
RateA	-0.12	-0.05	0.01	-0.06	0.02	0	0.05	0.01	0	0.03	0	0	-0.02	1	0.09	0.01

RateB	-0.11	0.01	-0.01	0.02	-0.03	-0.01	0.04	-0.01	-0.01	0.02	0.02	0.02	-0.02	0.09	1	0.03
Sex	0.01	0.04	0	0.01	0	0.02	-0.01	-0.03	0.02	-0.02	0.01	0.02	0.01	0.01	0.03	1

Note. Dur = word duration; Age = talker age; Base = baseline word duration; BigrA = Bigram probability of the target word, given the following word; BigrB = Bigram probability of the target word, given the previous word; Fam = Subjective familiarity rating; Freq = SUBTLEX word frequency; Len = orthographic length; ND = neighborhood density; PoS = part of speech; BiPhono = biphone positional probability; SPhono = single-phone positional probability; PrevMen = previous mention; RateA = speech rate following the target; RateB = speech rate preceding the target; Sex = talker sex (see text).

Table 4. Summary of the model of word durations.

Variable name	beta	SE	t	p_{MCMC}	AIC	Chisq	p(Chisq)
(Intercept)	0.1404	0.0295	4.759	0.0001	3652.1		
Frequency	-0.0281	0.0044	-6.407	0.0001	3575.2	78.86 (1)	<.0001
PoS					3538.9	42.28 (3)	<.0001
Adverb	-0.072	0.0653	-1.101	0.2334			
Noun	0.0202	0.02	1.009	0.2036			
Verb	-0.0896	0.0206	-4.357	0.0001			
BaselineDur	0.6442	0.0525	12.266	0.0001	3406.8	134.11 (1)	<.0001
Bigr_After	-0.0249	0.0014	-17.814	0.0001	3159.8	249.03 (1)	<.0001
Bigr_Bef	-0.0149	0.0016	-9.539	0.0001	3081.6	80.19 (1)	<.0001
Rate_After	-0.1382	0.0079	-17.514	0.0001	2756.2	327.39 (1)	<.0001
Rate_Bef	-0.0864	0.0075	-11.533	0.0001	2627.7	130.48 (1)	<.0001
Rate_Bef^2	-0.0263	0.011	-2.389	0.0156	2624	5.67 (1)	.0172
Neighb.Density	-0.0044	0.0009	-5.084	0.0001	2600.6	25.42 (1)	<.0001

Table 5. Random effects in the model of word durations.

Random effect	SD	MCMC median	HPD95lower	HPD95upper
Word (Intercept)	0.0983	0.0829	0.0750	0.0911
Speaker (Intercept)	0.0897	0.0874	0.0698	0.1096
Residual	0.2621	0.2632	0.2598	0.2666

Table 6. Summary statistics for the outcome variable and the numerical predictors in the model of vowel dispersion

	Median	Mean (SD)	Range
Degree of dispersion	-0.02	0.00 (1.0)	-3.9– 9.0
Bigram probability (Preceding)	.005	.026 (.079)	7.89e-5 - .75
Bigram probability (Following)	.005	.031 (.079)	7.90e-5 - .83
Consonant duration	137.8	145.0 (55.8)	0.0 – 632.1
Frequency	523.10	767.90 (699.04)	0.43 – 2610.0
Neighborhood density	21	21.15 (6.96)	3 – 40
Frequency-weighted neighborhood density	44.3	44.48	4.4-92.00
Orthographic length	4	4.005 (0.72)	3 -7
Phonotactic probability			
Single-phoneme probability	.049	.049 (.016)	.012 – .098
Biphone	.002	.003 (.002)	.000 - .016

probability			
Speech rate (Preceding) (ms/syl)	5.94	6.24 (2.29)	0.95 – 33.33
Speech rate (Following)	5.23	5.31 (1.68)	0.88 – 41.0
Vowel duration (ms)	92	103 (0.05)	25 – 490

Table 7. Summary statistics for categorical variables in the vowel dispersion database

Vowel type	[ɑ]: 1,193
	[æ]: 824
	[ɛ]: 1,263
	[eɪ]: 1,341
	[ɪ]:1,555
	[i]: 828
	[o]: 788
	[ʊ]: 918
	[u]: 365
Manner of articulation	Approximant ([l], [j], [w], [r]) : 1,643
(Preceding)	Nasal ([m], [n], [ŋ]): 1,092
	Obstruent (oral stop, fricative, affricate): 6,340
Manner of articulation	Approximant ([l], [j], [w], [r]) : 1,653
(Following)	Nasal ([m], [n], [ŋ]): 1,401
	Obstruent (oral stop, fricative, affricate): 6,021
Place of articulation	Front (bilabial, alveolar, labial dental, labial-alveolar): 7,137
(Preceding)	
	Back (velar, glottal): 1,938
Place of articulation	Front (bilabial, alveolar, labial dental, labial-alveolar): 6,643

(Following)	Back (velar, glottal): 2,432
Speaker sex	Female: 4,434 Male: 4,641
Speaker age	Young: 4,177 Old: 4,898
Part of speech	Adverb: 483
	Adjective: 1,994
	Noun: 2,618
	Verb: 3,980
Previous mention	True: 6,423
	False: 2,652

Table 8. Pairwise (Spearman) correlations between variables in the model of vowel dispersion

	Disp	BigrB	BigrA	CDur	Fq	ND	NDFq	Len	SPhon	BiPhon	RatB	RatA
Dispersion	1	0	.02	.08	-.07	.02	.04	.04	.07	-.01	0	-.03
BigrB	0	1	.01	-.07	.15	-.11	.01	.09	-.05	.02	-.02	.02
BigrA	.02	.01	1	-.05	.08	.06	.03	-.04	.03	.01	.03	-.06
CDur	.08	-.07	-.05	1	-.14	0	-.01	.02	.05	.07	-.09	-.1
Frequency	-.07	.15	.08	-.14	1	-.1	-.06	.1	-.15	-.08	.04	.05
ND	.02	-.11	.06	0	-.1	1	.76	-.22	.4	.28	-.02	-.01
NDFq	.04	.01	.03	-.01	-.06	.76	1	-.23	.6	.47	-.02	0
Len	.04	.09	-.04	.02	.1	-.22	-.23	1	-.41	-.28	-.01	.02
SPhon	.07	-.05	.03	.05	-.15	.4	.6	-.41	1	.56	0	-.01
BiPhon	-.01	.02	.01	.07	-.08	.28	.47	-.28	.56	1	0	0
RateBef	0	-.02	.03	-.09	.04	-.02	-.02	-.01	0	0	1	.08
RateAft	-.03	.02	-.06	-.1	.05	-.01	0	.02	-.01	0	.08	1
VDur	.03	0	-.06	.1	-.09	.08	.03	0	-.04	-.01	-.09	-.07
Vtype	0	-.07	-.02	.01	-.11	-.13	-.28	.03	-.16	-.27	-.01	-.03
MannerB	.06	.07	-.05	.09	.07	-.08	.02	-.02	.23	.24	.02	.03
MannerA	-.14	-.08	-.09	-.04	.23	-.03	-.08	-.17	-.08	-.11	.02	.01
PlaceB	-.18	.05	-.02	-.01	.03	-.07	-.02	-.02	-.18	-.02	-.03	.01
PlaceA	.03	-.09	-.05	.01	-.36	-.05	.08	-.15	.31	.17	0	-.03
Sex	-.05	0	0	0	-.02	.01	.02	-.02	.01	.02	.03	0
Age	0	-.02	.01	-.01	-.04	.01	0	-.01	0	-.03	0	-.05
PoS	-.05	-.23	-.15	.07	-.04	.08	.06	0	-.07	-.2	.02	.02
PrevMen	.02	-.08	-.03	.09	-.26	.12	.08	0	.06	.03	-.02	-.01

	VDur	Vtype	MnrB	MnrA	PlB	PlA	Sex	Age	PoS	PrevMen
Dispersion	.03	0	.06	-.14	-.18	.03	-.05	0	-.05	.02
Bi_Bef	0	-.07	.07	-.08	.05	-.09	0	-.02	-.23	-.08
Bi_Aft	-.06	-.02	-.05	-.09	-.02	-.05	0	.01	-.15	-.03
CDur	.1	.01	.09	-.04	-.01	.01	0	-.01	.07	.09
Frequency	-.09	-.11	.07	.23	.03	-.36	-.02	-.04	-.04	-.26
ND	.08	-.13	-.08	-.03	-.07	-.05	.01	.01	.08	.12
NDFq	.03	-.28	.02	-.08	-.02	.08	.02	0	.06	.08
Len	0	.03	-.02	-.17	-.02	-.15	-.02	-.01	0	0
SPhon	-.04	-.16	.23	-.08	-.18	.31	.01	0	-.07	.06
BiPhon	-.01	-.27	.24	-.11	-.02	.17	.02	-.03	-.2	.03
RateB	-.09	-.01	.02	.02	-.03	0	.03	0	.02	-.02
RateA	-.07	-.03	.03	.01	.01	-.03	0	-.05	.02	-.01
VDur	1	-.16	.03	.06	.06	0	.01	-.03	-.11	.05
Vtype	-.16	1	-.08	.03	-.15	.14	-.02	.01	.09	.01
MannerB	.03	-.08	1	.04	-.34	.04	.02	-.01	-.11	-.06
MannerA	.06	.03	.04	1	.17	-.17	-.02	-.02	.15	-.01
PlaceB	.06	-.15	-.34	.17	1	-.29	0	-.02	.01	.01
PlaceA	0	.14	.04	-.17	-.29	1	0	.01	-.03	.13
Sex	.01	-.02	.02	-.02	0	0	1	.08	-.02	.01
Age	-.03	.01	-.01	-.02	-.02	.01	.08	1	-.02	.05
PoS	-.11	.09	-.11	.15	.01	-.03	-.02	-.02	1	.09
PrevMen	.05	.01	-.06	-.01	.01	.13	.01	.05	.09	1

Note. Disp = vowel dispersion; BigrA = Bigram probability of the target word, given the following word; BigrB = Bigram probability of the target word, given the previous word; CDur = consonant duration; Frequency = Fq = target word frequency; ND = neighborhood density; NDFq = frequency-weighted neighborhood density; Len = orthographic length; SPhon = single-phone positional probability; BiPhon = biphone positional probability; RateB = speech rate preceding the target; RateA = speech rate following the target; VDur = vowel duration; Vtype = vowel type; MannerB = manner of articulation of consonant preceding the target vowel; MannerA = manner of articulation of consonant following the target vowel; PlaceB = place of articulation of consonant preceding the target vowel; PlaceA = place of articulation of consonant following the target vowel; Sex = talker sex; Age = talker age; PoS = part of speech; PrevMen = previous mention of target (see text).

Table 9. Summary of fixed effects in the model of vowel dispersion

Variable name	beta	SE	t	p_{MCMC}	AIC	Chisq	p(Chisq)
(Intercept)	0.3681	0.1024	3.596	0.0001	22895		
PlaceBeforefront	-0.4002	0.099	-4.042	0.0001	22883	14.24	0.0002
BigramBefore	-0.0125	0.0058	-2.156	0.0254	22876	8.91	0.0028
CDur	0.1797	0.0264	6.813	0.0001	22790	87.42	<.0001
SpeechRateAfter	-0.1026	0.029	-3.535	0.0004	22777	15.12	0.0001
VDur	0.1747	0.024	7.277	0.0001	22725	53.91	<.0001
Neighborhood density	-0.0086	0.0051	-1.695	0.0388	22723	4.91	0.0268
Neighborhood density, squared	-0.0015	0.0006	-2.687	0.0002	22717	7.12	0.0076

Table 10. Random effects in the model of vowel dispersion

Random effect	SD	MCMC median	HPD95lower	HPD95upper
Word (Intercept)	0.618	0.4286	0.3963	0.4628
Speaker (Intercept)	0.261	0.2538	0.2031	0.3193
Residual	0.807	0.8185	0.8064	0.8310

Table 11: Summary of simple linear regression models relating neighborhood density and phonotactic probability (N = 534)

Model	β (SE β)	R^2	Quantity represented by model residuals
ND ~ SPhono	13.77 (.75)	.39	rNDS = Variability in neighborhood density not attributable to single-phone positional probability
ND ~ BiPhono	6.33 (.39)	.33	rNDBi = Variability in neighborhood density not attributable to biphone positional probability
SPhono ~ ND	.03 (.002)	.39	rSPhono = Variability in single-phone positional probability not attributable to neighborhood density
BiPhono ~ ND	.05 (.003)	.33	rBiPhono = Variability in biphone positional probability not attributable to neighborhood density

Figure 1: Vowel space of a talker (s26) in the Buckeye corpus. Each observation represents F1 and F2 measurements for a single vowel token. Labels indicate the words the vowel token occurred in. The plus sign marks the center of the talker's F1/F2 space.

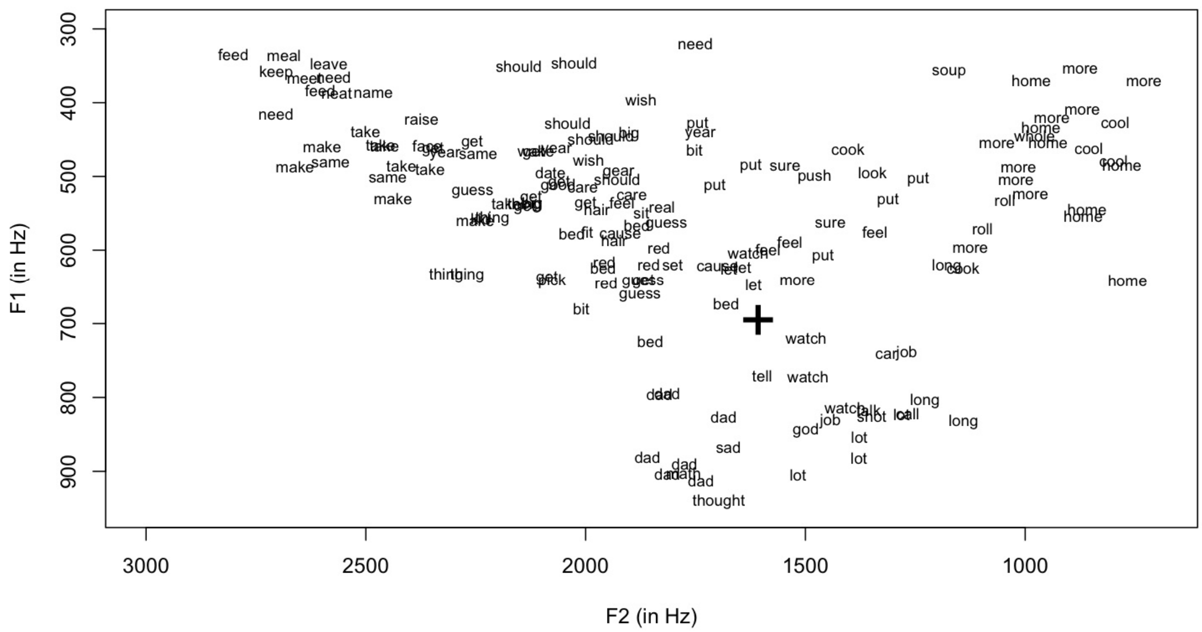

Figure 2: Partial effects, Word duration model

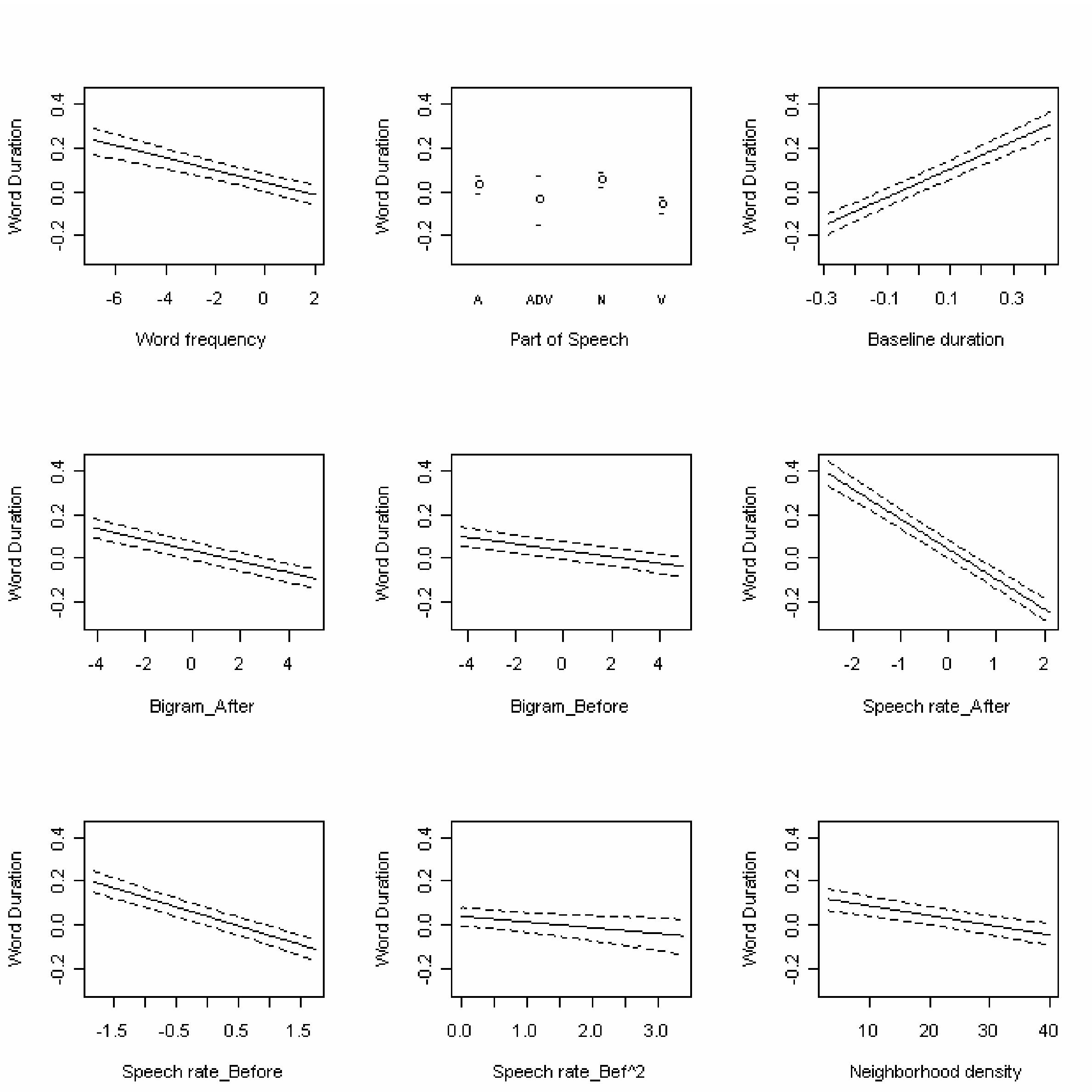

Figure 3: Partial effects, vowel dispersion model