



Published in final edited form as:

*J Cogn Neurosci*. 2011 March ; 23(3): 593–603. doi:10.1162/jocn.2010.21489.

## Phonological Neighborhood Effects in Spoken Word Production: An fMRI Study

D. Peramunage<sup>1</sup>, S. E. Blumstein<sup>1</sup>, E.B. Myers<sup>1</sup>, M. Goldrick<sup>2</sup>, and M. Baese-Berk<sup>2</sup>

<sup>1</sup> Brown University

<sup>2</sup> Northwestern University

### Abstract

The current study examined the neural systems underlying lexically conditioned phonetic variation in spoken word production. Participants were asked to read aloud singly presented words which either had a voiced minimal pair (MP) neighbor (e.g. *cape*) or lacked a minimal pair (NMP) neighbor (e.g. *cake*). The voiced neighbor never appeared in the stimulus set. Behavioral results showed longer voice-onset time for MP target words, replicating earlier behavioral results (Baese-Berk & Goldrick, 2009). fMRI results revealed reduced activation for MP words compared to NMP words in a network including the left posterior superior temporal gyrus, the supramarginal gyrus, inferior frontal gyrus, and precentral gyrus. These findings support cascade models of spoken word production and show that neural activation at the lexical level modulates activation in those brain regions involved in lexical selection, phonological planning, and ultimately motor plans for production. The facilitatory effects for words with minimal pair neighbors suggest that competition effects reflect the overlap inherent in the phonological representation of the target word and its minimal pair neighbor.

### Introduction

Speaking and understanding require that multiple sources of information be integrated in the service of communicating meaning. Most current models of the functional architecture of language propose that in accessing the words of a language, there are multiple stages of processing, each of which requires mapping from one level of representation to another. For example, in spoken word production, a word is selected from among all of the words in the mental lexicon to express a particular concept; this representation is mapped on to the sound shape of the word specifying its phonological form; and this abstract phonological representation in turn is mapped on to articulatory implementation processes which provide detailed information to the articulators about the ultimate phonetic realization of the word.

Current models also assume that at each level of processing, there is automatic activation not only of the target word, but also partial activation of other related representations that share structural properties with the word candidate (Gaskell & Marslen-Wilson, 1999; Dell, 1986). These representations compete with each other and the best fitting candidate is ultimately selected from the set of activated representations. The ultimate selection of a target from this set of activated presentations is typically called competition (Schnur et al., 2009). In some cases, selection among multiple activated representations leads to interference, resulting in increased processing difficulty (as shown behaviorally by longer processing times). In other

cases, it results in facilitation, resulting in enhanced processing (as shown by decreased processing times). For example, lexical decision latencies are slower for words that have many phonologically similar words or neighbors compared to words that have few phonologically similar neighbors (Luce & Pisoni, 1998), whereas naming latencies for pictures of words are faster for words that have many phonologically similar neighbors than for words which few neighbors (Vitevitch, 2002).

Recent research has suggested that the activation of multiple representations has consequences throughout the language processing system. In particular, information from one stage of processing cascades and influences other stages of processing downstream from it (see Goldrick, 2006, for a recent review). Thus, the selection of the phonological representation of a word is modulated by the number of words in the lexicon that share sound properties with it (Dell & Gordon, 2003), and this has a cascading effect on its articulatory implementation (Baese-Berk & Goldrick, 2009; Goldrick & Blumstein, 2006). For example, as described above, reaction-time latencies for naming pictures of words which have many phonological neighbors are faster than that for naming words which have few phonological neighbors (Vitevitch, 2002; but cf. Luce & Pisoni, 1998).

Neighborhood density effects also influence acoustic-phonetic patterns of speech output. Words with many phonological neighbors are produced with a larger vowel space than words from sparse neighborhoods (Wright, 2002; Munson, 2007; Munson & Solomon, 2004; Scarborough, in press). In a recent study, Baese-Berk & Goldrick (2009) also showed lexically conditioned phonetic variation for 'local' effects of neighbors, namely, the effects of a phonologically contrasting minimal pair lexical neighbor. In particular, the voice-onset time (VOT) productions of words with initial voiceless stop consonants are longer in words that have a contrasting initial voiced stop (*tart* with a contrasting voiced lexical item *dart*) than in words that do not have a contrasting initial voiced stop (e.g. *tar* does not have a voiced competitor *dar*).

This influence of lexical neighbors on articulatory processes reflects the cascading effects of lexical activation and selection processes on plans for articulation. Thus, the activation level of a target word is influenced by the phonological properties of the word's neighborhood. A lexical candidate from a dense neighborhood requires greater activation to override the activation of contrasting lexical items relative to a lexical candidate from a sparse neighborhood. Similarly, a lexical candidate will require greater activation if it must override a contrasting minimal pair neighbor. In both cases, this increased activation cascades throughout the system and influences processes downstream from lexical access - including the articulatory implementation of the lexical candidate itself. As a consequence, productions are 'hyperarticulated'. For example, vowels in words will be produced with wider vowel spaces and the voice-onset time (VOT) of voiceless stop consonants will be longer.

The finding that spoken word production is influenced by the number of potentially activated and hence competing phonological lexical competitors raises the question of the neural substrates of this effect. It is the goal of the current study to examine the neural systems underlying this 'lexically conditioned phonetic variation' (Baese-Berk & Goldrick, 2009) and to determine whether modulatory effects arising from the activation of phonologically similar words in the lexicon cascade throughout the spoken word production processing stream.

Recent studies exploring the neural systems underlying phonological/lexical competition in auditory word recognition show that posterior areas including the left supramarginal gyrus (SMG) and frontal areas including the inferior frontal gyrus (IFG) are modulated by lexical

competition (Prabhakaran et al, 2006; Righi et al., 2009). Okada and Hickok (2009) also showed activation in the bilateral superior temporal gyrus (STG) in a study exploring neural activation patterns for high density compared to low density words (although their analyses were restricted to the temporal lobes). Taken together, these findings suggest that the IFG and posterior STG, SMG are part of a network involved in accessing and maintaining the sound shape of a word from the mental lexicon (posterior STG, SMG) and ultimately selecting the word from among activated representations (IFG). Previous work by Gold and Buckner (2002) is consistent with this view. They showed coactivation of the SMG with frontal areas when subjects performed a controlled phonological task.

Less research has been conducted exploring the neural systems underlying the influence of phonologically related words (such as lexical neighbors) in spoken word production. One study (Schnur et al., 2009) using a blocked naming paradigm failed to show any neural areas sensitive to the presence of phonologically similar words. However, using a picture-word interference paradigm in which subjects were asked to name a picture with a written distractor presented within the target picture, De Zubicaray (2002) showed sensitivity to the presence of phonologically/orthographically related competitors in the left posterior superior temporal gyrus (STG) (see also de Zubicaray & McMahon, 2009) and a range of areas in the right hemisphere including the inferior temporal gyrus, inferior parietal lobule, superior and middle frontal gyrus and post-central gyrus. In a later study, Abel et al. (2009) also used a picture-word interference paradigm, but subjects were asked to name a picture presented 200 ms after the presentation of an auditory distractor. Results showed a broad network activated in the context of phonological distractors which shared the two initial phonemes of the stimulus to be named. This network encompassed posterior areas including the SMG and superior temporal gyrus (STG) and frontal areas including the IFG (BA44) and the post-central gyrus. These areas mirror those identified by Indefrey and Levelt (2004) as underlying the components involved in word production.

In the studies discussed above, both the target stimulus and its phonological neighbor are a part of the stimulus set. Thus, the speaker must select and produce the target word in the context of a strongly-activated phonological competitor. What is less clear is whether competition effects will also emerge when the competitor is inherent in the structure of the mental lexicon itself but not present in the stimulus array. More specifically, how does the existence of a phonologically similar neighbor in the lexicon affect the neural substrates underlying phonological processing in spoken word production?

The goal of the current study is to further investigate the influence of phonological neighbors in word production by examining the neural systems underlying lexically conditioned phonetic variation. Participants will be required to read singly presented words which either have or do not have minimal pair neighbors, e.g. *cape* with a voiced minimal pair *gape* vs. *cake* which does not have a voiced minimal pair. Thus, the target words in the minimal pair competitor condition will be maximally similar to their minimal pair competitor, sharing all phonemes except for the initial consonant. In contrast to previous fMRI word production studies, the competitor will never appear in the stimulus set. Thus, any effects of competition will arise implicitly from the phonological similarity of words in the mental lexicon and not from competition effects induced by the overt presence of a competitor in the stimulus array.

A reading task was selected rather than an auditory repetition task for several reasons. First, we wanted to avoid the possibility that subjects' productions would be shaped by the acoustic properties of the words to be repeated. Recent behavioral research has shown phonetic convergence between speakers in conversation (Pardo, 2006). In addition, we

wanted to assure that any speech output patterns were not influenced by potential misperceptions of the test stimuli.

We hypothesize that the effects of phonological competition on spoken word production will recruit a similar network to that shown for phonological competition in auditory word recognition. In particular, we expect activation in the SMG/posterior STG consistent with the view that these areas are involved in accessing and maintaining the lexical (sound shape form) of a word from the mental lexicon (Prabhakaran et al, 2006; Righi et al., 2009; Indefrey & Levelt, 2004; Paulesu et al., 1993). Activation of the SMG should be modulated by the extent to which there are words in the lexicon which share their sound shape with the lexical candidate.

Based on recent findings that the IFG is recruited in auditory word recognition when words share phonological onsets (Righi et al., 2009), we also expect activation in the IFG (and in particular, BA 45/44) since this area is recruited when a lexical candidate is selected from among a set of multiple activated representations. Such findings would support the claim that there is a common neural substrate for resolving competition at multiple levels of the language processing system (semantic and phonological) in both language production as well as comprehension and hence that the IFG plays a domain general role in cognitive control (cf. Thompson-Schill et al., 1997, 1999; Snyder et al. 2007; Badre & Wagner, 2007).

Of interest is whether in addition to the modulatory effects of phonological competition on the SMG and IFG there will be similar effects in regions involved in the planning of the motor gestures necessary for word production. In particular, the finding that regions such as the precentral gyrus are modulated by phonological competition would be consistent with those models of spoken word production in which access of a word has a cascading effect on the processes downstream from it (Baese-Berk & Goldrick, 2009; Goldrick & Blumstein, 2006). In this case, the selection of a word that has a competitor will affect not only its access and selection but also motor plans for production of that word.

In sum, in the current experiment, participants were visually presented with one word at a time in the scanner, and they read each word aloud. A sparse sampling design allowed for their productions to be recorded in relative silence. These recordings were analyzed off-line to measure the voice-onset time of the initial voiceless stop consonants of the test stimuli.

## Materials and methods

### Participants

Eighteen subjects, two of whom were males, participated in the MR portion of this study and all received payment for their involvement. Their ages ranged from 19 to 31 with a mean age of 25 years. All participants were native English speakers, were right-handed, as determined by the Oldfield handedness inventory (Oldfield 1971), reported normal hearing and had no known history of neurological disorders. Each subject was screened for MR safety before being placed in the scanner and gave written informed consent in accordance with the guidelines established and approved by the Human Subjects Committee of Brown University.

### Stimuli

Stimuli consisted of a subset of the stimuli from Baese-Berk and Goldrick (2009) (see Appendix). All of the twelve pairs of/k/target stimuli, eighteen out of 19/t/pairs (the pair tyke: tith was eliminated), and sixty filler words (from the original 128 fillers) were selected from their study and combined into one list for use in the current study. Target pairs were all monosyllabic words that shared both the same initial voiceless stop consonant and

vowel. Each minimal pair (MP) word, having a neighbor with a voiced initial consonant (e.g. *tart* with a voiced neighbor *dart*), was paired with a non-minimal pair (NMP) word, which lacked such a neighbor (e.g. *tar*). The /t, k/ target words were chosen because they showed the greatest mean VOT difference between minimal pair (MP) and non-minimal pair (NMP) target words in the Baese-Berk & Goldrick data. The test words differed as well by lexical density with MP words having more phonological neighbors (28.7) compared to NMP words (21.3) (Irvine Phonotactic Online Dictionary, [www.iphod.com](http://www.iphod.com)). Thus, MP words displayed greater local competition than NMP words by having a competing word which shares all attributes but initial stop consonant voicing and they had greater global competition having more words overall that share phonological properties with the target word. All pairs were matched across a number of parameters including sum segmental probability, sum biphone probability, and phoneme length (for details see Baese-Berk & Goldrick, 2009). Additionally, the form-level properties of the coda were controlled across the stimulus pairs. The length and phonological frequency (phonotactic probability) of the codas was taken into account; these did not significantly differ across the two sets. All words were low frequency (less than 20 per million) and were matched for frequency.

In addition to the target stimuli, the list of /t/ and /k/ filler words used by Baese-Berk & Goldrick were combined and reused in this study. Filler words were included to ensure that subjects did not become either implicitly or explicitly aware that half of the experimental stimuli had minimal pair word rhymes and also to provide a richer phonological set of stimuli for them to produce. The filler words were selected such that twenty-four of the fillers had initial stop consonants evenly distributed across the stop consonants which were not targets in the experiment, i.e. [p b d g]. The remaining fillers were selected so that they were distributed across the consonants of English and included fricatives, affricates, nasals, and glides. Finally, only filler words were selected which did not form a minimal pair or rhyme with the target word stimuli,

## Task

Subjects were asked to read each stimulus aloud as it appeared on a screen. Stimuli were presented via an LCD projector, which displayed stimuli on a back-projection screen in the scanner room. Subjects viewed this screen using the head coil mirror. Each trial consisted of the presentation of a word (black 24pt MS San Serif font on a white background) in the center of the screen for 2000 ms. Behavioral pilot work indicated that subjects could easily read and produce the stimuli for recording within the 2 second interval provided.

Presentation of stimuli was controlled by a laptop (IBM Thinkpad) running the BLISS software suite (Mertus, 2002). Subjects' responses to the stimuli were recorded using the built-in patient microphone of the Avotec SS-3100 Silent Scan audio system and an Edirol R-09 24bit Digital Recorder. Stimuli were recorded as 24bit uncompressed WAV files sampled at 44.1 kHz, and then down-sampled, using BLISS, to 16bit WAV before subsequent acoustic analysis.

Subjects participated in six experimental runs of an event-related design, each consisting of 60 stimulus presentations. The 120 stimuli were divided into 2 lists and each list was repeated three times, with stimuli in a pseudo-random order. Each run consisted of 6/k/MP, 6/k/NMP, 9/t/MP, and 9/t/NMP target stimuli and 30 fillers. Prior to running the six experimental runs, participants performed a short practice run during EPI data collection (thirty practice trials consisting of 7 MP and 7 NMP pairs and 16 fillers), so that they could accustom themselves to the timing of the stimuli and the scanner environment.

## Image acquisition

Both anatomical and functional images were acquired using a 3T TIM Trio scanner (Siemens Medical Systems, Erlangen, Germany). High-resolution 3D T1-weighted anatomical images were acquired for anatomical co-registration (TR=1900 ms, TE=4.15 ms, TI 1100 ms, 1 mm<sup>3</sup> isotropic voxels, 256 × 256 matrix). Each functional volume consisted of sixteen 5mm-thick echo planar (EPI) axial slices with a 3mm isotropic in-plane resolution, and slices were acquired in an ascending, interleaved order. Functional volumes were aligned to image the peri-sylvian cortex (TR=3sec, TE= 30ms, flip angle= 90 degrees, FOV= 192mm<sup>3</sup>, 64×64 matrix).

The sixteen slices in the EPI scan were acquired in the first 1000 ms of each 3000 ms TR, followed by 2000 ms of silence in which a stimulus was displayed and the subject's response was recorded. Stimuli presentation was jittered such that each stimulus was distributed across three trial onset asynchrony (TOA) bins (TOA = 3, 6, and 9 seconds). To account for T1 saturation effects, each of the six EPI runs were preceded by two 'dummy' volumes; these two volumes were discarded during analysis. In addition, five more volumes were added to the end of each EPI run to account for the decay in the hemodynamic response following the final stimulus. A total of 127 EPI volumes were acquired for each of the six runs.

## Analysis of Results

### Behavioral results

For each MP and NMP stimulus, voice onset time (VOT) of the initial voiceless stop consonant was measured by hand using the BLISS software program developed at Brown University (Mertus, 2002). To this end, the time (in ms) from the onset of the burst to the onset of the vowel was determined. Stimuli were excluded from analysis if the subject read the test stimulus incorrectly or if it was impossible to determine the burst onset from background noise. A total of 9.2% of the total productions were not included in the analysis. Analysis of the pattern of errors revealed no difference between MP and NMP words either in incorrect productions of words (average number across the subjects was .41 for MP words and .53 for NMP words) or in inability to measure VOT (average number across the subjects was 3.5 MP words and 3.9 (for NMP words).

The results of the acoustic analysis revealed that, as predicted, the MP condition showed longer VOTs than the NMP condition (96.2 vs. 94.1ms). Although the effect is smaller than the 4.5 ms effect shown by Baese-Berk and Goldrick (2009), it was nonetheless significant. Wilcoxon Matched-Pair sign rank tests revealed a significant difference between the VOT of the MP and NMP conditions effect by participant ( $W=-145$ ,  $p< 0.0016$ ) and by item ( $W=197$ ,  $p< 0.0434$ ).

Because the magnitude of the VOT effect was small and only occurred in some of the stimuli for some of the subjects, only those trials containing stimulus pairs showing a VOT difference greater than 1.0 ms across the three paired tokens were used in the fMRI analysis. This subset of stimulus pairs was still matched for sum segmental probability, sum biphone probability, and phoneme length. The remaining trials were modeled separately during deconvolution analysis (see MR Analysis: Statistical Analysis, below). Overall, stimuli were included in the analysis from all of the target pairs; however, for any given pair, there were different numbers of subjects contributing to that value in the MR analysis. Table 1 lists the target pairs, the differences in VOT between them, and the number of subjects whose data are included for that pair. Overall, the mean VOT of the MP stimuli included was 12 ms longer than that of the included NMP stimuli.

## MR Analysis

Imaging data were analyzed using the AFNI software package (Cox and Hyde 1997) on a cluster of eighteen Apple dual processor G5 XServe servers. The runs were concatenated and the EPI images were then corrected for head motion after aligning all the collected volumes with the fourth volume (Cox & Jesmanowicz, 1999), transformed to the Talairach-Tournoux space, resampled to 3-mm isotropic voxels, and smoothed with a 6mm full width half maximum Gaussian kernel. All subsequent analyses involving the EPI data were restricted to those voxels imaged for all the subjects and found inside the brain.

For one subject the data for only four trials were completed due to technical difficulties; as a result, there were only two repetitions of each stimulus for that subject.

## Statistical Analysis

Experimental stimuli not used in the fMRI analysis were placed in a separate 'bad stimulus' vector. These included stimuli for which a subject made an error as well as for stimulus pairs in which a subject made three errors on one member of the pair. Moreover, those stimulus pairs for which the computed VOT difference between the MP and NMP stimuli pairs was less than 1.0 ms were placed in separate 'bad MP' and 'bad NMP' vectors. Thus, the Good MP and Good NMP vectors referenced only those trials where MP and NMP pairs showed a mean VOT difference of 1 ms or more for a given subject. A total of 1509 out of the original 3240 productions were included in the final analysis of the data.

To estimate the hemodynamic response of each stimulus condition (Good MP, Good NMP, Filler, Bad MP, Bad NMP, Bad Stimuli), a deconvolution analysis was performed on the functional data using AFNI. Time-series files, which contained the time points at which stimuli were presented, were created for each condition. These were convolved with a gamma function to obtain the idealized hemodynamic response for each condition.

Multiple linear regressions were performed with AFNI's 3dDeconvolve program using the gamma function convolved time series files for each stimuli condition. In addition, the six parameters that were output by the motion correction process were also included as nuisance regressors. The 3dDeconvolve analysis returned by-voxel fit coefficients for each condition, which were used to calculate the percent signal change for each of the stimuli conditions for each subject. The data were then submitted to a mixed-factor ANOVA with subjects as a random factor and stimulus conditions as a fixed factor, and a planned comparison was made between the Good MP vs Good NMP (Good MP – Good NMP).

Monte Carlo simulations were performed to determine the number of contiguous voxels needed to achieve a correct significance level of  $p < 0.05$ . The simulations were run for 10,000 iterations on a small volume mask of the brain (Forman, et al, 1995). This mask consisted of bilateral areas previously implicated in language function including the IFG, SMG, MFG, AG, and STG. In addition to these areas, the mask also included other bilateral areas such as the TTG, cingulate gyrus, precuneus, IPL, precentral gyrus, insula, and posterior cingulate. At a voxel-level threshold of  $p < 0.05$ , a cluster size of 80 contiguous voxels achieved a corrected significance of  $p < 0.05$ . The maximum intensity point of the activated clusters was used to identify the location of the activated anatomical regions and the proportion of voxels within a particular cluster that fell within different anatomical regions using the N27 atlas (Eickhoff, 2005).

## fMRI Results

A summary of all the significant (cluster-threshold  $p < 0.05$ ) clusters activated in the Good MP vs Good NMP comparison are shown in Table 2. Four clusters emerged in this

comparison; three of which showed greater activation for the NMP stimuli condition (Figure 1).

The NMP stimuli condition showed greater activation than the MP condition in the left supramarginal gyrus (SMG), left inferior gyrus (IFG), and left precentral gyrus. The largest cluster (193 voxels), was in the SMG (59%) and extended into the STG (21%) and inferior parietal lobule (8%). The second largest cluster, 103 voxels in size, was found in the left IFG. The majority of this cluster fell in the left pars triangularis (49% of BA45), and extended into the pars orbitalis (16% of BA47), pars opercularis (6% of BA44), and insula. Finally, an 88 voxel cluster was found in the left precentral gyrus, which extended into the left postcentral gyrus. Only one cluster, 289 voxels, showed greater activation for the MP condition compared to the NMP condition. This medial cluster was located in the left precuneus and extended bilaterally into the right precuneus and both the left and right calcarine gyrus.

As noted in the methods section, the test words were distinguished by both local competition (MP words had a voiced lexical competitor and NMP words did not) and global competition (MP words had a higher lexical density than NMP words). In order to determine whether similar activation patterns would emerge solely due to local competition effects, we redid the cluster analysis as described above, controlling for global lexical density. To this end, we excluded 5 word pairs which differed in lexical density (Ted-tempt; tense-tenth; tile-tights; toe-toast; kit-kiln). Even with this reduced number of observations, results replicated the previous analysis; significant clusters emerged in the SMG and the precentral gyrus showing greater activation for the NMP compared to the MP words. A 63-voxel sub-threshold cluster in the IFG also emerged (equivalent to  $p < 0.15$ , corrected threshold).

## Discussion

The results of the current study show that lexically conditioned phonetic variation in spoken word production activates a network that includes the left posterior superior temporal gyrus, the supramarginal gyrus, inferior frontal gyrus, and precentral gyrus. In particular, the production of initial voiceless stop consonants is longer for words that have a voiced minimal pair than for words that do not (cf. Baese-Bark & Goldrick, 2009). The modulation of activation throughout the frontal-parietal network is consistent with a cascade model of language production, where lexically-driven differences in the activation of phonological representations modulate subsequent articulatory processing (Baese-Berk & Goldrick, 2009; Goldrick & Blumstein, 2006).

### The Frontal-Parietal Network

Activation of the SMG extending into the posterior STG is consistent with recent work showing activation in these areas in the perception of the phonological sound shape of words (Paulesu et al., 1993; Hickok & Poeppel, 2000) and in the perception of auditorily presented words under conditions of phonological competition (Prabhakaran, et al. 2006; Righi et al., 2009). Of importance, the current study shows that this area is recruited in spoken word production as well (cf. Indefrey and Levelt, 2004). Thus, this area appears to be modality independent and to be involved in accessing the sound shape of words from the mental lexicon for both auditory word recognition and for spoken word production.

Information from the SMG cascades to frontal areas including the IFG and the precentral gyrus for selecting the word from among the competing set of potential word candidates, for phonological planning processes, and ultimately for articulatory implementation. The literature has suggested that the IFG is involved not only in the selection of a word from among a set of competing alternatives (Thompson-Schill et al., 1997; Righi et al., 2009) but



also in phonological planning (Huang et al., 2001; Bookheimer et al., 1995; Guenther, 2006) and in grapheme-phoneme conversion (Indefrey and Levelt, Fiez et al., 1999; Pugh et al., 1996). The modulation of activation in the IFG is consistent with these results. Whether there is a functional division of the IFG as has been suggested in the literature (Burton, 2001; Poldrack et al., 1999; Fiez, 1997; Buckner, Raichle, & Petersen, 1995) with selection processes recruiting BA45 and phonological planning processes recruiting BA44 cannot be determined from the current data since the IFG cluster that emerged encompassed both of these areas.

Nonetheless, what this study does show is that the IFG is recruited when contrasting phonological neighbors become active. The competition induced by the contrasting elements of the target and its minimal pair neighbor (e.g., /t/ vs. /d/ for target *tart* and minimal pair neighbor *dart*) influences both selection and ultimately phonological planning stages for spoken word production. In all previous studies showing modulatory effects of the IFG as a function of phonological competition (Abel et al., 2009; de Zubicaray & McMahon, 2009; de Zubicaray et al., 2002), the competitor has been directly present in the stimulus array. The subject thus had to select the correct target from among other stimuli which directly and overtly competed with it. In contrast, the current study provides evidence that the competitor effect is determined by the phonological properties of the lexicon and not by the extent of competition present in the stimulus array. Participants had to read singly presented words; the competitor of the target, the voiced minimal pair, never appeared in the experiment. Thus, the competition effects that emerged were implicit; they reflected the representational properties inherent in the mental lexicon and the extent to which a particular lexical candidate shared phonological properties with other words in the lexicon.

That there was a *reduction* in activation in the IFG for minimal pair target words compared to non-minimal target words also indicates that selection processes in the IFG may reflect facilitatory as well as interference effects. As we discuss below (see *The Nature of Modulatory Effects*), the facilitatory effects of minimal pair words in spoken word production reflect the overlap in phonological properties of the activated target and its minimal pair neighbor. In either case, the IFG is recruited when multiple representations are activated and a candidate word must be selected from among these multiple representations.

In addition to competition effects emerging in the IFG, modulatory effects also emerged in the ventral precentral gyrus extending into the post-central gyrus. This modulation of activation in the precentral gyrus as a function of the lexical properties of words (i.e. whether or not a target stimulus had a minimal pair) suggests that information flow from those areas involved in lexical processing (SMG) and lexical selection (IFG) is retained and cascades to those areas involved in articulatory planning and articulatory implementation (precentral gyrus). Thus, these results suggest that spoken word production recruits a neural system in which the extent of neural activation at the lexical level modulates activation in those neural areas involved in post-lexical processes including articulatory implementation.

### The Nature of Modulatory Effects

Competition effects in the literature typically result in increased activation. Such results have been shown in auditory word recognition not only in the context of semantic competition (Thompson-Schill et al., 1999; Bilenko et al., 2008) but also in the context of phonological competition. In particular, increased activation has been shown in auditory word recognition when accessing words that share phonological onsets (Righi et al., 2009) and for words which have many phonological neighbors compared to those that have a few (Prabhakaran et al., 2006). Consistent with these findings, Abel et al. (2009) showed increased activation in naming the picture of a word with an auditory distractor presented 200 ms prior to the target that shared the initial consonant and vowel with the target.

In contrast to increased neural activation, the current study, as well as several other studies (Bles and Jansma, 2008; de Zubicaray & McMahon, 2009; de Zubicaray et al., 2002), have shown reduced neural activation under conditions of phonological competition. As noted in the Introduction, a similar contrast has been found behaviorally; interference from neighbors is found in speech perception tasks vs. facilitation of processing by neighbors in production. Dell & Gordon (2003) attribute these contrasting patterns to the differential demands of perception and production. In perception, phonologically related words are strongly activated by the incoming acoustic signal; the listener's task is made more difficult by the presence of many phonologically related words. In contrast, production is driven by meaning. Since the primary competitors for selection are semantically related words, phonologically related words do not substantially interfere with target encoding. In this context, target selection can benefit from the boost it receives from structure it shares with phonologically related words. Thus, the reduction in activation shown in the current study for target words which had a minimal pair is likely due to the overlap between the phonological representation of the target word and its minimal pair neighbor. In particular, minimal pairs share all phonological properties of the word except for the voicing of the initial consonant. This large overlap in the number of sound segments that the competitor shares with the target word primes or facilitates the production of the target word by increasing the activation of these shared segments in relation to the other sound segments in the lexicon and by facilitating those processes involved in both planning and implementing articulatory routines. Hence, fewer neural resources are required to access the sound shape of the target stimulus, leading to reduced neural activation (cf. also de Zubicaray & McMahon, 2009).

In our study, facilitatory effects emerged not only in terms of the neural response but also behaviorally. We compared the naming latencies for the target words with and without minimal pairs measuring from the onset of the visually presented target. Results showed a trend ( $p < .09$ ) for faster naming latencies for minimal pair targets (509 ms) compared to non-minimal pair targets (516 ms). Similar patterns were found in examining naming latencies for the subset of words used in the fMRI analyses taken from The English Lexicon Project (Balota et al., 2007; <http://elexicon.wustl.edu/default.asp>) (three pairs were excluded because one or both of the items was not in the database). Minimal pair words had a naming latency of 616 ms whereas non-minimal pair words had a naming latency of 628 ms.

### The Functional Architecture of Spoken Word Production

Taken together, the results of this study provide additional support for those models of spoken word production in which the extent of activation resulting from competition at the lexical phonological level affects the activation of phonetic representations and ultimately articulatory processes (Baese-Bark & Goldrick, 2009; Goldrick & Blumstein, 2006). The present pattern of results suggests that modulatory effects conditioned by the presence of phonological competition emerge in the SMG and cascade to frontal areas including the IFG and precentral gyrus. The effects of competition then are not simply resolved once the target word is selected or even planned for articulation. Rather, competition effects continue throughout the neural network leaving their signature in those areas involved not just in lexical access and the resolution of competition and selection but also those involved in phonetic processes in production.

These findings are consistent with other behavioral studies of speech production that have documented lexical influences on speech articulation. Studies of speech errors have shown "traces" of the phonetic properties of the target are present in both acoustic (Goldrick & Blumstein, 2006) and articulatory measures (McMillan, Corley, & Lickley, 2009). Critically, these phonetic traces are sensitive to the lexical properties of the produced

utterance (e.g., whether the utterance results in a word or nonword). This is consistent with the presence of cascading activation from lexical phonological to phonetic processes.

The results reported here augment these behavioral findings by demonstrating that brain areas involved in phonetic processing are influenced by lexical properties. This modulation of phonetic processing by lexical properties is consistent with theories of spoken word production that allow lexically-driven activation to cascade to phonetic processes. Functional theories postulating a discrete relationship between lexical and phonetic processes (e.g., Levelt, Roelofs, & Meyer, 1999) cannot account for such effects.

## Acknowledgments

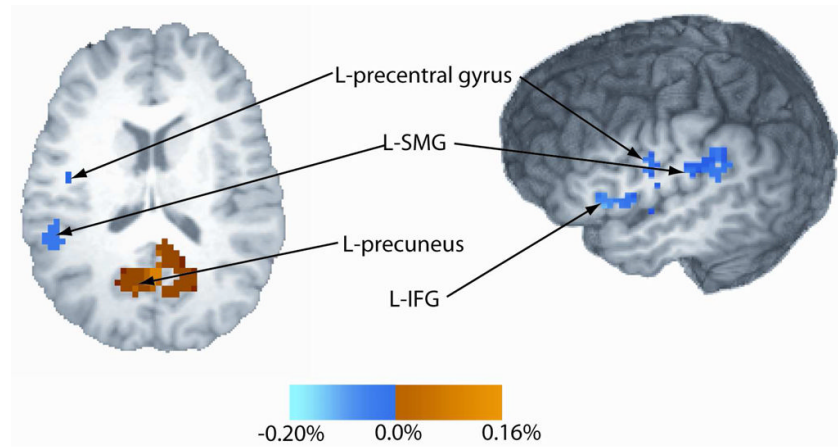
This research was supported in part by NIH NIDCD Grant RO1 DC006220 to Brown University and NIH NIDCD Grant R03 DC007977 to Northwestern University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

## References

- Abel S, Dressel K, Bitzer R, Kummerer D, Mader I, Weiller C, Huber W. The separation of processing stages in a lexical interference fMRI-paradigm. *Neuroimage* 2009;44:1113–1124. [PubMed: 19015036]
- Badre D, Wagner AW. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 2007;45:2883–2901. [PubMed: 17675110]
- Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, Loftis B, Neely JH, Nelson DL, Simpson GB, Treiman R. The English Lexicon Project. *Behavior Research Methods* 2007;39:445–459. [PubMed: 17958156]
- Baese-Berk M, Goldrick M. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 2009;24:527–554. [PubMed: 19946622]
- Bilenko N, Grindrod C, Blumstein SE. Neural Correlates of Semantic Competition during Processing of Ambiguous Words. *Journal of Cognitive Neuroscience* 2008;21:960–975. [PubMed: 18702579]
- Bles M, Jansma BM. Phonological processing of ignored distractor pictures: An fMRI investigation. *BMC Neuroscience* 2008;11:9–20.
- Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. *Neuroimage* 2006;32:821–841. [PubMed: 16730195]
- Bookheimer SY, Zeffiro TA, Blaxton T, Gaillard W, Theodore W. Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping* 1995;3:93–106.
- Buckner RL, Raichle ME, Petersen SE. Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. *Journal of Neurophysiology* 1995;74:2163–2173. [PubMed: 8592204]
- Burton MW. The role of inferior frontal cortex in phonological processing. *Cognitive Science* 2001;25:695–709.
- Cox RW, Hyde JS. Software tools for analysis and visualization of FMRI data. *NMR in Biomedicine* 1997;10:171–178. [PubMed: 9430344]
- Cox RW, Jesmanowicz A. Real-time 3-D image registration for functional MRI. *Magnetic Resonance in Medicine* 1999;42:1014–1018. [PubMed: 10571921]
- Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* 1996;29:162–173. [PubMed: 8812068]
- Dell, GS.; Gordon, JK. Neighbors in the lexicon: Friends or foes?. In: Schiller, NO.; Meyer, AS., editors. *Phonetics and phonology in language comprehension and production: Differences and similarities*. New York: Mouton de Gruyter; 2003.
- de Zubicaray GI, McMahon KL, Eastburn MM, Wilson SJ. Orthographic/phonological facilitation of naming responses in the picture-word task: an event-related fMRI study using overt vocal responding. *Neuroimage* 2002;16:1084–1093. [PubMed: 12202095]

- de Zubicaray GI, McMahon KL. Auditory context effects in picture naming investigated with event related fMRI. *Cognitive, Affective, & Behavioral Neuroscience* 2009;9:260–269.
- Dell GS. A spreading activation theory of retrieval in sentence production. *Psychological Review* 1986;93:283–321. [PubMed: 3749399]
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 2005;25:1325–1335. [PubMed: 15850749]
- Fiez JA. Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human Brain Mapping* 1997;5:79–83. [PubMed: 10096412]
- Fiez JA, Balota DA, Raichle ME, Petersen SE. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron* 1999;24:205–218. [PubMed: 10677038]
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine* 1995;33:636–647. [PubMed: 7596267]
- Gaskell MG, Marslen-Wilson WD. Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science* 1999;23:439–462.
- Gold B, Buckner R. Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron* 2002;35:803–812. [PubMed: 12194878]
- Goldrick M. Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes* 2006;21:817–855.
- Goldrick M, Blumstein SE. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 2006;21:649–683.
- Guenther FH. Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders* 2006;39:350–365. [PubMed: 16887139]
- Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science* 2000;4:131–138.
- Huang J, Varr TH, Cao Y. Comparing cortical activations for silent and overt speech using event-related fMRI. *Human Brain Mapping* 2001;15:39–53. [PubMed: 11747099]
- Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. *Cognition* 2004;92:101–144. [PubMed: 15037128]
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22:1–75. [PubMed: 11301520]
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 1998;19:1–36. [PubMed: 9504270]
- McMillan C, Corey M, Lickley R. Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes* 2009;24:44–66.
- Mertus, JA. BLISS: The Brown Lab interactive speech system. Providence, RI: Brown University; 2002.
- Munson, B. Lexical access, lexical representation, and vowel production. In: Cole, JS.; Hualde, JL., editors. *Laboratory Phonology*. Vol. 9. New York: Mouton de Gruyter; 2007. p. 201–228.
- Munson B, Solomon NP. The effects of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 2004;47:1048–1058.
- Okada K, Hickok G. Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *Neuroreport* 2006;17:1293–1296. [PubMed: 16951572]
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113. [PubMed: 5146491]
- Pardo JS. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119:2382–2393.
- Paulesu E, Frith CD, Frackowiak RSJ. The neural correlates of the verbal component of working memory. *Nature* 1993;362:342–345. [PubMed: 8455719]

- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrielli JD. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 1999;10:15–35. [PubMed: 10385578]
- Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, Britton B. An event related fMRI investigation of phonological-lexical competition. *Neuropsychologia* 2006;44:2209–2221. [PubMed: 16842827]
- Pugh KR, Shaywitz BA, Shaywitz SE, Constable RT, Skudlarski P, Fulbright RK, Bronen RA, Shankweiler DP, Katz L, Fletcher JM, Gore JC. Cerebral organization of component processes in reading. *Brain* 1996;119:1221–1238. [PubMed: 8813285]
- Righi G, Blumstein S, Mertus J, Worden M. Neural systems underlying lexical competition: An eyetracking and fMRI study. *Journal of Cognitive Neuroscience*. 2009 epub ahead of print.
- Scarborough, R. Lexical and contextual predictability: Confluent effects on the production of vowels. In: Fougerson, C.; D’Imperio, M., editors. *Papers in Laboratory Phonology*. Vol. 10. Berlin: Mouton de Gruyter; in press
- Schnur T, Schwartz M, Kimber D, Hirshorn E, Coslett H, Thompson-Schill S. Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca’s area. *Proceedings of the National Academy of Sciences* 2009;106:322–327.
- Snyder HR, Feigenson K, Thompson-Schill SL. Prefrontal cortical response to conflict during semantic and phonological tasks. *Journal of Cognitive Neuroscience* 2007;19:761–775. [PubMed: 17488203]
- Thompson-Schill SL, D’Esposito M, Aguirre G, Farah M. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences* 1997;94:14792–14797.
- Thompson-Schill SL, D’Esposito M, Kan IP. Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron* 1999;23:513–522. [PubMed: 10433263]
- Vitevitch MS. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2002;28(7):35–747.



**Figure 1.**

Clusters significant at a voxel threshold  $p < 0.05$  for the MP-NMP comparison. Activations are presented as percent signal changes. With the exception of the precuneus cluster, all clusters showed greater activation for the NMP condition. On the left, the axial slice ( $z=17$ ) shows a large medial cluster in the Precuneus (289 voxels). On the right, the Sagittal slice at  $x=43$ , shows three clusters in the Left SMG (193 voxels), L IFG (103 voxels) and L Precentral Gyrus (88 voxels).

Table 1

Mean voice-onset time (VOT), across all subjects of each item used in the fMRI analysis.

MP Stimuli	Average (ms)	STD Error	NMP Stimuli	Average (ms)	STD Error	MP-NMP	Number of subjects who showed an MP > NMP effect for that pair
tab	104.7	7.4	tat	90.7	7.0	14.0	11
tan	98.6	4.4	tag	88.8	3.9	9.9	10
tank	101.7	10.0	tap	89.8	8.4	11.9	9
teal	110.0	5.6	teat	89.2	5.3	20.8	13
teem	102.3	5.4	teethe	88.7	4.7	13.7	13
tick	96.1	11.6	tiff	84.0	7.3	12.1	10
tuck	80.4	8.2	tuft	68.2	5.6	12.2	9
ted	99.0	6.2	tempt	88.3	5.9	10.6	15
tense	109.4	11.6	tenth	95.8	12.1	13.5	8
tart	120.9	12.4	tar	110.8	7.2	10.1	5
taunt	96.7	6.0	torch	82.7	5.7	14.0	12
tore	106.8	6.5	taut	87.1	5.7	19.7	12
torque	108.4	12.5	torn	96.5	11.5	11.9	8
tomb	110.3	8.5	tooth	91.9	7.1	18.3	11
tame	97.5	6.8	taint	89.0	6.5	8.5	15
tile	96.3	5.4	tights	83.1	6.4	13.3	14
toe	107.4	6.9	toast	88.7	5.6	18.8	14
tote	103.4	13.9	toad	97.3	14.4	6.1	6
cob	108.1	8.5	cog	100.2	9.2	7.9	12
cod	94.8	5.9	cop	82.8	5.7	11.9	12
kilt	93.7	8.8	kin	82.6	7.4	11.1	4
kit	79.0	*	kiln	77.7	*	1.3	1
core	122.3	8.6	corn	108.1	7.7	14.2	14
cuss	101.8	7.6	cub	93.1	7.8	8.8	11
cuff	95.8	11.5	cud	89.3	10.6	6.4	6
curl	116.1	8.3	curb	105.7	9.1	10.4	11
coo	116.6	6.6	coot	97.1	6.2	19.4	12
cab	121.4	8.9	cad	108.5	7.4	12.8	8
cape	91.1	11.4	cake	79.3	8.9	11.9	9

MP Stimuli	Average (ms)	STD Error	NMP Stimuli	Average (ms)	STD Error	MP-NMP	Number of subjects who showed an MP > NMP effect for that pair
code	92.2	7.0	comb	90.2	9.3	2.0	7
MEAN	102.8			90.8		12.0	

A star (\*) indicates that there was only one observation for this stimulus and hence no std error could be computed.



Clusters thresholded at a cluster-level threshold of  $p < 0.05$  with a minimum of 80 contiguous voxels, and at a voxel-level threshold of  $p < 0.05$ ,  $t = 2.110$ .

**Table 2**

Cortical Region	Brodmann Areas	Cluster Size	t-stat (P value)	Talairach Coordinates		
				x	y	z
<i>Non Minimal Pair &gt; Minimal Pair</i>						
Left supramarginal gyrus	40, 22	193	4.331 (0.00045)	-53	-41	24
Left inferior frontal gyrus	44, 45, 47, 13	103	4.738 (0.00019)	-53	14	6
Left precentral gyrus	6, 2	88	4.323 (0.00046)	-53	-2	27
<i>Minimal Pair &gt; Non Minimal Pair</i>						
Left precuneus	31, 23	289	5.142 (0.000082)	-2	-59	18

The coordinates indicate the voxel with the largest intensity for that cluster. The t-stat column gives the t statistic at that maximum intensity point, and the corresponding voxel-level p value. The Brodmann Areas for anatomical regions where the clusters overlap are also given. The Eickhoff and Zilles Atlas (N27) was used to identify cortical regions.

**Table 3**

The list of filler words used in the experiment

bait	dump	jail	pond
bale	fierce	joke	rate
bear	fish	just	ripe
badge	fool	lamp	round
bike	fun	lane	sage
boss	gang	lint	self
charge	geese	look	share
chase	ghost	male	shirt
chip	gown	mint	they
church	guide	moth	vain
dance	gum	pack	vast
date	hand	peach	wait
deep	hedge	pest	west
dice	hole	pink	zip
duke	hunt	poise	zone