Sublexical Properties of Spoken Words Modulate Activity in Broca's Area but Not Superior Temporal Cortex: Implications for Models of Speech Recognition

Kenneth I. Vaden Jr.*, Tepring Piquado, and Gregory Hickok

Abstract

■ Many models of spoken word recognition posit that the acoustic stream is parsed into phoneme level units, which in turn activate larger representations [McClelland, J. L., & Elman, J. L. The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86, 1986], whereas others suggest that larger units of analysis are activated without the need for segmental mediation [Greenberg, S. A multitier theoretical framework for understanding spoken language. In S. Greenberg & W. A. Ainsworth (Eds.), *Listening to speech: An auditory perspective* (pp. 411–433). Mahwah, NJ: Erlbaum, 2005; Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312, 1979; Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 79,* 124–145, 1972]. Identifying segmental effects in the brain's response to speech may speak to this question. For example, if such effects were localized to relatively early processing stages in auditory cortex, this would support a model of speech recognition in which segmental units are explicitly parsed out. In contrast, segmental processes that occur outside auditory cortex may indicate that alternative models should be considered. The current fMRI experiment manipulated the phonotactic frequency (PF) of words that were auditorily presented in short lists while participants performed a pseudoword detection task. PF is thought to modulate networks in which phoneme level units are represented. The present experiment identified activity in the left inferior frontal gyrus that was positively correlated with PF. No effects of PF were found in temporal lobe regions. We propose that the observed phonotactic effects during speech listening reflect the strength of the association between acoustic speech patterns and articulatory speech codes involving phoneme level units. On the basis of existing lesion evidence, we interpret the function of this auditory–motor association as playing a role primarily in production. These findings are consistent with the view that phoneme level units are not necessarily accessed during speech recognition. ■

INTRODUCTION

The internal structure of words can be represented in part as an ordered sequence of phonemes that themselves are composed of a collection of feature bundles corresponding to speech articulation parameters (voicing, place, and manner of articulation). There is general agreement that such internal structure of words exists, as this level of analysis has clear import for speech production—the fact that the distinctive features are in articulator space attests to this—but more controversy exists over whether these sublexical units are recovered during speech recognition. Whereas most models assume that speech is recognized first by parsing the acoustic stream into phoneme level units and building from there (Poeppel, Idsardi, & van Wassenhove, 2008; Stevens, 2002; McClelland & Elman, 1986), others have suggested larger basic units of analysis (Greenberg, 2005; Klatt, 1979; Massaro, 1972; see Hickok & Poeppel, 2007, for a hybrid view). Research on the neuroscience of speech perception has successfully documented word level effects in auditory-related areas in the superior temporal gyrus (STG) but have not provided convincing evidence for sublexical effects (Okada & Hickok, 2006; Friedrich, 2005; Stockall, Stringfellow, & Marantz, 2004; Pylkkänen & Marantz, 2003; Pylkkänen, Stringfellow, & Marantz, 2002; Pulvermüller et al., 1996). This is potentially relevant to models of speech perception because, assuming a hierarchical organization of the auditory system (Okada et al., 2010; Davis & Johnsrude, 2003; Binder et al., 2000), a straightforward prediction of segmental models is that sublexical effects should be identifiable in auditory cortex, presumably upstream to regions showing lexical effects. Alternatively, if sublexical effects are not evident in the auditory processing stream, this may indicate a larger size basic unit of speech perception.

The current study aimed to functionally identify sublexical phonological activity during spoken word recognition. One metric that has been used to index sublexical processing in behavioral studies is phonotactic frequency (PF), a measure of the co-occurrence frequency of pairs of phonemes in the language. This variable is

University of California at Irvine
*Current affiliation: Medical University of South Carolina.

often manipulated along with another measure that indexes lexical–phonological processing, neighborhood density (ND), the number of similar sounding neighbors a word has. Using PF and ND manipulations in behavioral speech perception experiments, independent sublexical and lexical influences have been reported for many tasks, including word-likeness judgments (Bailey & Hahn, 2001; Frisch, Large, & Pisoni, 2000), same–different word decisions (Vitevitch, 2003; Luce & Large, 2001), nonword recall (Thorn & Frankish, 2005), and word learning (Storkel, Armbrüster, & Hogan, 2006). Although one might assume a serial model in explaining these effects (e.g., PF effects occur at an earlier stage of perceptual processing, which then feeds into the next level where ND effects occur), this need not be the case as noted by Vitevitch and Luce (1999), who write, "We … use the term 'level' … to refer to representations corresponding to lexical and sublexical representations. However, we do not assume that activation of sublexical units is a necessary prerequisite to activation of lexical units" (p. 376).

The goal of the current experiment was to localize sublexical activity during spoken word recognition, using PF manipulations of real words. We used real words because we are interested in studying processes leading to normal word recognition. Recent behavioral studies have successfully documented PF effects in real words (Mattys, White, & Melhorn, 2005; Vitevitch, Armbrüster, & Chu, 2004; Vitevitch, 2003). In addition to manipulating PF, we orthogonally manipulated ND as a means to identify lexical level processes (Luce & Large, 2001). Previous functional imaging studies covaried PF and ND and failed to find evidence of PF effects (Okada & Hickok, 2006; Prabhakaran, Blumstein, Myers, Hutchison, & Britton, 2006). We hoped that orthogonal manipulations would improve our ability to detect PF effects.

We predicted that lexical–phonological effects would be found in the STS and/or supramarginal gyrus (SMG) consistent with previous results (Okada & Hickok, 2006; Prabhakaran et al., 2006) and also that a distinct network would be sensitive to sublexical manipulations. If this network is located within auditory cortical regions, ideally in a region that could be interpreted as upstream from the lexical–phonological effects, this would support a serial model of speech perception, whereas if sublexical effects are found elsewhere, some alternative models must be considered.

METHODS

Participants

Twenty-one volunteer subjects, between the ages of 18 and 29 years old (mean = 21.95, SD = 3.22) participated in the experiment. There were 12 men and 9 women. All participants were right-handed, native English speakers, free of neurological disease, and had normal hearing by self-report. All subjects gave informed consent under the

protocol approved by the institutional review board of the University of California at Irvine.

Experimental Procedure

Each subject participated in a single experimental session lasting approximately 1 hr at the Phillips 3T scanner at the University of California at Irvine and was paid $30 for participation. Informed consent and health screening were obtained just before the session. Before the experiment began, the volume of auditory stimulation was adjusted to comfortable levels with feedback from the subject. The task was to monitor for the occasional presentation of pseudowords. This task was utilized to ensure lexical recognition of the word stimuli, which contained psycholinguistic manipulations. Subjects were instructed to listen to each wordlist and press a button only if that list contained one or more pseudowords. Pseudoword trials were excluded from the fMRI analysis. Between runs, we used the intercom to question subjects whether the volume was sufficient and to check whether they sounded alert.

Design

During the fMRI experiment, subjects listened to lists of four words selected from the same combination of ND (high or low) and PF (high or low). The experiment contained eight runs of blocked trials. In each of the eight runs, subjects were presented with six wordlists from each of the four combinations of density and phonotactics for a total of 24 experimental trials per run. During each run, there were also two catch trials containing pseudowords to verify that the subject could understand the words and was paying attention. There were 16 catch trials across the entire experiment. Table 1 summarizes trials and volumes collected per condition.

Trials consisted of speech stimulation followed by a jittered duration rest period, which was 8.4, 10.5, or 12.6 sec in length (4, 5, or 6 TRs). All conditions were presented in each of the jittered durations an equal number of times in each run. Words were presented with a silent ISI of 150 msec, so stimulus presentation lasted 2.55 sec, on average, equalized across conditions. The mean trial duration for each condition was, in milliseconds; high ND high PF = 2539 (SD = 32), high ND low PF = 2532 (SD = 34), low ND high PF = 2566 (SD = 34), and low ND low PF = 2573 (SD = 35). Trials were synchronized with serial bytes from the scanner, which signaled the stimulus computer that a functional image acquisition had started. The order that conditions are presented in may affect the resultant statistical efficiency of fMRI design, because BOLD responses typically overlap from one trial to another, and the number of overlaps between conditions is often unbalanced following trial order randomization. We modified the genetic algorithm (Wager & Nichols, 2003) to pseudorandomize trial order

Table 1. Summary of Experiment Conditions

| Condition	Trials per Run	Total Trials
Low density, low PF e.g., sniff, jolt, bribe, flag	6 trials, 30 volumes	48 trials, 240 volumes
Low density, high PF e.g., crib, blush, probe, spice	6 trials, 30 volumes	48 trials, 240 volumes
High ND, low PF e.g., belch, clot, sneak, fright	6 trials, 30 volumes	48 trials, 240 volumes
High ND, high PF e.g., crate, spill, fond, truce	6 trials, 30 volumes	48 trials, 240 volumes
Catch trials e.g., pinch, yorm, henth, fret	2 trials, 10 volumes	16 trials, 80 volumes
Total	26 trials, 130 volumes	208 trials, 1040 volumes

A summary of the data collected during each run and experiment session. Subjects heard lists from each of the four conditions with equal frequency. Catch trials appeared only twice per run.

for each subject on the basis of simulated hemodynamic response functions (HRFs), with the extra constraint that no condition occurred twice in a row.

Stimuli

Stimuli were recordings of spoken English CCVC or CVCC words. All items were selected using www.iphod.com (Vaden, Halpin, & Hickok, 2009) among orthogonal ranges of PFs and ND, as illustrated in Figure 1. PF ranges were defined using unweighted average biphoneme probability, which refers to the average frequency of each word's ordered phoneme pairs (mean high PF = 0.0034, low PF = 0.0015). ND ranges were defined with unweighted phonological ND, which counts all words that are only one phoneme different (mean high ND = 14.43, low ND = 6.22). A correlation test found no relationship between PF and ND among the stimuli, R^2 = .013, p = .90. Words were selected using bootstrapping procedures and ANOVAs to ensure that no other lexical or recording characteristics varied with density or phonotactic groups. Analyses of variance were used to determine that there were no differences in Kucera–Francis word frequency nor recording durations by condition and no interactions, $F(1, 96) < 0.75, p > .5$. There were no significant differences among high and low PF words or interactions with density groups in ND values, $F(1, 96) < 0.20, p > .65$. There were no significant biphoneme probability differences among density groups or interactions with PF conditions, $F(1, 96) < 0.10, p > .77$. Additional details concerning stimulus preparation and controls are found in Vaden (2009).

Because of the nature of the words, a post hoc analysis was performed to determine that there were no significant
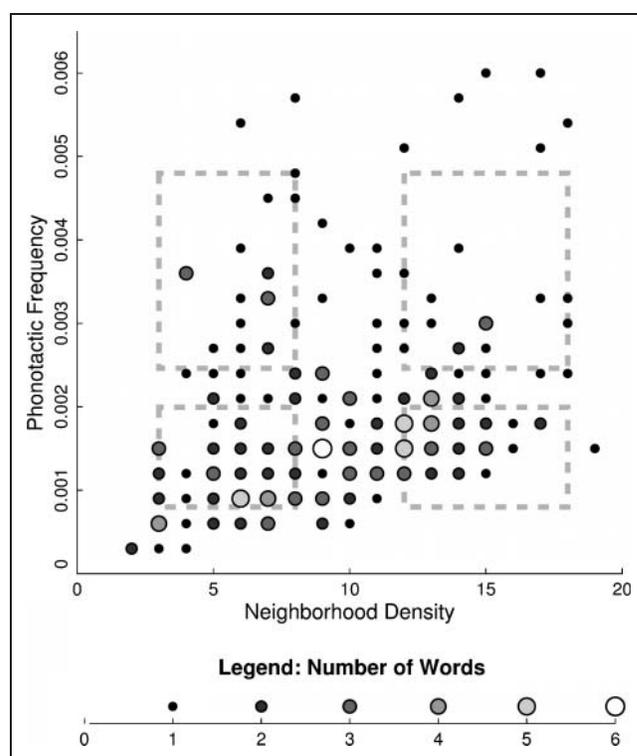
Figure 1. IPhOD words and item selection. The distribution of all monosyllabic, four-phoneme-long words in the IPhOD collection, either with CVCC or CCVC consonant–vowel structures. Each dot counts the number of words that occurred in a particular range of PF and ND values. The superimposed dashed squares show the selection ranges used to find 25 words for each cell in the 2 × 2 design: high or low PF and high or low ND. Despite the negative skew of the broader word population, this distribution allowed us to choose 100 words with statistically independent PF and ND values. We found that words with other consonant–vowel structures (such as CVC words) have stronger correlations between PF and ND, which challenged their independent manipulation.

relationships between ND or PF with concreteness. All 100 stimuli words were judged by 30 participants who did not participate in the fMRI study. Concreteness ratings were collected on a 1–7 scale using a method adapted from Cortese and Fugett (2004). Mean concreteness values were then entered as a predictor variable in our fMRI analysis and were found not to affect the results reported below.

Each item was recorded in an anechoic chamber, after practicing the pronunciation several times, to ensure natural speaking rate and clear pronunciation. Shure amplifier and Dell PC were used with Audacity software for recording and editing each item. We used a Matlab script to RMS normalize recordings to equalize the perceived loudness across all the words. Finally, ANOVAs performed on the recording durations showed no differences between lexicality or PF and ND conditions nor were there significant interactions, Fs < 0.75.

Scan Procedure and Preprocessing

The 3T Phillips MRI at the University of California at Irvine was used for this study. Cogent 2000 scripts (Romaya, 2003) synchronized sound delivery and response collection with the onsite button-box system and Resonance Technologies MR-compatible headphones (Resonance Technology, Inc., Northridge, CA). Functional volumes were acquired and analyzed in native 2.3 × 2.3 × 4 mm voxel dimensions, and 34 slices provided whole-brain coverage. Trial lengths were jittered to collect equal numbers of four, five, or six functional volumes across each condition of interest. Other specifications for the EPI sequence are TR = 2.1 sec, TE = 26 msec, flip angle = 90, field of view = 200; 130 volumes were acquired in 273 sec/run. Anatomical 1.0-mm^3 isomorphic images were collected using a T1-weighted sequence following all eight experiment runs. Four dummy scans were used in the beginning of the sequence. Subjects were asked to keep their eyes closed throughout each experiment run.

Preprocessing first and second level analyses were performed using SPM5 (Wellcome Department of Imaging Neuroscience). Data preprocessing included slice-timing correction, motion correction, and coregistration of the anatomical to the middle functional volume in the series. Anatomical and functional images were reoriented and normalized to a study-specific template in MNI space using Advanced Normalization Tools (www.picsl.upenn.edu/ ANTS; Klein et al., 2009). Spatial smoothing was performed in SPM5 using a 6-mm FWHM Gaussian kernel. Global mean signal fluctuations were detrended from the preprocessed functional images using voxel level linear model of the global signal (Macey, Macey, Kumar, & Harper, 2004).

After the functional images were preprocessed, we applied an algorithm described in Vaden, Muftuler, and Hickok (2010) to generate two nuisance regressors that identified extreme intensity fluctuations that occurred during each run on a per volume basis. The first vector detected volumes whose global intensity greatly exceeded the mean. The second vector labeled volumes that contained large numbers of voxels with higher-than-average intensity. Cutoff values were set to 2.5 SD from the mean. This algorithm identified 5.7 images per run, and only 11% of the volumes were shared by both vectors. The two outlier vectors were submitted to the general linear model (GLM) as nuisance variables, in addition to six motion vectors that were generated during realignment. Our approach did not censor or exclude functional volumes from analysis—instead the outlier and nuisance vectors that we used were able to account for extreme but attributable variability within voxel time courses.

fMRI Analyses

Preprocessed functional images were submitted to a parametric analysis in SPM5 at the individual level. Each word was modeled as a separate event with onset and duration from a main condition that had two parameters: PF and ND. All wordlist presentations were modeled in the GLM, except for catch trials that contained pseudowords. In this manner, functional time courses were fit using onsets and durations, convolved with the HRF, and parametric phonotactic and density values for each item that modulated the HRF. The GLM also included eight volume-wise nuisance regressors: six motion correction and two outlier vectors. Resultant individual level t statistic maps for the ND and PF regressions were submitted to a second level, random effects analysis to localize consistent phonotactic and density effects across subjects.

Three subjects were excluded from the group analyses because of behavioral errors that exceeded two standard deviations from the mean by error type. The first subject had excessive false alarms (FA = 0.41), the second had excessive misses (HR = 0.25), and the third had a high combination of misses and false alarms (FA = 0.25, HR = 0.56, A' = 0.74). A fourth was excluded from the second level analyses because there were no responses recorded because of a technical error. Group level t statistical maps (n = 17) were initially thresholded at a more lenient value to increase sensitivity to borderline significant results, t = 2.92 (df = 16, p = .005) with cluster size extent > 20 voxels (p = .05, uncorrected). Results are also reported using a stricter threshold, t = 3.69 (df = 16, p = .001). All reported results were corrected for multiple comparisons at the cluster level, with corrected p = .05.

We performed two additional planned submodel analyses to examine whether activity correlated with the density or phonotactic regressor differently when both were not entered into GLM. Similarly, Wilson, Isenberg, and Hickok (2009) used submodels to detect correlations that may have been obscured by performing multiple regressions on inherently colinear lexical variables. The post hoc analyses used t statistic and cluster extent thresholds that were identical to the main model. Because our variables were manipulated orthogonally, we did not expect to

see different patterns of results between the main model and submodels.

RESULTS

Behavioral Performance

We began the analysis by examining response data from all of the subjects. The task was to press a button whenever a list contained pseudowords. Subjects responded correctly in 85.6% of the trials, on average ($SD = 10.9\%$). The average hit rate was 71.1% ($SD = 20.0\%$), and average false alarm rate was 14.0% ($SD = 12.4\%$). Correcting for bias, the average proportion correct is estimated by $A' = 76.9\%$ ($SD = 5.26\%$); A' ranged from 74.0% to 99.4%.

A logistic regression analysis was performed using the R system for statistical computing (www.r-project.org) to determine if false alarm responses were systematically related to PF and ND. The same subjects were excluded as in the group fMRI analysis described in Methods. Specifically, we performed a logistic regression analysis across subjects[1] (Baayen, Davidson, & Bates, 2008) to identify significant correlations between PF and ND factors and false alarm responses. The association between PF and false alarms was significant, $Z = 3.89, p = .0001$. The association between ND and false alarms was $Z = 3.62, p = .0003$. Using ANOVAs, we found that PF ($\chi^2 = 13.22, p = .0003$) and ND ($\chi^2 = 15.21, p = .00001$) factors significantly improved the model, whereas including their interaction term did not ($\chi^2 = 1.74, p = .19$). Subjects made more false alarms on high PF words than low PF words, and high-density words resulted in more false alarms than low-density words. The direction of this effect

was not expected (high-frequency and high-density items causing more errors). We speculate that this is related to the fact that the overall lexical frequency of our items was low. It may be that low-frequency words containing high-frequency sequences (e.g., trot) or from high-density neighborhoods (e.g., dank) are more often judged as nonwords because there is a tendency to judge lexical status relative to high-frequency cohorts. Nonetheless, these results indicate that our PF manipulation was successful in modulating behavioral responses to word stimuli.

Functional Image Results

In the main analysis, we initially identified clusters that were significantly correlated with PF or ND by thresholding t statistic maps at $t(16) = 2.92, p = .005$, and cluster size > 20 voxels ($p = .05$, uncorrected). This analysis found significant parametric effects of PF, but not ND, on activity during word listening. The left inferior frontal gyrus (IFG) demonstrated a positive correlation with PF, with peak $t(16) = 6.32, p < .001$, corrected at the cluster level. The MNI coordinates of the peak were [−46, 19, −8], and the cluster contained 174 voxels in MNI space including portions of Brodmann's areas 47, 48, and 45. Following the stricter t statistic cutoff ($t > 3.69, p = .001$), the same cluster contained 36 voxels (BA 47, 45) and cluster size corrected $p = .021$. Words that consisted of more common phoneme sequences elicited greater activity in left IFG than words with unusual sequences (Figure 2).

The submodel analysis included only the ND or PF parameter to determine whether activation was sensitive to either variable when manipulated in isolation. The significant result of the submodel was consistent with the
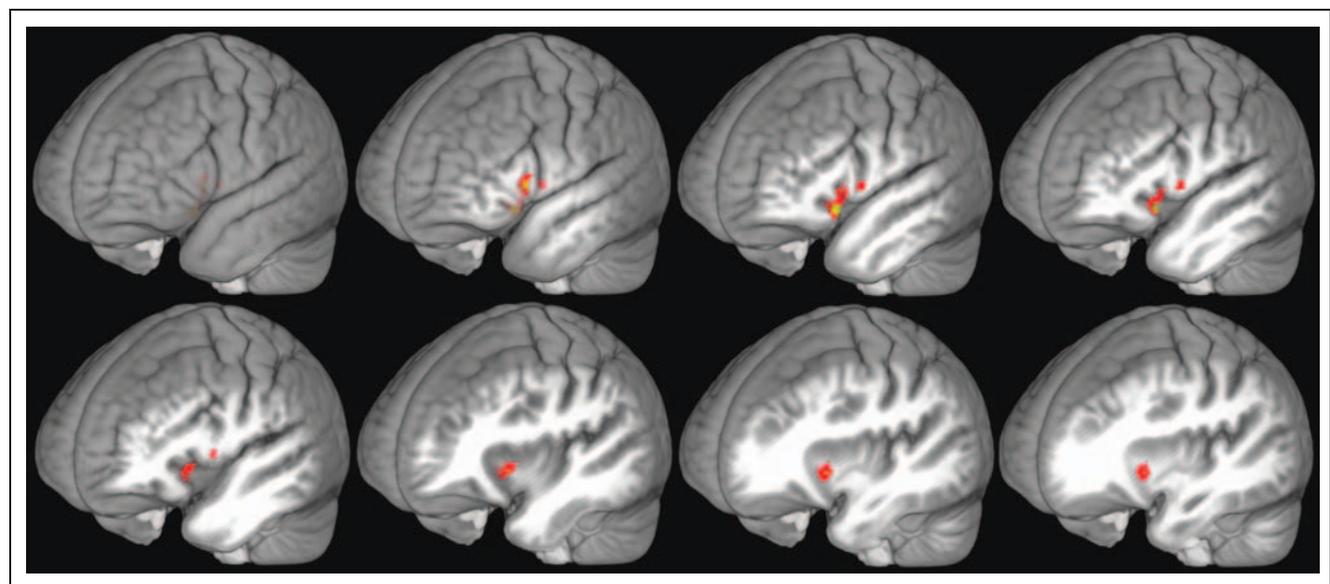
Figure 2. Positive phonotactic effect in left IFG. The left IFG increased response to words with higher PFs. Shown in red, the parametric PF modulation and monotonic trend analysis identified voxels that passed a t statistic threshold of $t(16) = 2.92, p = .005$ uncorrected, and cluster extent (174 voxels) yielded a corrected $p < .001$ at the cluster level. Yellow voxels also passed a stricter t statistic threshold of $t(16) = 3.69, p = .001$ uncorrected, and cluster size (36 voxels) yielded a corrected $p = .021$.

main model: The only significant cluster was in left IFG, positively correlated with PF but not density, even when density was the only explanatory factor. This supports that the orthogonal density and phonotactic manipulations modulated activity independently despite their computational similarity (Vitevitch, Luce, Pisoni, & Auer, 1999).

DISCUSSION

PF manipulations are thought to modulate processing at the sublexical level (Luce & Large, 2001; Vitevitch & Luce, 1999). In the present fMRI study, we found that PF manipulations in spoken word recognition resulted in robust modulation of neural activity, not in auditory-related cortex, as one might expect, but in motor-related cortex in the left IFG, a portion of Broca's area, where activity was positively correlated with PF. To the best of our knowledge this is the first time that fMRI has detected sublexical processing during spoken word recognition as a result of PF manipulations. We did not observe a main effect of ND, which is somewhat surprising given that we found behavioral effects of ND and that ND effects have been observed previously in auditory-related areas using different stimuli, tasks, and imaging modalities (Okada & Hickok, 2006; Stockall et al., 2004; Pylkkänen & Marantz, 2003). We did find a weak negative correlation (significant only at a relaxed threshold) between ND and activation in the right STG that was exaggerated for items with lower PF values. Furthermore, informal examination of individual subject data revealed ND effects in several subjects in various locations within the STS; it is possible that our sample was particularly variable in the location of ND activation, thus precluding group level significance. As our focus in this study was on PF effects, we did not pursue ND effects further.

We suggested in the Introduction that detecting PF effects at a relatively early stage of auditory cortical processing would provide evidence that is at least consistent with a hierarchical model of lexical access in which segmental information is first extracted from the acoustic stream and then subsequently used to build up or access lexical level phonological forms. Consistent with previous studies (Papoutsi et al., 2009; Okada & Hickok, 2006; Burton, Small, & Blumstein, 2000), we did not find evidence of segmental processing (i.e., PF effects) in auditory cortical fields. Instead, we found sublexical effects in motor speech-related areas. This finding raises important questions about the role of the motor system and sublexical (segmental level) processes in speech perception. In what follows, we will first consider the role of the motor system in speech perception and then discuss the implications of this for models of speech recognition.

It is relatively uncontroversial that frontal motor circuits including portions of Broca's region, the pars opercularis (BA 44) in particular, play a role in sublexical processing during production. For example, Blumstein (1995), arguing

from lesion data, has suggested that Broca's area plays a critical role in phonetic level processes, and a number of functional imaging studies have shown that activity in portions of Broca's area and surrounding regions (premotor cortex, anterior insula) is modulated by sublexical frequency manipulations similar to those we used here (Papoutsi et al., 2009; Riecker, Brendel, Ziegler, Erb, & Ackermann, 2008; Bohland & Guenther, 2006; Carreiras, Mechelli, & Price, 2006; although cf. Majerus et al., 2003). Consistent with Blumstein's proposal, Papoutsi et al. (2009) have interpreted results such as these as evidence that Broca's area, the ventral pars opercularis in particular, plays a role in phonetic encoding during speech production.

The present result indicates that these sublexical circuits are also active to some extent during perception. But what role does this motor-related sublexical information play in speech recognition processes? We will consider three possibilities: (1) that activation of sublexical articulatory speech information is critical to speech recognition, (2) that such information exerts a modulatory influence on recognition systems in auditory areas, and (3) that it is epiphenomenal to speech recognition.

Sublexical Articulatory Speech Information Is Critical to Speech Recognition

Some theorists have argued that motor-related areas comprise a critical node in the speech perception network— an idea that is typically couched in motor–theoretical terms and inspired by claims from the mirror neuron literature (D'Ausilio et al., 2009; Pulvermuller, Hauk, Nikulin, & Ilmoniemi, 2005). On this school of thought, our findings may be interpreted as evidence for the role of motor articulatory processes in building up the phonological representation of a word during spoken word recognition. This theory could maintain a hierarchical model of speech recognition, in which sublexical units are represented not in early auditory areas but in motor cortex. However, there is strong neuropsychological evidence against this view. Damage to or underdevelopment of the motor speech system, including its complete functional disruption, does not cause similar deficits in speech recognition, indicating that motor speech systems are not necessary for speech recognition (Hickok, 2009a, 2010; Lotto, Hickok, & Holt, 2009; Hickok et al., 2008). The fact that one does not need a motor speech system to recognize speech indicates that a strong version of a motor theory of speech perception, possibility (1), is incorrect.

Sublexical Articulatory Speech Information Modulates Speech Recognition

A more moderate view that has been promoted by some authors is that the motor system may at most provide a modulatory influence on perceptual processes carried out in auditory cortical fields under adverse listening conditions,

such as when the acoustic signal is degraded (Hickok, 2009b, 2010; Hickok, Holt, & Lotto, 2009; Wilson, 2009). According to this proposal, motor information provides a top–down influence on perceptual processes, perhaps in the form of forward models (Hickok & Saberi, in press; Rauschecker & Scott, 2009; Poeppel et al., 2008; Hickok & Poeppel, 2007; van Wassenhove, Grant, & Poeppel, 2005). Our subjects were listening to words against the background of scanner noise, which may have resulted in motor system recruitment. Consistent with this, we observed activation of a portion of Broca's area in the contrast, listen versus rest, which is what one would expect if the motor system is recruited during perception of speech in noise generally. However, because the significant PF effects occurred mainly in voxels that were not significantly activated by the listening–rest contrast, it is unclear whether sublexical information contributed to speech perceptual processes. For reference, Figure 3 shows that the extent of overlap occurred in the left anterior insula, whereas the majority of significantly PF correlated voxels occurred in the IFG.

Motor Speech-related Activity Is Epiphenomenal to Speech Recognition

It is possible that listening to speech "passively" activates motor articulatory systems via associative links between perception and production systems. According to this view, associative links between perception and production exist primarily for the purpose of auditory guidance or feedback control of speech production (Hickok & Saberi, in press; Hickok, 2009c; Rauschecker & Scott, 2009; Guenther, Hampson, & Johnson, 1998; Houde & Jordan, 1998). This view emphasizes a kind of sensory

theory of speech production as opposed to a motor theory of speech perception. Because perceived speech regularly interfaces with the motor system during production (e.g., in auditory feedback control), perceiving speech may activate this circuit via spreading activation even when the task demands don't require it. This spreading activation may be modulated by PF as the frequency of sublexical patterns is known to affect speech production (e.g., Goldrick & Larson, 2008; Munson, 2001). One idea is that frequently used syllables are stored in a syllabic-lexicon or syllabary (Levelt & Wheeldon, 1994), and articulating those syllables is easier than low-frequency syllables, because the former are simply retrieved as overlearned motoric sequences whereas the latter must be assembled from smaller pieces (Aichert & Ziegler, 2004). This provides a natural explanation for the direction of the PF effect in the present experiment: Higher PF words yield more activation in Broca's region because they have stronger auditory–motor associations. One would predict the reverse effect, however, during speech production— that is, more activation during production of lower-frequency sequences because of the increased assembly requirements—and this prediction appears to hold as low-frequency items generate more activity in Broca's area and surrounding fields during speech production (Papoutsi et al., 2009; Riecker et al., 2008; Bohland & Guenther, 2006; Carreiras et al., 2006; although cf. Majerus et al., 2003).

It is difficult to adjudicate between the second (modulatory) and third (epiphenomenal) possibilities regarding the role of the motor system in speech recognition, and indeed these are not mutually incompatible as motor speech activity may play a modulatory role under some circumstances and may be epiphenomenal in others. What
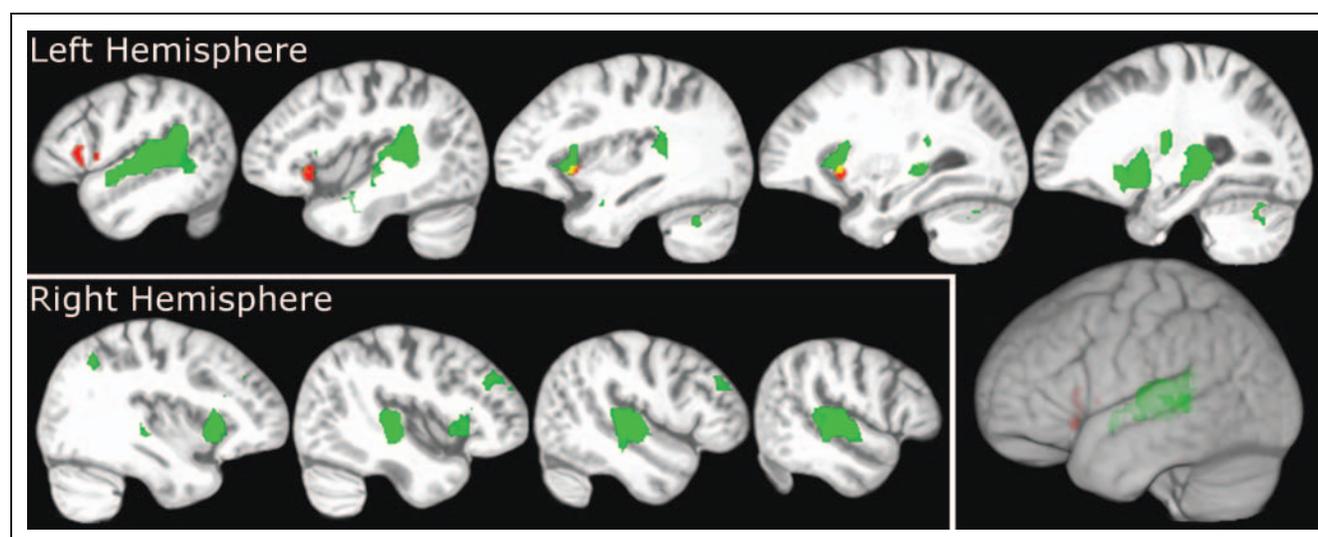
Figure 3. Auditory activity during speech perception and overlap with sublexical effects. Areas relevant to speech perception responded robustly to word presentations in the current experiment. Bilaterally, superior temporal gyrus responded to auditory stimuli, when contrasting the response to words with rest, shown in green. PF effects were found in left IFG, shown in red. We found that the contrasts only overlapped in left anterior insula (yellow). All active voxels in the two contrasts passed the statistic threshold, $t(16) = 2.92$, and a cluster extent threshold = 75 voxels, which yielded a corrected $p = .05$. The cluster extent threshold did not affect the region where the two contrasts overlapped.

is clear from much research though is that the first possibility is not viable.

The present findings have potentially important implications for models of speech recognition. Unlike previous attempts, we were successful in documenting robust sublexical effects during speech recognition, but consistent with these previous studies, we failed to find evidence of such effects in auditory regions and found them in motor speech-related regions instead. This result, coupled with the evidence from other sources indicating at most a modulatory role of the motor system in speech perception, questions the role of segmental information in speech recognition. Our findings are more in line with the view that segment level information is only represented explicitly on the motor side of speech processing and that segments are not explicitly extracted or represented as a part of spoken word recognition as some authors have proposed (Massaro, 1972). One challenge for this view comes from research showing apparent perceptual effects of transitional probabilities and PF in prelingual infants and nonhuman primates, neither of which have speech production abilities (Hauser, Newport, & Aslin, 2001; Mattys & Jusczyk, 2001; Saffran, Newport, Johnson, & Aslin, 1999; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Saffran, Newport, & Aslin, 1996; Jusczyk, Luce, & Charles-Luce, 1994; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). However, effects like these may stem from an analysis of syllable frequency rather than a fully segmented speech stream.

The idea that segmental information may be explicitly represented in the motor articulatory system but not within the auditory perceptual system explains a long-standing puzzle in the neuroscience of language. Performing so-called sublexical tasks on heard speech, such as deciding whether two syllables end with the same phoneme, yields strong activation in and around Broca's area (Burton, Paul, LoCasto, Krebs-Noble, & Gullapalli, 2005; Callan, Jones, Callan, & Akahane-Yamada, 2004; Heim, Opitz, Müller, & Friederici, 2003; Siok, Jin, Fletcher, & Tan, 2003; Burton et al., 2000; Zatorre, Evans, Meyer, & Gjedde, 1992). However, as noted previously, damage to this region does not cause substantial speech recognition deficits (Hickok & Poeppel, 2000, 2004, 2007). This is a paradox because such tasks are typically viewed as a measure of early (phonemic) perceptual processing, that is, those that feed into higher level word recognition systems, yet poor performance on sublexical tasks does not result in poor word recognition. Hickok and Poeppel (2000, 2004, 2007) attempted to resolve this paradox on the assumption that the frontal recruitment in sublexical tasks involved vaguely defined metalinguistic processes (e.g., working memory) that are not required during normal speech recognition. However, if segmental information is only explicitly represented in frontal motor-related circuits and if this information primarily serves production not recognition, as we are suggesting here, then tasks that require access to such information will necessarily involve activation of motor-related information, although the tasks are nominally "perceptual" tasks. This provides a more principled explanation of the paradox noted by Hickok and Poeppel (2000).

Conclusion

PF manipulations during auditory word recognition were found to modulate neural activity in motor speech-related systems in Broca's area but not in auditory-related areas in the superior temporal region. This finding, together with the observation that damage to Broca's area does not substantially disrupt speech recognition, is more consistent with speech perception models in which segmental information is not explicitly accessed during word recognition. We propose that the observed phonotactic effects during speech listening reflect the strength of the association between acoustic speech patterns and sublexical articulatory speech codes. This auditory–motor network functions primarily to support auditory guidance of speech production but may also be capable of modulating auditory perceptual systems via predictive coding under some circumstances.

Note

1. Because there was only one response for each four word trial, item analyses or mixed-model regressions were not performed.

REFERENCES

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language, 88,* 148–159.

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

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44,* 568–591.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10,* 512–528.

Blumstein, S. (1995). The neurobiology of the sound structure of language. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 913–929). Cambridge, MA: MIT Press.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage, 32,* 821–841.

Burton, M. W., Paul, T., LoCasto, C., Krebs-Noble, D., & Gullapalli, R. P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage, 26,* 647–661.

Burton, M. W., Small, S., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience, 12,* 679–690.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage, 22,* 1182–1194.

Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping, 27,* 963–972.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavioral Research Methods, Instruments & Computers, 36,* 384–387.

D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19,* 381–385.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience, 23,* 3423–3431.

Friedrich, C. K. (2005). Neurophysiological correlates of mismatch in lexical access. *BMC Neuroscience, 6.* Accessed on-line May 6, 2010.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42,* 481–496.

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition, 107,* 1155–1164.

Greenberg, S. (2005). A multi-tier theoretical framework for understanding spoken language. In S. Greenberg & W. A. Ainsworth (Eds.), *Listening to speech: An auditory perspective* (pp. 411–433). Mahwah, NJ: Erlbaum.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* 611–633.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78,* B53–B64.

Heim, S., Opitz, B., Müller, K., & Friederici, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research, 16,* 285–296.

Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes, 25,* 749–776.

Hickok, G. (2009a). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience, 21,* 1229–1243.

Hickok, G. (2009b). *Speech perception does not rely on motor cortex.* Available at: www.cell.com/current-biology/comments/S0960-9822(09)00556-9.

Hickok, G. (2009c). The functional neuroanatomy of language. *Physics of Life Reviews, 6,* 121–143.

Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Response to Wilson: What does motor cortex contribute to speech perception? *Trends in Cognitive Sciences, 13,* 330–331.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., et al. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language, 107,* 179–184.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92,* 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402.

Hickok, G., & Saberi, K. (in press). Redefining the functional organization of the planum temporale region: Space, objects, and sensory-motor integration. In D. Poeppel, T. Overath, A. N. Popper, & R. R. Ray (Eds.), *Springer handbook of auditory research: Human auditory cortex.* Berlin: Springer.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279,* 1213–1216.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerund, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32,* 402–420.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630–645.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage, 46,* 786–802.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239–269.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13,* 110–114.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Langauge and Cognitive Processes, 16,* 565–581.

Macey, P. M., Macey, K. E., Kumar, R., & Harper, R. M. (2004). A method for removal of global effects from fMRI time series. *Neuroimage, 22,* 360–366.

Majerus, S., Collette, F., Van der Linden, M., Peigneux, P., Laureys, S., Delfiore, G., et al. (2003). A PET investigation of lexicality and phonotactic frequency in oral language processing. *Cognitive Neuropsychology, 19,* 343–360.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 79,* 124–145.

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27,* 644–655.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134,* 477–500.

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

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44,* 778–792.

Okada, K., & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *NeuroReport, 17,* 1293–1296.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex, 20,* 2486–2495.

Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. *Cerebral Cortex, 19,* 2156–2165.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 363,* 1071–1086.

Prabhakaran, R., Blumstein, S. E., Myers, E. B., Hutchison, E., & Britton, B. (2006). An event-related fMRI investigation of phonological-lexical competition. *Neuropsychologia, 44,* 2209–2221.

Pulvermüller, F., Eulitz, C., Pantev, C., Mohr, B., Feige, B., Lutzenberger, W., et al. (1996). High-frequency cortical responses reflect lexical processing: An MEG study. *Electroencephalography and clinical Neurophysiology, 98,* 76–85.

Pulvermuller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21,* 793–797.

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7,* 187–189.

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81,* 666–678.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12,* 718–724.

Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language, 107,* 102–113.

Romaya, J. (2003). Cogent 2000, Version 1.25. [Software]. Retrieved from the Laboratory of Neurobiology Web page, June 2007: www.vislab.ucl.ac.uk/Cogent.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101–195.

Saffran, J. R., Newport, E. L., Johnson, E. K., & Aslin, R. N. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Siok, W. T., Jin, S., Fletcher, P., & Tan, L. H. (2003). Distinct brain regions associated with syllable and phoneme. *Human Brain Mapping, 18,* 201–207.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111,* 1872–1891.

Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language, 90,* 88–94.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49,* 1175–1192.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects of serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 729–735.

Vaden, K. I. (2009). *Phonological processes in speech perception* Doctoral thesis, University of California, Irvine. Retrieved from proquest.umi.com/pqdweb?did=1781083831&sid=1&Fmt=6&clientId=48051&RQT=309&VName=PQD.

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 1.4. [Data file]. Available from www.iphod.com.

Vaden, K. I., Muftuler, L. T., & Hickok, G. (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *Neuroimage, 49,* 1018–1023.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 1181–1186.

Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *Clinical Linguistics and Phonetics, 17,* 487–499.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 514–529.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68,* 306–311.

Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: A general framework using a genetic algorithm. *Neuroimage, 18,* 293–309.

Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences, 13,* 329–330.

Wilson, S. M., Isenberg, A. L., & Hickok, G. (2009). Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. *Human Brain Mapping,* Published On-line: April 2009.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256,* 846–849.