



Published in final edited form as:

Lang Cogn Process. 2009 May 1; 24(4): 527–554. doi:10.1080/01690960802299378.

Mechanisms of interaction in speech production

Melissa Baese-Berk and Matthew Goldrick

Department of Linguistics, Northwestern University

Abstract

Many theories predict the presence of interactive effects involving information represented by distinct cognitive processes in speech production. There is considerably less agreement regarding the precise cognitive mechanisms that underlie these interactive effects. For example, are they driven by purely production-internal mechanisms (e.g., Dell, 1986) or do they reflect the influence of perceptual monitoring mechanisms on production processes (e.g., Roelofs, 2004)? Acoustic analyses reveal the phonetic realization of words is influenced by their word-specific properties—supporting the presence of interaction between lexical-level and phonetic information in speech production. A second experiment examines what mechanisms are responsible for this interactive effect. The results suggest the effect occurs on-line and is not purely driven by listener modeling. These findings are consistent with the presence of an interactive mechanism that is online and internal to the production system.

A great deal of research has documented the presence of interactive effects—the interaction of distinct types of information represented by different cognitive processes—within both speech production and perception. The Ganong effect (Ganong, 1980) is a prototypical example; it illustrates the interaction of lexical and phonetic category information in speech perception. When presented with syllables whose initial consonants vary along a voice-onset time (VOT) continuum, listeners' identification of the initial phoneme is sensitive to whether the resulting syllable is a word or nonword. This is an interactive effect because the two types of information are assumed to be encoded by distinct processing stages in speech perception. The lexicality of a sound sequence is represented within lexical level processes, while the structure of phonetic categories is represented by pre-lexical processes (see McClelland, Mirman, and Holt, 2006, for a recent review of architectures based around this perspective; but see Gaskell and Marslen-Wilson, 1997, for an alternative perspective).

Within the domain of speech production, many studies have focused on interactive effects involving semantic and phonological information. Most speech production theories assume that at least two distinct stages (representing different types of information) are involved in mapping semantic representations onto abstract, long-term memory representations of word form (e.g., mapping 'flying nocturnal mammal' to "bat"). The first processing stage involves selection of a word to express an intended concept, and the second involves retrieval of sound information corresponding to the selected word from long-term memory (Garrett, 1980). We refer to the former as lexical selection and the latter as lexical phonological processing. Critically, semantic information is represented within lexical selection processes while phonological information is represented within lexical phonetic processes. Chronometric studies, spontaneous speech error analyses, and studies of individuals with neurological impairment have all documented interactive effects in speech production involving semantic and phonological information (see Goldrick, 2006, for a recent review). This large body of

evidence suggests that debates between speech production theories can shift from the question of whether interactive effects exist to the question of how such effects arise.

This paper aims to add to the growing body of evidence for interactive effects in speech production, particularly in the area of lexically conditioned phonetic variation— how whole-word properties can affect the fine-grained phonetic realization of that word. Within this context, the study seeks to differentiate between three mechanisms of interaction which could all potentially account for the observation of lexically conditioned phonetic variation. The remainder of the paper is organized as follows. First, we will present recent studies of lexically conditioned phonetic variation. Next we will more precisely define the mechanisms of interaction examined in this study, and demonstrate that the existing results do not differentiate between these mechanisms of interaction. We will then present results from two experiments that aim to distinguish these various accounts. Finally, we discuss the implications of these results for current theories and future work regarding interaction.

Lexically conditioned phonetic variation

As noted above, production theories typically assume at least two processing stages are involved in the retrieval of sound information from long term memory: selection of a word to express an intended concept (lexical selection) and retrieval of phonological information from long term memory (lexical phonological processing). Form-related properties specific to lexical items—e.g., how frequent the word is, its formal relationships to other words in the lexicon—are encoded by these latter processes. Most theories further assume that the phonological information stored in long term memory is relatively abstract and must be elaborated prior to accessing motor processes for production (see Goldrick and Rapp, 2007, for a recent review). We refer to this third stage as phonetic processing. Recently, several studies have suggested the presence of interactive effects involving information represented in these processes. Specifically, word-specific phonological properties interact with phonetic properties in speech production.

This work focuses on one aspect of word-specific information encoded by lexical-phonological processes: neighborhood density, the number of words in a speaker's lexicon phonologically related¹ to a target word. Several studies have suggested that this lexical property affects the fine-grained phonetic realization of words (see Pluymaekers, Ernestus, and Baayen, 2005; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; for examinations of how other lexical properties affect phonetic realization).

Wright (1997, 2004) examined the vowel spaces of 'hard' and 'easy' words. Roughly defined, hard words are from dense neighborhoods with many high-frequency neighbors. Easy words are words from sparse neighborhoods with primarily low frequency neighbors. Wright demonstrated that vowels in hard vs. easy words are realized with more extreme phonetic properties (specifically, they are more dispersed in F1-F2 space). These results suggest that lexical phonological properties (i.e., neighborhood density and frequency) are affecting the phonetic realization of words. Munson and Solomon (2004) examined the effect of neighborhood density separate from that of lexical frequency and found vowel space expansion for words in high- vs. lowdensity neighborhoods (see also Munson, 2007). Finally, Scarborough (2003, 2004) also examined vowels of 'hard' and 'easy' words. Instead of vowel space measures, she focused on coarticulation (e.g., anticipatory nasalization of the vowel in

¹The studies reviewed below assume a categorical definition of phonologically related— specifically, all words related by the addition, substitution, or deletion of a single phoneme (see Vitevitch, 2002, for additional discussion of this measure in the context of speech production). In contrast, spreading activation theories tend to assume a more graded definition of neighborhood structure—where various non-target words are partially activated based on their degree of phonological overlap with the target. (see Goldrick and Rapp, 2001, for further discussion).

'ban'). She found that vowels in hard words were produced with a higher degree of coarticulation than those in easy words.

While these studies do suggest an influence of neighborhood density on phonetic properties of words, there are both empirical and theoretical issues remaining². Empirically, the studies did not control for phonotactic probability (i.e., the likelihood a particular segment occurs in a particular position or together with another particular segment). Bailey and Hahn (2001) demonstrated that phonotactic probability is highly correlated with neighborhood density. Vitevitch, Ambrüster and Chu (2004) showed that phonotactic probability influences production independent of neighborhood density. Finally, Goldrick and Rapp (2007) present evidence that phonotactic probability effects arise specifically at the phonetic level of processing. The apparently "lexically" conditioned effects reviewed above may therefore be conditioned not by high neighborhood density but rather by high phonotactic probability. Rather than reflect interaction between lexical and phonetic level properties, this variation could simply reflect effects limited to phonetic level processing itself. Experiment 1 aims to deconfound these two possibilities by examining whether lexically conditioned phonetic variation occurs when phonotactic probability is controlled. Additionally, Experiment 1 extends previous work on lexically-conditioned phonetic variation by examining words with a minimal pair neighbor rather than the effects of neighborhood density more generally.

Even if the effect is truly due to the interaction of lexical and phonetic properties, the data discussed above cannot adjudicate between three possible accounts of these interactive effects. As outlined below, each of these proposes contrasting types of mechanisms to account for lexically conditioned phonetic variation.

Possible mechanisms underlying lexically conditioned phonetic variation

Speaker internal account

The speaker internal account of lexically conditioned phonetic variation postulates direct interaction between speaker-internal production processes, by which one production process influences the organization and structure of processing within another production process. Cascading activation, for instance, allows for multiple activated representations – not just the output of a process – to exert an influence on subsequent processes (Goldrick, 2006; Peterson and Savoy, 1998). Feedback allows processes occurring relatively late in the production process to influence those occurring earlier (e.g., Dell, 1986; Rapp and Goldrick, 2000).

In the case of lexically conditioned phonetic variation, we can consider the influence of cascade and feedback at two levels of the system: lexical phonological and phonetic. At the lexical phonological level, cascading activation allows multiple active lexical representations to influence processing—not just the target (e.g., CAT) but also its formally-related lexical neighbors which have been activated via feedback from the target's phonology (e.g., feedback from /ae/ and /t/ activates BAT, HAT, etc.). These lexical neighbors activate not only phonological structure they share with the target (e.g., BAT, HAT provide additional support for the /ae/ and /t/ of CAT) but also non-target phonological representations (e.g., /b/, /h/). These non-target representations compete with the target's phonological representation for selection—a competition that is particularly acute for segments with a high degree of featural similarity (e.g., /k/ and /g/ compete with one another more strongly than /k/ and /d/; Meyer and Gordon, 1985; Yaniv, Meyer, Gordon, Huff, and Sevald, 1990). If we assume that this competitive process requires the target representation to become highly active (so as to

²Note that Goldrick and Rapp (2007) failed to find any influence of lexical variables on categorical measures of phonetic processing (accuracy and categorical properties of error outcomes). It is possible that interaction between these processes is limited such that interactive effects are limited to sub-categorical modulations of fine-grained phonetic aspects of production.

effectively inhibit its competitors), we would predict a boost in the activation level of the lexical phonological representation of words in high density neighborhoods. In contrast, in low density neighborhoods non-target phonological representations would receive less activation. Due to this reduction in competition, there would be less of a need to boost the phonological representations of the target, leading to lower activation levels compared to words in high density neighborhoods.

Turning to the phonetic level, cascade allows this competition-induced contrast in activation at the lexical phonological level to affect the activation of phonetic representations. The higher activation level for words in dense neighborhoods will lead to more active phonetic representations and consequently more extreme articulatory realizations. In contrast, when a target has very few neighbors its lexical phonological and phonetic representations will be less active and consequently produced with a relatively less extreme articulation. In this way, the combined effects of cascading activation, feedback, and competition between phonological representations provides a production-internal interaction account of lexically conditioned phonetic variation. We emphasize here the *production-internal* aspect of this account; according to this hypothesis, lexically conditioned phonetic variation is solely a reflection of the structure of processes within the production system.

Perceptual monitoring

Alternatively, interactive effects could be explained by monitoring. Rather than focusing on the interrelationship of two processes internal to the production system, a monitoring account relies on interaction between perception and production processes. For example, speech errors are generally more likely to result in word errors (e.g., hat->cat) than in nonword errors (e.g., hat->zat; Dell, 1986). To account for this tendency, a perceptual process could monitor the output of lexical phonological processing in order to ensure that all outputs are words. If a nonword is perceived, the monitor could halt and restart the production process (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 2004).

The discussion here uses the term monitoring to encompass not only ‘supervising’ one’s own speech during production, but any on-line listener modeling process. For example, as a speaker is talking, she could model what the listener knows and attempt to alter her speech in order to best aid her listener (e.g., making it easier to perceive; Wright, 1997, 2004). Critically, both senses of monitoring rely on interaction between the perceptual and production systems (as proposed by Levelt, Roelofs, and Meyer, 1999).

According to a monitoring account, if a speaker knows a listener is potentially confused by an utterance he will amend his production in order to make it minimally confusing. In the case of lexically conditioned phonetic variation, words in dense neighborhoods are similar to many other words; this makes perception of such targets more difficult than that of words in sparse neighborhoods (see Luce and Pisoni, 1998, for a review). According to a monitoring account, this perceptual difficulty motivates lexically conditioned phonetic variation. Additionally, it is the case that listeners are sensitive to very small phonetic differences in perception. McMurray, Tanenhaus, and Aslin (2002) used an eye-tracking paradigm to demonstrate that listeners are sensitive to very small within-category differences in VOTs. Participants looked longer at competitor pictures with the opposite voicing specification than the target when the target was produced with a slightly less distinct VOT. This suggests that small, within-category variations in VOT may modulate lexical competition effects within perception³. (For

³As noted by a reviewer, the phonetic variation reported in this study (5% of VOT) is substantially smaller than that typically examined in the perceptual studies cited above. Although no study has examined the perceptual consequences of VOT variation of the specific type and magnitude we report here, we believe that the substantial body of results cited above make the perceptual monitoring account a plausible hypothesis. (Although, as shown by our results, ultimately incorrect; see below.)

converging evidence that relatively small degrees of variation along this and other phonetic dimensions have measurable consequences for lexical processing in perception, see Andruski, Blumstein, & Burton, 1994; Salverda, Dahan & McQueen, 2003; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, & McDonough, 2007.) Therefore, in order to make words more distinct from their neighbors, and hence less confusable, monitoring processes cause the speaker to hyperarticulate. In contrast to the speaker internal interaction account, then, this particular hypothesis claims that lexically conditioned phonetic variation is motivated by perceptual processes—specifically, the need to modulate perceptual competition between lexical neighbors in the listener.

Perceptual restructuring

Interactivity could also be the result of perceptual restructuring. In these types of accounts, a perceptual process actually shapes and reorganizes production processes. For instance, in some exemplar approaches, listeners update their production targets based on specific perceptual episodes of lexical items (Pierrehumbert, 2002). Since these are perceptual episodic memories, they can be affected by properties of the perceptual system. As these perceptually-modulated memories are then used as targets for production, they provide an off-line mechanism supporting an interactive effect. Through learning, information represented within the perceptual system interacts with the production system.

Pierrehumbert (2002) describes an exemplar model of production and perception that can account for lexically conditioned phonetic variation. In this account, as learners are listening to speech, they store every exemplar of a particular word with its specific phonetic properties. Critically, a word will only be correctly stored if it is perceived correctly by the listener. For instance, if a word in a dense neighborhood is produced with a very centralized vowel it may fall in the distribution of normal pronunciations of other words in its neighborhood. Therefore, in the presence of a neighbor with a similar vowel, only those exemplars with extreme vowels would be stored correctly. As this process repeats over many instances of perception and production, most of the stored exemplars associated with words in dense neighborhoods would be extreme examples of vowels, maximally separating this word's phonetic distribution from that of neighboring words. In contrast, words in sparse neighborhoods would be much less susceptible to such confusions; therefore, there would be little pressure for them to be stored with extreme vowel exemplars. This yields the observed lexically-conditioned phonetic variation effect: a word in a dense neighborhood would be stored and produced with a more extreme articulation than a word in a sparse neighborhood (see Pierrehumbert, 2002, for further discussion). Like the perceptual monitoring account, then, this account assumes that the ultimate driving force for lexically conditioned phonetic variation is perceptual (i.e., the need to modulate lexical competition effects in the listener). Both of these accounts therefore stand in contrast to the speaker internal account which does not accord any role to the perceptibility of the phonetic variation.

To summarize, the existing evidence on lexically conditioned phonetic variation can be accounted for by all three mechanisms of interaction. In Experiment 2, we attempt to disambiguate between these three possibilities.

Distinguishing mechanisms of interaction

The first dimension that distinguishes these three theories is whether lexically conditioned phonetic variation is an online or offline process. The perceptual restructuring hypothesis predicts that this variation occurs primarily offline, meaning it does not rely on online interactions between production and/or perceptual processes. Instead, the theory relies on a permanent restructuring of production processes. In contrast, the perceptual monitoring and

the speaker-internal interaction hypotheses both rely on online processes to account for lexically conditioned phonetic variation.

In order to differentiate between these two types of hypotheses, we placed target words in a context with their closely related neighbor (see Table 1 for examples of contexts in which words were produced). Specifically, participants were asked to provide instructions to click on a particular word within a set of three items containing a closely related neighbor (e.g., produce “Click on the *cod*” in the Context Condition in Table 1). The perceptual restructuring hypothesis predicts that even when the words are placed in context with closely related neighbors their productions should be no more hyperarticulated compared to productions of the same target in contexts without the neighbor (e.g., in the No Context Condition in Table 1). According to this account, participants have simply learned to always hyperarticulate these forms. In contrast, the other two hypotheses both predict that productions of targets should be more extreme when they are placed in a context with their neighbor. The monitoring hypothesis predicts that contexts containing two very similar words are more confusing to listeners (Luce and Pisoni, 1998). Therefore, in order to help listeners understand their speech, speakers hyperarticulate to maximally differentiate the target from the neighbor. The production-internal interaction model predicts that the target’s neighbor will become more highly active when the neighbor is present vs. absent in the visual context. This will increase competition between target and non-target phonological representations, leading to a further boost in the activation of the target. Via cascade, this increased activation of lexical phonological representations will result in even greater hyperarticulation relative to cases where the neighbor is not explicitly present in the context.

These theories also differ along a second dimension: whether lexically conditioned phonetic variation can be attributed to listener modeling or not. The monitoring hypothesis predicts that listener modeling plays a critical role in producing the interactive effects discussed above. In contrast, the perceptual restructuring hypothesis and the speaker-internal production hypothesis do not predict a role for online listener modeling.

In order to differentiate between the theories along this dimension, we placed items with and without a very closely related neighbor in small closed-set contexts of three items. Specifically, participants were asked to provide instructions to click on a particular word within a set of three items (e.g., produce “Click on the *cod*” in the No Context Condition in Table 1). Clopper, Pisoni, and Tierney (2006) demonstrate that in perceptual tasks where potential targets are limited to a very small closed set, listener confusion caused by lexical competition decreases compared to large closed set and open set tasks (see also Sommers, Kirk, and Pisoni, 1997, for similar findings). This leads us to predict that in a small closed set task, listeners would not be confused if the neighbor is absent from the context. Therefore, according to the monitoring hypothesis, speakers would not need to hyperarticulate; they will predict that listeners will not be confused. On the other hand, the speaker-internal production and perceptual restructuring theories predict that speakers will still hyperarticulate in this context. The latter hypothesis predicts that production processes have been permanently restructured such that words with a closely related neighbor will always be hyperarticulated compared to words without such a neighbor. Therefore, regardless of the context, these words will be hyperarticulated. The speaker-internal production hypothesis also predicts hyperarticulation in closed sets. Feed-forward and feedback activation mechanisms will automatically operate even within a closed set context. The heightened activation of the target induced by this pattern of activation flow will result in hyperarticulation.

These two experimental contexts should allow us to distinguish the various mechanisms that could account for lexically-conditioned phonetic variation.

- If lexically conditioned phonetic variation is primarily an offline process, a word that is presented with its neighbor should be produced with the same VOT as when it is presented without the neighbor. In contrast, if there are online contributions to this variation, context should modulate it. Words presented with a neighbor should be produced with longer VOTs than the same words presented without a neighbor. This effect would be contrary to the predictions of the perceptual restructuring hypothesis, but in line with the speaker-internal production and listener-modeling hypotheses.
- If this variation can be attributed solely to listener modeling, it should not appear when the words are presented without their neighbor in a closed set. However, if listener modeling is not the only cause of this effect, words with minimal pairs should be hyperarticulated, even when they are presented without their neighbor. This effect would be contrary to the predictions of the listener-modeling hypothesis, but in line with the speaker-internal production and perceptual restructuring hypotheses.

The remainder of the paper will focus on the results from two production experiments aiming to examine how lexically conditioned phonetic variation behaves along these two dimensions.

Experiments 1a and 1b

As previously mentioned, Experiment 1 aimed to alleviate some of the methodological issues in previous studies of lexically conditioned phonetic variation. Specifically, it examined whether lexically conditioned phonetic variation occurs independent of phonotactic probability and other phonetic level factors such as length, which previous studies failed to control for. Additionally, we extended previous studies of this phenomenon by examining consonant production, rather than vowel production. We contrasted words with initial voiceless stops that have a minimal pair neighbor (e.g., ‘cod,’ with neighbor ‘god’) and matched words that do not have a minimal pair neighbor (e.g., ‘cop,’ with no neighbor ‘*gop’). Because the words with a minimal pair neighbor vary from their neighbor only by the voicing of their initial consonant, examining the voice onset time (VOT) of the initial consonant allowed us to determine whether there were any differences, or lexically conditioned phonetic variation, between these groups of words. Experiment 1a examined /p/-initial targets and Experiment 1b used different subjects to examine /t/- and /k/-initial targets.

Experiment 1a: /p/-initial targets

Method

Participants—Twelve Northwestern University undergraduates (6 males and 6 females) participated in this study. All participants were native English speakers. Ten participants were right-handed and two were left-handed. Two participants were bilingual; however, their performance was within the normal range of the monolingual participants. All participants received partial course credit for their participation in this experiment.

Materials—The stimuli used in this experiment were pairs of monosyllabic words beginning with voiceless stop /p/. All pairs were matched for initial consonant and vowel. Pairs were additionally matched for sum segmental probability, sum biphone probability, and phoneme length (see Table 2). All words were low frequency (i.e., less than 20 per million) and each pair was matched for lexical frequency. Phonotactic probability measures were determined using the Phonotactic Probability Calculator (Vitevitch and Luce, 2004). Lexical frequency was determined using the CELEX database (Baayen, Piepenbrock, and Gulikers, 1995).

Within each pair, one member had a minimal pair neighbor when the initial consonant was voiced (e.g., ‘pox’ with neighbor ‘box’). The other member of the pair had no such neighbor

(e.g., 'posh', no neighbor *'bosh'). A total of 16 /p/-initial pairs were included as target items in this experiment.

Critical words were embedded in a list with monosyllabic fillers. One third of each list consisted of target items (e.g., 'pox,' 'posh'). The minimal pair neighbor of the target items (e.g., 'box') was not presented to participants. One quarter of the list consisted of non-stop initial words (e.g., 'fish', 'lamp', 'yell'). Another quarter consisted of words beginning with non-labial voiced and voiceless stops (e.g., 'gang', 'terse', 'dirt'). The remainder of the list consisted of 8 /p/-/b/ minimal pairs (e.g., 'pelt'-'belt').

Procedure—Stimuli were presented in a self-paced reading task. One word appeared in the center of a computer screen. Participants were asked to read each word aloud and to press a button on a button box to advance to the next trial. The list was presented three times in a different random order. The three repetitions were included in order to increase the number of tokens and reduce the influence of random variance within participants. Because the tokens were embedded in a large list of fillers, and because the lists were randomized for each repetition, it is unlikely that this repetition would encourage any strategic effects. There was a short break between each repetition of the list.

The data were recorded using a headset microphone and a flash memory recorder at 22.05 kHz in a large single-walled booth. The sound files were then transferred from flash memory to hard disk for phonetic analysis.

Acoustic and statistical analysis—Praat (Boersma and Weenink, 2002) was used for all measurements. Voice onset time was measured for each token from the stop burst on the waveform to the first zero crossing after the onset of periodicity.

Tokens were not included in the analysis if the word was mispronounced or disfluent. If the participant coughed or yawned during recording of the token, or if any external noise (e.g., clicking the button on the button box) disrupted recording, the token was also not used in the analysis. 1.5% of tokens were excluded from the analysis. For each participant, the average of each of three repetitions for each token was calculated. This average across repetitions was used in the analysis. (The effects reported below were consistent across each repetition.)

All tokens were measured by a research assistant and 37.5% of the tokens were checked for reliability by the first author. The differences in measurements between the two experimenters were not significant ($V=1170$, $p>0.5$). Voice onset time results were submitted to Wilcoxon Matched-Pairs sign rank tests, due to the non-normal distribution of our data. Two analyses were performed. A by-participant analysis compared the mean VOT for each participant across words with a minimal pair neighbor to the mean VOT for the same participant for words without a minimal pair neighbor. A by-items analysis compared the mean VOT across participants for each word with a minimal pair neighbor (e.g., the mean VOT for 'pox' across participants) to the mean for the matched word with a non-minimal pair neighbor (e.g., the mean VOT for 'posh' across participants).

Results and Discussion

A by-participants Wilcoxon Matched-Pairs sign rank test demonstrated that the VOT of words with minimal pair neighbors was significantly longer than words without minimal pair competitors (words with minimal pair competitors 68.9 msec; words without minimal pair competitors 65.7 msec; $V=87$, $p<0.002$). A by-items Wilcoxon test demonstrated a similar pattern, but was not significant ($V=91$, $p=0.17$). While the by-items analysis is not significant, the general pattern does trend in the same direction as the by-participant results.

These results provide some preliminary evidence that lexically conditioned phonetic variation occurs independent of variables that influence phonetic processing (e.g., phonotactic probability, length). In the next experiment, we examined if these results would generalize to other places of articulation (coronal and dorsal). In addition, to minimize possible strategic effects, we eliminated minimal pairs from the list of fillers.

Experiment 1b: /t/ and /k/-initial targets

Method

Participants—Ten Northwestern University undergraduates (8 males and 2 females) participated in this study. All participants were native English speakers. Eight participants were right-handed and two were left-handed. One participant was bilingual; however, his performance was within the range of the monolingual participants. The order of each condition was balanced across participants. As in Experiment 1a, all participants received partial course credit for their participation in this experiment.

Materials—As in Experiment 1a, the stimuli used in this experiment were pairs of monosyllabic words varying as to whether they had a minimal pair neighbor when the initial consonant was voiced. Separate lists of stimuli were created for words beginning with voiceless stops /t/ and /k/. All pairs were matched for initial consonant and vowel, as in Experiment 1a. Pairs were additionally matched for sum segmental probability, sum biphone probability, and phoneme length (see Table 3 and Table 4). All words were low frequency (i.e., less than 20 per million) and each pair was matched for lexical frequency. A total of 19 /t/-initial and 12 /k/-initial pairs were included as target items in this experiment.

Critical words were embedded in a list with approximately twice as many monosyllabic fillers (/t/ condition: 78 fillers; /k/ condition: 50 fillers). Separate lists were created for each initial stop. Fillers were chosen separately for each list. Approximately one half of the fillers had initial stops, evenly distributed across 5 non-target stop consonants (/p, b, d, g/ plus either /t/ or /k/; 25 fillers for the /k/ list and 40 fillers for the /t/ list). The remainder of the fillers began with non-stop consonants (e.g., heat, thumb). Minimal pair neighbors for target words were not included in any list. No fillers rhymed with targets⁴. Additionally, in this experiment, none of the stop initial fillers had minimal pair neighbors.

Procedure—Stimuli were presented and recorded as described in Experiment 1a.

Acoustic and statistical analysis—Results were measured and analyzed as described in Experiment 1a.

Again, tokens were not included in the analysis if the word was mispronounced or disfluent. If the participant coughed or yawned during recording of the token, or if any external noise (e.g., clicking the button on the button box) disrupted recording, the token was also not used in the analysis. 2.1% of tokens in the /t/-initial condition and 0.1% of tokens in the /k/-initial condition were excluded.

For each participant, the average of each of three repetitions for each token was calculated. This average across repetitions was used in the analysis. (The effects reported below were consistent across each repetition.)

⁴Three participants were run with two fillers in the /t/-initial list that rhymed with targets in the /k/-initial list.; these were replaced for subsequent participants.

All tokens in Experiment 1b were measured by the first author and 6% of the tokens were checked for reliability by another experimenter. The differences in measurements between these two experimenters were also not significant ($V=412$, $p>.98$).

Statistics were performed as in Experiment 1 above.

Results

A by-participants Wilcoxon Matched-Pairs sign rank test demonstrated that for each place of articulation, the VOT of words with minimal pair neighbors was significantly longer than words without minimal pair competitors (/t/-initial words with minimal pair competitors 84.3 msec; words without minimal pair competitors 80.3 msec; $V=51$, $p<0.02$; /k/-initial words with minimal pair competitors 95.5 msec; words without minimal pair competitors 90.6 msec; $V=50$, $p<0.02$). A by-items analysis for each place of articulation demonstrated similar results though the Wilcoxon test failed to reach significance for the /k/-initial list (/t/-initial list $V=161$, $p<0.007$; /k/-initial list $V=62$, $p=0.07$). Although the by-items analyses in the /k/-initial list and in Experiment 1 do not reach significance, the results do trend in the correct direction.

Considering both Experiments 1a and 1b, the VOT difference between words with minimal pairs and words without minimal pairs can be visualized both in terms of differences in raw values (Figure 1) as well as ratios (Figure 2). Although the raw values suggest contrasting patterns across place of articulation, these appear to reflect differences in baseline VOTs. As shown in Figure 2, VOT increases by roughly the same proportion for each place of articulation between words without minimal pairs and words with minimal pairs.

Additionally, although the by-items analyses did not reach significance at every place of articulation, when we consider the results across both Experiments 1a and 1b the effect is consistent across items. A by-item analysis including all items across these two experiments was significant ($V=901$, $p<0.0005$, mean VOT of words with minimal pair competitors: 83 msec; mean VOT of words without minimal pair competitors: 78.9 msec)⁵. Additionally, a by-subject analysis including all subjects across both experiments is also significant ($V=222$, $p<0.01$, mean VOT of words with minimal pair competitors: 79 msec; mean VOT of words without minimal pair competitors: 76 msec).⁶

Discussion

Words with minimal pair neighbors were produced with longer VOTs than words without. This specific form of lexically conditioned phonetic variation occurs independent of phonotactic probability and other lexical-phonological and phonetic level properties, which previous studies did not control for. Additionally, these results extend findings from previous studies of lexically conditioned phonetic variation in vowels by demonstrating that phonetic properties of consonants are also sensitive to lexical properties of words. Finally, the proportional increase in VOT is approximately the same across all places of articulation. This consistency would be unexpected if the effect was driven by purely phonetic properties.

This pattern of results provides further support for the claim that some aspects of phonetic variation are indeed conditioned by lexical properties of words (in particular, neighborhood density). However, it does not differentiate between the three aforementioned mechanisms of interaction. Experiment 2 addresses this question.

⁵It should be noted that this analysis is slightly unconventional as it combines across subjects from two different experiments.

⁶As noted with the previous analysis, this analysis is unconventional as it combines subjects across two experiments. Additionally, the mean VOTs for participants in Experiment 1b collapse across both the /t/- and /k/-initial lists, whereas the mean VOTs for participants in Experiment 1a are from the /p/-initial list alone.

Experiment 2

Experiment 2 embedded the target stimuli in a conversational context. Target words with minimal pairs were presented in contexts that either included or did not include the minimal pair neighbor in order to differentiate the accounts outlined in the introduction.

If lexically conditioned phonetic variation is produced because of offline, learned effects, words presented with their minimal pair neighbor (e.g., target ‘cod’ presented with its neighbor ‘god’) should be produced with the same VOT as these words when they are presented without their minimal pair neighbor (e.g., target ‘cod’ presented in a closed set without its neighbor ‘god’). If this variation is produced online during lexical access, context should modulate the effect. Therefore, under online hypotheses, words presented with their minimal pair neighbor should be produced with longer VOTs than the same words presented without their neighbor.

If lexically conditioned phonetic variation is due to listener modeling, the variation should disappear when the words in minimal pairs are presented without their neighbor. That is, when the word ‘cod’ is presented in a closed-set without its neighbor ‘god’ and there is no potential confusion between the minimal pair and its competitor, the VOT of words with and without minimal pairs should be equal. However, if listener modeling does not drive this effect, the effect from Experiment 1 should remain, even in the condition where words with minimal pairs are presented without their neighbor.

Method

Participants—Twelve Northwestern University undergraduates (4 males and 8 females) participated as speakers in this experiment for partial course credit. One participant was left-handed. All participants were native English speakers, and two were bilingual. The bilinguals did not differ significantly from the monolinguals. Listeners in the experiment were undergraduate research assistants who acted as confederates.

Materials—Of the stimuli in Experiment 1, the 12 best-matched pairs were chosen as targets for each place of articulation, for a total of 36 pairs of targets (see Table 5 for matching statistics). This was done in order to equalize the number of targets for each place of articulation. These targets were embedded in a list of monosyllabic fillers, described in more detail below.

Procedure—Speakers and listeners sat at a table facing each other. Each also faced a computer monitor. Three words simultaneously appeared on each screen. After 1000 ms, a box appeared around one of the words on the speaker’s screen only. The speaker instructed the listener to ‘Click on the *target word*’ (where the target word stands for the word in the box). After the listener clicked the mouse, the trial advanced. The speaker was recorded during the entire experiment. Additionally, the speakers and listeners were asked not to talk to each other, other than the speaker using the command ‘Click on the *target word*.’

Trials were presented in three conditions: Context, No Context and No Competitor. In the Context condition, both a target and its minimal pair neighbor were presented on the screen with a filler word as the third word on screen. In the No Context condition, the targets from the Context Condition were presented on screen without the minimal pair. Instead, two unrelated fillers were presented on screen with the target. In the No Competitor condition, words with no competitors were presented with two unrelated fillers. Examples of each trial type are presented in Table 1.

Words without competitors were always presented in the No Competitor condition, and were presented to every participant. Words with competitors were presented either in the Context

condition or the No Context condition. Assignment of words with competitors to the two conditions was counterbalanced across participants. Participants saw half of the targets with competitors in the Context condition and half in the No Context condition. Target trials were mixed with filler trials. The location of the target item on the screen was also counterbalanced. Each speaker completed the list three times in three random orders. As in Experiment 1, the lists were repeated three times in order to reduce the effect of random variance across productions of the same word (the results were the same across each repetition).

Acoustic and statistical analysis—VOT was measured in the same way as in Experiment 1. Tokens were excluded for mispronunciations and other disruptions to recording. 1% of tokens in the /p/- and /t/-initial conditions and 1.4% of tokens in the /k/-initial condition were excluded. 1.2% of all tokens were excluded from the analysis. As in Experiment 1, the participant's three repetitions of each token were averaged together during analysis. VOT results were submitted to Wilcoxon Matched-Pairs sign rank tests for both by-participants and by-items analyses, as in Experiment 1.

Results

When collapsing across places of articulation, both by-participant and by-items analyses demonstrates that words presented in the Context condition (82.5 msec) were produced with significantly longer VOTs than the same words presented in the No Context condition (77.4 msec; by-participant Wilcoxon test, $V=68$, $p<.025$; by-items Wilcoxon test, $V=479$, $p<.007$). Additionally, words in the No Context condition were produced with significantly longer VOTs than words in the No Competitor condition (72.4 msec; by-participant Wilcoxon test $V=3$, $p<.003$; by-items Wilcoxon test $V=510$, $p<.005$). These results are shown in Figure 4. Within each place of articulation, the same numerical pattern appears. These results are shown in terms VOT Ratios (VOT in Context or No Context/No Competitor) in Figure 5.

Figure 5 suggests that the increase between conditions gets smaller across place of articulation. While the increase is not numerically as large for /k/-initial words as for /p/-initial words, there is no significant difference in the ratios in the /p/-initial and /k/-initial conditions (Context Competitor ratio comparison: $t(11)=1.91$, $p=.07$; No Competitor ratio comparison: $t(11)=1.21$, $p=.20$). It is possible that this numerical difference is due to a ceiling effect. Participants may only be capable of lengthening their VOTs to a certain value given the phonetic structure of English consonantal categories. Since /p/ has a shorter intrinsic VOT than /k/ (Volaitis and Miller, 1992) participants may have more room to lengthen their VOTs, before reaching this maximum value.

Discussion

When words are presented with their minimal pair neighbor, they are produced with longer VOTs than when they are presented in a limited context without their minimal pair neighbor. This implies that lexically conditioned phonetic variation is modulated online. The perceptual restructuring account alone cannot explain this result, since it predicts the same amount of hyperarticulation regardless of the context.

Additionally, when words in minimal pairs are presented in a limited context without their neighbors, they are produced with longer VOTs than words without minimal pairs. This implies speakers are not modulating their speech simply to avoid listener confusion. A purely perceptually-based, listener-modeling account cannot explain this finding.

The speaker internal production hypothesis, on the other hand, can account for both results. When words with neighbors are produced, they are more active than words without neighbors. This results in hyperarticulated productions, even in a limited context where no neighbors are

present. Because words with neighbors become even more active when their neighbors are present, they are further hyperarticulated within this condition. Critically, in contrast to the two alternative accounts, this hypothesis makes no appeal to the perceptual consequences of phonetic variation to account for these effects. Lexically conditioned phonetic variation is purely a reflex of interaction between processes within the production system.

Before the general discussion, we consider two possible objections to this conclusion. It is possible that a combined perceptual monitoring and restructuring account could be used to explain this data. However, we disprefer this explanation as it would posit two separate mechanisms to account for an effect. In contrast, the speaker internal production account utilizes a single mechanism to explain both effects; it is clearly the most parsimonious account of these results.

The second point is more concrete. It is possible that even within a closed-set listeners may still experience lexical competition effects (albeit at a reduced level relative to an open-set), thus undermining our argument against the listener modeling account. For example, using an eye-tracking paradigm Magnuson, Dixon, Tanenhuas, and Aslin (2007) demonstrate lexical effects in a closed set that excludes strong lexical competitors. However, we do not believe this interpretation to be correct because the size of the hyperarticulation effect is identical across closed and open sets. Given the results of Clopper et al. (2006) and Sommers et al. (1997) we know that perceptual lexical competition effects are at the very least reduced in closed-set contexts. Therefore, if the effects in Experiment 2 were to be attributed to listener modeling, they should be reduced compared to those effects in an open set in Experiment 1. Comparing across only the items used in both experiments, there is no significant difference across experiments. The difference in the VOT of words with vs. without minimal pair neighbors in the closed-set condition in Experiment 2 (no context vs. no competitor condition; mean difference 4.4 msec) is not statistically different than that observed in Experiment 1 (3.4 msec; if anything, the effect trends in the direction opposite that predicted by the listener modeling account)⁷. This lack of reduction of hyperarticulation in the closed- vs. open-set conditions provides further evidence against an on-line listener-modeling account of lexically conditioned phonetic variation.

General Discussion

The goal of this paper was to examine what mechanisms of interaction could produce lexically conditioned phonetic variation. Both Experiment 1 and Experiment 2 lend support to the growing body of evidence that lexically conditioned phonetic variation occurs in a number of phonetic contexts. Experiment 2 demonstrates that there are online aspects of lexically conditioned phonetic variation, and that this variation cannot be attributed solely to listener modeling. This pattern is most consistent with an on-line interaction mechanism that is internal to the production system.

The results of Experiment 1 and Experiment 2 are both congruent with and expand upon previous studies involving lexically conditioned phonetic variation. Wright (1997, 2004), Munson and Solomon (2004) and Munson (2007) demonstrate that vowels of words in dense neighborhoods, or ‘hard’ words, are hyperarticulated compared to vowels of words in sparse neighborhoods, or ‘easy’ words. The current study demonstrates similar hyperarticulation can be caused by a very specific form of neighborhood density — namely, minimal pair relationships. Furthermore, previous studies focused exclusively on various phonetic aspects

⁷It should be noted that, although we only compared items used in both Experiment 1 and Experiment 2 in our analysis, the two experiments were not strictly comparable as the experimental context differed. Therefore, the comparison should not be considered conclusive.

of vowels. By demonstrating similar findings for the phonetic properties of consonants, our results demonstrate that lexical effects are not restricted to vocalic segments. Furthermore, as noted above, unlike many previous studies Experiment 1 controlled for phonotactic probability as well as a number of other phonetic properties; this suggests that this type of phonetic variation is truly lexically conditioned.

Experiment 2 examines an issue unaddressed by previous work: what mechanisms drive this interactive effect? Three possible accounts were outlined: a speaker-internal production hypothesis, a listener-modeling hypothesis, and a perceptual restructuring hypothesis. Hyperarticulation of words with minimal pair competitors compared to words without minimal pairs was seen in closed sets. Additionally, when the minimal pair competitor was present, hyperarticulation was increased even further. The perceptual restructuring hypothesis can account for the former, but not the latter result. The listener-modeling hypothesis can account for the latter, but not the former result. The speaker-internal production hypothesis, however, can account for both types of hyperarticulation using the same mechanism. We take these results to be support for a speaker-internal production hypothesis for lexically conditioned phonetic variation.

Interaction between lexical phonological and phonetic processes

These findings are congruent with previous results suggesting that cascading activation between lexical phonological and phonetic processes can influence the phonetic properties of consonants (Goldrick and Blumstein, 2006). Interestingly, this work suggested that such activation can also induce interference or reduction in articulation. Goldrick and Blumstein (2006) found that relative to correctly produced targets, initial consonants in speech errors showed less extreme VOT values (e.g., voiceless initial stops in errors had shorter VOTs than voiceless stops in correct productions). In contrast to this hypoarticulation effect, both experiments in this study found hyperarticulation effects (e.g., longer voiceless initial stops for words with vs. without minimal pair neighbors).

It is possible that these different patterns result from the processing dynamics associated with errorful vs. non-errorful speech. When naming proceeds in a non-errorful fashion (as in the productions analyzed here), the target word dominates processing from its very initial stages. Non-target words—in particular, purely formally related words such as minimal pair neighbors—are never strong competitors for output. The target is therefore able to quickly inhibit non-target phonological representations. In contrast, when an error is produced (as in Goldrick and Blumstein, 2006), the error outcome has to overcome an initial processing disadvantage to suppress production of the target; it may therefore have less time to suppress activation of the competing (target) representation. Hyper- vs. hypo-articulation effects may therefore reflect the contrast in initial processing advantage vs. disadvantage in lexical phonological and phonetic processing.

Listener-oriented phonetic variation

Although a strong listener-modeling hypothesis cannot account for the particular type of phonetic variation observed here, we do not deny the existence of perceptual monitoring processes. As discussed in the introduction, speech production data at higher levels of the production system strongly support an influence of monitoring on production (e.g., Hartsuiker, Corley, and Martensen, 2005). With respect to phonetic variation specifically, other studies have shown that hyperarticulation can be caused by some form of listener modeling. Clear speech—an “intelligibility-enhancing mode of speech production” (Smiljanic and Bradlow, 2005: 1677)—is an obvious example of this type of effect. This speech mode reflects the speaker’s intention to be better understood by listeners, and is therefore by definition listener-oriented⁸. Similar to lexically conditioned phonetic variation discussed above, clear speech

induces hyperarticulation of vowels (e.g., Bradlow, 2002, demonstrates vowel space expansion in clear speech) as well as consonants (e.g., Smiljanic and Bradlow, 2005, demonstrate increased VOT of voiceless stops in clear speech). Similar types of phonetic variation can therefore be induced by the speaker's intention to enhance the ability of listeners to perceive the speaker's productions.

The interaction of perception and production processes

Though listener-modeling processes, such as clear speech, and automatic processes, such as lexically conditioned phonetic variation, may not be attributed to the same mechanisms, it is clear that their effects are often similar (i.e., hyperarticulation). What can account for this similarity?

One possibility is the account outlined above that combines the listener-monitoring and perceptual restructuring accounts. As noted above, this can account for the results here, albeit at the cost of being less parsimonious than the production-internal account. However, it could explain the similarity between lexically conditioned phonetic variation and clear speech; under this account, the listener-modeling processes that are used in clear speech are also exploited in contexts inducing lexically conditioned phonetic variation. This hypothesis is consistent with the findings of Hartsuiker et al. (2005), who suggest that both production-internal and monitoring processes contribute to the lexical bias effect in speech production.

We would advocate an alternative account that preserves the parsimonious, production-internal account of the results reported here by inverting the perceptual restructuring hypothesis. It is possible that interactions within the production system may reshape the structure of the perceptual system. Specifically, the perceptual system could tune itself to those contrasts that automatically occur during the course of production. MacDonald (1999) proposes a similar account in the context of sentence processing. She attributes a 'short before long' phrase preference in speech production to the incremental nature of production planning. Critically, she claims this production preference induces a similar bias in comprehension. That is, a mechanism completely internal to the production system—incremental planning—induces a change in the structure of perceptual processes (see also Arnold, 2008, for a similar proposal for the production and perception of referential expressions). A similar account could be constructed with respect to lexically conditioned phonetic variation. Hyperarticulation, induced by interactions internal to the production system, could cause perceptual processes to be restructured. Due to this restructuring, enhancing speaker-internal interactive effects naturally leads to speech which is more easily perceived. For example, since hyperarticulated exemplars are always produced when a highly similar competitor is present in the speaker's lexicon, the perception system may learn to increase inhibition of highly similar neighbors whenever it encounters a hyperarticulated production. The similarity between lexically conditioned variation and clear speech may not reflect perception leading production, but a case where perception exploits the variation that is naturally present within the outputs of speech production processes.

This discussion illustrates how moving beyond questions of "is interaction present or absent?" to "what is the nature of mechanisms that support interactive effects?" opens a broad set of interesting issues in theories of language processing. Exploring these new questions will

⁸Other instances of phonetic variation seem to be hybrids of listener modeling and more automatic processes. For example, Fowler and Housum (1987) examined second mention reduction. When speakers produce a word for the second time in a discourse it is shorter and less intelligible than the first mention. While this is sensitive to the communicative context, it is clear that it is a more automatic process than clear speech—which appears to be more under talker control. Similarly, the more general effects of articulatory reduction in predictable speech contexts (see Aylett and Turk, 2004, for a recent review) also appear to be due to more automatic processes that may or may not reflect explicit listener modeling.

illuminate not only the internal architecture of speech production but also its relationship to other components of the cognitive system.

Acknowledgments

This research was supported by National Institutes of Health grant DC 0079772. Portions of this work were presented at the 10th Conference on Laboratory Phonology (Paris, France, 2006), the 3rd International Workshop on Language Production (Chicago, IL, 2006), and the 48th Annual Meeting of the Psychonomic Society (Long Beach, CA, 2007). We thank Arim Choi, Yaron McNabb and Jaro Pylypczak for help in running the experiments, Eva Tomczyk and Kristin Van Engen for assistance in acoustic measurements and Ann Bradlow for extremely helpful discussions of this research.

References

1. Andruski JE, Blumstein SE, Burton M. The effect of subphonetic differences on lexical access. *Cognition* 1994;52:163–187. [PubMed: 7956004]
2. Arnold J. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes* 2008;23:495–527.
3. 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]
4. Baayen, R.; Piepenbrok, R.; Gulikers, L. The CELEX lexical database. Philadelphia, PA: Linguistic Data Consortium; 1995.
5. Bailey TM, Hahn U. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 2001;44:568–591.
6. Bell A, Jurafsky D, Fosler-Lussier E, Girard C, Gregory M, Gildea D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversational speech. *Journal of the Acoustical Society of America* 2003;113:1001–1024. [PubMed: 12597194]
7. Boersma, P.; Weenink, D. Praat v. 4.0.8. A system for doing phonetics by computer. Institute of Phonetic Sciences of the University of Amsterdam; 2002.
8. Bradlow, AR. Confluent talker- and listener-related forces in clear speech production. In: Gussenhoven, C.; Warner, N., editors. *Laboratory Phonology*. Vol. 7. Berlin & New York: Mouton de Gruyter; 2002. p. 241-273.
9. Clopper CG, Pisoni DB, Tierney AT. Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology* 2006;17:331–349. [PubMed: 16796300]
10. Dell GS. A spreading activation theory of retrieval in language production. *Psychological Review* 1986;93:283–321. [PubMed: 3749399]
11. Ferreira VS, Slevc LR, Rogers ES. How do speakers avoid ambiguous linguistic expressions? *Cognition* 2005;96:263–284. [PubMed: 15996561]
12. Fowler CA, 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.
13. Ganong WF. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 1980;6:110–125. [PubMed: 6444985]
14. Garrett, MF. Levels of processing in sentence production. In: Butterworth, B., editor. *Language production*. Vol. 1. New York: Academic Press; 1980. p. 177-220.
15. Gaskell MG, Marslen-Wilson WD. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 1997;12:613–656.
16. Goldrick M. Limited interaction in speech production: Chronometric, speech error and neuropsychological evidence. *Language and Cognitive Processes* 2006;21:817–855.
17. Goldrick M, Blumstein SE. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 2006;21:649–683.
18. Goldrick, M.; Rapp, B. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. Poster presented at the Psychonomic Society Annual Meeting; 2001.

19. Goldrick M, Rapp B. Lexical and post-lexical phonological representations in spoken production. *Cognition* 2007;102:219–260. [PubMed: 16483561]
20. Hartsuiker RJ, Corley M, Martensen H. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al 1975. *Journal of Memory and Language* 2005;52:58–70.
21. Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22:1–75. [PubMed: 11301520]
22. Luce PA, Pisoni DB. Recognizing spoken words; the neighbourhood activation model. *Ear & Hearing* 1998;19:1–36. [PubMed: 9504270]
23. MacDonald, MC. Distributional information in language comprehension, production and acquisition: Three puzzles and a moral. In: MacWhinney, B., editor. *Emergence of Language*. Mahwah, NJ: Erlbaum; 1999. p. 177-196.
24. McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? *Trends in Cognitive Sciences* 2006;10:363–369. [PubMed: 16843037]
25. McMurray B, Tanenhaus M, Aslin R. Gradient effects of within-category phonetic variation on lexical access. *Cognition* 2002;86(2):B33–B42. [PubMed: 12435537]
26. Meyer DE, Gordon PC. *Speech Production: Motor Programming of Phonetic Features*. *Journal of Memory and Language* 1985;24:3–26.
27. Munson, B. Lexical access, lexical representation, and vowel production. In: Cole, JS.; Hualde, JI., editors. *Laboratory Phonology*. Vol. 9. New York: Mouton de Gruyter; 2007. p. 201-228.
28. Munson B, Solomon PN. The effect of phonological density on vowel articulation. *Journal of Speech, Language and Hearing Research* 2004;47:1048–1058.
29. 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.
30. Pierrehumbert, JB. Word-specific phonetics. In: Gussenhoven, C.; Warner, N., editors. *Laboratory Phonology*. Vol. 7. Berlin: Mouton de Gruyter; 2002. p. 101-139.
31. Pluymaekers M, Ernestus M, Baayen R. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 2005;118:2561–2569. [PubMed: 16266176]
32. Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. *Psychological Review* 2000;107:460–499. [PubMed: 10941277]
33. Roelofs A. Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review* 2004;111:561–572. [PubMed: 15065924]
34. Salverda AP, Dahan D, McQueen JM. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 2003;90:51–89. [PubMed: 14597270]
35. Salverda AP, Dahan D, Tanenhaus MK, Crosswhite K, Masharov M, McDonough J. Effects of prosodically-modulated sub-phonetic variations on lexical competition. *Cognition* 2007;105:466–476. [PubMed: 17141751]
36. Scarborough, RA. Lexical confusability and degree of coarticulation. *Proceedings of the 29th annual meeting of the Berkeley Linguistics Society*; 2003.
37. Scarborough, RA. Doctoral dissertation. Los Angeles: UCLA; 2004. Coarticulation and the structure of the lexicon.
38. Smiljanic R, Bradlow AR. Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America* 2005;118:1677–1688. [PubMed: 16240826]
39. Sommers MS, Kirk KI, Pisoni DB. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear & Hearing* 1997;18:89–99. [PubMed: 9099558]
40. Vitevitch MS. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2002;28:735–747.
41. 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.

42. 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.
43. Vitevitch MS, Sommers MS. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition* 2003;31:491–504.
44. 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;92:723–735. [PubMed: 1506527]
45. Wright, RA. Speech Research Lab, Psychology Department. Bloomington, IN: Indiana University; 1997. Lexical competition and reduction in speech: a preliminary report. Research on spoken language processing: Progress report.
46. Wright, RA. Factors of lexical competition in vowel articulation. In: Local, JJ.; Ogden, R.; Temple, R., editors. *Laboratory Phonology*. Vol. Vol. 6. Cambridge: Cambridge University Press; 2004. p. 26-50.
47. Yaniv I, Meyer DE, Gordon PC, Huff CA, Sevald CA. Vowel Similarity, Connectionist Models, and Syllable Structure in Motor Programming of Speech. *Journal of Memory and Language* 1990;29:1–26.

Appendix

Experimental Stimuli

Table A1

Labial Stop Stimuli

Minimal pair	No minimal pair
Pie	pipe
*peek	*peel
*palm	*pomp
Pore	pork
*punk	*pulp
*punch	*pulse
*pun	*pup
*pad	*pal
Pall	paunch
*peat	*peal
*pare	*pep
Pill	pinch
*pig	*pith
*poll	*poach
*pox	*posh
*putt	*pub

Note: * stimuli were used in Experiment 2; all stimuli were used in Experiment 1.

Table A2

Alveolar Stop Stimuli

Minimal pair	No minimal pair
*tab	*tat
*tan	*tag
Tank	tap
*teal	*teat
*teem	*teethe
*tick	*tiff
*tuck	*tuft
Ted	tempt
*tense	*tenth
Tart	tar
Taunt	torch
Tore	taut
*torque	*torn
*tomb	*tooth
*tame	*taint
Tyke	tithe
*tile	*tights
Toe	toast
*tote	*toad

Note: * stimuli were used in Experiment 2; all stimuli were used in Experiment 1.

Table A3

Velar stop stimuli

Minimal pair	No minimal pair
*cob	*cog
*cod	*cop
*kilt	*kin
*kit	*kiln
*core	*corn
*cuss	*cub
*cuff	*cub
*cuff	*cud
*curl	*curb
*coo	*coot
*cab	*cad
*cape	*cake
*code	*comb

Note: * stimuli were used in Experiment 2; all stimuli were used in Experiment 1.

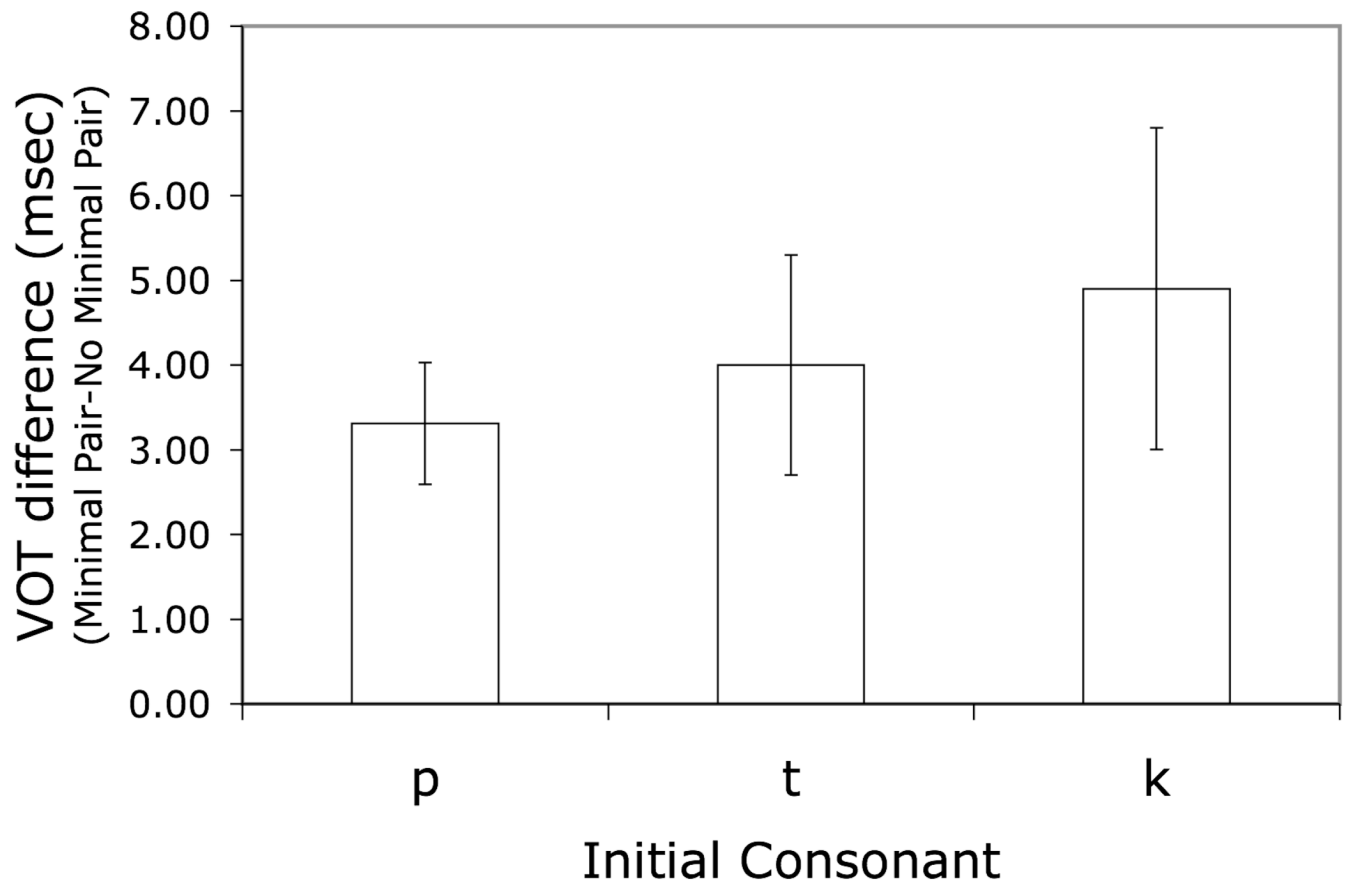


Figure 1. Mean VOT differences (word with minimal pair neighbor – word with no minimal pair neighbor), Experiment 1. Error bars represent standard error.

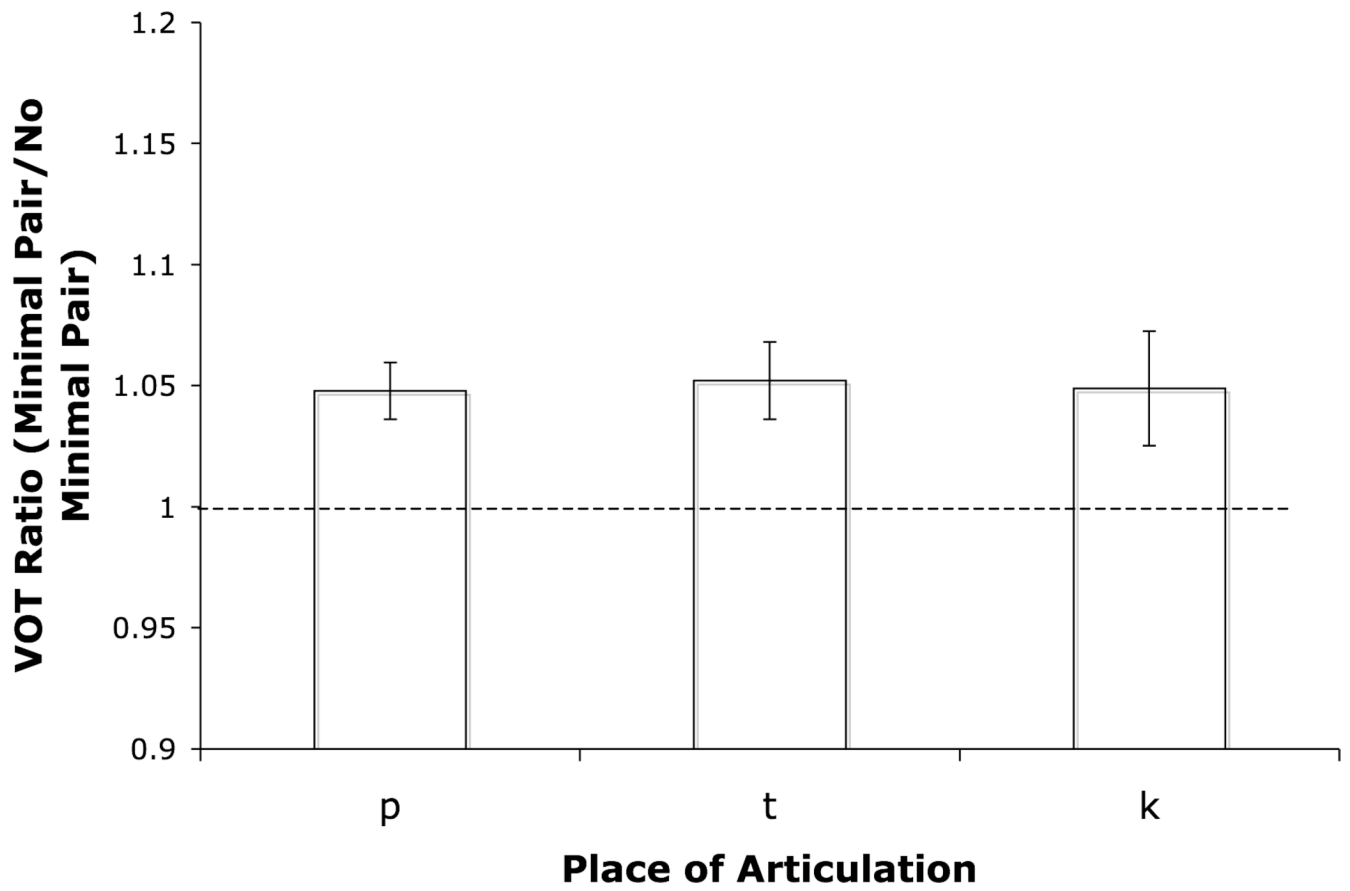


Figure 2. VOT ratio (word with minimal pair neighbor / word with no minimal pair), Experiment 1. Error bars represent standard error. The dashed line represents equivalent VOTs across conditions.

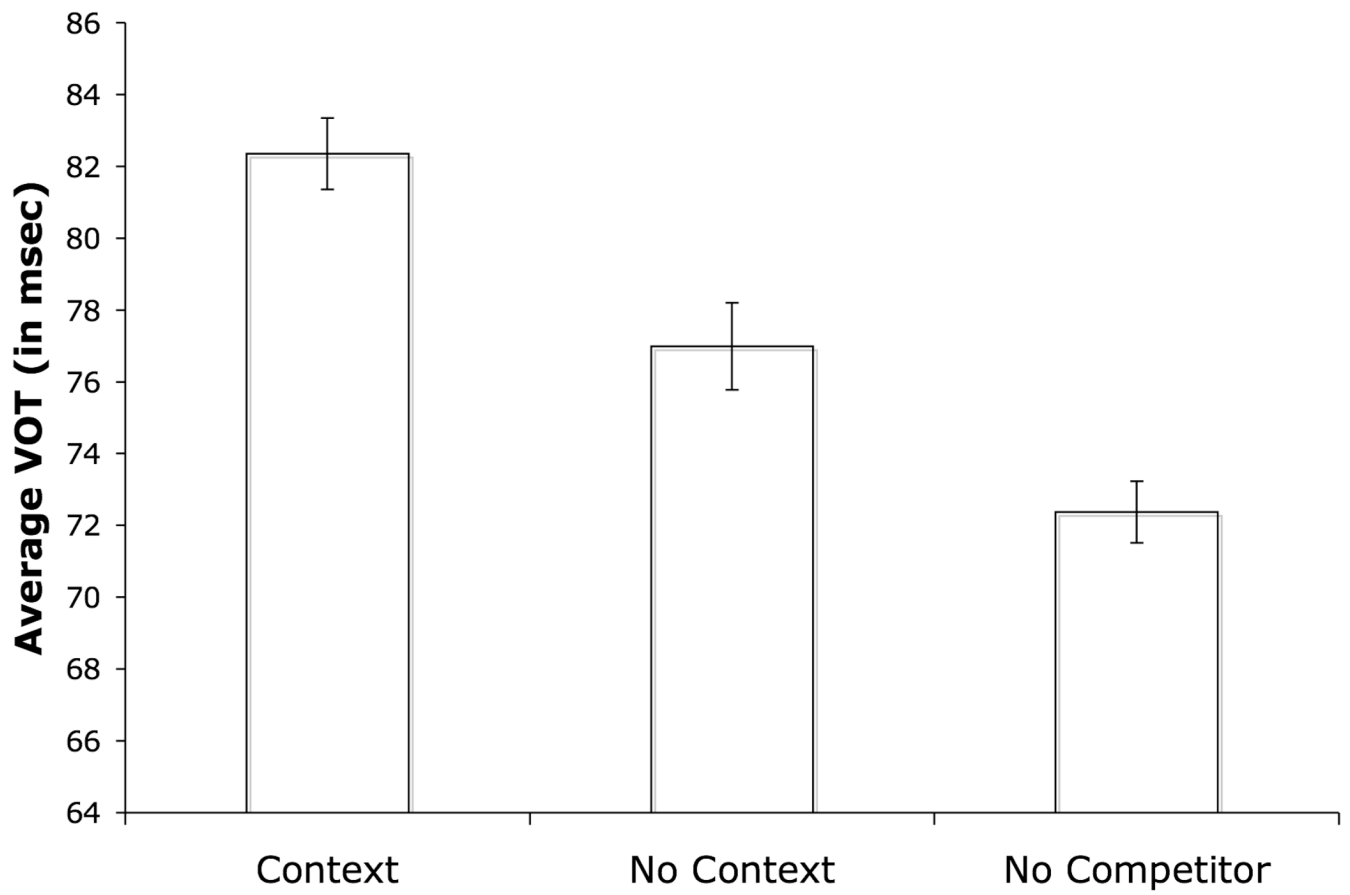


Figure 3. Raw VOTs for words in each condition, Experiment 2. Error bars represent standard error.

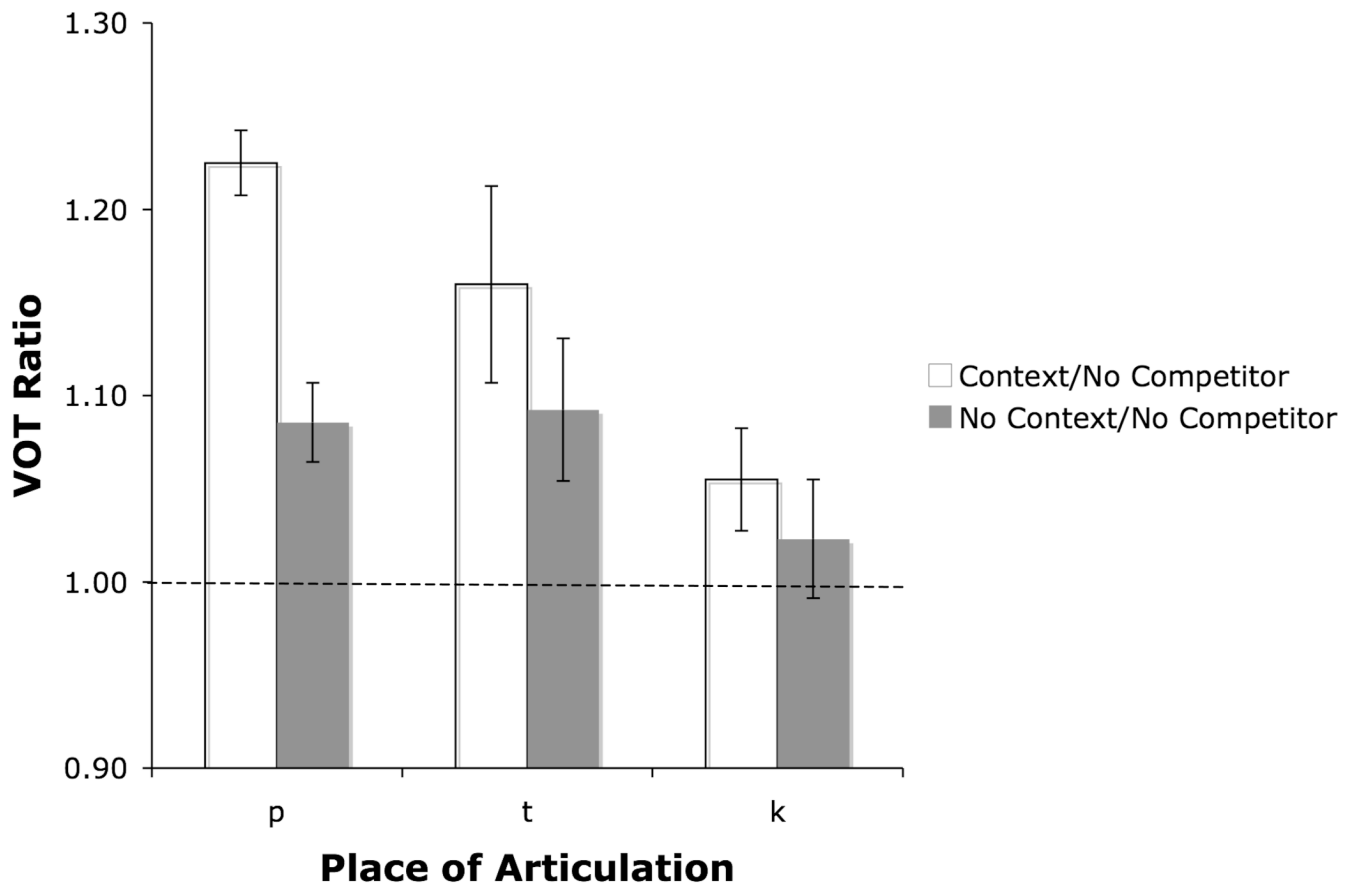


Figure 4. VOT ratio (relative to no competitor condition) for Experiment 2. Error bars represent standard error. The dashed line represents equivalent VOTs across conditions.

Table 1

Sample trial setup for Experiment 2

Condition			
Context Condition	cod	god	vell
No Context Condition	cod	lamp	vell
No Competitor Condition	cop	lamp	vell

Table 2

Experiment 1a Matching Statistics

	Minimal Pair	Non-Minimal Pair	t-test	P value
Sum Segmental Probability	1.104	1.199	t(15)=0.31	p>0.75
Average Biphone Probability	0.004	0.003	t(15)=1.91	p>0.07
Average Phoneme Length	3.125	3.375	t(15)=-1.73	p>0.10
Average Lexical Frequency	4.875	4.125	t(15)=0.47	p>0.65

Table 3

Experiment 1b Matching Statistics for /t/-initial Stimuli

/t/-initial stimuli	Minimal Pair	Non-Minimal Pair	t-test	P value
Sum Segmental Probability	1.2	1.2	t(18)=0.0137	p>0.95
Average Biphone Probability	0.005	0.004	t(18)=0.61	p>0.50
Average Phoneme Length	3.2	3.4	t(18)=-1.42	p>0.15
Average Lexical Frequency	5.76	6.55	t(18)=-0.51	p>0.60

Table 4

Experiment 1b Matching Statistics for /k/-initial Stimuli

/k/-initial stimuli	Minimal Pair	Non-Minimal Pair	t-test	P value
Sum Segmental Probability	1.2	1.2	t(11)=-0.33	p>0.75
Average Biphone Probability	0.004	0.004	t(11)=-.042	p>0.65
Average Phoneme Length	3.08	3.25	t(11)=-1.0	p>0.30
Average Lexical Frequency	8.3	6.2	t(11)=1.0	p>0.30

Table 5

Experiment 2 Matching Statistics

	Minimal Pair	Non-Minimal Pair	t-test	P value
Sum Segmental Probability	1.18	1.19	$t(35)=-0.41$	$p>0.85$
Average Biphone Probability	0.004	0.004	$t(35)=0.68$	$p>0.68$
Average Phoneme Length	3.17	3.31	$t(35)=-0.92$	$p>0.09$
Average Lexical Frequency	5.9	6.2	$t(35)=0.32$	$p>0.9$