Published in final edited form as:

J Mem Lang. 2015 August 1; 83: 97–117. doi:10.1016/j.jml.2015.04.002.

Phonological Neighborhood Competition Affects Spoken Word Production Irrespective of Sentential Context

Neal P. Fox^a, Megan Reilly^a, and Sheila E. Blumstein^{a,b}

Megan Reilly: megan_reilly@brown.edu; Sheila E. Blumstein: sheila_blumstein@brown.edu

^aBrown University, Department of Cognitive, Linguistic & Psychological Sciences; Box 1821; Providence, RI, USA 02912

^bBrown University, Brown Institute for Brain Science; Box 1953; Providence, RI, USA 02912

Abstract

Two experiments examined the influence of phonologically similar neighbors on articulation of words' initial stop consonants in order to investigate the conditions under which *lexically-conditioned phonetic variation* arises. In Experiment 1, participants produced words in isolation. Results showed that the voice-onset time (VOT) of a target's initial voiceless stop was predicted by its overall neighborhood density, but not by its having a voicing minimal pair. In Experiment 2, participants read aloud the same targets after semantically predictive sentence contexts and after neutral sentence contexts. Results showed that, although VOTs were shorter in words produced after predictive contexts, the neighborhood density effect on VOT production persisted irrespective of context. These findings suggest that global competition from a word's neighborhood affects spoken word production independently of contextual modulation and support models in which activation cascades automatically and obligatorily among all of a selected target word's phonological neighbors during acoustic-phonetic encoding.

Keywords

speech production; voice-onset time (VOT); phonological competition; neighborhood density; sentence context; hyperarticulation

1. Introduction

The phonetic realization of a segment varies considerably depending on many factors, including properties of the word in which the segment occurs. Investigating the sources of such systematic variation can reveal features of the cognitive architecture underlying spoken language production. Recent research has explored the ways in which a word's acoustic properties are influenced by lexical-level information such as its frequency or its relation to

Corresponding author: Neal P. Fox, Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Box 1821, Providence, RI 02912, Phone: (401) 863-2727, Fax: (401) 863-2255, neal_fox@brown.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof, before it is published in its final citable form. Please note that during the production, process errors may be discovered which could affect the content, and all legal disclaimers, that apply to the journal pertain.

other words, a phenomenon known as *lexically-conditioned phonetic variation*. For instance, Baese-Berk and Goldrick (2009) showed that words that have word-initial voicing minimal pairs (MP words; e.g., *cape*, because *gape* is an English word) were produced with significantly longer voice-onset times (VOTs) for their initial stops compared to words that do not have word-initial voicing minimal pairs (NMP words; e.g., *cake*, **/geɪk/* is not an English word). This result exemplifies a body of work indicating that how a target word is pronounced depends on its phonological relationships to other words in the lexicon; the mere existence of similar words in a language can influence acoustic dimensions of a spoken target (Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010, 2012, 2013; Wright, 2004).

While this general class of findings is robust, more work is needed to delineate the parameters that govern the influence of lexical-level information on acoustic-phonetic encoding during spoken word production. Put another way, it is clear that phonologically similar neighbors influence the production of a target word, but less well understood are the conditions that determine whether and/or when neighbors can influence different aspects of spoken word production. A better understanding of this issue would provide important constraints for models of spoken word production. The current study aimed to examine this issue by probing two questions regarding the conditions under which phonological neighbors can influence a target word's initial voicing.

Firstly, to what extent is a neighbor's influence on a target's production contrast-specific? In particular, can the degree of lengthening of the VOT of the initial /k/ in *cape* (cf. Baese-Berk & Goldrick, 2009) be attributed to the existence of a voicing contrast in the lexicon (*gape*), or, alternatively, to the combined influence of its entire neighborhood (*gape*, *tape*, *cane*, *cope*...)? If VOT is associated with the presence or absence of a minimal pair competitor that contrasts on initial voicing, this could lend support to models of spoken word production in which articulatory implementation of a target word is highly dependent on its particular phonological relationships to each of its neighbors. On the other hand, the latter case would support models in which all phonologically related neighbors (rather than just a particularly proximal competitor) contribute to the processes that drive variation in the articulation of voicing.

Secondly, to what extent is a neighbor's influence on a target's production context-sensitive? The phonological relationships between any given target word and its neighbors might be considered a relatively static influence on the target's articulation – for adult speakers of a language, the network structure of the lexicon is relatively stable. However, speech production outside the laboratory occurs within the context of dynamic patterns of language use. When speakers produce a word in context, how might that context modulate a neighbor's influence on the target's ultimate acoustic realization? One possibility is that a predictive semantic context (e.g., *Superman tried to fly away but the phone booth door shut on his long red...*) will boost the activation of a target (*cape*) relative to less contextually supported phonological competitors (e.g., *gape*, *tape*), or the context might reduce or restrict activation of words that are unlikely, semantically incongruent, or ungrammatical. In such a case, the influence of phonologically similar neighbors might be reduced (or even eliminated) in a highly semantically predictive context because unlikely neighbors are not

competing as strongly. Alternatively, if contextual cues do not interfere with the dynamics of phonological neighborhood activation patterns during production processes, then neighborhood-conditioned influences on phonetic variation will remain strong irrespective of the context. That is, even if a target is highly predictable in context, the influence of the word's phonological neighbors on its acoustic realization may persist.

The present work investigates this pair of questions with two experiments. Following Baese-Berk and Goldrick (2009) and others (Bullock-Rest et al., 2013; Fricke, 2013; Goldinger & Summers, 1989; Goldrick, Vaughn & Murphy, 2013; Kilanski, 2009; Kirov & Wilson, 2012; Peramunage, Blumstein, Myers, Goldrick & Baese-Berk, 2011), we examined how lexical and contextual information could influence the production of voiceless stop consonants' VOTs.

1.1. Lexically-Conditioned Phonetic Variation

Experimental findings and corpus studies support the view that whole-word properties of a lexical item selected for production and its formal (phonological) relationships with other words in the lexicon can influence its acoustic-phonetic processing and production (for review, see Sadat, Martin, Costa & Alario, 2014). For instance, more frequent words are produced with shorter durations than less frequent words, their vowel spaces are more spectrally reduced (i.e., less dispersed from the center of a speaker's vowel space), and deletion of their final segment is more common (Aylett & Turk, 2004, 2006; Bell, Brenier, Gregory, Girand & Jurafsky, 2009; Fidelholz, 1975; Gahl, 2008; Jurafsky, Bell & Girand, 2002; Jurafsky, Bell, Gregory & Raymond, 2001; Munson & Solomon, 2004). Other lexical characteristics, such as grammatical class, also drive phonetic reduction: closed-class function words are more acoustically reduced compared to content words (Bell et al., 2009). Thus, information encoded in the lexicon about a "whole word," rather than only information about the segments that make up its abstract phonological form, can influence the articulatory encoding of segments within that word.

Production of the segments within a word is not only influenced by lexical-level information about the target word itself; the production of a word's component segments can also be influenced by information encoded about its phonological relationships with other words in the lexicon. Such relationships between a target and other lexical items can be characterized by measures of *neighborhood density*. Here, a word's neighborhood is defined by the set of words that differ from the target via the substitution, addition or deletion of a single segment in any position (Luce & Pisoni, 1998), and a word's density is computed by summing the total size of its neighborhood (Vitevitch & Luce, 1999; Vitevitch, 2002).

Studies examining the relationship between phonological neighborhood density and acoustic characteristics of spoken words have generally shown that the F1/F2 vowel space associated with words from sparse neighborhoods is reduced compared to the vowel space associated with words from denser neighborhoods (Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010, 2012, 2013; Wright, 2004; but cf. Gahl, Yao & Johnson, 2012). These results suggest that the presence of phonologically similar neighbors in the lexicon is associated with *hyperarticulation* (i.e., production with more extreme phonetic properties) of a target's vowel. Nonetheless, it remains unclear to what extent such neighborhood-driven

hyperarticulation is dependent on specific phonological relationships between targets and neighbors.

1.2. Lexically-Conditioned Phonetic Variation: Contrast-Specificity?

As noted earlier, Baese-Berk and Goldrick (2009) examined lexically-conditioned phonetic variation in VOTs of word-initial voiceless stop consonants, showing that initial stops of MP words were produced with longer VOTs than those of NMP words. They proposed that their experiments could extend previous findings on lexically-conditioned phonetic variation by “examining words with a minimal pair neighbour rather than effects of neighbourhood density more generally” (p. 530). Baese-Berk and Goldrick’s results, then, could be taken as evidence for highly specific lexically-conditioned effects in speech production. That is, the results suggest that the kinds of acoustic effects a competitor has on the production of a target is a function of the nature of the formal relationship between those lexical items: existence of a voicing competitor in the lexicon appeared to modulate subjects’ articulatory implementation of voicing.

However, given that more general effects of neighborhood density have been noted for other acoustic parameters (e.g., vowel dispersion; Wright, 2004), attributing the modulation of this particular acoustic parameter (VOT) to a unique, more specific mechanism (the presence of one phonologically related competitor, a voicing MP) may be premature. If the observed VOT effects could be explained by more general properties of the lexicon (i.e., phonological neighborhood density), then it would challenge the view that hyperarticulatory effects are contrast-specific and link a particular type of phonetic variation to the presence of a particular local phonemic contrast. In fact, neither Baese-Berk and Goldrick’s study nor other studies using their stimuli (Bullock-Rest et al., 2013; Peramunage et al., 2011) were designed to examine this issue since the MP stimuli were from denser neighborhoods than the NMP stimuli.¹

The present study’s Experiment 1 aimed to tease apart these two possible sources of hyperarticulation by directly testing the influence of the existence of an initial voicing minimal pair and the influence of a word’s neighborhood density. If the VOTs of words’ initial stop consonants are found to vary as a function of lexical density but not the presence of a voicing minimal pair neighbor, then this pattern of results would suggest that global characteristics of a word’s phonological similarity network drive generalized hyperarticulatory effects in spoken word production. In such a case, even the presence of non-voicing competitors would influence the articulatory implementation of voicing during spoken word production.

1.3. Lexically-Conditioned Phonetic Variation: Context-Sensitivity?

In addition to the aforementioned evidence for lexically-conditioned and neighborhood-conditioned effects on the acoustic properties of speech, there are also well-known effects of

¹Although Baese-Berk and Goldrick (2009) do not directly report neighborhood density values for their stimuli, later studies employing their stimuli (Peramunage et al., 2011) acknowledge this fact. Using the stimulus list in their Appendix, density values were obtained from IPhOD (Vaden, Halpin & Hickok, 2009) and submitted to two-tailed paired *t*-tests, confirming that their MP and NMP lists differed significantly in their densities ($NHD_{MP} > NHD_{NMP}$) for stimuli in the /p/ and /t/ places of articulation (each $p < 0.02$) and combining all of their stimuli ($p < 0.0001$).

linguistic context on the acoustic realization of words. A word that is highly likely given its preceding context will tend to exhibit a reduced vowel space and a shorter duration (and be less intelligible) than that same word in a less predictive context (Aylett & Turk, 2004; Bell et al., 2009; Bolinger, 1963; Bradlow & Alexander 2007; Buz & Jaeger, 2012; Chafe, 1974; Clopper & Pierrehumbert, 2008; Gahl & Garnsey, 2004, 2006; Gahl, Yao & Johnson, 2012; Hawkins & Warren, 1994; Hunnicut, 1985; Jurafsky et al., 2001; Kuperman & Bresnan, 2012; Lieberman, 1963; Lindblom, 1990; Pickett & Pollack, 1963; Scarborough, 2010; Tily et al., 2009).

Thus, phonetic variation in spoken word production can be driven both by the presence of phonological neighbors and by linguistic context. However, relatively little attention has been paid to the question of whether these factors interact; in other words, is neighborhood-conditioned phonetic variation modulated by context? In particular, will the semantic and syntactic constraints of a sentence context eliminate or alter the production effects that a target's phonological neighbors have in isolation? Relatedly, will any such effects on word production in context be modulated by how semantically predictive or constraining the context is?

Such context-sensitive modulation of neighborhood-conditioned effects on spoken word production is compatible with two classes of models. The first are those models in which speakers' utterances are optimal in their balance of efficiency and informativity (e.g., Aylett & Turk, 2004; Levy & Jaeger, 2007; Lindblom, 1990; Jaeger, 2010). Under such models, hyperarticulation (e.g., lengthening of a voiceless stop's VOT; Smiljanic & Bradlow, 2005) should fail to emerge if that expenditure of effort is unlikely to facilitate disambiguation of the target from a particular competitor. In a context that is highly predictive of a target word, then, hyperarticulation of the target should be suboptimal. A second class of models that would predict context-sensitivity of neighborhood-conditioned effects are those in which only semantically plausible or grammatically appropriate words are accessed (see, e.g., Dell, Oppenheim & Kittredge, 2008; Mahon, Costa, Peterson, Vargas & Caramazza, 2007; Rodriguez-Ferreiro, Davies & Cuetos, 2014). For instance, in one model, the selection of a lexical item for production extinguishes competition from neighbors that are syntactically inappropriate continuations of the sentence (Dell, Oppenheim & Kittredge, 2008). In contrast, neither of these model architectures (in their strictest forms) could explain the emergence of similar patterns of neighborhood-conditioned phonetic variation irrespective of context. Rather, such results would suggest that a target's phonological neighbors are accessed automatically and obligatorily during production (i.e., independent of their relative likelihoods given the target word's context).

Recent studies testing for an interaction between context and neighborhood-conditioned phonetic variation have not yielded clear support for contextual modulation of neighborhood-driven effects on spoken word production (Heller, 2014; Heller & Goldrick, 2014, 2015; Scarborough, 2010), but a number of issues complicate the interpretation of these results. In one study (Scarborough, 2010), words from either dense or sparse neighborhoods (e.g., *home* vs. *stone*) were read aloud after sentence contexts in which the targets were either semantically predictable or unpredictable (e.g., *There's no place like [HOME]* vs. *I think Paul has decided to buy a [HOME]*). Scarborough (2010) found that

high neighborhood density and low predictability each independently predicted greater hyperarticulation (longer and more dispersed vowels), but found no evidence for an interaction between these variables, thus concluding that neighborhood-conditioned effects were just as strong when the target word was in a predictive context as a neutral one. However, Flemming (2010) raises several concerns about the stimuli and analyses presented by Scarborough (2010). For instance, given that all of the low-density stimuli and none of the high-density stimuli in Scarborough's study had onset clusters (e.g., *plump* vs. *pump*), the fact that obstruent-liquid clusters (which compose 75% of Scarborough's high-density list) tend to shorten subsequent vowels compared to singleton onsets (van Santen, 1992) may present a confound for the reported effect of density on vowel duration (*cf.* studies with only CVC words that have not found vowel duration effects of density: e.g., Wright, 2004; Munson & Solomon, 2004). Moreover, Flemming (2010) argues that Scarborough's reliance on (potentially non-linearly) correlated measurements of vowel dispersion and duration without appropriate phonetic and statistical controls could obscure her results, especially when testing for an interaction between density and context.

These critiques notwithstanding, studies by Heller (2014) and Heller and Goldrick (2014, 2015) that included phonetic and statistical controls also found no evidence for an interaction between neighborhood-conditioned competition and contextual constraints. However, while these studies all failed to find evidence for contextual modulation of neighborhood effects on the dispersion and duration of spoken words' vowels, that is where the similarities in their results end. While Scarborough (2010) found more extreme articulations for high-density words (see also, e.g., Munson, 2007; Munson & Solomon, 2004; Wright, 2004; Scarborough & Zellou, 2013), Heller (2014) found the opposite result (see also, e.g., Gahl, Yao & Johnson, 2012; Heller & Goldrick, 2014). As a further complication, although Scarborough (2010) found less extreme articulations for words that are more predictable given their context (see also, e.g., Bell et al., 2009; Gahl, Yao & Johnson, 2012; Heller, 2014), Heller and Goldrick (2014) found no evidence for phonetic reduction in more constrained contexts.

It is difficult to interpret these divergent patterns of results, but such considerable differences raise questions about the status of evidence for or against the interaction of neighborhood density and context in spoken word production. Differences in how studies operationalize neighborhood competition (e.g., overall neighborhood density vs. within-syntactic category density) and context (e.g., word-reading in isolation vs. picture-naming in isolation vs. word-reading in context vs. spontaneous speech production; for review, see Chen & Mirman, 2012) may represent a partial explanation of these disparate effects on the same acoustic variables. However, such inconsistencies between studies also recommend caution in drawing conclusions based on those studies' mutual lack of evidence for contextual modulation of neighborhood competition.

Given the lack of clarity regarding potential contextual modulation of neighbors' effects on target word production, Experiment 2 was designed to examine whether a sentential context interacts with either local or global phonological neighborhood structure to drive systematic variation in spoken word production. To this end, we investigated the effects of contextual information on VOTs of the same words' initial stop consonants by embedding Experiment

l's stimuli in sentence contexts that were highly semantically constraining (thereby making the target words highly predictable continuations) or less semantically constraining (thereby making the target words less predictable continuations).

Finally, it bears noting that nearly all laboratory studies of lexically-conditioned phonetic variation have divided their stimuli into binary splits of low- and high-density words (*cf.*, e.g., Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010, 2012, 2013; Scarborough & Zellou, 2013; Wright, 2004; Heller, 2014; Heller & Goldrick, 2014). However, this treatment ignores the fact that density is a continuous variable. At worst, this could result in insufficient detection power (Cohen, 1983), thereby complicating interpretations of the null interactions discussed above. Thus, unlike prior work, we consider lexical density a continuous variable in the current study and use words' actual density values in all analyses.

To summarize, Experiments 1 and 2 were designed to investigate the conditions under which a word's phonological neighbors can influence the articulation of voicing during spoken word production. Experiment 1 serves as a test of contrast-specificity in lexically-conditioned phonetic variation, examining whether phonological neighbors that differ from a target along dimensions other than word-initial voicing might still drive variation in word-initial voicing. Experiment 2 serves as a test of context-sensitivity in lexically-conditioned phonetic variation, examining whether neighbors still drive variation in word-initial voicing even when the target word is highly likely given the preceding sentence context.

2. Experiment 1

2.1. Methods

2.1.1. Materials—Stimuli included 12 pairs of monosyllabic nouns beginning with voiceless stops (4 pairs beginning with each of: /p/, /t/ and /k/). All pairs were matched on their initial consonant and subsequent vowel. As in previous studies (Baese-Berk & Goldrick, 2009; Bullock-Rest et al., 2013; Peramunage et al., 2011), one member of each pair (e.g., *coat*) had a minimal pair differing only in voicing of the initial stop consonant (e.g., *goat*) while the other member of the pair (e.g., *comb*) did not (e.g., **/gom/*). Following previous studies, we refer to the former set of 12 words as MP targets and the latter set of 12 words as NMP targets. Appendix A contains a list of the target stimuli.

Pairs were matched for frequency (calculated using the SUBTLEXUS database; Brysbaert & New, 2009), length (in phonemes), and several measures of phonotactic probability (mean segmental probability, sum segmental probability, and mean bigram frequency, calculated using the Phonotactic Probability Calculator; Vitevitch & Luce, 2004). Crucially, and in contrast to previous studies, MP and NMP words were also matched for neighborhood density (an unweighted density measure and a frequency-weighted measure, computed according to Vitevitch & Luce, 1999; retrieved using the IPhOD database of Vaden, Halpin & Hickok, 2009). Table 1 provides a summary of the parameters used to match MP and NMP words. Additionally, MP words did not differ in frequency or neighborhood density from their voiced competitors. That is, *coat*-type targets (MP words) did not differ from

goat-type non-target competitors. These voiced minimal pair neighbors of the target words (e.g., *goat*) never appeared during the experiment.

As noted earlier, we employed neighborhood density as a continuous variable. This element of our design stands in contrast to nearly all published laboratory studies of lexically-conditioned phonetic variation, which have divided stimuli into binary splits of low- and high-density words (or, e.g., low/high frequency; easy/difficult). A continuous operationalization of neighborhood density provides greater detection power (Cohen, 1983) and more accurately characterizes the variable of interest. Overall, this design allowed us to examine the effects of two independent variables on the VOTs of initial stop consonants:

NEIGHBORHOOD-DENSITY, a continuous factor indexing global characteristics of a target's phonological neighborhood structure, and COMPETITOR-STATUS, a discrete factor indicating the presence or absence of a lexical item differing from a target on word-initial voicing.

In addition to the critical target words, the stimulus materials included 24 monosyllabic filler nouns beginning with either voiced stops (e.g., *book*, *gun*) or consonants other than stops (e.g., *fish*, *hand*). These filler targets (listed in Appendix A) did not differ in length, frequency, or neighborhood density from either the MP or NMP targets. Only one of the 24 filler targets (*sea*) had a minimal pair that contrasted in voicing (*zee*).

2.1.2. Task and Recording—Individual words were presented to participants on a computer monitor in a sound-dampened booth at a comfortable viewing distance. Each word appeared for 2.5 seconds during which the subject could produce the word. This period was followed by a 1.5-second inter-trial interval. Subjects did not have difficulty producing the target in the time provided; the average time between visual onset of the target and speech onset was 544 ms (sd = 128 ms). Each of the 48 target words (12 MP targets + 12 NMP targets + 24 filler targets) was presented three times in random order, yielding 144 word presentations. An AKG D900 microphone and Edirol digital recorder (model R09-HR; sampled: 44,100 Hz / 24 bits / mono) were used to record subjects' speech, and BLISS speech-editing software (Mertus, 1989) was used to resample the signal prior to phonetic analysis (22,050 Hz / 16 bits / mono). Because of slight differences in recording volumes between subjects, each subject's maximum volume level was scaled to a consistent maximum volume prior to analysis.

2.1.3. Participants—Twenty members of the Brown University community were paid for their participation. Participants were self-reported right-handed monolingual native speakers of American English with no hearing deficit or history of neurological disorder. All were at least 18 years old and gave written informed consent in accordance with Brown University Institutional Review Board policies.

2.2. Acoustic Analysis

A computer-automated method for measuring VOT was designed and implemented in BLISS (Mertus, 2014). Details of the analysis procedure, software validation methodology, and data inclusion criteria are described in Appendix B. According to these procedures, 1418 out of 1440 trials (>98%) across all 20 subjects were included in the dataset. Overall, our mean VOT values were 86 ms for /p/ targets, 95 ms for /t/ targets, and 96 ms for /k/

targets. These were somewhat higher than those reported by Baese-Berk and Goldrick (2009) (67 ms, 82 ms, and 93 ms respectively).²

2.3. Statistical Analysis

All tokens that met the inclusion criteria as described in Appendix B were used in the primary analyses, which employed linear mixed effects regression (**lmer**). Appendix C motivates and describes our implementation of this statistical technique (e.g., coding of fixed effects, random effects structures, computation of p-values).

In a second set of non-parametric analyses (henceforth: *type analyses*), we pre-processed the data according to the same technique used in previous studies (cf. Baese-Berk & Goldrick, 2009; Peramunage et al., 2011): all of a subject's tokens of the same target that met the inclusion criteria were averaged to yield a single mean for that subject/word. Because the VOTs were not normally distributed (nor were our model residuals), these analyses also had the benefit of confirming the key results of the omnibus **lmer** analyses. Appendix C provides a detailed discussion of this issue. For simplicity when discussing the type analyses, we refer to the mean VOT for a given cell (subject/word pair) as a VOT.

2.4. Results

Results are displayed in Figure 1. To examine the effects of *NEIGHBORHOOD-DENSITY* (a continuous variable), *COMPETITOR-STATUS* (a categorical variable), and their interaction (*COMPETITOR-STATUS* × *NEIGHBORHOOD-DENSITY*) on VOT, we specified a **lmer** model with these three fixed effects. A significant main effect was found for *NEIGHBORHOOD-DENSITY* ($\beta = 0.283$, $SE = 0.092$, $|t| = 3.09$, $\chi^2(1) = 8.99$, $p < 0.003$). As Figure 1 shows, the effect of *NEIGHBORHOOD-DENSITY* was positive: the denser the neighborhood in which a word resides, the longer the VOT of its initial stop. There was also an interaction between *COMPETITOR-STATUS* and *NEIGHBORHOOD-DENSITY* ($\beta = -0.412$, $SE = 0.162$, $|t| = 2.55$, $\chi^2(1) = 5.52$, $p < 0.02$) such that the VOTs of NMP words' initial stops were more strongly affected by neighborhood density than MP words. We found no evidence for a main effect of *COMPETITOR-STATUS* on VOT ($|t| < 1$, $\chi^2(1) = 0.03$, $p > 0.85$).

A follow-up analysis included *TRIAL* as a factor in the model (along with all of its interactions with the model's other fixed effects) to examine whether repeating each target word three times influenced the VOTs subjects produced (cf. Fowler & Housum, 1987). The results were nearly identical, with significant effects of *NEIGHBORHOOD-DENSITY* ($\beta = 0.276$, $SE = 0.092$, $|t| = 3.00$, $\chi^2(1) = 8.60$, $p < 0.004$) and a *COMPETITOR-STATUS* × *NEIGHBORHOOD-DENSITY* interaction ($\beta = -0.425$, $SE = 0.172$, $|t| = 2.48$, $\chi^2(1) = 5.35$, $p < 0.03$). There was also an effect of *TRIAL* ($\beta = 7.563$, $SE = 2.931$, $|t| = 2.58$, $\chi^2(1) = 6.15$, $p < 0.02$) such that VOTs tended to lengthen over the course of the experiment by about 8 ms. No effect of *COMPETITOR-STATUS* was detected ($|t| < 1$), and *TRIAL* did not interact with any fixed effects (all $|t| < 1$).

²Baese-Berk and Goldrick (2009) reported mean VOTs for MP and NMP lists at each place of articulation (POA). To compare their VOTs at each POA to our data, we averaged their MP and NMP means at each POA and computed values for our data in the same way: (a) compute mean VOT for each subject/word; (b) average over subjects for each word, (c) compute average VOT for MP and NMP lists at each POA, (d) average the MP and NMP means at each POA.

Because the VOTs and model residuals were not normally distributed (see Appendix C), the data were subjected to nonparametric type analyses. First, Wilcoxon matched-pairs sign rank tests compared the VOTs in MP and in NMP words with both an F1-analysis (by-subject) and an F2-analysis (by-item). As shown in Table 2, and as was observed in the **lmer** analyses above, results failed to show a significant difference in VOT as a function of **COMPETITOR-STATUS**. Next, we used Spearman's correlation test to evaluate the 24 words for a relationship between **NEIGHBORHOOD-DENSITY** and VOT (the mean VOT of the 20 subjects), obtaining a significant correlation (Spearman's $\rho = 0.47$; $p = 0.02$).

Recall that items (e.g., *coat/comb*) were matched between MP and NMP lists both in initial consonant and subsequent vowel, two factors known to influence VOT (Lisker & Abramson, 1964; Volaitis & Miller, 1992; Yao, 2009), but density ranged continuously over each list (having controlled for density among the pairs, overall). Thus, **COMPETITOR-STATUS** controlled better for these phonetic properties than did the **NEIGHBORHOOD-DENSITY** manipulation. A further test was conducted to ensure that the influence of density could not be attributed to these segmental factors (*cf.* Flemming, 2010). Twelve item-defined neighborhood density difference scores ($NHD_{diff} = NHD_{MP} - NHD_{NMP}$) were computed (e.g., by subtracting the neighborhood density value for *comb* from that of *coat*: $NHD_{coat} - NHD_{comb}$). We then specified another **lmer** to examine the influence of NHD_{diff} (a continuous variable), **COMPETITOR-STATUS**, and their interaction on VOT. Because both members of an item-pair had the same NHD_{diff} value, both NHD_{diff} and **COMPETITOR-STATUS** were controlled for initial consonant and subsequent vowel. The key prediction for this analysis is a $NHD_{diff} \times \text{COMPETITOR-STATUS}$ interaction: for items with negative NHD_{diff} values (i.e., items for which the NMP word resides in a denser neighborhood than the MP word; $NHD_{MP} < NHD_{NMP}$), the NMP word should tend to be produced with a longer VOT than the MP word, and for items with positive NHD_{diff} values (i.e., items for which the MP word resides in a denser neighborhood than the NMP word; $NHD_{MP} > NHD_{NMP}$), the opposite should be true. Crucially, the magnitude of the difference between the MP and NMP words' VOTs for an item should be related to the magnitude of that item's NHD_{diff} .

The results of this analysis are summarized in Figure 2, which plots the difference between each item's mean VOT (averaged over subjects) for the MP word and its mean VOT for the NMP word ($VOT_{diff} = VOT_{MP} - VOT_{NMP}$) against each item's NHD_{diff} . As predicted, a significant interaction between NHD_{diff} and **COMPETITOR-STATUS** was found ($\beta = 0.175$, $SE = 0.056$, $|t| = 3.10$, $\chi^2(1) = 8.07$, $p < 0.005$). There was still no main effect of **COMPETITOR-STATUS** on VOT ($|t| < 1$). Finally, as with the first analysis, we included **TRIAL** and all of its interactions as factors in a follow-up model. The results confirmed a significant $NHD_{diff} \times \text{COMPETITOR-STATUS}$ interaction ($\beta = 0.174$, $SE = 0.059$, $|t| = 2.95$, $\chi^2(1) = 7.50$, $p < 0.007$) and found a main effect of **TRIAL** ($\beta = 7.290$, $SE = 2.887$, $|t| = 2.53$, $\chi^2(1) = 5.90$, $p < 0.02$). Again, no effect of **COMPETITOR-STATUS** was found ($|t| < 1$), and **TRIAL** did not interact with any factors (all $|t| < 1$).

2.5. Discussion

The results of Experiment 1 showed that, controlling for neighborhood density, the presence of a voiced competitor in the lexicon failed to result in longer initial VOTs for words produced in isolation. Rather, such hyperarticulatory effects emerged as a function of the

overall density of the neighborhood in which a spoken word resides. While no analysis provided evidence that MP and NMP words' initial stops differed (overall) in their VOTs, a significant effect of density (modeled as a continuous linear predictor) indicated that the articulatory encoding of a target word's initial voicing is influenced by lexical competitors that differ from the target on phonological parameters other than word-initial voicing. These results, therefore, suggest that global (rather than local) features of a word's phonological neighborhood drive lexically-conditioned phonetic variation in stop consonant production. The relationship between neighborhood density and VOT supports spoken word production models in which activation of a target word's phonological form co-activates its entire phonological neighborhood. It is this co-activation that drives phonetic variation in the production of the target word. The null effect of COMPETITOR-STATUS suggests that local, contrast-specific competition cannot explain the observed hyperarticulatory effects.

While Experiment 1 established a role of global neighborhood competition in lexically-conditioned phonetic variation, local competition may not be entirely immaterial. A COMPETITOR-STATUS × NEIGHBORHOOD-DENSITY interaction indicated that density had a stronger impact on the VOTs of the initial consonants of NMP words than those of MP words. One interpretation of this interaction is that the effects of global neighborhood characteristics on phonetic properties of speech are modulated or constrained by the effect of "local" neighborhood characteristics. That is, if a target has a word-initial voicing minimal pair, then all of its other neighbors have a diminished influence in driving phonetic variation. Broadly, this interpretation of the interaction would support a model that includes a competitive network of local-global dynamics at play within the phonological neighborhood architecture of the lexicon.

However, we believe that this interaction should be interpreted with caution. Although the MP and NMP lists were balanced with respect to neighborhood density (*cf.* Table 1; $p = 0.35$), they were not perfectly matched (i.e., it was *not* the case that the words making up each item-pair had identical neighborhood densities). In fact, the three words from the sparsest neighborhoods (*pants*, *toast*, *teeth*) all happened to come from the NMP list and the two words from the densest neighborhoods (*toe*, *pan*) both came from the MP list (*cf.* Figure 1). If ceiling effects are present in the production of VOTs (i.e., if subjects only hyperarticulate to a certain extent and there is a maximum VOT that they approach as their VOTs lengthen), then the linearity assumption of the regression may be violated, with subjects' slopes decreasing as they approach a maximum VOT. Thus, with higher-density MP words and lower-density NMP words, a steeper slope for NMP-words might be explained by a VOT ceiling effect. Some support for this explanation comes from previous work in which interactions emerging for VOT effects have been attributed to ceiling effects (see, e.g., Baese-Berk & Goldrick, 2009; Cho, Lee & Kim, 2011) and from the observation that the VOTs in our data were somewhat longer than the VOTs in comparable work (*cf.* Baese-Berk & Goldrick, 2009).

Whether the interaction we observe is theoretically meaningful or an artifact, finding evidence for a model with the kinds of local-global dynamics suggested under the first interpretation would not be altogether surprising. Clearly, more research examining well-matched stimuli would be needed to characterize the implications of a COMPETITOR-STATUS × NEIGHBORHOOD-DENSITY interaction. Nevertheless, this issue does not affect Experiment 1's central

theoretical conclusion – that phonological neighbors besides voicing competitors modulate the articulation of voicing in a target.

3. Experiment 2

Although the results of Experiment 1 suggest that a target's entire phonological neighborhood conditions its articulatory implementation when produced in isolation, it is not clear whether those effects still emerge when the target is produced in a more constrained environment. Since words are typically uttered in context rather than in isolation, a more complete characterization of the processes underlying spoken word production should also consider whether neighborhood effects are dynamic (such that a neighbor's influence depends on the context in which the target is produced). That is, while Experiment 1 provides some insight into the rules that govern which neighbors influence the realization of VOTs, it did not investigate the extent to which such influence varied as a function of the targets' context. Experiment 2 was designed to test this. Speakers produced the same target words as in Experiment 1 after reading sentence contexts that were either highly constraining (so as to make the target very likely) or less constraining (so as to make the target one of many possible continuations to the sentence).

3.1. Methods

3.1.1. Materials

3.1.1.1. Target Words: All 48 words that served as stimuli in Experiment 1 were included as targets in Experiment 2: 12 MP words, 12 NMP words, and 24 filler words that did not begin with voiceless stop consonants. Subjects produced these words aloud after silently reading context sentences, the characteristics of which are described below.

3.1.1.2. Sentence contexts: Each target word appeared as the final word of six carrier sentences. Three of these sentences were designed to be highly predictive of the target (biased contexts), while the other three sentences were less semantically constraining (neutral contexts). Table 3 provides examples of biased and neutral sentences for each target word type (see Appendix D for the complete list of stimuli).

For the biased sentence contexts, stimuli were designed to have a relatively high Cloze probability. To confirm this, the biased sentences were used as materials for an online norming task. The biased sentence contexts were divided into three lists, each list having one context designed for each critical and filler target (for a total of 48 sentence contexts in each list). One hundred fifty participants were recruited via Mechanical Turk and divided into three groups of fifty subjects, with each group seeing one of the three lists. Participants read each of the 48 sentence contexts in their list and were instructed to type the word (free response) that they felt best completed each sentence. Each participant was paid \$0.50 (expected rate: \$6/hour) and only saw one of the lists (*cf.* TurkGate; Goldin & Darlow, 2013). Results indicated that participants completed the biased contexts with the target word 73% of the time. Thus, the biased sentence contexts were considered to be (on average) highly predictive of the target words. Additionally, for sentence contexts that were designed

to be biased toward a target MP word (e.g., a sentence biased toward *coat*), no participant ever completed a sentence with the target's voicing competitor (*goat*).

For the neutral sentence contexts, stimuli were designed to have relatively low Cloze probability (i.e., be generally non-predictive of any specific word). Additionally, for neutral contexts that were followed by MP targets (e.g., *coat*), we tried to ensure that the MP target's voicing competitor (*goat*) was also a plausible continuation given the context.

All biased and neutral sentences were between 7 and 12 words long (mean: 8.47 words), and no condition contained significantly longer sentences than any other condition (all $p > 0.1$). Target words in each of the critical conditions were preceded by function words (e.g., articles, possessive pronouns, prepositions) in half of the stimuli and content words (e.g., adjectives) in the other half. No critical target words ever appeared in any sentence contexts, nor did any critical target's voicing minimal pair (e.g., *goat* never appeared in a sentence).

In all, 48 different words (12 MP targets + 12 NMP targets + 24 filler targets) occurred in 3 unique sentences in the biased condition and 3 other unique sentences in the neutral condition, yielding $48 \times [3 \times 2] = 288$ stimulus sentences. Thus, Experiment 2's design crossed three factors: two levels of CONTEXT (biased and neutral sentence contexts), two levels of COMPETITOR-STATUS (MP and NMP targets), and NEIGHBORHOOD-DENSITY.

3.1.2. Task and Recording—Sentences were presented visually on a computer monitor in a sound-dampened booth at a comfortable viewing distance. Each word prior to the target appeared in white against a black background, left-to-right, for 400 ms before the next word appeared and remained visible to the participant on the screen until the trial was over. The (sentence-final) target word appeared in yellow and remained on the screen for 2.5 seconds, at which point the entire screen went black for 1.5 seconds before the next trial began. Subjects were instructed to read the sentences silently until the final word (in yellow) appeared on the screen, at which point the subjects were asked to read the target word aloud. Subjects did not have difficulty producing the target in the time provided; the average time between visual onset of the target and speech onset was 613 ms (sd = 141). Subjects took a break after completing half of the trials.

The stimuli were randomly ordered within 3 blocks, each consisting of 96 of the 288 sentences. Each block included two presentations of each of the 48 unique target words (one biased and one neutral sentence context for each). The same procedures for recording and data processing were used as in Experiment 1.

3.1.3. Participants—The same 20 subjects who completed Experiment 1 also participated in Experiment 2. Experiment 2 always preceded Experiment 1 during a participant's experimental session in order to ensure that effects on production in context could not be attributed to targets' having already been produced three times in isolation.

3.2. Acoustic Analysis

The same analysis technique used in Experiment 1 was employed for Experiment 2. Using the same data inclusion criteria (see Appendix B), 2792 out of 2880 trials' VOTs (97%)

were included in the analysis. Overall, mean VOTs (computed as in Experiment 1) were shorter in Experiment 2: 80 ms, 88 ms, and 91 ms for /p/, /t/, and /k/ targets, respectively.

3.3. Statistical Analysis

As in Experiment 1, all tokens that met the inclusion criteria were used in the **lmer** analyses (see Appendix C), and non-parametric type analyses used subject/word/condition means.

3.4. Results

Results are displayed in Figure 3. To examine the effects of NEIGHBORHOOD-DENSITY, COMPETITOR-STATUS, CONTEXT and their potential interactions on VOT, we specified a **lmer** model with these three fixed main effects, three two-way interactions and a three-way interaction. Results showed a marginal main effect of CONTEXT ($\beta = -0.745$, $SE = 0.393$, $|t| = 1.89$, $\chi^2(1) = 3.50$, $p = 0.06$) such that subjects' VOTs were longer in neutral sentence contexts than in the biased sentence contexts, and a main effect of NEIGHBORHOOD-DENSITY ($\beta = 0.320$, $SE = 0.120$, $|t| = 2.67$, $\chi^2(1) = 7.10$, $p < 0.008$). The density effect was virtually identical in both direction and magnitude to that obtained in Experiment 1 ($\beta = 0.283$ in isolation vs. $\beta = 0.320$ in sentence contexts). As in Experiment 1, we found no evidence for a main effect of COMPETITOR-STATUS ($|t| = 1.16$, $\chi^2(1) = 1.40$, $p > 0.23$). However, unlike Experiment 1, there was no evidence for an interaction between COMPETITOR-STATUS and NEIGHBORHOOD-DENSITY ($|t| < 1$, $\chi^2(1) = 1.04$, $p > 0.30$). Crucially, neither the COMPETITOR-STATUS $\times$ CONTEXT interaction nor the NEIGHBORHOOD-DENSITY $\times$ CONTEXT interaction was significant (each $|t| < 1$, each $p > 0.30$).

A follow-up analysis included TRIAL and all of its interactions as factors in the model to examine possible influences of repetition and/or experiment length on the results. The effect of CONTEXT, which was marginal ($p = 0.06$) before accounting for variance related to TRIAL, reached significance in this analysis ($\beta = -0.817$, $SE = 0.407$, $|t| = 2.01$, $\chi^2(1) = 3.69$, $p = 0.05$). Again, there was a significant effect of NEIGHBORHOOD-DENSITY ($\beta = 0.318$, $SE = 0.120$, $|t| = 2.65$, $\chi^2(1) = 7.04$, $p < 0.008$). There was no effect of TRIAL on VOT ($|t| = 1.50$, $\chi^2(1) = 2.25$, $p > 0.13$), nor did TRIAL interact with any fixed effects (all $p > 0.14$).

As in Experiment 1, VOTs and model residuals were not normally distributed (see Appendix C), so the data were also subjected to nonparametric type analyses. First, Wilcoxon matched-pairs sign rank tests compared the VOT for MP and NMP words in both an F1-analysis (by-subject) and an F2-analysis (by-item). These tests were conducted within both sentence context conditions as well as collapsing across conditions. Table 4 displays mean VOTs (by condition) and results of the Wilcoxon matched-pairs sign rank tests. All analyses showed null effects, confirming the **lmer** analyses above, and replicating Experiment 1's findings.

Next, by-participant, by-item, and by-word Wilcoxon matched-pairs sign rank tests compared words' VOTs in biased versus neutral sentence contexts. Table 5 presents mean VOTs by condition and the results of those tests. Results confirmed the **lmer** analyses above, showing that word-initial VOTs were shorter after biased sentence contexts than neutral contexts.

Finally, we used Spearman's correlation test to evaluate the 24 words for a relationship between the word's NEIGHBORHOOD-DENSITY and its onset's mean VOT (the mean of the 20 subjects). The test was conducted within only biased sentence contexts, within only neutral sentence contexts, and collapsing across both. Each test revealed a significant correlation (biased: Spearman's $\rho = 0.44$, $p < 0.04$; neutral: Spearman's $\rho = 0.41$, $p < 0.05$; overall: Spearman's $\rho = 0.44$, $p < 0.04$).

As in Experiment 1, a further test was conducted to leverage item-by-item phonetic controls in the analysis of NEIGHBORHOOD-DENSITY. Neighborhood density difference scores ($NHD_{diff} = NHD_{MP} - NHD_{NMP}$) for each item replaced the NEIGHBORHOOD-DENSITY factor in the original analysis of Experiment 2 in order to specify a **lmer** that included fixed effects of NHD_{diff}, COMPETITOR-STATUS, CONTEXT, and all of their interactions on VOT. The prediction for this analysis is a NHD_{diff} × COMPETITOR-STATUS interaction, as well as a significant effect of CONTEXT. Figure 4 summarizes the results of this analysis, plotting (at each level of CONTEXT) the difference in mean VOT (averaged over subjects) between the MP and NMP word in each item ($VOT_{diff} = VOT_{MP} - VOT_{NMP}$) against NHD_{diff}. As predicted, a significant interaction between NHD_{diff} and COMPETITOR-STATUS was found ($\beta = 0.167$, $SE = 0.063$, $|t| = 2.64$, $\chi^2(1) = 6.53$, $p < 0.02$), as was a significant main effect of CONTEXT ($\beta = -0.804$, $SE = 0.385$, $|t| = 2.09$, $\chi^2(1) = 4.11$, $p < 0.05$). No other effects reached significance (all $|t| < 1$).

Finally, as with the other analyses, we included TRIAL and all of its interactions as factors in a follow-up analysis model. The results were nearly identical, with a significant NHD_{diff} × COMPETITOR-STATUS interaction ($\beta = 0.166$, $SE = 0.064$, $|t| = 2.61$, $\chi^2(1) = 6.43$, $p < 0.02$), and a main effect of CONTEXT ($\beta = -0.822$, $SE = 0.383$, $|t| = 2.15$, $\chi^2(1) = 4.28$, $p < 0.04$). The effect of TRIAL on VOT was still not significant ($|t| = 1.53$, $\chi^2(1) = 2.31$, $p > 0.12$), nor did any fixed effect's interaction with TRIAL reach significance (all $p > 0.10$).

Finally, follow-up tests were conducted to examine effects of NEIGHBORHOOD-DENSITY and COMPETITOR-STATUS on VOT at each level of CONTEXT. The two statistical models specified for the follow-up tests (one for words produced in biased contexts; a second for words produced in neutral contexts) were identical to the **lmer** analyses used in Experiment 1: each model included fixed effects for NEIGHBORHOOD-DENSITY, COMPETITOR-STATUS, and their interaction. Results demonstrated an effect of NEIGHBORHOOD-DENSITY on VOT in both types of sentence contexts (after biased contexts: $\beta = 0.365$, $SE = 0.122$, $|t| = 3.00$, $\chi^2(1) = 8.46$, $p < 0.004$; after neutral contexts: $\beta = 0.283$, $SE = 0.132$, $|t| = 2.14$, $\chi^2(1) = 4.81$, $p < 0.03$). Neither model suggested an effect of COMPETITOR-STATUS (both $p > 0.1$) nor did either suggest a COMPETITOR-STATUS × NEIGHBORHOOD-DENSITY interaction (both $p > 0.3$).

Following previous analyses, we also replaced the NEIGHBORHOOD-DENSITY factor with NHD_{diff} (neighborhood density difference scores for each item) at each level of CONTEXT, obtaining significant NHD_{diff} × COMPETITOR-STATUS interactions in both models (after biased contexts: $\beta = 0.189$, $SE = 0.065$, $|t| = 2.92$, $\chi^2(1) = 7.59$, $p < 0.006$; after neutral contexts: $\beta = 0.146$, $SE = 0.069$, $|t| = 2.12$, $\chi^2(1) = 4.49$, $p < 0.04$); no other effects were significant (all $|t| < 1$).

3.5. Discussion

Experiment 2 was designed to investigate the extent to which phonological neighbors can drive phonetic variation in a target preceded by a biased versus neutral sentence context. If neighborhood-conditioned effects on phonetic variation are modulated as a function of context, then the effect of neighborhood density should be reduced or abolished by a highly constraining (biased) context in which a single word (the target) is strongly predicted. Alternatively, if a word's neighborhood density has comparable hyperarticulatory effects on VOT after both highly predictive (biased) contexts and relatively uninformative (neutral) contexts, this may indicate that semantic information does not interfere with the internal dynamics of phonological neighborhood activation during spoken word production.

The main effect of density in Experiment 2 mirrored the results of Experiment 1, suggesting that the total number of neighbors in a word's phonological neighborhood influences the way the word is articulated, whether it is produced in context or in isolation. Interestingly, the effect size (slope) across the two experiments (Experiment 1: $\beta = 0.284$; Experiment 2: $\beta = 0.320$) was nearly identical, suggesting that the hyperarticulatory influence exerted by a word's neighbors is not diminished when the word is preceded by a sentence context. Additionally, no interaction between NEIGHBORHOOD-DENSITY and CONTEXT emerged in Experiment 2, indicating that the influence of number-of-neighbors on VOT was not diminished for biased sentences compared to neutral sentences. These results are consistent with Scarborough's (2010) finding that neighborhood-conditioned effects on vowel duration and dispersion are not modulated by the contextual predictability of the target. The lack of evidence for contextual modulation of neighborhood-conditioned phonetic variation in Experiment 2 is also convergent with Fricke's (2013) unpublished findings with respect to VOTs and vowel durations in spontaneous speech. Finally, robust effects of NEIGHBORHOOD-DENSITY emerged in follow-up tests of both biased and neutral sentence contexts (after biased contexts: $\beta = 0.365$; after neutral contexts: $\beta = 0.283$), further suggesting that the effect of NEIGHBORHOOD-DENSITY was no weaker in biased contexts than in neutral contexts. Taken together, Experiment 2's results strongly support a model in which a selected lexical item's phonologically related neighbors modulate its articulatory implementation independent of constraints from a preceding sentence context.

Nonetheless, a word's context did affect articulation: participants produced words' initial voiceless stops with shorter VOTs after predictive sentence contexts than after neutral contexts. This finding is consistent with previous work examining effects of context on the production of vowels (e.g., Bradlow & Alexander 2007; Gahl & Garnsey, 2004, 2006; Gahl, Yao & Johnson, 2012; Jurafsky et al., 2001; Kuperman & Bresnan, 2012; Scarborough, 2010; Tily et al, 2009), and provides further support for the well-known capacity of informative contextual cues to drive online changes in articulatory-motor planning during speech production (see, e.g., Aylett & Turk, 2004; Bell et al., 2009; Lieberman, 1963; Lindblom, 1990). Thus, while sentential context does not appear to modulate the influence of phonological neighborhood competition on articulation of a consonant, context does independently affect articulation.

It is worth noting that the obtained main effect of CONTEXT also dispels an alternative explanation for the parallel effects of neighborhood density in biased and neutral contexts. If

subjects had ignored the sentence contexts in Experiment 2, then consistent effects of density irrespective of contextual predictability (i.e., lack of an interaction) would be a trivial finding; however, the main effect of CONTEXT shows that subjects were attending to targets' contexts.

Finally, as in Experiment 1, no evidence was found in Experiment 2 that the VOTs of initial consonants of MP words differ from those of NMP words independent of density, further indicating that hyperarticulation results from competition with many phonological neighbors and not the presence of a voicing minimal pair.

Although both Experiments 1 and 2 showed a significant effect of NEIGHBORHOOD DENSITY but no effect of COMPETITOR-STATUS, their results differed in that Experiment 2 found no evidence for an interaction between COMPETITOR-STATUS and NEIGHBORHOOD-DENSITY, whereas, in Experiment 1, density appeared to have a stronger influence on NMP words than on MP words. Two possible interpretations of this interaction were provided above: local constraints on the global influence of less phonologically similar neighbors, and/or nonlinear effects of NEIGHBORHOOD-DENSITY such that the effect of density is increasingly moderated by a ceiling effect. While our data cannot adjudicate between these two interpretations, it is unclear why local-global interactions would disappear in Experiment 2. On the other hand, if the interaction arose due to a ceiling effect, then the lower mean VOTs in Experiment 2 relative to Experiment 1 might explain why no such interaction arose in Experiment 2. Most importantly for the present work, the interaction's presence or absence is not problematic for our main conclusions, nor is it directly related to our key questions.

4. General Discussion

Lexically-conditioned phonetic variation describes systematic variation in speech production in which lexical properties of a word (e.g., frequency, syntactic category) or its phonological similarity network (e.g., neighborhood density) influence the acoustic-phonetic characteristics of its articulatory implementation. Past research has demonstrated that the size of the neighborhood in which a word resides is associated with measures of hyperarticulation such that a word residing in a dense neighborhood is produced with more extreme phonetic properties (e.g., greater vowel dispersion; Munson, 2007; Munson & Solomon, 2004; Scarborough, 2003, 2010, 2012, 2013; Wright, 2004). Less work has examined the specific conditions under which a word's phonological neighbors affect its production.

The present study investigated two questions regarding the rules that govern such lexically-conditioned phonetic variation occurs. First, Experiment 1 considered whether hyperarticulation of a target word's initial VOT is modulated by a neighbor that differs from the target on initial voicing, or, alternatively, whether the target's entire neighborhood contributes to modulation of initial voicing irrespective of any specific voicing relationship with one neighbor. Experiment 2 examined whether sentential context might influence either or both of these neighborhood-conditioned effects on the production of voicing.

Our results showed that, whether produced in isolation, in a neutral sentence context, or in a biased sentence context, the duration of a word-initial stop's VOT was correlated with the target word's neighborhood density. Importantly, the effect on voicing realization emerged as a function of lexical density, not the presence of a local voicing contrast. Thus, higher levels of global phonological similarity between a target word and its neighbors predicted hyperarticulation along the voicing dimension. Additionally, although sentential context (biased vs. neutral) did affect implementation of voicing in spoken word production overall, it did not interact with either neighborhood-conditioned effect we examined. Together, these results support a generalized influence of phonological neighbors on articulation in spoken word production irrespective of preceding sentential context.

4.1. Generalized Hyperarticulation in Speech Production

From one theoretical perspective, it may, at first, seem counterintuitive that the existence of an initial voicing contrast had no main effect on initial VOTs while a more global measure of phonological similarity (neighborhood density, which treats all phonetic contrasts at all word positions equally) would drive variation in initial VOTs. After all, it is not necessarily clear *why* a target's initial stop's VOT, specifically, would be modulated by "any old" lexical neighbor. Importantly, as noted earlier, global measures of neighborhood density have often been shown to modulate other acoustic qualities of speech, and although these results similarly fail to directly answer the question of why certain *specific* acoustic parameters (e.g., vowel formants) are modulated by *global* characteristics of a word's phonological similarity network, they also highlight the fact that there appear to be several hyperarticulatory effects modulated by global neighborhood structure (see also Scarborough & Zellou, 2013). This, in turn, calls into question the idea that specific types of phonetic variation are driven by specific relationships (e.g., phonetic contrasts) between lexical neighbors. Rather, our results, when considered in light of these global neighborhood density effects on other acoustic parameters, suggest that hyperarticulation in speech production occurs along many dimensions and that the extent to which hyperarticulation of a target is realized is primarily conditioned by global features of a target's phonological neighborhood structure.

Notably, Baese-Berk and Goldrick (2009) described their study as one investigating neighborhood effects on phonetic variation. However, their stimuli precluded differentiation between two possible sources of the observed systematic variation in VOT: global neighborhood competition versus contrast-specific (local) neighbor competition. Nevertheless, their "minimal pair effect" has been widely interpreted as evidence that particular neighbors drive hyperarticulation in predictable ways (see, e.g., Gahl, Yao & Johnson, 2012). Our findings, however, suggest that their behavioral effect was driven by global lexical neighborhood characteristics rather than by the presence or absence of a single specific competitor, and thus serve to demonstrate the (global) locus of the original finding. Unpublished results by Fricke (2013) similarly indicate a strong effect of NEIGHBORHOOD-DENSITY on VOT and an unreliable effect of COMPETITOR-STATUS both in a reanalysis of Baese-Berk and Goldrick's (2009) data and in a spontaneous speech corpus (Pitt et al., 2007). VOT, then, might be one more example of a dimension along which neighborhood-conditioned phonetic

variation can be measured (just like vowel dispersion and coarticulation; see Scarborough, 2013; Scarborough & Zellou, 2013).

4.2. Theoretical Implications for Models of Speech Production

Taken together, the results of Experiments 1 and 2 are consistent with a model of spoken word production in which accessing the phonological form of a target word results in the automatic and obligatory co-activation of the target's neighborhood. This theoretical claim has two key implications for modeling spoken word production: (1) that a target's entire phonological neighborhood is co-activated upon accessing the target's phonological form, and (2) that the activation dynamics within the target's phonological neighborhood are impervious to interference from information external to the lexicon such as sentential context.

4.2.1. Cascading Activation Models of Speech Production—A cascading activation framework can account for the current findings. In this view, accessing a lexical target activates multiple phonological forms (the target as well as its neighbors) through spreading activation (Goldrick, 2006; Peterson & Savoy, 1998). Competition among phonologically related word-forms (e.g., Dell, 1986; Rapp & Goldrick, 2000) results from feedback dynamics (Vitevitch, 2002) and lateral inhibition (Chen & Mirman, 2012). Targets with more neighbors receive more feedback and face more competition; more competition with neighbors drives higher levels of activation in the target's phonological form; activation cascades downstream during phonetic encoding and motor planning, yielding greater degrees of hyperarticulation (e.g., longer VOTs; Baese-Berk & Goldrick, 2009).

While we found no evidence that sentential context *mediated* neighborhood effects on VOT, we did find support for the claim that contextual cues can modulate the acoustic-phonetic encoding of a target. That is, systematically shorter VOTs occurred when a word was in a biased context compared to a neutral context. This finding, too, may be accommodated within the cascading activation framework: as more evidence accumulates for the eventual target word over the course of a sentence, the target's threshold of activation is incrementally lowered, thereby facilitating articulatory programming of the target word and resulting in shorter VOTs. Crucially, though, accessing a target word necessarily triggers the co-activation of the target's entire phonological neighborhood no matter what threshold is determined by contextual cues. Consequently, context modulates lexical access in production, but once access takes place, activation dynamics are determined by characteristics of the phonological similarity network irrespective of preceding context.

It is important to note that cascading activation, as we present it here, is a *framework* for a model of spoken word production, not a model itself (see Baese-Berk & Goldrick, 2009). As such, invoking this framework does not (and cannot) constitute a complete explanation for the findings of this study. Although it is beyond the scope of the present work to present such a model, a successful cascading activation model of spoken word production would need to fully characterize how potential mechanisms (e.g., activation levels, thresholds, feedback, competition, lateral inhibition) could explain the observed patterns of facilitation/inhibition (in, e.g., production latency, naming accuracy, speech errors; see, e.g., Chen &

Mirman, 2012; Sadat, Martin, Costa & Alario, 2014) and hyperarticulation/hypoarticulation (in, e.g., vowel space, vowel/word duration, VOT). Further, any such model must not only consider effects of phonological neighborhood density, but also of frequency, predictability, phonotactic probability, and other lexical variables.

4.2.2. Listener-Oriented Models of Speech Production—Listener-oriented models of spoken word production constitute a set of computational theories which propose that utterances are tailored to enhance communicative efficiency by balancing opposing imperatives: (1) minimize confusability with other possible utterances, and (2) minimize effortful hyperarticulation (e.g., Aylett & Turk, 2004; Lindblom, 1990; Levy & Jaeger, 2007; Jaeger, 2010). Under this computational framework, when a selected target word is less likely to occur or is more confusable, a cooperative speaker will tend to hyperarticulate the word's segments to aid the listener in comprehending the intended message. Thus, “ideal speakers” expend more effort only to the extent that such expenditure is likely to benefit the listener. Under this class of models, then, hyperarticulation serves a functional role. Importantly, context-driven hyperarticulation (or hypoarticulation/reduction) and neighborhood-driven hyperarticulation need not necessarily result from identical mechanisms. Scarborough (2010) describes both as forms of competition (increased semantic competition in low-predictability contexts vs. increased phonological competition in high-density words), but the current results cannot shed light on how related these two implementations of competition are from the pattern of VOTs alone.

If producing a longer VOT incurs a greater cost (e.g., in time or in energy), then a speaker should only resort to that strategy when it enhances his/her chances of being understood by the listener. Wright (2004) suggests that this is likely to be the cause of hyperarticulation as measured in vowel quality in “difficult” words (low-frequency words from high-density neighborhoods) relative to “easy” words (high-frequency words from low-density neighborhoods) (*cf.* Gahl, 2015). By producing a vowel that is more dissimilar from all other vowels in a speaker's vowel space *only to the extent* that there are many other words that sound relatively similar to the target, Wright argues, the speaker is selectively paying the costs associated with hyperarticulation in order to reduce the confusability of his/her speech. Consistent with this view are the results from studies that have demonstrated that words from dense neighborhoods are recognized more slowly and less accurately than words from sparse neighborhoods (Goldinger, Luce & Pisoni, 1989; Luce & Pisoni, 1998; Strand & Sommers, 2011; Vitevitch & Luce, 1998).

Listener-oriented efficient communication models of spoken word production have no difficulty accounting for the main effect of CONTEXT in Experiment 2 (shorter VOTs in biased contexts compared to neutral contexts). When a word is very probable given its context (e.g., in Experiment 2's biased sentences), acoustic reduction saves the speaker costs in articulatory implementation without compromising the listener's ability to correctly perceive the word. On the other hand, words produced in less informative contexts (e.g., in Experiment 2's neutral sentences), are more likely to be misperceived, so the speaker hyperarticulates to the extent that it is likely to aid the listener in accurately and quickly recognizing the speaker's message (see, e.g., Arnold, Kahn & Pancani, 2012; Anderson, Bard, Sotillo, Newlands & Doherty-Sneddon, 1997; Aylett & Turk, 2004; Bell, Brenier,

Gregory, Girand & Jurafsky, 2009; Bolinger, 1963; Bradlow & Alexander 2007; Buz & Jaeger, 2012; Chafe, 1974; Clopper & Pierrehumbert, 2008; Hawkins & Warren, 1994; Lindblom, 1990).

However, it is less clear that listener-oriented models can account for our other findings. Firstly, we observed that the existence of a voicing competitor does not increase VOT, whereas increasing neighborhood density incrementally drives up VOT for a word's initial consonant. Strict obedience to an ideal speaker model would only predict hyperarticulation inasmuch as the associated costs "buy" the listener something (e.g., making the target word less confusable), but it is unclear how hyperarticulation of the initial VOT of the word *cone* (which comes from a dense neighborhood, but does not have a voicing minimal pair) could make the target less confusable, especially when a speaker might lengthen an already 80ms+ VOT.

Secondly, perhaps the greatest challenge for listener-oriented models of speech production presented by our study's results is that hyperarticulation appears even when the target is produced at the end of sentences where there is essentially no contextually plausible competitor (biased sentences with high Cloze probabilities). An efficient speaker would certainly prefer *not* to hyperarticulate a word when no other competitors are plausible continuations. Thus, the lack of a $\text{NEIGHBORHOOD-DENSITY} \times \text{CONTEXT}$ interaction and robust effect of $\text{NEIGHBORHOOD-DENSITY}$ in the follow-up test examining only targets produced in biased contexts further challenge this category of models; they predict that the presence of additional competitors in a word's neighborhood should only warrant incremental lengthening of VOT to the extent that those competitors might be consistent with the context. Minimally, our results suggest that there are limits on the efficiency with which speakers pay the costs of hyperarticulation in such a model.

4.3. Lexically-Conditioned Phonetic Variation in Context

Our results suggest that all of a target's neighbors compete during phonological encoding irrespective of contextual constraints. As already noted, this is consistent with work by Scarborough (2010), who found longer, more dispersed vowels in words from high-density neighborhoods and after unpredictable sentences, but found no evidence that context modulated neighborhood-conditioned hyperarticulation (see also Fricke, 2013). It is also consistent with Heller (2014) and with Heller and Goldrick (2014, 2015) who found no difference in neighborhood-driven competition effects in grammatically constraining contexts compared to unconstrained contexts (picture-naming after reading a sentence frame vs. picture-naming in isolation). This convergent lack of any evidence for an interaction between context and neighborhood effects on spoken word production suggests largely independent roles for context and neighborhood competition in lexically-conditioned phonetic variation.

Notably, some models predict that different contextual information sources (e.g., syntactic vs. semantic constraints) may differentially modulate lexical-phonological encoding during spoken word production (Dell, Oppenheim & Kittredge, 2008; for review, see Heller, 2014). In particular, these models predict that grammatically incongruent words do not compete with the target during production while grammatical-but-implausible words do, allowing

these models to explain why lexical speech error patterns (more specifically, phonologically related lexical error patterns) tend to preserve the syntactic category of the target word (Berndt, Mitchum, Haendiges & Sandson, 1997; Fay & Cutler, 1977; Goldrick, Folk & Rapp, 2010; Nootboom, 1969), and why there is reduced interference from semantically-related distractors when the distractor is of a different syntactic category than the target (Alario, Matos & Segui, 2004; Pechmann & Zerbst, 2002; Vigliocco, Vinson & Siri, 2005). We neither directly tested nor controlled for a potential role for grammatical relationships among neighbors in mediating lexically-conditioned phonetic variation, but unpublished results by Heller (2014) suggest that the degree of hyperarticulation in a target's vowel is influenced by how many of the target's neighbors are members of the same syntactic category. However, neither our manipulation of words produced in isolation (as in Experiment 1) versus context (as in Experiment 2) nor work by Heller and Goldrick (2014, 2015) support Dell, Oppenheim and Kittredge's (2008) prediction that a preceding context should affect the extent to which a target's neighbors compete with it for production.

4.4. Neighborhood Structure in Spoken Word Production: Zooming In

One remaining question regards the relative contributions of different neighbors to systematic variation in spoken word production. Baese-Berk and Goldrick's (2009) work can be seen as an attempt to "zoom in" on neighborhood effects in spoken word production by proposing a special role for one individual neighbor's contribution to a target's articulation (see also Goldrick, Vaughn & Murphy, 2013). To be sure, the present results cannot identify the relative contributions of individual neighbors to hyperarticulation, but we believe that it follows in the spirit of Baese-Berk and Goldrick (2009) by directly confronting questions about the conditions under which neighbors might influence a target word's voicing. However, a full understanding of the rules governing whether, when, and to what extent particular neighbors influence target word production will only be possible once we articulate explicit models capable of characterizing these differences – a task that will require a more theoretically nuanced approach to phonological neighborhood density (e.g., Strand & Sommers, 2011).

Thus, although we argue that the modulation of VOT by neighbors other than voicing minimal pairs illustrates generalization of hyperarticulation processes in speech production, other types of specificity might well be built into the system. For instance, Fricke (2013) presents data suggesting that phonetic variation in the duration of the first segment of a word is modulated by the number of neighbors a target has which contrast in their onset, whereas variation in the duration of a target's vowel exhibits hyperarticulation correlated with the number of neighbors which contrast on that segment. Fricke (2013) cites this as evidence for position-specificity in lexically-conditioned phonetic variation (see also Vitevitch, Ambrüster & Chu, 2004). As noted earlier, Heller (2014) and Heller and Goldrick (2014) posit that neighbors that are members of the same syntactic category as the target will contribute more to hyperarticulation of the target. In related work, Gahl (2008) argues that frequency-driven hyperarticulation depends on lexeme frequency in the production of homophones (e.g., *thyme* vs *time*), a result that calls for another layer of specificity in lexically-conditioned phonetic variation. It remains unclear to what extent these factors and

others might interact and/or whether they influence articulatory processes via the same mechanism(s).

Finally, it is worth noting that all of these factors relate to static features of a word and its relationships to other words in the lexicon. While this body of research may suggest that specific neighbors differentially modulate spoken word production, neither the present work nor previous work (Fricke, 2013; Heller, 2014; Heller & Goldrick, 2014, 2015; Scarborough, 2010) has offered evidence that context-dependent factors modulate the influence of phonological neighbors on target word production. This points toward a key theoretical claim: that neighborhood effects on spoken word production depend on static features of the architecture of the lexicon and not on dynamic factors in language production. Of course, much work remains to investigate this issue more fully, but we hope that the present work calls attention to this category of questions, while also placing constraints on two proposed sources of such specificity.

5. Conclusion

The present results suggest that the articulatory implementation of a target word's VOT is influenced by its phonological neighbors whether or not they differ from the target on voicing and irrespective of the target's context. This lends support for models of spoken word production in which all of a target's phonological neighbors' word-forms are co-activated when the target's is accessed, irrespective of their plausibility in context. The more words there are in a target's neighborhood, the greater the degree of competition, the greater the target word-form's activation level, and the greater the extent of hyperarticulation. Thus, global competition from a word's neighborhood affects the implementation of voicing, suggesting that hyperarticulatory effects are generalized in spoken word production. Finally, contextual information can influence a word's articulatory encoding, but it appears not to interfere with the influence of that word's neighbors on its production. Together with previous work, these findings constrain models of spoken word production, first by placing limits on listener-oriented models of spoken word production that present hyperarticulation as a functional tool for reducing confusability (e.g., Lindblom, 1990) and, second, by challenging those cascading activation models which hold that phonological neighborhood dynamics are context-dependent (e.g., Dell, Oppenheim & Kittredge, 2008).

Acknowledgements

We thank Sahil Luthra and Corey Cusimano for assistance in conducting these experiments and Jordana Heller for technical advice on the statistical tests included here. John Mertus designed and implemented the automated VOT measurement software. We are also grateful for feedback provided by Matthew Goldrick, two anonymous reviewers, and attendees of the Spring 2014 Meeting of the Acoustical Society of America, where we presented initial results. This research was supported in part by NIH NIDCD Grant R01DC006220 to Brown University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health. NF was supported by a Graduate Research Fellowship (DGE 0228243) from the National Science Foundation and a Graduate Fellowship from the Brown Institute for Brain Science at Brown University.

References

- Alario FX, Matos RE, Segui J. Gender congruency effects in picture naming. *Acta Psychologica*. 2004; 117(2):185–204. [PubMed: 15464013]

- Anderson AH, Bard EG, Sotillo C, Newlands A, Doherty-Sneddon G. Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception & Psychophysics*. 1997; 59(4):580–592. [PubMed: 9158332]
- Arnold JE, Kahn JM, Pancani GC. Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*. 2012; 19(3):505–512. [PubMed: 22419403]
- Aylett M, Turk A. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech*. 2004; 47:31–56. [PubMed: 15298329]
- Aylett M, Turk A. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*. 2006; 119:3048–3058. [PubMed: 16708960]
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59(4):390–412.
- Baese-Berk M, Goldrick M. Mechanisms of interaction in speech production. *Language and Cognitive Processes*. 2009; 24(4):527–554. [PubMed: 19946622]
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68:255–278.
- Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014 <URL: <http://CRAN.Rproject.org/package=lme4>>.
- Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*. 2009; 60(1):92–111.
- Berndt RS, Mitchum CC, Haendiges AN, Sandson J. Verb retrieval in aphasia. 1. Characterizing single word impairments. *Brain and language*. 1997; 56(1):68–106. [PubMed: 8994699]
- Bolinger D. Length, vowel, juncture. *Linguistics*. 1963; 1:5–29.
- Bradlow AR, Alexander JA. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-Ynative listeners. *Journal of the Acoustical Society of America*. 2007; 121:2339–2349. [PubMed: 17471746]
- Brysbaert M, New B. Moving beyond Kucera 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*. 2009; 41:977–990. [PubMed: 19897807]
- Bullock-Rest N, Cerny A, Sweeney C, Palumbo C, Kurowski K, Blumstein SE. Neural systems underlying the influence of sound shape properties of the lexicon on spoken word production: Do fMRI findings predict effects of lesions in aphasia? *Brain and Language*. 2013; 126(2):159–168. [PubMed: 23743183]
- Buz, E.; Jaeger, TF. Effects of phonological confusability on speech duration. Poster at the 25th Annual CUNY Conference on Human Sentence Processing; New York, NY. 2012.
- Chafe W. Language and consciousness. *Language*. 1974; 50:111–133.
- Chen Q, Mirman D. Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*. 2012; 119:417–430. [PubMed: 22352357]
- Cho T, Lee Y, Kim S. Communicatively driven versus prosodically driven hyperarticulation in Korean. *Journal of Phonetics*. 2011; 39(3):344–361.
- Clopper CG, Pierrehumbert JB. Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*. 2008; 124:1682–1688. [PubMed: 19045658]
- Cohen J. The Cost of Dichotomization. *Applied Psychological Measurement*. 1983; 7(3):249–253.
- Dell GS. A spreading Y activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Dell GS, Oppenheim GM, Kittredge AK. Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*. 2008; 23:583–608. [PubMed: 20622975]

- Fay D, Cutler A. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*. 1977; 8:505–520.
- Fidelholtz JL. Word frequency and vowel reduction in English. In Chicago Linguistic Society. 1975 Apr.11:200–213.
- Flemming, E. Modeling listeners: Comments on Pluymaekers et al. and Scarborough. In: Fougeron, C.; Kuehnert, B.; Imperio, M.; Vallee, N., editors. *Laboratory Phonology*. Vol. 10. New York: Mouton de Gruyter; 2010. p. 587–606.
- Fowler C, Housum J. Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*. 1987; 26:489–504.
- Fricke, M. Ph.D. Thesis. Berkeley, CA: University of California, Berkeley; 2013. Phonological encoding and phonetic duration.
- Gahl S. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*. 2008; 84(3):474–496.
- Gahl S. Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics*. 2015; 49:96–116.
- Gahl S, Garnsey SM. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation. *Language*. 2004; 80(4):748–775.
- Gahl S, Garnsey SM. Knowledge of grammar includes knowledge of syntactic probabilities. *Language*. 2006; 82(2):405–410.
- Gahl S, Yao Y, Johnson K. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*. 2012; 66:789–806.
- Goldin G, Darlow A. TurkGate (Version 0.4.0). 2013 <URL: <http://gideongoldin.github.com/TurkGate/>>.
- Goldinger SD, Luce PA, Pisoni DB. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*. 1989; 28:501–518. [PubMed: 24465086]
- Goldinger S, Summers W. Lexical neighborhoods in speech production: A first report. Research on Speech Perception Progress Report. Psychology Department, Speech Research Laboratory, Indiana University. 1989; 15:331–342.
- Goldrick M. Limited interaction in speech production: Chronometric, speech error and neuropsychological evidence. *Language and Cognitive Processes*. 2006; 21:817–855.
- Goldrick M, Folk JR, Rapp B. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*. 2010; 62:113–134. [PubMed: 20161591]
- Goldrick M, Vaughn C, Murphy A. The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*. 2013; 134(2):EL172–EL177. [PubMed: 23927221]
- Hawkins S, Warren P. Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*. 1994; 22:493–511.
- Heller, JR. Ph.D. Thesis. Evanston, IL: Northwestern University; 2014. Contextual constraints on phonological activation during sentence production.
- Heller JR, Goldrick M. Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin and Review*. 2014; 21(6):1576–1582. [PubMed: 24687733]
- Heller JR, Goldrick M. Corrigendum to "Grammatical constraints on phonological encoding in speech production". *Psychonomic Bulletin and Review*. 2015 (submitted).
- Hunnicutt S. Acoustic correlates of redundancy and intelligibility. *Speech, Music and Hearing Quarterly Progress and Status Report*. 1987; 28:7–14.
- Jaeger TF. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*. 2010; 61:23–62. [PubMed: 20434141]
- Jurafsky, D.; Bell, A.; Girand, C. The Role of the Lemma in Form Variation. In: Gussenhoven, C.; Warner, N., editors. *Laboratory Phonology*. Vol. 7. New York: Mouton de Gruyter; 2002. p. 3–34.
- Jurafsky, D.; Bell, A.; Gregory, M.; Raymond, WD. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Bybee, J.; Hopper, P., editors. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins; 2001. p. 229–254.

- Kirov, C.; Wilson, C. The Specificity of Online Variation in Speech Production. 34th Annual Meeting of the Cognitive Science Society; Sapporo, Japan. 2012.
- Kuperman V, Bresnan J. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*. 2012; 66(4):588–611.
- Levy, R.; Jaeger, TF. Speakers optimize information density through syntactic reduction. In: Schölkopf, B.; Platt, JC.; Hoffman, T., editors. *Advances in Neural Information Processing Systems*. Vol. 19. 2007. p. 841–848.
- Lieberman P. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*. 1963; 6:172–187.
- Lindblom, B. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, WJ.; Marchal, A., editors. *Speech Production and Speech Modelling*. The Netherlands: Kluwer; 1990. p. 403–439.
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. *Word: Journal Of The International Linguistic Association*. 1964; 20(3):384–422.
- Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*. 1998; 19(1):1–36. [PubMed: 9504270]
- Mahon BZ, Costa A, Peterson R, Vargas KA, Caramazza A. Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of experimental psychology. Learning, memory, and cognition*. 2007; 33(3): 503–535.
- Mertus, J. BLISS User Manual. Providence, RI: Brown University; 1989.
- Mertus, J. Automated VOT Measurement for Isolated Words in English: A Novel Tool for BLISS. Providence, RI: Brown University; 2014. <URL: <http://www.mertus.org/Bliss/AutoVOT.pdf>>.
- Munson, B. Lexical access, lexical representation, and vowel production. In: Cole, JS.; Hualde, JL., editors. *Laboratory Phonology*. Vol. 9. New York: New York: Mouton de Gruyter; 2007. p. 201–228.
- Munson B, Solomon NP. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*. 2004; 47:1048–1058.
- Nooteboom, SG. The tongue slips into patterns. In: Sciarone, AG.; van Essen, AJ.; Avan Raad, A., editors. *Leyden studies in linguistics and phonetics*. The Hague: Mouton; 1969. p. 114–132.
- Pechmann T, Zerbst D. The activation of word class information during speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:233–243.
- Peramunage D, Blumstein SE, Myers EB, Goldrick M, Baese-Berk M. Phonological neighborhood effects in spoken word production: an fMRI study. *Journal of cognitive neuroscience*. 2011; 23(3): 593–603. [PubMed: 20350185]
- Peterson RR, Savoy P. Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1998; 24:539–557.
- Pickett JM, Pollack I. Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*. 1963; 6:151–164.
- Pitt, MA.; Dilley, L.; Johnson, K.; Kiesling, S.; Raymond, W.; Hume, E.; Fosler-Lussier, E. *Buckeye Corpus of Conversational Speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University (Distributor); 2007. <URL: www.buckeyecorpus.osu.edu>.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. <URL: <http://www.R-project.org/>>.
- Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. *Psychological Review*. 2000; 107:460–499. [PubMed: 10941277]
- Rodríguez-Ferreiro J, Davies R, Cuetos F. Semantic domain and grammatical class effects in the picture-word interference paradigm. *Language and Cognitive Processes*. 2014; 29(1):125–135.
- Sadat J, Martin CD, Costa A, Alario FX. Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*. 2014; 68:33–58. [PubMed: 24291531]

- Scarborough, RA. Lexical confusability and degree of coarticulation. Proceedings of the 29th annual meeting of the Berkeley Linguistics Society; 2003. 2003.
- Scarborough, RA. Lexical and contextual predictability: Confluent effects on the production of vowels. In: Fougerson, C.; Kuehnert, B.; Imperio, M.; Vallee, N., editors. *Laboratory Phonology*. Vol. 10. New York: Mouton de Gruyter; 2010. p. 557-586.
- Scarborough RA. Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua*. 2012; 122(2):164–176.
- Scarborough RA. Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*. 2013; 41(6):491–508.
- Scarborough RA, Zellou G. Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *Journal of the Acoustical Society of America*. 2013; 134(5):3793–3807. [PubMed: 24180789]
- Smiljanic R, Bradlow AR. Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*. 2008; 36(1):91–113. [PubMed: 19122747]
- Strand JS, Sommers MS. Sizing up the competition: Quantifying the influence of the mental lexicon on auditory and visual spoken word recognition. *Journal of the Acoustical Society of America*. 2011; 130(3):1663–1672. [PubMed: 21895103]
- Tily H, Gahl S, Armon I, Snider N, Kothari A, Bresnan J. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*. 2009; 1(2):147–165.
- Vaden KI, Halpin HR, Hickok GS. Irvine Phonotactic Online Dictionary, Version 2.0. 2009 <URL: <http://www.iphod.com>>.
- Van Santen JPH. Contextual effects on vowel duration. *Speech Communication*. 1992; 11:513–546.
- Vigliocco G, Vinson DP, Siri S. Semantic similarity and grammatical class in naming actions. *Cognition*. 2005; 94:B91–B100. [PubMed: 15617670]
- Vitevitch MS. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:735–747.
- Vitevitch MS, Ambrüster J, Chu S. Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2004; 30:514–529.
- Vitevitch MS, Luce PA. When words compete: Levels of processing in spoken word perception. *Psychological Science*. 1998; 9:325–329.
- Vitevitch MS, Luce PA. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*. 1999; 40:374–408.
- Vitevitch MS, Luce PA. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*. 2004; 36:481–487.
- Volaitis LE, Miller JL. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*. 1992; 92(2):723–735. [PubMed: 1506527]
- Wright, R. Factors of lexical competition in vowel articulation. In: Local, JJ.; Ogden, R.; Temple, R., editors. *Laboratory Phonology*. Vol. VI. Cambridge, UK: Cambridge University Press; 2004. p. 75-87.
- Yao, Y. Ph.D. Thesis. Berkeley, CA: University of California, Berkeley; 2009. The effect of phonological neighborhoods on pronunciation variation in conversation speech.

Appendix A: Target Word Stimuli – Experiments 1 and 2

MP words	NMP words	Filler words	
peak (beak)	peel	bait	hole
pan (ban)	pants	book	jail
park (bark)	part	bag	job

MP words	NMP words	Filler words	
peach (beach)	peace	church	land
tent (dent)	test	dance	mom
teen (dean)	teeth	date	noise
toe (dough)	toast	dice	rum
time (dime)	type	fish	sand
coat (goat)	comb	gun	sea
coast (ghost)	cone	guest	ship
card (guard)	car	hair	shirt
cash (gash)	cast	hand	zone

Appendix B: Automated Acoustic Analysis – Algorithm & Validation

I. Details of automated measurement algorithm

For each target stimulus, the VOT of the initial consonant was calculated by determining the elapsed time between two critical zero-crossings in the word's recorded waveform using the algorithm below. Further information is available in online documentation (Mertus, 2014).

1. **Onset of burst:** the zero-crossing immediately preceding the burst of the initial consonant
 - a. Specify a search window ("*burst-search window*") of 200 ms prior to the onset of the extracted pitch contour.
 - b. Compute the root-mean-square energy (RMSE) of the entire speech waveform. This RMSE value is smoothed (exact smoothing methods described online; Mertus, 2014) and normalized (minimum = 0; maximum = 1), yielding a "*transformed RMSE*."
 - c. Identify the earliest *time* t_b within the burst-search window at which the transformed RMSE exceeds 0.15.
 - d. Mark the latest zero-crossing occurring before t_b as the *onset of the burst*.
2. **Onset of voicing:** the zero-crossing closest to the onset of voicing
 - a. *Identify a candidate set of zero-crossings* to consider. Extract pitch contour; identify peaks of waveform within the vowel; mark zero-crossings immediately preceding each of the peaks; choose 5 zero-crossings as candidates.
 - b. *Three independent methods* each select the zero-crossing from among this candidate set that it most confidently predicts is the onset of voicing.
 - i. **Energy:** Selects the candidate zero-crossing closest to the peak rise in energy, or the highest rate of change in the waveform's energy.
 - ii. **Correlation:** Selects the candidate zero-crossing that best separates frication noise from periodicity. Specifically, it computes the

correlation between waveform values of two adjacent pitch periods (of normalized window-lengths) and selects the zero-crossing with the smallest correlation of the values for the pitch period immediately before it and the pitch period immediately after it.

- iii. **Count:** Selects the earliest candidate zero-crossing with fewer than 3 other zero-crossings within a 2ms time range centered around the candidate zero-crossing.

- c. **Enforce agreement criteria:** exclude trials for which no two methods (b.i., ii., or iii) converge on one candidate with a positive VOT measurement (duration separating onset of burst and onset of voicing).

II. Details of discarded data

For a small number of trials (Experiment 1: 22 of 1440 = 1.5%; Experiment 2: 88 of 2880 = 3.1%), no VOT measurement was recorded because the automated system could not converge on a single positive VOT value, as described above. These trials were discarded prior to analysis.

Each word was repeated by a subject three times in each condition (3× total in Experiment 1; 6× total in Experiment 2). Take subject X in Experiment 1, for example. If one of X's three productions of *coat* was discarded, that trial was not linked to any specific one of X's three productions of *comb*. In such a case, X would simply have 2 VOT values for *coat* and 3 for *comb*. All tokens (that were not discarded) were included in the **lmer** analyses. It is worth noting that **lmer** analyses are robust to unequal numbers of observations in each cell (Baayen et al., 2008).

For all non-parametric type analyses, mean VOTs for every subject/word/condition cell were computed. It was never the case that all three trials of a given word (in a given condition) for a given subject were discarded. Therefore, when computing these mean VOTs, most cells had three VOTs contributing to that mean, a few had only one or two, but no cell had zero values. Table B.1 shows the distribution of discarded data in our experiments.

III. Automated software validation methodology

In order to establish the reliability of the automated measures, the VOTs for all critical targets in Experiment 1 from seven subjects (35%) were measured by hand to compare against the automated VOT measurements. Out of the 504 manually measured VOTs (7 subjects × 24 words × 3 trials per word), 501 had corresponding automated measurements after discarding trials as described above. All validation analyses included manual and automated measurements for these 501 trials.

To determine whether MEASUREMENT-METHOD had an effect on VOT measurements, a linear mixed effects regression was fit for each subject's VOTs. Each model included a fixed effect of MEASUREMENT-METHOD (a discrete factor with 2 levels: automated and manual), as well as random intercepts and slopes for each word. The results, summarized in Table B.2, showed that

automated measurements of each subject's VOTs were not significantly different from manual measurements of the same VOTs.

All seven subjects data were combined for an overall test of $\text{MEASUREMENT-METHOD}$'s effect on VOT measurements. This model included a fixed effect of $\text{MEASUREMENT-METHOD}$ and random intercepts and slopes for each word and each subject. The results showed that the method used to measure a VOT did not significantly influence the value obtained. It is notable that the standard error of the measurements of the automated system was not larger than the standard error of manual measurements.

Finally, to test whether differences emerged by-condition, a linear mixed effect model identical to the one used in Experiment 1 was fit separately to the automated and manual VOT measurements for the 7 subjects. Results are summarized in Table B.3. Overall, the methods did not differ with respect to the pattern of effects that were found to be significant. This analysis confirmed that both measurement methods yielded similar statistical inferences.

Of the 501 pairs of automated/manual measurements, 53% of the automated measurements were within 2 milliseconds of the manual measurement; 89% of them were within 6 milliseconds. Automated measurements were longer than their corresponding manual measurements 47% of the time, shorter than them 50% of the time and identical for 3% of VOTs.

Appendix C: Linear Mixed Effects Regression

I. Motivation and Distributional Assumptions

Our analyses make use of linear mixed effects regression (**lmer**) (see, e.g., Baayen, Davidson & Bates, 2008), a parametric statistical technique that can simultaneously model both discrete and continuous fixed effects, crossed random effects (e.g., subjects and items), and can accommodate an unequal number of repeated observations in each condition. Although our model residuals were not normally distributed (as confirmed by Shapiro tests), and log-transforming the VOTs did not coerce the data into a normal distribution, simulations by Jacqmin-Gadda, Sibillot, Proust, Moline and Thiebaut (2007) suggest that inference on fixed effects parameters in linear mixed models is generally robust to non-Gaussian error distributions, especially when the random effects structure of the model includes random slopes, as is the case for our analyses (see below). Furthermore, all key results were confirmed with non-parametric type analyses (see main text).

II. Implementation Details, Computation of p -values, and Random Effects Structures

Models were implemented using **lmer**, part of the lme4 package (Bates, Maechler, Bolker & Walker, 2014) in R (R Core Team, 2014).

FIXED EFFECTS

All continuous fixed effects ($\text{NEIGHBORHOOD-DENSITY}$, NHD_{diff} , TRIAL) were centered (mean = 0). TRIAL was normalized (making each experiment unit length). Discrete fixed effects (COMPETITOR-STATUS , CONTEXT) were deviation-coded (contrasts = $[-1, 1]$; negative contrasts corresponded to NMP words and neutral sentence contexts). Each word's density was retrieved using the IPhOD database (Vaden, Halpin & Hickok, 2009; *cf.* Vitevitch and Luce; 1999). Models were fit with restricted maximum likelihood (REML). The *maxfun* argument had to be increased for some models to achieve convergence. For each model, we report the following for each statistically significant effect: coefficient estimate (β), standard error of the estimate (SE), t -value ($= \beta / SE$), χ^2 statistic from likelihood-ratio test (nested model comparison; described below), and associated p -value (*cf.* Barr, Levy, Scheepers & Tily, 2013).

RANDOM EFFECTS

All models' random effect structures included every random intercept/slope/interaction for subjects and for items that was motivated by the design (Barr et al, 2013). Specifically, as the design was within-subjects, by-subject random intercepts were included, as well as random slopes for all fixed main effects and interactions. Although the design was also within-items, "items" in our analyses represented item-pairs (e.g., *coat / comb*), not words (i.e., *coat* and *comb* shared a random intercept). Thus, by-item variance in the $\text{NEIGHBORHOOD-DENSITY}$ effect was unidentifiable from by-item variance in the COMPETITOR-STATUS effect. Therefore, by-item random intercepts were always included, but some by-item random slopes were never included: $\text{NEIGHBORHOOD-DENSITY}$, $\text{NEIGHBORHOOD-DENSITY} \times \text{CONTEXT}$, and their interactions with TRIAL . Similarly, because NHD_{diff} was item-defined, by-item slopes were excluded for NHD_{diff} and any interactions involving NHD_{diff} . Finally, random effects were not allowed to covary by-subject or by-item. Thus, models' random effects structures were not maximal, but simulations suggest that this adjustment to achieve convergence does not hinder inference (Barr et al, 2013).

SIGNIFICANCE TESTING

Significance levels (i.e., p -values) for fixed effects were computed via nested model comparison (Barr et al, 2013) implemented using **anova** in *R* (resulting χ^2 statistics are also reported). According to this technique, the p -value associated with a fixed effect indicates whether a model with that fixed effect accounts for significantly more variance in the dependent measure than the model without that effect. This is accomplished by comparing the two models' fits (i.e., the log-likelihood of each model producing the observed data), and the improvement achieved by the model with the extra parameter is weighed against the penalty incurred for increased model complexity. The p -value we report reflects the significance level of a χ^2 test ($df = 1$) comparing the residuals of the two models. We encountered occasional convergence difficulties during model comparison due to refitting (with maximum-likelihood (ML) instead of REML; *cf.* Bates et al, 2014); these issues were resolved by fitting both models with ML prior to calling the **anova** function.

Appendix D: Sentence Context Stimuli for Critical Items – Experiment 2

tent:

Biased contexts:

During the camping trip, we slept in a
 He needed help holding the poles to pitch the
 The fancy wedding reception was held under a

Neutral contexts:

The elderly woman thought she could see an enormous
 The worried traveller yelled upon seeing the other
 The lady caught a glimpse of the strange

test:Biased contexts:

After failing the quiz, he needed to pass the
 A study group crammed the night before the big
 The SATs are a very important college admissions

Neutral contexts:

The young boy was excited about the
 The popular gentleman did not know about the
 Nobody wanted to talk with her about that

peak:Biased contexts:

The exhausted climber had finally reached the
 The mountaineer yearned to stand on the
 The outdoorsman proudly planted his flag at the

Neutral contexts:

The curious scientist rigorously studied the interesting
 The textbook chronicled the gradual changes in the
 The class was captivated learning about the intriguing

peel:Biased contexts:

A ripe banana has a bright yellow
 When he finished eating his orange, he discarded the
 In the movie, the silly man slipped on a

Neutral contexts:

The foreigner was unsure what to do with the
The worker neglected to throw away the old
It had been years since he had eaten a

coat:*Biased contexts:*

The model sported her stylish new fur
The geeky professor wore his favorite tweed
His finicky mother buttoned the young boy's warm

Neutral contexts:

Only rarely had the fictional character touched the unique
The artist had a lot of trouble drawing the
The detective suddenly noticed the suspiciously missing

comb:*Biased contexts:*

The little girl gingerly held the barber's
The elderly gentleman managed his hair with the
The editor examined the draft with a fine-toothed

Neutral contexts:

The traveler brought with her a new
Surprisingly, his friend owned a very similar
The smart storeowner sold him a second

teen:*Biased contexts:*

The punk rocker was just another angsty
Between childhood and adulthood, someone is a
The elderly woman's slow driving annoyed the impatient

Neutral contexts:

She desperately wanted to sit down with the
The professor was offended by the actions of the
She apologized profusely after bumping into the irritated

teeth:*Biased contexts:*

The dentist announced that he would remove some
 Children do anything possible to avoid brushing their
 Evolving to eat meat, dogs have sharp canine

Neutral contexts:

The archaeologist dug up some apparently ancient
 He carefully examined a number of artificial
 The quirky collector became fascinated by the

toe:

Biased contexts:

She screamed in agony after stubbing her
 Preschoolers count to eleven using ten fingers and one
 The dancer's clumsy partner stepped on her

Neutral contexts:

The woman was surprised to see the same
 The photographer snapped a shot of the
 The golden retriever puppy was eager to sniff the

toast:

Biased contexts:

For breakfast, the chef slathered jam on the
 The best man clinked his glass before making his
 The new appliance kept burning her wheat

Neutral contexts:

The furious patron loudly and angrily demanded more
 The fussy child refused to even touch the
 The frightened busboy hurried to take away the

time:

Biased contexts:

The hurried commuter checked his watch for the sixth
 She anxiously twiddled her thumbs to pass the
 The broken clock consistently told the wrong

Neutral contexts:

While reading aloud he mispronounced the word

The young boy could not find the
The child didn't know how to spend the extra

type:

Biased contexts:

The editor set the title in boldface
He decided that the red-headed girl just wasn't his
The patient had diabetes, but the doctor didn't know which

Neutral contexts:

The pastor was unsure if he had the right
The forgetful uncle proclaimed to be open to any
She doubted that he was really the

pan:

Biased contexts:

The cook fried an egg in the
The homemaker poured vegetable oil on the non-stick
Miners sifted for gold in the river using a metal

Neutral contexts:

The young mother was totally displeased with the new
The consumer wanted to find out more about the
The popular magazine had an entire page about the

pants:

Biased contexts:

Though the boy preferred shorts, for church he wore
For the interview, he purchased new khaki
The neurotic businessman cautiously ironed his pleated

Neutral contexts:

The store's grand opening offered deals on typically pricey
Her eyes were drawn to the flamboyant, brightly colored
The picky customer was unimpressed by the mail-ordered

park:

Biased contexts:

The happy lovers took an evening stroll through the

Dozens of kids enjoyed a sunny day at the
There were beautiful vistas all through the national

Neutral contexts:

Researchers were interested in the organisms living in the
The cute puppy was not intimidated by the loud
The new invasive beetle was rapidly destroying the

part:

Biased contexts:

The mechanic couldn't fix the truck without the missing
While building his bike, the cyclist ordered a needed
The final scene of the romantic comedy was her favorite

Neutral contexts:

The frazzled bartender looked everywhere for the lost
The poor family struggled to afford the essential
It is rare that someone would find a loose

peach:

Biased contexts:

The most famous fruit grown in Georgia is the
The boy thought the nectarine was just an unripened
The cobbler would have been perfect with one more juicy

Neutral contexts:

The mayor posed for a picture next to the award-winning
The local journalist wrote an article about the now-famous
The chemist took some measurements while studying the

peace:

Biased contexts:

Armies are generally downsized during periods of
The warring countries finally signed an accord for
Rather than disturb the grave, she let the body rest in

Neutral contexts:

The engineer desperately just wanted to have some
The clinician spent many hours praying fervently for

They had forgotten what it was like to experience

coast:

Biased contexts:

Boston, Philadelphia and Orlando are along the eastern
The boat threw down its anchor a mile off the
Fishing and seaside tourism bolster local economies on the

Neutral contexts:

The artist's painting was his rendition of the well-known
He was interested in what townspeople thought of the
Late at night, the young boy dreamed about the forgotten

cone:

Biased contexts:

The child ordered his ice cream in a waffle
To guide drivers into the lane, there was an orange
A pointed, three-dimensional shape with a round base is a

Neutral contexts:

Standing in line, the child decided he would order a
The construction worker brought along with him a single
The office building downtown is shaped like a

card:

Biased contexts:

The referee had unfairly given the soccer player a red
The Hallmark store had the perfect Mother's Day
The ambitious consultant handed the analyst his business

Neutral contexts:

A shopper in the aisle of the supermarket saw a
The woman immediately ran out to go find another
She was embarrassed not to have noticed the

car:

Biased contexts:

A sleazy salesman sold the gullible new driver the used
Because she was speeding, the policeman pulled over her

With some difficulty, he learned to drive the

Neutral contexts:

She sat gazing out her window at the

He wasn't quite sure how to work the brand new

Forty research participants filled out a survey about the

cash:

Biased contexts:

The robber waited by the ATM to steal unfortunate people's

Some patrons had credit and debit, but couldn't pay with

Game show winners can take the prize or its value in

Neutral contexts:

He hesitated to admit that he had the well-hidden

The women around him were impressed by the

The immature 23-year-old showed off his

cast:

Biased contexts:

The television drama had a new actor added to its

Surgeons immobilized the worker's broken arm in a heavy

The movie premiere featured the director and the entire

Neutral contexts:

Nobody could remember if he was in the winning

For over three months, the old farmer wore the same

The high school students all wanted to join the

Highlights

- VOT production driven by neighborhood density, not specific voicing competitors
- Contextual predictability affects VOTs of words produced in sentences
- Neighborhood density effects persist irrespective of sentential context
- Results support cascading activation models of spoken word production

Figure 1.

Mean VOTs for target stimuli plotted as a function of $\text{NEIGHBORHOOD-DENSITY}$ (squares = MP words; circles = NMP words). The two gray traces indicate best-fit lines for regressions predicting VOT based on COMPETITOR-STATUS and $\text{NEIGHBORHOOD-DENSITY}$.

Figure 2.

For each item-pair, the magnitude of the difference in mean VOTs as a function of the difference of the neighborhood densities of the words comprising that item-pair.

Figure 3.

Mean VOTs for target stimuli plotted as a function of $\text{NEIGHBORHOOD-DENSITY}$, grouped by COMPETITOR-STATUS (squares = MP words; circles = NMP words) and by CONTEXT (+ = biased contexts; $\times$ = neutral contexts). Each target word appears twice in this figure (once each for the mean VOT when produced after biased and neutral sentence contexts). The two gray traces indicate best-fit lines for linear regressions predicting VOTs based on $\text{NEIGHBORHOOD-DENSITY}$ at each level of CONTEXT (collapsing over MP and NMP words).

Figure 4.

For each item-pair in each condition (biased vs. neutral sentence contexts), the magnitude of the difference in mean VOTs as a function of the difference of the neighborhood densities of the words comprising that item-pair.

Table 1

Parameters used to match critical target words in the MP and NMP conditions. Rows contain each variable's mean (standard error) for each list and a p-value (two-tailed paired t-tests comparing lists by item; $df = 11$).

Target Stimulus Variable	MP stimuli	NMP stimuli	<i>p</i> -value
Avg. Lexical Frequency	193.02 (160.73)	94.70 (40.56)	0.57
Avg. Log Lexical Frequency	1.47 (0.21)	1.58 (0.19)	0.66
Avg. Length (in phonemes)	3.25 (0.18)	3.5 (0.19)	0.39
Avg. Mean Segmental Probability	0.062 (0.004)	0.066 (0.005)	0.39
Avg. Sum Segmental Probability	0.207 (0.021)	0.238 (0.026)	0.26
Avg. Mean Bigram Frequency	0.006 (0.001)	0.007 (0.002)	0.46
Avg. Neighborhood Density	32.42 (2.28)	27.92 (2.86)	0.35
Avg. Neighborhood Density (weighted)	36.68 (5.95)	29.60 (4.09)	0.44

Table 2

Mean VOTs (and standard errors) for MP and NMP word lists with results of F_1 and F_2 Wilcoxon matched-pairs sign rank tests of the effect of COMPETITOR-STATUS on VOT.

	MP list	NMP list	Wilcoxon tests
F₁ (by subjects)	93 ms (5 ms)	92 ms (4 ms)	$V = 135, p > 0.27$
F₂ (by items)	93 ms (3 ms)	92 ms (3 ms)	$V = 44, p > 0.73$

Table 3

Sample sentence contexts in Experiment 2.

	BIASED	NEUTRAL
<i>coat</i>	The model sported her stylish new fur The geeky professor wore his favorite tweed His finicky mother buttoned the young boy's warm	Only rarely had the fictional character touched the unique The artist had a lot of trouble drawing the The detective suddenly noticed the suspiciously missing
<i>comb</i>	The little girl gingerly held the barber's The elderly gentleman managed his hair with the The editor examined the draft with a fine-toothed	The traveler brought with her a new Surprisingly, his friend owned a very similar The smart storeowner sold him a second

Table 4

Mean VOT (standard error) for MP vs. NMP stimuli, overall and by condition.

	F₁ (by subjects)		
	MP list	NMP list	Wilcoxon tests
OVERALL	86 ms (4 ms)	87 ms (4 ms)	V = 82, $p > 0.40$
BIASED	85 ms (4 ms)	86 ms (4 ms)	V = 73, $p > 0.24$
NEUTRAL	87 ms (4 ms)	88 ms (4 ms)	V = 96, $p > 0.75$
	F₁ (by subjects)		
	MP list	NMP list	Wilcoxon tests
OVERALL	86 ms (2 ms)	87 ms (2 ms)	V = 37, $p > 0.90$
BIASED	85 ms (2 ms)	86 ms (2 ms)	V = 34, $p > 0.73$
NEUTRAL	87 ms (3 ms)	88 ms (2 ms)	V = 35, $p > 0.79$

Table 5

Mean VOT (standard error) for words produced in biased vs. neutral contexts.

	BIASED	NEUTRAL	Wilcoxon tests
F₁ (by subjects)	86 ms (4 ms)	87 ms (4 ms)	$V = 51, p < 0.05$
F₂ (by items)	86 ms (2 ms)	87 ms (2 ms)	$V = 10, p < 0.03$
F₂ (by words)	86 ms (2 ms)	87 ms (2 ms)	$V = 67, p < 0.02$

Table B.1

Distribution of discarded data in Experiments 1 and 2. A cell in the experimental design summarizes the data for a subject/word/condition (20 subjects $\times$ 24 target words = 480 total cells per condition), and each cell could have 3, 2, 1 or 0 trials contributing to the cell's mean, depending on the number of discarded trials.

	<i>number of trials contributing to cell mean:</i>			
	3	2	1	0
Experiment 1 (Isolation: 480 cells)	460 (95.8%)	18 (3.8%)	2 (0.4%)	0 (0%)
Experiment 2 (Overall: 960 cells)	880 (91.7%)	72 (7.5%)	8 (0.8%)	0 (0%)
Experiment 2 (Biased Contexts: 480 cells)	437 (91.0%)	38 (7.9%)	5 (1.0%)	0 (0%)
Experiment 2 (Neutral Contexts: 480 cells)	443 (92.3%)	34 (7.1%)	3 (0.6%)	0 (0%)

Table B.2

Automated VOT measurement validation via comparison with manual measurements. Rows contain manually and automatically computed mean VOTs (standard error) for 7 subjects. Overall mean VOTs for each method average over subjects' mean VOTs (standard error of subject means). The results of nested model comparisons testing the effect of MEASUREMENT-METHOD on VOT measurements for each subject and overall appear on the right.

Comparison	Manual	Automated	Different?
Subject 1	94 ms (2 ms)	94 ms (2 ms)	$\chi^2(1) < 0.01, p > 0.99$
Subject 2	94 ms (2 ms)	94 ms (2 ms)	$\chi^2(1) = 0.03, p > 0.87$
Subject 3	77 ms (2 ms)	74 ms (2 ms)	$\chi^2(1) = 0.80, p > 0.37$
Subject 4	97 ms (2 ms)	94 ms (2 ms)	$\chi^2(1) = 1.53, p > 0.21$
Subject 5	95 ms (2 ms)	95 ms (2 ms)	$\chi^2(1) = 0.09, p > 0.76$
Subject 6	81 ms (2 ms)	80 ms (2 ms)	$\chi^2(1) = 0.20, p > 0.65$
Subject 7	89 ms (2 ms)	89 ms (2 ms)	$\chi^2(1) = 0.05, p > 0.81$
Overall	89 ms (3 ms)	89 ms (3 ms)	$\chi^2(1) = 0.53, p > 0.46$

Table B.3

Fixed effects parameter estimates (β) and p-values for identical linear mixed effects models fit to VOT data that was measured automatically and manually. Fixed and random effects structures were identical to the model described in Experiment 1's Results.

	Automated		Manual	
	β	p	β	p
COMPETITOR-STATUS	-0.597	$p > 0.48$	-1.088	$p > 0.27$
NEIGHBORHOOD-DENSITY	0.283	$p < 0.005$	0.381	$p < 0.001$
COMPETITOR-STATUS $\times$ NEIGHBORHOOD-DENSITY	-0.534	$p < 0.02$	-0.440	$p < 0.04$