



Published in final edited form as:

J Cogn Neurosci. 2018 March ; 30(3): 267–280. doi:10.1162/jocn_a_01208.

LIFG sensitivity to phonetic competition in receptive language processing: a comparison of clear and conversational speech

Xin Xie¹, Emily Myers^{2,*}

¹University of Rochester, Department of Brain and Cognitive Sciences, 323 Meliora Hall, Rochester, NY, 14627

²University of Connecticut, Department of Speech, Language, and Hearing Sciences, 850 Bolton Road, Unit 1086, Storrs, CT, 06269

Abstract

The speech signal is rife with variations in phonetic ambiguity. For instance, when talkers speak in a conversational register, they demonstrate less articulatory precision, leading to greater potential for confusability at the phonetic level compared to a clear speech register. Current psycholinguistic models assume that ambiguous speech sounds activate more than one phonological category, and that competition at prelexical levels cascades to lexical levels of processing. Imaging studies have shown that the left inferior frontal gyrus (LIFG) is modulated by phonetic competition between simultaneously activated categories, with increases in activation for more ambiguous tokens. Yet these studies have often used artificially manipulated speech and/or metalinguistic tasks, which arguably may recruit neural regions that are not critical for natural speech recognition. Indeed, a prominent model of speech processing, the Dual Stream Model, posits that the LIFG is not involved in prelexical processing in receptive language processing. In the current study, we exploited natural variation in phonetic competition in the speech signal in order to investigate the neural systems sensitive to phonetic competition as listeners engaged in a receptive language task. Participants heard nonsense sentences spoken in either a clear or conversational register as neural activity was monitored using fMRI. Conversational sentences contained greater phonetic competition, as estimated by measures of vowel confusability, and these sentences also elicited greater activation in a region in the LIFG. Sentence-level phonetic competition metrics uniquely correlated with LIFG activity as well. This finding is consistent with the hypothesis that the LIFG responds to competition at multiple levels of language processing, and that recruitment of this region does not require an explicit phonological judgment.

Keywords

phonetic competition; LIFG; speech register; clear speech; conversational speech

Speech recognition involves continuous mapping of sounds onto linguistically meaningful categories that help to distinguish one word from another (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Psycholinguistic models of spoken word recognition share a

*Corresponding Author: emily.myers@uconn.edu.

common assumption that acoustic-phonetic details of speech incrementally activate multiple candidates (phonetic categories and words in a language), which compete for selection and recognition (e.g., Gaskell & Marslen Wilson, 1997; McClelland & Elman, 1986; Norris, 1994). Supporting this assumption, human listeners show sensitivity to acoustic-phonetic variation in spoken words to the extent that word recognition is not determined just by the goodness of fit between incoming speech and one particular lexical entry, but also the fit between speech and multiple phonetically-similar words as well (Andruski, Blumstein, & Burton, 1994; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008), which jointly casts a gradient effect on word recognition (e.g., Warren & Marslen-Wilson, 1987). Despite ample behavioral evidence, it is still poorly understood how the brain resolves phonetic competition (i.e., competition between similar sounds like ‘cat’ and ‘cap’) and arrives at the correct linguistic interpretation. In the present study, we address this question by probing the neural sensitivity of multiple brain regions in response to phonetic competition in connected speech.

Recent research on the cortical organization of speech perception and comprehension has generated a few hypotheses about the neural structures that support the speech-to-meaning mapping. A prominent neuroanatomical model, the Dual Stream Model proposed by Hickok and Poeppel (Hickok & Poeppel, 2004, 2007; Hickok, 2012), argues for two functionally-distinct circuits that are critical for different aspects of speech processing. According to this model, cortical processing of speech signal starts at the temporal areas (dorsal STG and mid-post STS) where the auditory input is analyzed according to its spectro-temporal properties and undergoes further phonological processing. From there, information about incoming speech is projected to other parts of the temporal lobe as well as fronto-parietal regions via two separate streams, depending on the specific task demands. The dorsal stream, which consists of several left-lateralized frontal areas and the temporal-parietal junction region, is responsible for mapping speech sounds onto articulatory representations; the ventral stream, which includes bilateral middle and inferior temporal lobes, is critical for mapping the acoustic signal to meaning.

The involvement of the bilateral superior temporal gyri (STG) and Heschl’s gyri in speech perception is uncontroversial. Involvement of these areas is seen across a wide range of speech perception and comprehension tasks such as passive listening, segmentation, syllable discrimination/identification and sentence comprehension, etc. (Chang et al., 2010; Davis & Johnsrude, 2003; Myers, 2007; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007; see Leonard & Chang, 2014; Rauschecker & Scott, 2009 for reviews). A number of functional imaging studies have reported intelligibility-sensitive regions within the temporal lobe (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Obleser, Wise, Alex Dresner, & Scott, 2007; Okada et al., 2010; Scott, Rosen, Lang, & Wise, 2006; Wild, Davis, & Johnsrude, 2012). Further, the posterior STG in particular exhibit fine-grained sensitivity to phonetic category structure (Chang et al., 2010; Myers, 2007), showing graded activation that scales with the degree of fit of a token to native language phonetic categories.

In contrast, the exact role of the frontal areas, and in particular, the left inferior frontal gyrus (LIFG), in speech perception has been vigorously debated. LIFG is recruited under conditions of phonetic ambiguity, for instance, when a token falls between two possible

phonetic categories (e.g., midway between /da/ and /ta/; Binder, Liebenthal, Possing, Medler, & Ward, 2004; Myers, 2007; Rogers & Davis, 2017). LIFG responses are often more categorical (that is, less sensitive to within-category variation) than responses in superior temporal areas (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Lee, Turkeltaub, Granger, & Raizada, 2012; Myers, Blumstein, Walsh, & Eliassen, 2009), suggesting a role for these regions in accessing phonetic category identity. In general, studies have shown increased involvement of LIFG under conditions of perceptual difficulty, including increased recruitment when listeners are confronted with accented speech (Adank, Rueschemeyer, & Bekkering, 2013), and increased activity in noisy or degraded stimulus conditions (Binder et al., 2004; D'Ausilio, Craighero, & Fadiga, 2012; Davis & Johnsrude, 2003; Eisner et al., 2010). The general observation that LIFG is recruited under these “unusual” listening conditions has led to proposals that LIFG activity either (a) reflects executive or attentional control processes that are peripheral to the computation of phonetic identity and/or (b) only are necessary for speech perception under extreme circumstances that involve significant perceptual difficulty. Indeed, studies of people with aphasia (PWA) with inferior frontal damage have often struggled to find a speech-specific deficit in processing, as opposed to a higher-level deficit in lexical retrieval or selection (Rogalsky, Pitz, Hillis, & Hickok, 2008). In the Dual Stream Model, the LIFG, as part of the dorsal stream, does not have an essential role in speech recognition (Hickok & Poeppel, 2004, 2007; Hickok, 2012). A challenge to this view would be to discover that LIFG is recruited for the type of phonetic competition that exists naturally in the listening environment (i.e., in the case of hypo-articulated speech) even when intelligibility is high and using a task that emphasizes lexical access rather than metalinguistic identification, discrimination, or segmentation tasks.

In the present study, we investigated the neural organization of speech processing with respect to the perception of phonetic category competition, an integral component of spoken word recognition. Specifically, we are interested in the division of labor between temporal speech processing areas such as STG and frontal areas such as LIFG during the online processing of phonetic competition. It is of interest to note that when appearing in the context of real words, increased phonetic competition unavoidably leads to increased lexical competition among phonologically similar words, as assumed in current psycholinguistic models of spoken word recognition (e.g. Luce & Pisoni, 1998; McClelland & Elman, 1986) and evident in numerous behavioral studies (Allopenna, Magnuson, & Tanenhaus, 1998; McMurray et al., 2008). Given that our primary goal concerns whether LIFG is recruited for speech recognition at all, we do not make a distinction between *phonetic competition* and *lexical competition* at this point insofar as they are both essential (sub)components of word recognition processes. For now, with respect to our major hypothesis, we use the term ‘phonetic competition’ in reference to competition that exists between similar sounds of a language (e.g., /i/ and /ɪ/), but also any lexical competition that may ensue as activation cascades from the phonetic level to the lexical level. We hypothesized that LIFG is functionally recruited to resolve phonetic competition as part of natural speech recognition. In addition to recruitment of the LIFG for challenging listening conditions, Hickok and Poeppel (2007) also noted that studies that show prefrontal engagement in speech perception have used sublexical tasks that do not require contact with lexical representations and hence do not inform the neural realization of speech recognition, for which the ultimate target is

word meaning (although see Dial & Martin, 2017 for evidence that sublexical tasks tap a level that is a precursor to lexical processing in aphasia). In light of these discussions, our foremost goal was to create a testing situation that reflects challenges faced by listeners in the real world and yet allows comparison of brain activation patterns across speech utterances varying in the degree of phonetic competition. To this end, we exploited a sentence listening task, in which participants were presented with a set of semantically anomalous sentences, produced in two styles of natural speech: *clear speech* (hyper-articulated, careful speech) vs. *conversational speech* (hypo-articulated, casual speech).

A major part of real-world speech communication occurs among friends, family and co-workers where speech is spontaneously and casually articulated, whereas a clear speech register is often adopted in noisy acoustic environments or when the addressed listeners have perceptual difficulty (e.g., non-native or hearing-impaired listeners). It is well documented that clear speech is perceptually more intelligible relative to conversational speech (see Smiljanic & Bradlow, 2010 for review). A variety of acoustic factors have been associated with enhanced intelligibility in clear speech, including slower speaking rate, higher pitch level and greater pitch variation, as well as spectro-temporal changes in the production of consonants and vowels. In terms of phonetic competition, phonemes vary in the degree to which they are confusable with other tokens (Miller & Nicely, 1955). Vowels may be especially vulnerable to confusion in English, given that English has a dense vowel space, with vowel categories that overlap acoustically (Hillenbrand, Getty, Clark, Wheeler, & others, 1995; Peterson & Barney, 1952). For instance, the “point” vowels (e.g., /i/ & /u/) are less likely to have near vowel neighbors in F1 and F2 space, whereas mid and central vowels (e.g., /ɪ/, /ə/, /ɛ/) are likely to fall in a dense vowel neighborhood, and thus be subject to increased competition from other vowels. Indeed, vowel space expansion is reported to lead to significant improvements in intelligibility and is a key characteristic of clear speech cross-linguistically (Ferguson & Kewley-Port, 2007; Liu, Tsao, & Kuhl, 2005; Picheny, Durlach, & Braida, 1985; Smiljani & Bradlow, 2005). In theory, vowel tokens that are more dispersed in the acoustic-phonetic space will be more distant from acoustic territory occupied by competing vowels, and should elicit reduced phonetic competition (McMurray et al., 2008). We thus expect clear speech to result in a lesser amount of phonetic competition than conversational speech.

Hence, the stimulus set offers an opportunity to examine brain changes that are associated with naturally-occurring differences in phonetic confusability that exist even in unambiguous speech. In addition, the current experiment was designed to isolate the effect of phonetic competition on brain activation. First, we chose a probe verification task which does not require any metalinguistic decision about the speech stimuli, nor does it impose a working memory load any more than it is necessary for natural speech recognition, to avoid additional load posed by sublexical identification tasks. Second, sentences were semantically anomalous, which avoids evoking extensive top-down influences from semantic prediction (Davis, Ford, Kherif, & Johnsrude, 2011). This manipulation is in place to isolate effects of phonetic/lexical ambiguity resolution from the top-down effects of semantic context. Although left IFG is suggested to support both semantic and syntactic processing of speech sentences (see Friederici, 2012 for review), the two sets of speech stimuli are identical on these dimensions and differ only in their acoustic-phonetic patterns. Third, in light of

previous findings of the intelligibility-related activation in IFG, especially for degraded speech or noise-embedded speech, we equated the auditory intelligibility between the two sets of speech stimuli: clear vs. conversational speech (see details under Methods).

By comparing naturally-varying phonetic competition present in different speech registers, we can investigate phonetic competition in a situation that reflects the perceptual demands of the real-life environment. We predicted that increased phonetic competition would result in increased activation in the LIFG driven by additional demands on the selection between activated phonetic categories. We thus expect greater activation in the LIFG for conversational speech relative to clear speech. We predicted an opposite pattern in the temporal lobe given findings that superior temporal lobe encodes fine-grained acoustic detail. Because clear speech is expected to contain speech tokens that have better goodness-of-fit to stored phonological representations (Johnson, Flemming, & Wright, 1993), we expect the temporal areas to be more responsive to clear speech relative to conversational speech (Myers, 2007). Furthermore, by characterizing the degree of potential phonetic competition in each sentence, we can ask whether natural variability in phonetic competition is associated with modulation of activity in LIFG.

Methods

Participants

Sixteen adults (8 women) between the ages of 18 and 45 years old from the University of Connecticut community participated in the study. One female participant was excluded from following behavioral and fMRI analyses due to excessive head movement in multiple scanning sessions, leaving $n=15$ in all analyses. All participants were right-handed native speakers of American English, with no reported hearing or neurological deficits. Informed consent was obtained, and all participants were screened for ferromagnetic materials according to guidelines approved by the Institutional Review Board of University of Connecticut. Participants were paid for their time.

Stimuli

Ninety-six semantically anomalous sentences (consisting of real words) were adapted from Herman and Pisoni (2003) and were used in both the behavioral and fMRI testing sessions. All sentences were produced by the second author, a female native speaker of English. Three repetitions of each sentence were recorded in each speaking style: Clear speech and Conversational speech. Recordings were made in a sound-proof room using a microphone linked to a digital recorder, digitally sampled at 44.1 kHz and normalized for root mean square (RMS) amplitude to 70 dB SPL. The tokens were selected to minimize the duration differences between the two speaking styles. Detailed acoustic analyses were conducted in Praat (Boersma & Weenink, 2013) on the selected sentence recordings. Consistent with past research, preliminary analyses revealed many acoustic differences between Clear and Conversational speech, with differences manifested in speaking rate, pitch height and variation, among other characteristics. Clear and Conversational sentence sets were equated on three measures: duration, mean pitch, standard deviation of F0 variation within a sentence, using a resynthesis of all sentences and were implemented in the GSU Praat Tools

(Owren, 2008)¹. After equating the two sets of sentences on these measures, 84 sentences were selected as critical sentences and 12 sentences served as fillers (presented as target trials) for the in-scanner listening task. The critical sentences ranged between 1322ms and 2651ms, with no mean difference between the two speaking styles (Clear: 1986ms vs. Conversational: 1968ms, $t(83) = 1.24$, $p = .22$); filler sentences had a mean duration of 2000ms (SD = 194ms).

Stimulus Properties.—A number of acoustic and lexical properties of each sentence were measured for experimental control and for use the fMRI analysis. First, we analyzed all stressed vowels (Table 1). Critically, the mean F1 and F2 of all vowels, with the exception of /e/ and /ʌ/, differed significantly between the two speaking styles. In general, vowel space was more expanded in Clear speech relative to Conversational speech and there was considerably greater overlap between vowel categories in Conversational speech (see Fig. 1A).

In order to estimate the degree of phonetic competition inherent in each trial sentence, an additional analysis was performed on each stressed vowel token. Although Clear sentences differ from Conversational sentences on several phonetic dimensions (e.g., longer closure durations for voiceless stops, more release bursts for stops), we chose vowel density as a way to approximate the phonetic competition in each sentence, given that multiple vowel measurements could be made in every sentence. We adopted a measure (elsewhere termed “repulsive force”, see Wright, 2004 and McCloy et al., 2015 for details, here called “Phonetic Competition,” PC) that represents the mean of the inverse squared distances between this vowel token and all other vowel tokens that do not belong to the same vowel category. A token that is close to only vowels of the same identity (e.g., an /i/ vowel surrounded only by other /i/ tokens and far away from other vowel types) would have lower values on this measure and would be deemed to have low PC, whereas a token surrounded by many vowels of different identities (e.g., an /i/ with near-neighbors that are /e/ or /æ/) would score high on measures of PC (Figure 1C). Given the same target vowels across Clear and Conversational sentences, vowels from Clear sentences had significantly lower scores ($t(392) = 7.18$, $p < .0001$) on measures of PC (Figure 1B), although there was substantial overlap in these measures.

As noted in the Introduction, for any given word, changes in phonetic competition inevitably cascade to the lexical level and create competition among phonologically similar words. Although it is not our primary interest to distinguish between neural activation patterns responsive to phonetic competition versus that to lexical competition, it is possible to gain some insight into this question by linking BOLD signal to variation in lexical properties. To this end, we calculated lexical frequency (LF) and neighborhood density (ND) for each content word in the critical sentences. Sentence-level measures were then obtained by

¹We digitally adjusted the length of each sentence to equal the corresponding mean duration of the two original productions. In the case where excessive lengthening or shortening renders unnatural sounding of the sentences, both versions (clear vs. conversational) of the same sentence were re-adjusted and resynthesized such that the lengths were as close as possible without creating unnatural acoustic artifacts (as judged by the experimenters). All sentences were highly intelligible and deemed to be natural by an independent group of listeners, according to post-experiment survey questions in a pilot study (see Stimulus Norming section).

averaging across all content words within a sentence. Neither of these lexical measures correlated significantly with the PC values ($ps > .10$) at the sentence-level.

Stimulus Norming.—A pilot study was conducted to ensure that the manipulated sentences were highly intelligible and sounded natural. An independent group of ten native-English listeners transcribed all sentences, with each participant transcribing half of the Clear sentences and half of the Conversational sentences, such that no sentence was repeated within a participant. All participants reported the sentences to be natural and of high perceptual clarity in a post-experiment survey. The critical sentences were equated on their intelligibility, as assessed by listeners' transcription accuracy (Clear: 93.7% (SE = 0.8%) vs. Conversational 92.4% (SE = 0.8%), $t(83) = 1.45$, $p = .15$). None of the ten participants participated in the main experiment (fMRI and post-scanning behavioral tasks).

FMRI Design and Procedure

The fMRI experiment consisted of six separate runs presented in a fixed order across participants with trials within the runs presented in a fixed, pseudorandom order. The 84 Clear and Conversational sentences and 12 Target trials (filler sentences) were evenly distributed in a non-repetitive fashion across the first three runs and were repeated with a different set of order in the last three runs. Each run consisted of 14 Clear, 14 Conversational and 4 target trials. For each critical sentence, if the Clear version was presented in the first three runs, then the Conversational version appeared in one of the last three runs and vice versa. Stimuli were delivered over air-conduction headphones (Avotech Silent Scan SS-3300) that provide an estimated 28 dB of passive sound attenuation. Stimuli were assigned to SOAs of 6 and 12 s. Accuracy data were collected for the infrequent Target trials. Stimulus presentation and response collection were performed using PsychoPy v1.83.01.

Participants were told to pay attention to the screen and the auditory stimuli and to keep their heads as still as possible. In order to focus participants' attention on the content of the auditory stimuli, on target trials, a probe word appeared on the screen at the offset of the auditory sentence. Participants were asked to judge whether that word had appeared in the previous sentence, and indicated their response via an MRI-compatible button box (Current Designs, 932) held in the right hand. For half of the Target trials, the target word was contained in the previous sentence. Imaging data from Target trials was modeled in the subject-level analyses, but did not contribute to the group-level analysis.

FMRI Acquisition

Anatomical and functional MRI data was collected with a 3T Siemens Prisma scanner. High-resolution 3D T1-weighted anatomical images were acquired using a multi-echo MPRAGE sequence (Repetition Time [TR] = 2300 ms; Echo Time [TE] = 2.98 ms; Inversion Time [TI] = 900ms; 1-mm³ isotropic voxels; 248 × 256 matrix) and reconstructed into 176 slices. Functional EPI images were acquired in ascending, interleaved order (48 slices, 3 mm thick, 2 mm² axial in-plane resolution, 96×96 matrix, 192 mm³ field of view, flip angle = 90°), and followed a sparse sampling design: each functional volume was acquired with a 3000 msec acquisition time, followed by 3000 msec of silence during which

auditory stimuli were presented (effective TR = 6000ms). Stimuli were always presented during the silent gap (see Figure 2A).

FMRI Data Analysis

Images were analyzed using AFNI (Cox, 1996). Preprocessing of images included transformation from oblique to cardinal orientation, motion correction using a six-parameter rigid body transform aligned with each participant's anatomical dataset, normalization to Talairach space (Talairach & Tournoux, 1988), and spatial smoothing with a 4-mm Gaussian kernel. Masks were created using each participant's anatomical data to eliminate voxels located outside the brain. Individual masks were used to generate a group mask, which included only those voxels imaged in at least 14 of 15 participants' functional datasets. The first two TRs of each run were removed to allow for T1 equilibrium effects. Motion outliers and signal fluctuation outliers were removed following standard procedures.

In-scanner behavioral results indicated that all participants responded to all target trials and there were no inadvertent button presses in response to Clear or Conversational sentences. We generated time series vectors for each of the three trial conditions (Clear, Conversational and Target) for each participant in each run. These vectors contained the onset time of each stimulus and were convolved with a stereotypic gamma hemodynamic function. The three condition vectors along with six additional nuisance movement parameters were submitted to a regression analysis. This analysis generated by-voxel fit coefficients for each condition for each participant.

The above by-subject by-voxel fit coefficients were taken forward to group level t -test (@3dttest++, AFNI) analysis, comparing Clear speech to Conversational speech. We masked the t -test output with a small volume-corrected group mask that included anatomically defined regions that are typically involved in language processing: bilateral IFG, MFG, the insula, STG, HG, SFG, MTG, SMG, IPL, SPL and AG. Cluster-level correction for multiple comparisons was determined by running ten thousand iterations of Monte Carlo simulations (@3dClustSim, AFNI) on the small-volume corrected group mask. Specifically, we used -acf option in 3dFWHMx and 3dClustSim (AFNI) to estimate the spatial smoothness and generate voxelwise and clusterwise inference. These methods, consistent with recent standards for second-level correction (Eklund, Nichols, & Knutsson, 2016), estimated the spatial autocorrelation function of the noise using a mixed ACF model instead of the pure Gaussian-shaped model and have been reported to be effective in overcoming the issue of high false positive rates in cluster-based analysis. Data were corrected at a cluster-level correction of $p < 0.05$ (voxel-level threshold of $p < 0.005$, 59 contiguous voxels)^{2,3}.

²In an exploratory analysis, we also tested the possibility that by-subject variability in difficulty drove activation differences across subjects. To this end, we fitted a mixed-effects model (@3dLME, AFNI): fixed effects included Condition (Clear vs. Conversational) as a within-subject factor and by-subject, by-condition RT in the behavioral task as a continuous covariate; random effects included by-subject intercept and slope for RT. This model did not reveal a main effect of the covariate or an interaction between the covariate and Condition in any brain regions that survived the cluster-level correction for multiple comparisons.

³In order to test the replicability of these effects, a jackknifing procedure was used in which separate analyses were conducted, leaving one subject out in succession. All of the clusters reported here save one are robust to this test, emerging in all combinations of 14 subjects. The exception is the STG cluster reported for the Clear vs. Conversational contrast, which emerged in 5/15 simulations, an

A second analysis was conducted to search for relationships in the hemodynamic response and by-item measures of Phonetic Competition (PC), Reaction Time (RT), Neighborhood Density (ND) and Lexical Frequency (LF). PC, ND, and LF measures were calculated for each sentence using methods described above (Stimulus Properties). By-item mean RT was estimated for each sentence in the post-scanning behavioral test. For this analysis, Clear and Conversational tokens were collapsed, and relationships between hemodynamic response to each sentence and that sentence's by-item factors were analyzed in one analysis. Factors were mean-centered by run, and the stereotypic hemodynamic response was entered together with an amplitude-modulated version of this stereotypic timecourse. This analysis allows us to look for regions in which the by-item measures correlate with by-trial differences in BOLD above and beyond those accounted for by the base timecourse. By-subject beta coefficients were extracted, entered into a t-test vs. zero via 3dtttest++, and corrected for multiple comparisons using the same method as the standard group-level analysis.

Post-Scanning Behavioral Design and Procedure

After scanning, the same group of participants completed a 20–30 minute behavioral experiment to test by-participant sensitivity to the Clear vs. Conversational sentence distinction. During this test, participants completed a probe verification listening task concurrently with a visual search task (see Figure 2B for a schematic). In a behavioral pilot study where the probe verification listening task was used in isolation, standard behavioral measures (RT and accuracy) revealed no differences in responses to Clear versus Conversational speech. This result suggests that the variation in phonetic competition may be too subtle to transform into observable behavioral changes. One way of revealing subtle differences in processing load is to increase the cognitive load more generally. Previous findings have shown that a higher cognitive load degrades fine acoustic-phonetic processing of speech signal and causes poorer discrimination between similar speech tokens, especially for tokens near the category boundaries (e.g., Mattys & Wiget, 2011; Mattys et al., 2009). In particular, increased domain-general cognitive effort (i.e., the presence of a concurrent non-linguistic task) deteriorates the precision of acoustic-phonetic encoding, resulting in more mishearing of words (Mattys et al, 2014). In light of such findings, we reasoned that the inclusion of a concurrent cognitive task would negatively affect listeners' differentiation of subtle phonetic variation, especially where the amount of phonetic competition is high (Conversational). A second behavioral pilot study confirmed this hypothesis. We thus kept the visual search task as a secondary task in the post-scanning behavioral test.

Speech stimuli for the listening task were the 96 sentences used in the imaging session. The test was presented using Eprime 2.0.10. On each trial, an auditory sentence was delivered via headphones and a visual word was presented at the offset of the sentence. Participants were asked to listen carefully to the sentence and verify whether the visual word matched part of the auditory sentence with a 'yes' or 'no' button press. For half the trials, the visual word was part of the auditory sentence. Coincident with the onset of the auditory sentence, participants saw a visual array each consisting of a 6 column \times 6 row grid. In half the trials

indicant that this difference is weaker than the other findings. Notably, at a slightly reduced threshold ($p < 0.01$, 59 contiguous voxels), the STG cluster emerged in every simulation, which rules out the possibility that one outlier participant drives this result.

18 black squares and 18 red triangles were randomly arranged; in the other half of the grids, there was a red square with its position randomly assigned (See examples in Figure 2C). Following the sentence probe, participants were asked to press the ‘yes’ button if a red square was present and the ‘no’ button otherwise. After a practice phase with each task separately, participants were instructed to complete the two tasks simultaneously. For both tasks, they were instructed to respond with two labeled buttons ‘yes’ and ‘no’ as quickly as possible. Accuracy and reaction time (RT) data were collected for both tasks.

Post-Scanning Behavioral Data Analysis and Results

The visual search task was administered solely to impose a cognitive load on the participants and the results did not reveal any differences as a function of the sentence types. We thus omitted the results for this task. We analyzed the accuracy and RT data separately for the 84 critical sentences in the listening task. Participants showed no significant difference in accuracy between the Clear ($M = .90$, $SD = .06$) and Conversational sentences ($M = .90$, $SD = .06$; $F(1,14) = 0.085$, $p = .78$). RT results of correct trials revealed a main effect of condition ($F(1,14) = 4.435$, $p = .05$), with faster responses to Clear sentences ($M = 978$ ms, $SD = 180$ ms) than to Conversational sentences ($M = 994$ ms, $SD = 192$ ms). As expected, while both types of sentences were highly intelligible, the RT differences indicated greater perceptual difficulty for the Conversational sentences compared to the Clear sentences. Note that the participants already heard the whole set of sentences in the scanner before they were tested in this listening task. If anything, repetition of these sentences should attenuate any perceptual difference between Clear vs. Conversational speech. In order to factor out changes in activation due to differences in perceptual difficulty, we calculated the mean RT of each condition for each participant and included them as covariates in the group analysis of imaging data.

Imaging Results

Comparison of *Clear* trials to *Conversational* trials (Figure 3) showed differential activation in functional clusters within left IFG (pars triangularis, pars opercularis), left IPL extending into superior parietal lobule (SPL), left posterior STG and a small portion of Heschl’s gyrus (see Table 2). Specifically, greater activation was found for Clear speech than for Conversational speech in the left superior temporal gyrus, extending into Heschl’s gyrus, whereas the opposite patterns were observed in left IFG and IPL regions.

A secondary analysis was conducted to examine several variables that differ across sentences. In particular, we wished to examine the hypothesis that phonetic competition (which is hypothesized to be greater for Conversational than Clear sentences) drives activation in frontal regions. A wide variety of regions showed increases in activation as PC increased, including bilateral IFG (pars triangularis and pars opercularis) extending on the left into the middle frontal gyrus (Table 3, Figure 4). Notably, there was overlap between this activation map and that identified by the Conversational vs. Clear contrast in the left IFG, pars triangularis (43 voxel-overlap) and the left IPL (43 voxel-overlap). Of interest, there was no correlation between BOLD responses and PC within the left or right superior temporal lobes. A similar analysis was conducted using by-item RT estimates, but showed no significant correlation at the corrected threshold. Finally, to rule out the possibility that

areas that correlated with PC are explained by the overall ‘difficulty’ of stimuli, PC was entered into the same analysis with RT. This did not change the overall pattern of results, which is perhaps unsurprising given that by-item PC measures show no significant correlation with RT ($r = .08, p > .10$). Taken together, this suggests that PC measures account for variance that is not shared with RT.⁴

Discussion

Using a receptive listening task that requires no metalinguistic judgment, we have shown that LIFG is recruited for resolving phonetic competition in speech recognition. First, LIFG showed greater activation for conversational speech, which presents more reduced forms of articulation and consequently a greater level of phonetic competition than clear speech. Increased activity for increased phonetic competition was also found in the inferior parietal cortex. Importantly, the opposite pattern was observed within the superior temporal lobe, demonstrating a functional dissociation between the frontal-parietal regions and temporal language regions. Second, by associating trial-by-trial variability in the amount of phonetic competition as well as lexical properties of words within a sentence with BOLD signal changes, we found that variation in activation within bilateral inferior frontal areas was predicted by sentence-to-sentence changes in phonetic competition. A similar pattern was observed in the left inferior parietal area and bilateral middle frontal gyri (MFG). Temporal regions showed no such selective sensitivity to phonetic competition on a trial-by-trial basis. Crucially, the modulatory effect of phonetic competition on LIFG activity persisted after controlling for difficulty (measured by RT in post-scanning task) and other lexical factors (frequency and frequency-weighted neighborhood density). The results provide clear evidence that LIFG activity is driven by the confusability between speech sounds, suggesting a critical role in the resolution of phonetic identity in a naturalistic, receptive speech task. Below we discuss the separate functional roles of frontal and temporo-parietal regions in a highly distributed network that map sounds onto words.

A number of studies have identified a critical role of LIFG in the encoding of phonetic identity (e.g., Myers et al., 2009; Poldrack et al., 2001). Because many of these studies have employed sublexical tasks such as category identification, discrimination or phoneme monitoring, what remains debatable is whether the recruitment of LIFG is essential in natural speech recognition. The DSM model, for instance, has argued explicitly that these sublexical tasks engage functions that are dissociable from spoken word recognition; hence, they are not relevant for the discussion on the neural bases of speech recognition, for which explicit attention to sublexical units is not required (Hickok & Poeppel, 2007). In the present study, we overcome such task-dependent confounds by utilizing a sentence listening task in which listeners perceive natural continuous speech, and presumably, access the mental lexicon as they do in normal speech communication, a function that has been ascribed to the ventral pathway that does not include frontal regions in the DSM.

⁴We also asked whether the BOLD signal correlated trial-by-trial fluctuation in frequency weighted neighborhood density or lexical frequency. No clusters survived correction for multiple comparisons for frequency-weighted neighborhood density. By-trial measures of lexical frequency positively correlated with activation in the left IFG (pars triangularis, $x=-47, y=25, z=16$). Neither inclusion of lexical frequency nor frequency-weighted neighborhood density in the model affected the outcome of the phonetic competition analysis.

Another functional role associated with LIFG in the literature is that it facilitates effortful listening (Adank, Nuttall, Banks, & Kennedy-Higgins, 2015; Eisner et al., 2010; Obleser, Zimmermann, et al., 2007). Unlike past studies that have shown increased LIFG activity in the presence of degraded listening conditions or an ambiguous sound signal (e.g., accented speech), we exposed listeners to highly intelligible speech in two types of typically heard registers: clear and conversational. As shown by large corpus studies (Johnson, 2004), conversational speech is a frequently (arguably, the most frequently) heard speaking register in daily life, and exhibits massive reduction and hypoarticulation of pronunciations. Vowel reduction in conversational speech is a particularly widely acknowledged and well-studied phenomenon (e.g. Gahl, Yao, & Johnson, 2012; Johnson et al., 1993). We argue that the phonetic competition that listeners are exposed to in the current study closely resembles the phonetic ambiguity that listeners hear in daily life, with the caveat that the lack of semantic context in the current study prevents top-down resolution of ambiguity. In this sense, resolution of phonetic competition is viewed as an inherent part of speech perception, rather than an unusual or exceptional case.

It is of theoretical interest to ask whether the LIFG activation in the current study reflects a specific function in the processing of phonetic categories, or a more general role in resolving conflict between competing lexical or semantic alternatives. As noted in the Introduction, a direct consequence of competition at the phonological level is competition at the lexical level (Andruski et al., 1994; McMurray et al., 2008). Indeed, lexical factors (e.g., word frequency, and neighborhood density) that have direct consequences on the dynamics of lexical access (Luce & Pisoni, 1998) are reported to modulate activity in a number of brain regions, spanning across frontal-temporal-parietal pathways (Minicucci, Guediche, & Blumstein, 2013; Okada & Hickok, 2006; Prabhakaran, Blumstein, Myers, Hutchison, & Britton, 2006; Zhuang, Randall, Stamatakis, Marslen-Wilson, & Tyler, 2011; Zhuang, Tyler, Randall, Stamatakis, & Marslen-Wilson, 2014). In particular, LIFG shows elevated activity for words with larger phonological cohort density and is thus argued to be responsible for resolving increased phonological-lexical competition (Minicucci et al., 2013; Prabhakaran et al., 2006; Righi, Blumstein, Mertus, & Worden, 2010; Zhuang et al., 2011, 2014). Of particular interest, Minicucci et al. (2013) manipulated pronunciations of a word such that it sounded more similar to a phonological competitor. For instance, reducing the voice onset time of /t/ in the word ‘time’ makes it more similar to ‘dime’. They found greater responses in LIFG for modified productions that lead to greater activation for a phonological competitor than when the modification did not lead to greater lexical competition. Similarly, Rogers and Davis (2017) showed that LIFG was especially recruited when phonetic ambiguity led to lexical ambiguity, e.g., when listeners heard a synthesized blend of two real words (e.g., ‘blade’-’glade’) compared to a blend of two non-words (e.g., ‘blem’-’glem’). In sum, evidence is consistent with the interpretation that phonetic competition, especially as it cascades to lexical levels of processing, modulates frontal regions.

While we did not observe any modulatory effect of phonological neighborhood structure on the activity in LIFG or any other typically implicated areas, a theoretically interesting possibility is that LIFG (or its subdivisions) serves multiple functional roles that help to resolve competition across various levels of linguistic processing. In the present study, the posterior and dorsal regions of LIFG (pars opercularis and pars triangularis; ~ BA44/45)

were modulated by phonetic competition. These regions have been posited to serve a domain-general function in competition resolution (see Badre & Wagner, 2007; Thompson-Schill, Bedny, & Goldberg, 2005 for reviews), with evidence coming predominantly from studies that investigate competing scenarios in semantic-conceptual representations. In a few recent studies on lexical competition, pars triangularis (BA45) has consistently been shown to be sensitive to phonological cohort density (Righi et al., 2010; Zhuang et al., 2011, 2014). Our findings suggest that to the extent that LIFG is crucial for conflict resolution, this function is not limited to higher-level language processing. In light of past research on phonetic category encoding using other paradigms (e.g. Myers et al., 2009), we take the current results as strong evidence for a crucial role of posterior LIFG regions in the phonological processing of speech sounds. Notably, we did not find any modulatory effects of phonetic competition on other language regions (left-lateralized MTG and STG) that have been previously reported to be responsive to word frequency and/or neighbor density manipulations (Kocagoncu, Clarke, Devereux, & Tyler, 2017; Prabhakaran et al., 2006; Zhuang et al., 2011). Therefore, it is plausible that different neural networks are engaged for the resolution of phonetic versus lexical competition. We suggest that it is particularly important for future research to determine the extent to which the recruitment of LIFG in phonetic competition is dissociable from lexical and/or semantic selection, and from more general-purpose mechanisms for competition resolution.

In addition to LIFG, we found a relationship between phonetic competition and activation in left IPL. Not only did this region show a Conversational > Clear pattern, its activation was gradiently affected by the degree of phonetic competition, as shown by the amplitude-modulated analysis. Anatomically and functionally connected with Broca's area (see Friederici, 2012; Hagoort, 2014 for reviews), left IPL has been reliably implicated in phonological processing, showing a similar pattern to that of LIFG across a range of speech perception tasks (Joanisse et al., 2007; Turkeltaub & Branch Coslett, 2010). At the lexical level, this region has been hypothesized to be the storage site for word form representations (Gow, 2012) and has emerged in studies that examined lexical competition effects in spoken word recognition and production (Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011; Prabhakaran et al., 2006). The shared similarities between left-lateralized IFG and IPL in response to changes in phonetic competition across sentences are highly compatible with a frontal-parietal network that is often engaged in sound-to-word mapping processes.

It is worth noting that the use of semantically anomalous sentences could have increased working memory demands and consequentially engaged IFG and IPL to a greater extent, relative to the listening of semantically meaningful sentences (e.g., Buchsbaum & D'Esposito, 2008; Eriksson et al., 2017; Smith et al. 1998; Buchsbaum et al. 2011; Venezia et al., 2012; Rogalsky & Hickok, 2011). However, since the same set of sentences were used for clear and conversational speech, an overall elevated level of working memory demands associated with semantic anomaly cannot explain the recruitment of LIFG for clear vs. conversational sentences. Another concern is that working memory load may increase with the amount of phonetic competition on a trial-by-trial basis. Our data cannot rule out the possibility that the working memory systems do modulate as a function of variability in phonetic competition, for example, by maintaining the acoustic-phonetic information until the category membership is resolved. For now, whether or not this is true is inconsequential

to our interpretation that left frontal and parietal regions are involved in processing phonetic competition. Working memory may be one of the core cognitive processes on which the resolution of phonetic competition is dependent. A theoretically-relevant question for future studies is to what extent the engagement of the identified regions is functionally separable from their role in supporting domain-general working memory components (see Smith & Jonides, 1997, 1998) that are not required for the resolution of phonetic competition.

Similarly, it is possible that the absence of reliable semantic cues may push listeners to use bottom-up, phonetic pathways to a greater degree than in typical language comprehension, much in the same way that listeners in noisy conditions show greater use of top-down information, and listeners with compromised hearing benefit more from semantic context than typical-hearing individuals (e.g., Wagner et al., 2016, Lash et al., 2013). Although semantically anomalous sentences are rare in the listening environment, challenging listening conditions — that is, hearing fragments of sentences, speech occluded intermittently by noise—are not rare. All of these conditions weaken available semantic and contextual cues available to the listener. It is an empirical question whether these same effects would emerge in a more predictive and naturalistic context, a topic worthy of future study. However, to the extent that these results replicate findings from several different task paradigms (Rogers & Davis, 2017; Myers et al., 2007), there is no inherent reason to suspect that the patterns seen here are specific to anomalous sentences.

Interestingly, in comparison to the activation patterns in frontal-parietal regions, left STG and Heschl's gyrus exhibited greater response for clear speech than for conversational speech. With respect to the perception of specific speech sounds, studies have shown that graded activation in bilateral STG as a function of token typicality as members of a particular sound category (Myers, 2007; Myers et al., 2009). To the extent that overall, carefully articulated speech tokens are further away from category boundaries and better exemplars (see Figure 1) compared to casually articulated speech tokens, greater activity in response to clear speech was expected.

Another interesting finding is that beyond the typically implicated fronto-temporo-parietal network in the left hemisphere, we also observed modulatory effects of phonetic competition in the right inferior frontal gyrus (RIFG). This finding is consistent with previous reports on the effects of phonetic competition in phonetic categorization tasks. In Myers (2007), bilateral IFG areas show increased activation to exemplar pairs of speech sounds that straddle across a category boundary (greater competition) versus those are within a category (lesser competition). Beyond speech and language processing, bilateral IFG are implicated in tasks that broadly engage cognitive control resources (e.g., Aron et al., 2004, 2014; Badre & Wagner, 2007; Novick et al., 2005; Levy & Wagner, 2004, 2011; Robbins, 2007). It is possible that phonetic/lexical competition recruits domain-general cognitive control mechanisms that are more bilaterally organized. This does not mean that LIFG and RIFG are engaged for the same purpose. In particular, greater RIFG activity has been suggested to reflect increased response uncertainty (e.g., in a Go/No Go task; Levy & Wagner, 2011) or inhibitory control (e.g., Aron et al., 2014). While our study does not speak to the specific division of labor between the two hemispheres, it might be an interesting avenue for future research to compare differences and similarities between the response patterns of LIFG and

RIFG to phonetic competition across a variety of tasks. For instance, the RIFG might be differentially engaged in more passive tasks (e.g., eye-tracking) versus those that require motor responses (phonetic categorization), whereas the LIFG might be less sensitive to task demands that are external to the resolution of phonetic competition itself. We suggest such investigations might further elucidate the nature of LIFG's role in processing phonetic competition.

In sum, our results add important evidence to an understanding of the functional roles of LIFG and the inferior parietal cortex in sentence comprehension. The clear dissociation between the temporal regions and the frontal-parietal regions in processing conversational versus clear speech is consistent with their respective roles implicated in the literature of speech perception. We suggest that elevated responses for clear speech relative to conversational speech are compatible with the view that STG regions have graded access to detailed acoustic-phonetic representations (Scott, Blank, Rosen, & Wise, 2000), whereas the greater engagement of LIFG and LIPL are consistent with their roles in encoding abstract category information. In the context of sentence processing, the notion that LIFG and LIPL are responsible for resolving phonetic competition is also consistent with a view that these regions may deliver top-down feedback signal to temporal regions to facilitate acoustic-phonetic analyses of distorted sound signal (Davis & Johnsrude, 2003; Evans & Davis, 2015) or to guide perceptual adaptation (e.g., Sohoglu, Peelle, Carlyon, & Davis, 2012). Importantly, while fMRI is useful for identifying regions that are recruited for speech perception processes, a true confirmation of the proposed role of the LIFG in resolving phonetic ambiguity awaits confirmation by data from people with aphasia with left IFG lesions. Taken together, these findings support the notion that resolution of phonetic competition is inherent to receptive language processing and is not limited to unusual or exceptional cases.

Acknowledgments

This work was supported by NIH NIDCD grant R01 DC013064 to EBM. The views expressed here reflect those of the authors and not the NIH or the NIDCD. We would like to thank Sahil Luthra, David Saltzman, Pamela Fuhrmeister, and Kathrin Rothermich and three anonymous reviewers for helpful comments on an earlier version of this manuscript.

References

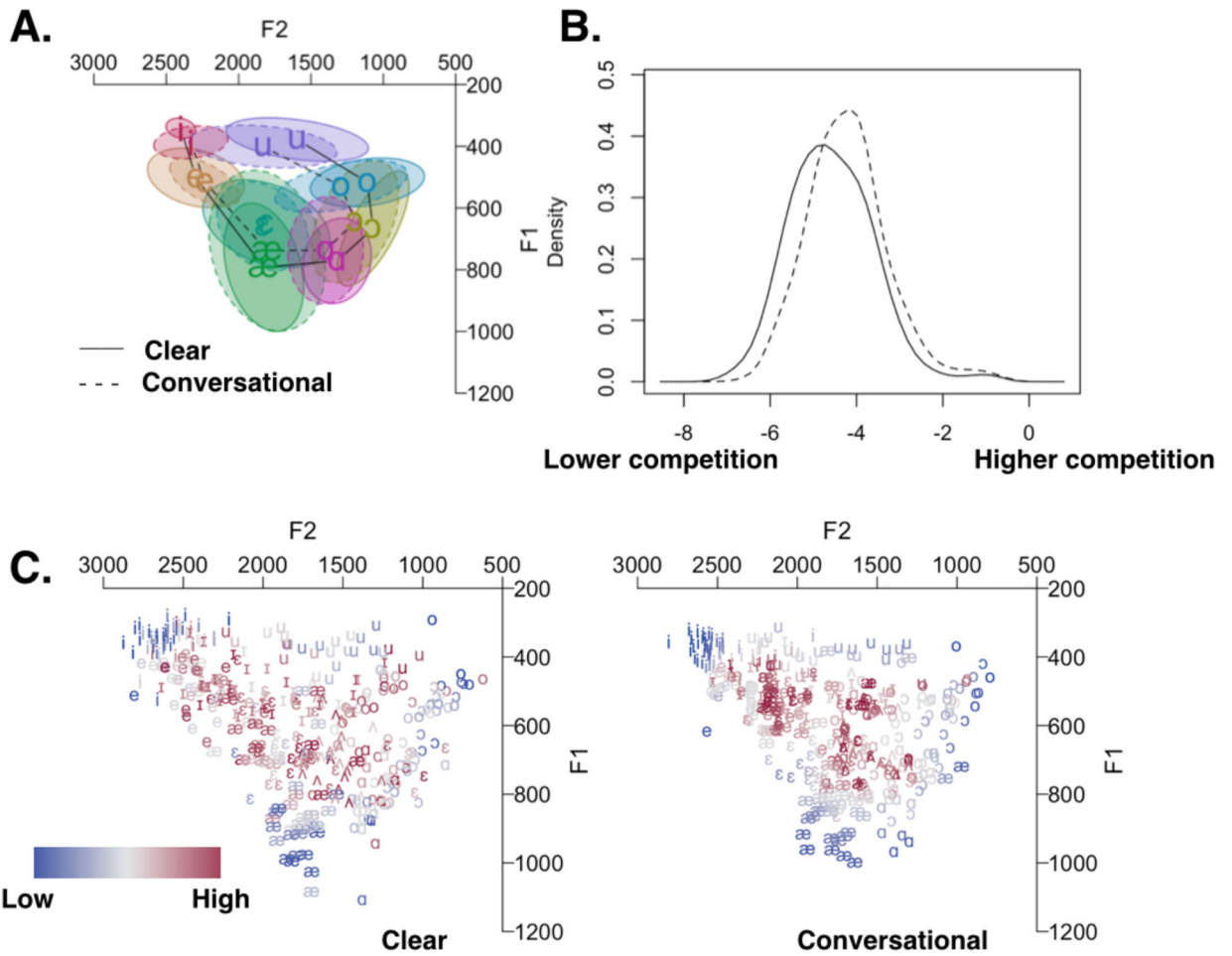
- Adank P, Nuttall HE, Banks B, & Kennedy-Higgins D (2015). Neural bases of accented speech perception. *Frontiers in Human Neuroscience*, 9, 558. 10.3389/fnhum.2015.00558 [PubMed: 26500526]
- Adank P, Rueschemeyer S-A, & Bekkering H (2013). The role of accent imitation in sensorimotor integration during processing of intelligible speech. *Frontiers in Human Neuroscience*, 7, 634. 10.3389/fnhum.2013.00634 [PubMed: 24109447]
- Allopenna PD, Magnuson JS, & Tanenhaus MK (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38(4), 419–439. 10.1006/jmla.1997.2558
- Andruski JE, Blumstein SE, & Burton MW (1994). The effects of subphonetic differences on lexical access. *Cognition*, 52(3), 163–187. [PubMed: 7956004]
- Badre D, & Wagner AD (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13), 2883–2901. 10.1016/j.neuropsychologia.2007.06.015 [PubMed: 17675110]

- Binder JR, Liebenthal E, Possing ET, Medler DA, & Ward BD (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295–301. [PubMed: 14966525]
- Blumstein SE, Myers EB, & Rissman J (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17(9), 1353–66. [PubMed: 16197689]
- Boersma P, & Weenink D (2013). Praat: doing phonetics by computer (Version 5.3.57). Retrieved from <http://www.praat.org/>
- Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, & Knight RT (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428–1432. 10.1038/nn.2641 [PubMed: 20890293]
- Chevillet MA, Jiang X, Rauschecker JP, & Riesenhuber M (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(12), 5208–5215. 10.1523/JNEUROSCI.1870-12.2013 [PubMed: 23516286]
- Connine CM, Blasko DG, & Wang J (1994). Vertical similarity in spoken word recognition: multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics*, 56(6), 624–636. [PubMed: 7816533]
- Cox RW (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(162–173). [PubMed: 8812068]
- D’Ausilio A, Craighero L, & Fadiga L (2012). The contribution of the frontal lobe to the perception of speech. *Journal of Neurolinguistics*, 25(5), 328–335. 10.1016/j.jneuroling.2010.02.003
- Davis MH, Ford MA, Kherif F, & Johnsrude IS (2011). Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, 23(12), 3914–3932. 10.1162/jocn_a_00084 [PubMed: 21745006]
- Davis MH, & Johnsrude IS (2003). Hierarchical processing in spoken language comprehension. *J Neurosci*, 23(8), 3423–31. [PubMed: 12716950]
- Dial H, & Martin R (2017). Evaluating the relationship between sublexical and lexical processing in speech perception: Evidence from aphasia. *Neuropsychologia*, 96, 192–212. 10.1016/j.neuropsychologia.2017.01.009 [PubMed: 28093277]
- Eisner F, McGettigan C, Faulkner A, Rosen S, & Scott SK (2010). Inferior Frontal Gyrus Activation Predicts Individual Differences in Perceptual Learning of Cochlear-Implant Simulations. *Journal of Neuroscience*, 30(21), 7179–7186. 10.1523/JNEUROSCI.4040-09.2010 [PubMed: 20505085]
- Eisner F, Melinger A, & Weber A (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, 4, 148. 10.3389/fpsyg.2013.00148 [PubMed: 23554598]
- Eklund A, Nichols TE, & Knutsson H (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, 113(28), 7900–7905. 10.1073/pnas.1602413113 [PubMed: 27357684]
- Evans S, & Davis MH (2015). Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cerebral Cortex*, 25(12), 4772–4788. [PubMed: 26157026]
- Ferguson SH, & Kewley-Port D (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research: JSLHR*, 50(5), 1241–1255. 10.1044/1092-4388(2007/087) [PubMed: 17905909]
- Friederici AD (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262–268. 10.1016/j.tics.2012.04.001 [PubMed: 22516238]
- Gahl S, Yao Y, & Johnson K (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789–806. 10.1016/j.jml.2011.11.006
- Gaskell Mg., & Marslen Wilson WD (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12(5–6), 613–656.

- Gow DW Jr. (2012). The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain and Language*, 121(3), 273–288. 10.1016/j.bandl.2012.03.005 [PubMed: 22498237]
- Hagoort P (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, 28, 136–141. 10.1016/j.conb.2014.07.013 [PubMed: 25062474]
- Herman R, & Pisoni DB (2003). Perception of “Elliptical Speech” Following Cochlear Implantation: Use of Broad Phonetic Categories in Speech Perception. *The Volta Review*, 102(4), 321–347. [PubMed: 21625300]
- Hickok G (2012). The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6), 393–402. 10.1016/j.jcomdis.2012.06.004 [PubMed: 22766458]
- Hickok G, & Poeppel D (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99. [PubMed: 15037127]
- Hickok G, & Poeppel D (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. 10.1038/nrn2113 [PubMed: 17431404]
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K, & others. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099–3111.
- Hutchison ER, Blumstein SE, & Myers EB (2008). An event-related fMRI investigation of voice-onset time discrimination. *NeuroImage*, 40(1), 342–352. [PubMed: 18248740]
- Joanisse MF, Zevin JD, & McCandliss BD (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(9), 2084–2093. 10.1093/cercor/bhl124
- Johnson K (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st Sessio of the 10th International Symposium*. Tokyo, Japan.
- Johnson K, Flemming E, & Wright R (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language*, 69(3), 505–528.
- Kocagoncu E, Clarke A, Devereux BJ, & Tyler LK (2017). Decoding the Cortical Dynamics of Sound-Meaning Mapping. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 37(5), 1312–1319. 10.1523/JNEUROSCI.2858-16.2016 [PubMed: 28028201]
- Lee Y-S, Turkeltaub P, Granger R, & Raizada RDS (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(11), 3942–3948. 10.1523/JNEUROSCI.3814-11.2012 [PubMed: 22423114]
- Leonard MK, & Chang EF (2014). Dynamic speech representations in the human temporal lobe. *Trends in Cognitive Sciences*, 18(9), 472–479. 10.1016/j.tics.2014.05.001 [PubMed: 24906217]
- Levy BJ, & Wagner AD (2011). Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences*, 1224(1), 40–62. 10.1111/j.1749-6632.2011.05958.x [PubMed: 21486295]
- Lieberman AM, Cooper FS, Shankweiler DP, & Studdert-Kennedy M (1967). Perception of the speech code. *Psychological Review*, 74(6), 431. [PubMed: 4170865]
- Liu H-M, Tsao F-M, & Kuhl PK (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, 117(6), 3879–3889. [PubMed: 16018490]
- Luce PA, & Pisoni DB (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear*, 19(1), 1–36. [PubMed: 9504270]
- Mattys SL, & Wiget L (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2), 145–160. 10.1016/j.jml.2011.04.004
- McClelland JL, & Elman JL (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [PubMed: 3753912]
- McMurray B, Aslin RN, Tanenhaus MK, Spivey MJ, & Subik D (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology. Human Perception and Performance*, 34(6), 1609–1631. 10.1037/a0011747 [PubMed: 19045996]

- Miller G, & Nicely P (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America*, 27(2), 338–352. 10.1121/1.1907526
- Minicucci D, Guediche S, & Blumstein SE (2013). An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. *Neuropsychologia*, 51(10), 1980–1988. 10.1016/j.neuropsychologia.2013.06.016 [PubMed: 23816958]
- Myers EB (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, 45(7), 1463–73. [PubMed: 17178420]
- Myers EB, Blumstein SE, Walsh E, & Eliassen J (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol Sci*, 20(7), 895–903. [PubMed: 19515116]
- Norris D (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234.
- Obleser J, Wise RJS, Alex Dresner M, & Scott SK (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(9), 2283–2289. 10.1523/JNEUROSCI.4663-06.2007 [PubMed: 17329425]
- Obleser J, Zimmermann J, Van Meter J, & Rauschecker JP (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(10), 2251–2257. 10.1093/cercor/bhl133
- Okada K, & Hickok G (2006). Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *Neuroreport*, 17(12), 1293–1296. 10.1097/01.wnr.0000233091.82536.b2 [PubMed: 16951572]
- Okada K, Rong F, Venezia J, Matchin W, Hsieh I-H, Saberi K, ... Hickok G (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex (New York, N.Y.: 1991)*, 20(10), 2486–2495. 10.1093/cercor/bhp318
- Owren MJ (2008). GSU Praat Tools: scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, 40(3), 822–829. [PubMed: 18697678]
- Peramunage D, Blumstein SE, Myers EB, Goldrick M, & Baese-Berk M (2011). Phonological neighborhood effects in spoken word production: an fMRI study. *Journal of Cognitive Neuroscience*, 23(3), 593–603. 10.1162/jocn.2010.21489 [PubMed: 20350185]
- Peterson GE, & Barney HL (1952). Control methods used in a study of vowels. *J Acoust Soc Am*, 24, 175.
- Picheny MA, Durlach NI, & Braida LD (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28(1), 96–103. [PubMed: 3982003]
- Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, & Gabrieli JDE (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *J Cognitive Neurosci*, 13(5), 687–697.
- Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, & Britton B (2006). An event-related fMRI investigation of phonological-lexical competition. *Neuropsychologia*, 44(12), 2209–21. [PubMed: 16842827]
- Raizada RD, & Poldrack RA (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56(4), 726–40. [PubMed: 18031688]
- Rauschecker JP, & Scott SK (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. 10.1038/nn.2331 [PubMed: 19471271]
- Righi G, Blumstein SE, Mertus J, & Worden MS (2010). Neural systems underlying lexical competition: an eye tracking and fMRI study. *Journal of Cognitive Neuroscience*, 22(2), 213–224. 10.1162/jocn.2009.21200 [PubMed: 19301991]
- Rogalsky C, Pitz E, Hillis A, & Hickok G (2008). Auditory word comprehension impairment in acute stroke: Relative contribution of phonemic versus semantic factors. *Brain and Language*, 107(2), 167–169. 10.1016/j.bandl.2008.08.003 [PubMed: 18823655]

- Rogers JC, & Davis MH (2017). Inferior Frontal Cortex Contributions to the Recognition of Spoken Words and Their Constituent Speech Sounds. *Journal of Cognitive Neuroscience*, 29(5), 919–936. 10.1162/jocn_a_01096 [PubMed: 28129061]
- Scott SK, Blank CC, Rosen S, & Wise RJ (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123 Pt 12, 2400–6. [PubMed: 11099443]
- Scott SK, Rosen S, Lang H, & Wise RJS (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. *The Journal of the Acoustical Society of America*, 120(2), 1075–1083. [PubMed: 16938993]
- Smiljanic R, & Bradlow A (2010). Teaching and Learning Guide for: Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, 4(3), 182–186. 10.1111/j.1749-818X.2009.00184.x
- Smiljani R, & Bradlow AR (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, 118(3 Pt 1), 1677–1688. [PubMed: 16240826]
- Sohoglu E, Peelle JE, Carlyon RP, & Davis MH (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(25), 8443–8453. 10.1523/JNEUROSCI.5069-11.2012 [PubMed: 22723684]
- Talairach J, & Tournoux P (1988). Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.
- Thompson-Schill SL, Bedny M, & Goldberg RF (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, 15(2), 219–24. [PubMed: 15831406]
- Turkeltaub PE, & Branch Coslett H (2010). Localization of sublexical speech perception components. *Brain and Language*, 114(1), 1–15. 10.1016/j.bandl.2010.03.008 [PubMed: 20413149]
- Utman JA, Blumstein SE, & Sullivan K (2001). Mapping from sound to meaning: reduced lexical activation in Broca's aphasics. *Brain and Language*, 79(3), 444–472. 10.1006/brln.2001.2500 [PubMed: 11781053]
- Warren P, & Marslen-Wilson W (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41(3), 262–275. [PubMed: 3575084]
- Wild CJ, Davis MH, & Johnsrude IS (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, 60(2), 1490–1502. 10.1016/j.neuroimage.2012.01.035 [PubMed: 22248574]
- Zevin JD, Yang J, Skipper JI, & McCandliss BD (2010). Domain General Change Detection Accounts for “Dishabituation” Effects in Temporal-Parietal Regions in Functional Magnetic Resonance Imaging Studies of Speech Perception. *Journal of Neuroscience*, 30(3), 1110. [PubMed: 20089919]
- Zhuang J, & Devereux BJ (2017). Phonological and syntactic competition effects in spoken word recognition: evidence from corpus-based statistics. *Language, Cognition and Neuroscience*, 32(2), 221–235. 10.1080/23273798.2016.1241886
- Zhuang J, Randall B, Stamatakis EA, Marslen-Wilson WD, & Tyler LK (2011). The interaction of lexical semantics and cohort competition in spoken word recognition: an fMRI study. *Journal of Cognitive Neuroscience*, 23(12), 3778–3790. 10.1162/jocn_a_00046 [PubMed: 21563885]
- Zhuang J, Tyler LK, Randall B, Stamatakis EA, & Marslen-Wilson WD (2014). Optimally efficient neural systems for processing spoken language. *Cerebral Cortex (New York, N.Y.: 1991)*, 24(4), 908–918. 10.1093/cercor/bhs366

**Figure 1.**

Acoustic measures for content words taken from Clear and Conversational sentences. A. Geometric centers for vowels from Clear (connected by solid line) and Conversational (dotted line) sentences. B. Probability density function for Phonetic Competition measures on vowels drawn from Clear (solid line) and Conversational (dotted line) sentences. Units are expressed in terms of the log-transformed mean of the inverse squared distances to all tokens that are not of the same type, with lower values showing fewer different-neighbor tokens (lower competition), and positive values indicating more different-neighbor tokens (higher competition). C. Individual tokens from Clear (left) and Conversational (right) sentences, coded according to the degree of Phonetic Competition each token is subject to, from Low (blue) to High (red).

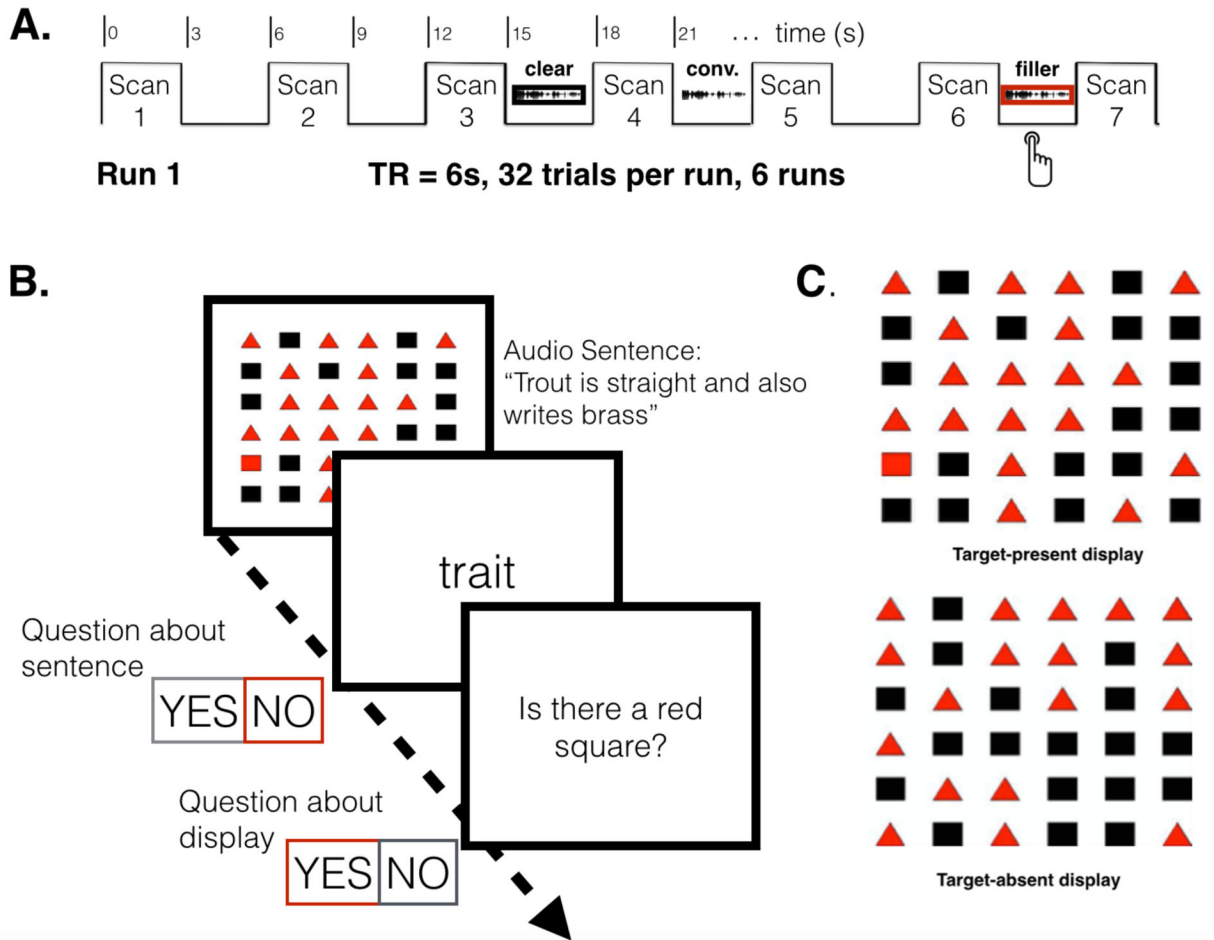


Figure 2.
 A. Schematic showing stimulus presentation timing with respect to EPI scans. B. Post-scanning behavioral study schematic. Listeners perform a visual target detection task, searching for a red square in the array. Simultaneously, they hear a sentence. Immediately after the sentence, participants see a visual probe on the screen and are asked to indicate whether that word was in the sentence. Then they are queried about the presence of the visual target. C. Example arrays for the visual target detection.

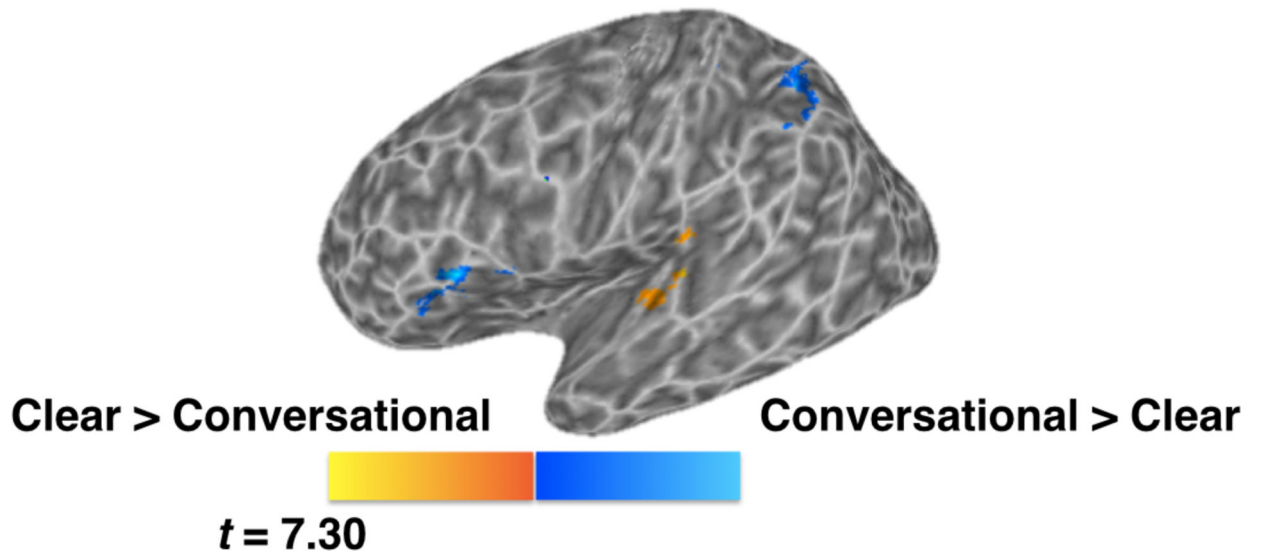


Figure 3.

Blue shows areas that show greater activation for Conversational than Clear, yellow shows areas that are greater for Clear than Conversational. Clusters at a corrected $p < 0.05$ (voxel-wise $p < 0.005$, minimum 59 voxels per cluster).

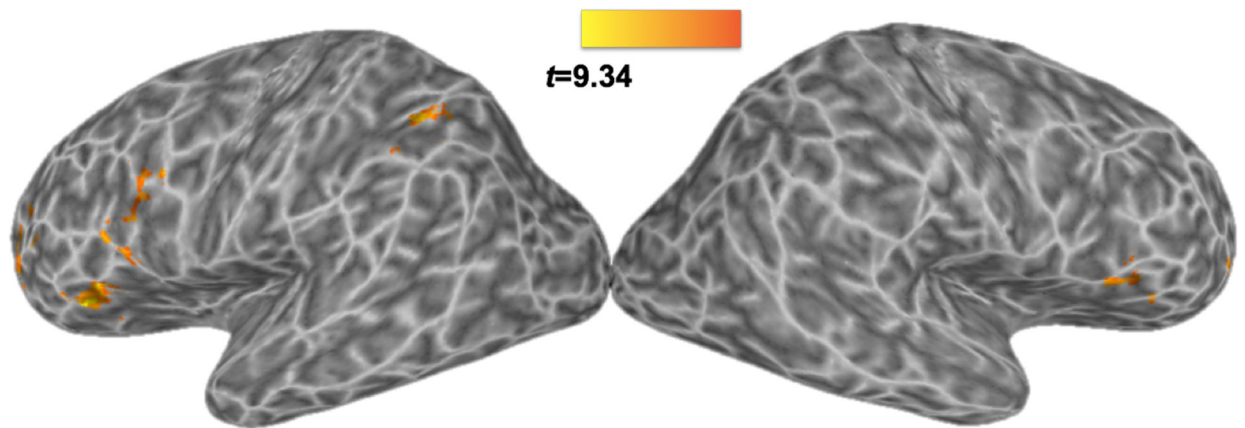


Figure 4. Results of the amplitude-modulated analysis, showing areas in which by-trial activation fluctuates with by-trial measures of phonetic competition. All regions show a positive correlation between phonetic competition and activation. Clusters at a corrected $p < 0.05$ (voxel-wise $p < 0.005$, minimum 59 voxels per cluster).

Table 1.

Acoustic analysis of the first and second formants of stressed vowels in Clear and Conversational speech sentences. Group means and standard deviations (in parentheses) are presented for F1 and F2 separately.

Vowel	No. of tokens	F1 Mean (SD) in Hz		F2 Mean (SD) in Hz		F1 Diff.	F2 Diff.	Paired <i>t</i> -test (2-tailed)	
		Conversational	Clear	Conversational	Clear			F1	F2
i	50	380 (35)	347 (42)	2480 (171)	2588 (141)	-34	108	$p < .00001$	$p < .00001$
ɪ	43	514 (52)	495 (61)	1962 (292)	2042 (350)	-19	80	$p < .01$	$p < .05$
e	46	517 (51)	485 (61)	2260 (171)	2390 (233)	-32	130	$p < .001$	$p < .00001$
ɛ	52	659 (93)	651 (92)	1835 (203)	1809 (283)	-8	-26	$p = .45$	$p = .39$
æ	49	738 (168)	803 (141)	1804 (259)	1821 (176)	65	18	$p < .00001$	$p = .54$
ʌ	27	665 (87)	683 (92)	1576 (145)	1565 (129)	18	-11	$p = .17$	$p = .60$
ɑ	35	737 (111)	781 (104)	1399 (167)	1320 (149)	44	-78	$p < .05$	$p < .01$
ɔ	32	644 (126)	666 (119)	1195 (192)	1071 (158)	23	-124	$p < .05$	$p < .00001$
o	31	530 (54)	510 (64)	1291 (291)	1105 (248)	-20	-186	$p < .05$	$p < .00001$
u	28	401 (44)	378 (45)	1833 (324)	1596 (312)	-23	-237	$p < .05$	$p < .00001$

Table 2.

Results of t-test comparing BOLD responses to Clear and Conversational sentences. Clusters corrected at voxel level $p < 0.005$, 59 contiguous voxels, corrected threshold of $p < 0.05$.

Area	Cluster size in voxels	Maximum intensity coordinates			Maximum t value
		x	y	z	
Conversational > Clear					
left IPL, left SPL	109	-37	-51	56	3.86
left IFG (p. Triangularis, p. Opercularis)	133	-39	21	4	3.44
Clear > Conversational					
left posterior STG, left Heschl's gyrus	78	-45	-23	10	3.97

Table 3.

Results of the amplitude-modulated analysis. In clusters reported below, by-item variability in Phonetic Competition correlated significantly with activation beyond that attributable to the event time course. No clusters correlated significantly with Reaction Time at this threshold. Clusters corrected at $p < 0.05$ (voxel-level $p < 0.005$, 59 contiguous voxels).

Area	Cluster size in voxels	Maximum intensity coordinates			Maximum t value
		x	y	z	
LIFG, pars opercularis, pars triangularis	160	-49	7	26	2.49
LMFG	139	-39	47	16	5.22
LIFG pars triangularis	133	-37	25	6	9.34
RIFG pars triangularis, pars opercularis	85	51	15	4	5.82
LIPL	80	-31	-51	40	7.11
RMFG	66	37	49	14	5.21