Left Inferior Frontal Gyrus Sensitivity to Phonetic Competition in Receptive Language Processing: A Comparison of Clear and Conversational Speech

Xin Xie[1] and Emily Myers[2]

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 with 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 to investigate the neural systems sensitive to phonetic competition as listeners engage 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. ■

INTRODUCTION

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 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; Norris, 1994; McClelland & Elman, 1986). Supporting this assumption, human listeners show sensitivity to acoustic–phonetic variation in spoken words to the extent that word recognition is determined by not only 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 (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; Andruski, Blumstein, & Burton, 1994), 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 (PC; i.e., competition

between similar sounds like "cat" and "cap") and arrives at the correct linguistic interpretation. In this study, we address this question by probing the neural sensitivity of multiple brain regions in response to PC 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 (DSM) proposed by Hickok and Poeppel (Hickok, 2012; Hickok & Poeppel, 2004, 2007), 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 superior temporal gyrus [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 TPJ, is responsible for mapping speech sounds onto articulatory representations; the ventral stream, which includes bilateral middle

---

[1]University of Rochester, [2]University of Connecticut

and inferior temporal lobes, is critical for mapping the acoustic signal to meaning.

The involvement of the bilateral STGs and Heschl's gyri (HGs) 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, sentence comprehension, and so forth (Chang et al., 2010; Myers, 2007; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007; Davis & Johnsrude, 2003; see Leonard & Chang, 2014; Rauschecker & Scott, 2009, for reviews). A number of functional imaging studies have reported intelligibility-sensitive regions within the temporal lobe (Wild, Davis, & Johnsrude, 2012; Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Okada et al., 2010; Obleser, Wise, Alex Dresner, & Scott, 2007; Scott, Rosen, Lang, & Wise, 2006). Furthermore, 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/; Rogers & Davis, 2017; Myers, 2007; Binder, Liebenthal, Possing, Medler, & Ward, 2004). LIFG responses are often more categorical (i.e., 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 (D'Ausilio, Craighero, & Fadiga, 2012; Eisner et al., 2010; Binder et al., 2004; Davis & Johnsrude, 2003). 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 is necessary for speech perception under extreme circumstances that involve significant perceptual difficulty. Indeed, studies of people with aphasia 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 DSM, the LIFG, as part of the dorsal stream, does not have an essential role in speech recognition (Hickok, 2012; Hickok & Poeppel, 2004, 2007). A challenge to this view would be to discover that LIFG is recruited for the type of PC that exists naturally in the lis-

tening environment (i.e., in the case of hypoarticulated speech) even when intelligibility is high and using a task that emphasizes lexical access rather than metalinguistic identification, discrimination, or segmentation.

In this 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 PC. It is of interest to note that, when appearing in the context of real words, increased PC 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 (McMurray et al., 2008; Allopenna, Magnuson, & Tanenhaus, 1998). Given that our primary goal concerns whether LIFG is recruited for speech recognition at all, we do not make a distinction between PC 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 "PC" 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 PC 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 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 yet allows comparison of brain activation patterns across speech utterances varying in the degree of PC. To this end, we used 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 (hyperarticulated, careful speech) versus conversational speech (hypoarticulated, casual speech).

A major part of real-world speech communication occurs among friends, family, and coworkers 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., nonnative listeners or listeners with hearing impairment). It is well documented that clear speech is perceptually more intelligible relative to conversational speech (see Smiljanic & Bradlow, 2010, for a 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 PC, 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, 1995; Peterson & Barney, 1952). For instance, the "point" vowels (e.g., /i/ and /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; Smiljanić & Bradlow, 2005; Picheny, Durlach, & Braida, 1985). 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 PC (McMurray et al., 2008). We thus expect clear speech to result in a lesser amount of PC 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 PC on brain activation. First, we chose a probe verification task that does not require any metalinguistic decision about the speech stimuli, nor does it impose a working memory load any more than 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 LIFG is suggested to support both semantic and syntactic processing of speech sentences (see Friederici, 2012, for a 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 versus conversational speech (see details under Methods).

By comparing naturally varying PC present in different speech registers, we can investigate PC in a situation that reflects the perceptual demands of the real-life environment. We predicted that increased PC 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 PC in each sentence, we can ask whether natural variability in PC is associated with modulation of activity in LIFG.

METHODS

Participants

Sixteen adults (eight women) between the ages of 18 and 45 years from the University of Connecticut community participated in the study. One female participant was excluded from the behavioral and fMRI analyses because of 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 soundproof room using a microphone linked to a digital recorder, digitally sampled at 44.1 kHz and normalized for a root mean square amplitude of 70 dB of sound pressure level. 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, and 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 in length between 1322 and 2651 msec, with no mean difference between the two speaking styles (clear vs. conversational: 1986 vs. 1968 msec, respectively; $t(83) = 1.24$, $p = .22$); filler sentences had a mean duration of 2000 msec ($SD = 194$ msec).

Stimulus Properties

A number of acoustic and lexical properties of each sentence were measured for experimental control and for use of the fMRI analysis. First, we analyzed all stressed vowels (Table 1). Critically, the mean F1 and F2 of all vowels, with the exception of /ɛ/ 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 Figure 1A).

To estimate the degree of PC inherent in each trial sentence, an additional analysis was performed on each stressed vowel token. Although clear sentences generally 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 PC in each sentence, given that multiple vowel measurements could be made in every sentence. We adopted a measure (elsewhere termed "repulsive force"; see McCloy, Wright, & Souza, 2015, and Wright, 2004, for details; here called "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 /ɪ/ 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 PC 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 PC versus that to lexical competition, it is possible to gain some insight into this question by linking BOLD signal to variation in the 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 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 10 native English listeners transcribed all sentences, with each participant transcribing half of the clear sentences and half of the

Table 1. Acoustic Analysis of the First and Second Formants of Stressed Vowels in Clear and Conversational Speech Sentences

| Vowel | No. of Tokens | F1 Mean (SD) in Hz | | F2 Mean (SD) in Hz | | F1 Diff. | F2 Diff. | Paired t Test (Two-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	$p < .05$	$p < .00001$

Group means and standard deviations (in parentheses) are presented for F1 and F2 separately.
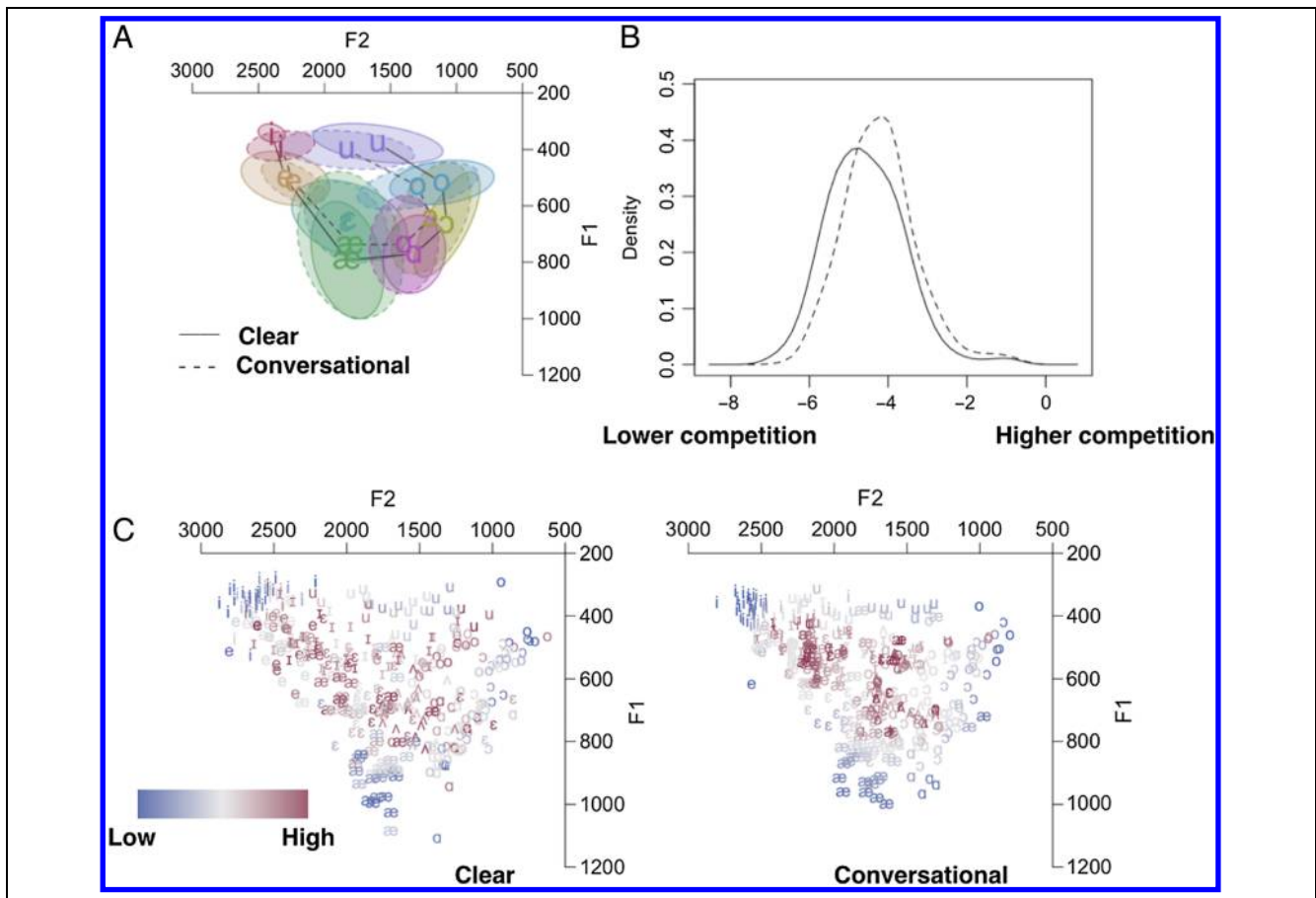
Figure 1. Acoustic measures for content words taken from clear and conversational sentences. (A) Geometric centers for vowels from clear (connected by solid lines) and conversational (dotted lines) sentences. (B) Probability density function for PC measures on vowels drawn from clear (solid line) and 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 PC each token is subject to, from low (blue) to high (red).

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 postexperiment survey. The critical sentences were equated on their intelligibility, as assessed by listeners' transcription accuracy (clear vs. conversational: 93.7% (SE = 0.8%) vs. 92.4% (SE = 0.8%), respectively; $t(83)$ = 1.45, p = .15). None of the 10 participants participated in the main experiment (fMRI and postscanning 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 nonrepetitive 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 tar-

get 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, Stuart, FL) that provide an estimated 28 dB of passive sound attenuation. Stimuli were assigned to SOAs of 6 and 12 sec. 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. 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, Philadephia, PA) 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 were modeled in the participant level analyses but did not contribute to the group level analysis.

fMRI Acquisition

Anatomical and functional MRI data were collected with a 3-T Siemens Prisma scanner (Erlangen, Germany). High-resolution 3-D T1-weighted anatomical images were acquired using a multiecho magnetization prepared rapid gradient echo sequence (repetition time [TR] = 2300 msec, echo time = 2.98 msec, inversion time = 900 msec, 1-mm^3 isotropic voxels, 248 × 256 matrix) and reconstructed into 176 slices. Functional EPIs were acquired in ascending, interleaved order (48 slices, 3-mm thick, 2-mm^2 axial in-plane resolution, 96 × 96 matrix, 192-mm^3 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 = 6000 msec). 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 data set, 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 data sets. The first two TRs of each run were removed to allow for T1 equilibrium effects. Motion 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-participant by-voxel fit coefficients were taken forward to group-level t test (@3dttest++, AFNI) analysis, comparing clear speech with 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, middle frontal gyrus (MFG), insula, STG, HG, superior frontal gyrus, middle temporal gyrus, supramarginal gyrus (SMG), inferior parietal lobule (IPL), superior parietal lobe (SPL), and angular gyrus (AG). Cluster level correction for multiple comparisons was determined by running 10,000 iterations of Monte Carlo simulations (@3dClustSim, AFNI) on the small-volume-corrected group

Figure 2. (A) Schematic showing stimulus presentation timing with respect to EPI scans. (B) Postscanning 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.
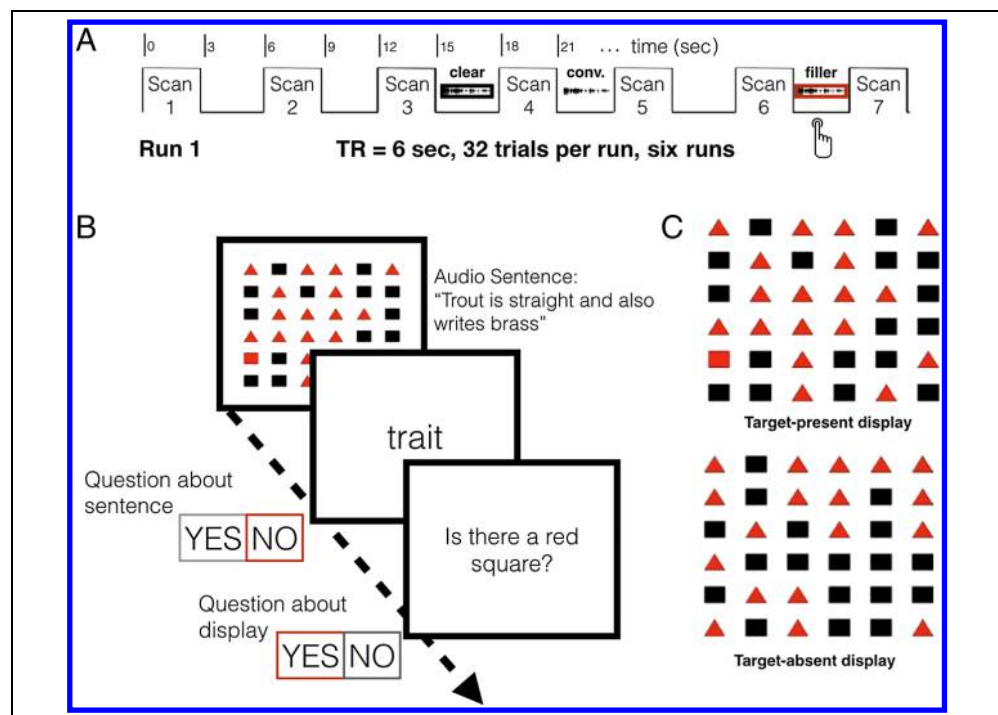
mask. Specifically, we used -acf option in *3dFWHMx* and *3dClustSim* (AFNI) to estimate the spatial smoothness and generate voxelwise and clusterwise inference. These methods, consistent with recent standards for second-level correction (Eklund, Nichols, & Knutsson, 2016), estimated the spatial autocorrelation function of the noise using a mixed autocerralation function 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 $p < .05$ (voxel level threshold of $p < .005$, 59 contiguous voxels).[2,3]

A second analysis was conducted to search for relationships in the hemodynamic response and by-item measures of PC, reaction time (RT), ND, and LF. PC, ND, and LF measures were calculated for each sentence using methods described above (Stimulus properties section). By-item mean RT was estimated for each sentence in the postscanning 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 time course. 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 time course. By-participant beta coefficients were extracted, entered into a *t* test versus zero via *3dttest++*, and corrected for multiple comparisons using the same method as the standard group-level analysis.

Postscanning Behavioral Design and Procedure

After scanning, the same group of participants completed a 20- to 30-minute behavioral experiment to test by-participant sensitivity to the clear versus 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 the responses to clear versus conversational speech. This result suggests that the variation in PC may be too subtle to transform into observable behavioral changes. One way of revealing subtle differences in processing load is to increase the overall cognitive load. 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, Brooks, & Cooke, 2009). In particular, increased domain-general cognitive effort (i.e., the presence of a concurrent nonlinguistic task) deteriorates the precision of acoustic–phonetic encoding, re-

sulting in more mishearing of words (Mattys, Barden, & Samuel, 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 PC is high (conversational). A second behavioral pilot study confirmed this hypothesis. We thus kept the visual search task as a secondary task in the postscanning behavioral test.

Speech stimuli for the listening task were the 96 sentences used in the imaging session. The test was presented using E-Prime 2.0.10 (Psychology Tools, Inc., Pittsburgh, PA). 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 of 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 × 6-row grid. In half of the trials, 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). After the sentence probe, participants were asked to press the "yes" button if a red square was present and the "no" button if 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 RT data were collected for both tasks.

RESULTS

Postscanning 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 ($M = 0.90$, $SD = 0.06$) and conversational sentences ($M = 0.90$, $SD = 0.06$; $F(1, 14) = 0.085$, $p = .78$). RT results of correct trials revealed a main effect of condition ($F(1, 14) = 4.435$, $p = .05$), with faster responses to clear sentences ($M = 978$ msec, $SD = 180$ msec) than to conversational sentences ($M = 994$ msec, $SD = 192$ msec). As expected, although both types of sentences were highly intelligible, the RT differences indicated greater perceptual difficulty for the conversational sentences compared with 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
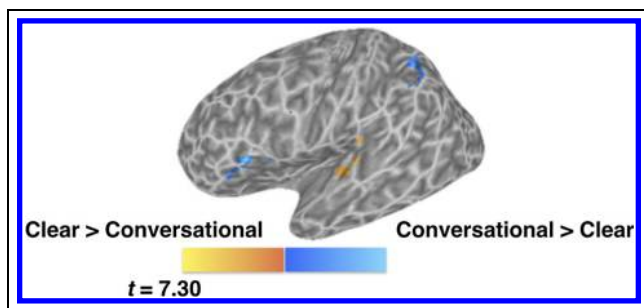
Figure 3. Blue shows areas that show greater activation for conversational speech than clear speech; yellow shows areas that are greater for clear speech than conversational speech. Clusters at a corrected $p < .05$ (voxelwise $p < .005$, minimum = 59 voxels per cluster).

between clear versus conversational speech. 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 with conversational trials (Figure 3) showed differential activation in functional clusters within LIFG (pars triangularis and pars opercularis), left IPL (LIPL) extending into SPL, left posterior STG, and a small portion of HG (see Table 2). Specifically, greater activation was found for clear speech than for conversational speech in the left STG, extending into HG, whereas the opposite patterns were observed in LIFG 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 PC (which is hypothesized to be greater for conversational than for clear sentences) drives activation in the 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 MFG

(Table 3, Figure 4). Notably, there was overlap between this activation map and that identified by the conversational versus clear contrast in LIFG, pars triangularis (43-voxel overlap), and LIPL (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 the 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 PC in speech recognition. First, LIFG showed greater activation for conversational speech, which presents more reduced forms of articulation and, consequently, a greater level of PC than clear speech. Increased activity for increased PC 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 PC, 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 PC. A similar pattern was observed in the left inferior parietal area and bilateral MFGs. Temporal regions showed no such selective sensitivity to PC on a trial-by-trial basis. Crucially, the modulatory effect of PC on LIFG activity persisted after controlling for difficulty (measured by RT in the postscanning task) and other lexical factors

Table 2. Results of t Test Comparing BOLD Responses to Clear and Conversational Sentences

| Area | Cluster Size in Voxels | Maximum Intensity Coordinates | | | Maximum t Value |
		x	y	z	
Conversational > clear					
LIPL, left SPL	109	−37	−51	56	3.86
LIFG (pars triangularis, pars opercularis)	133	−39	21	4	3.44
Clear > conversational					
Left posterior STG, left HG	78	−45	−23	10	3.97

Clusters corrected at the voxel level of $p < .005$, with 59 contiguous voxels and a corrected threshold of $p < .05$.

Table 3. Results of the Amplitude-modulated Analysis

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

In the clusters reported above, by-item variability in PC correlated significantly with activation beyond that attributable to the event time course. No clusters correlated significantly with RT at this threshold. Clusters were corrected at $p < .05$ (voxel level $p < .005$, 59 contiguous voxels).

(frequency and frequency-weighted ND). 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, 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 this 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

Figure 4. Results of the amplitude-modulated analysis, showing areas in which by-trial activation fluctuates with by-trial measures of PC. All regions show a positive correlation between PC and activation. Clusters at a corrected $p < .05$ (voxelwise $p < .005$, minimum = 59 voxels per cluster).

function that has been ascribed to the ventral pathway that does not include frontal regions in the DSM.

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 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 PC 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 PC 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 (McMurray et al., 2008; Andruski et al., 1994). 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 (Zhuang, Tyler, Randall, Stamatakis, & Marslen-Wilson, 2014; Minicucci, Guediche, & Blumstein, 2013; Zhuang, Randall, Stamatakis, Marslen-Wilson, & Tyler, 2011; Okada
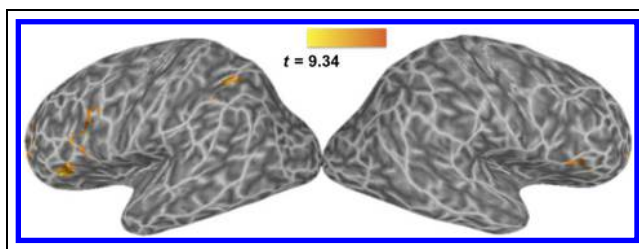
& Hickok, 2006; Prabhakaran, Blumstein, Myers, Hutchison, & Britton, 2006). In particular, LIFG shows elevated activity for words with a larger phonological cohort density and is thus argued to be responsible for resolving increased phonological–lexical competition (Zhuang et al., 2011, 2014; Minicucci et al., 2013; Righi, Blumstein, Mertus, & Worden, 2010; Prabhakaran et al., 2006). 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 VOT 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, for example, when listeners heard a synthesized blend of two real words (e.g., "blade"–"glade") compared with a blend of two nonwords (e.g., "blem"–"glem"). In summary, evidence is consistent with the interpretation that PC, especially as it cascades to lexical levels of processing, modulates frontal regions.

Although 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 this study, the posterior and dorsal regions of LIFG (pars opercularis and pars triangularis; ~BA 44/45) were modulated by PC. 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 (BA 45) has consistently been shown to be sensitive to phonological cohort density (Zhuang et al., 2011, 2014; Righi et al., 2010). 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 PC on other language regions (left-lateralized middle temporal gyrus and STG) that have been reported to be responsive to word frequency and/or neighbor density manipulations (Kocagoncu, Clarke, Devereux, & Tyler, 2017; Zhuang et al., 2011; Prabhakaran et al., 2006). 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 resolving PC 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 PC and activation in LIPL. Not only did this region show a conversational > clear pattern, its activation was gradiently affected by the degree of PC, as shown by the amplitude-modulated analysis. Anatomically and functionally connected with Broca's area (see Hagoort, 2014; Friederici, 2012, for reviews), LIPL has been reliably implicated in phonological processing, showing a similar pattern to that of LIFG across a range of speech perception tasks (Turkeltaub & Coslett, 2010; Joanisse, Zevin, & McCandliss, 2007). 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 PC 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., Eriksson, Vogel, Lansner, Bergström, & Nyberg, 2015; Venezia, Saberi, Chubb, & Hickok, 2012; Buchsbaum et al., 2011; Rogalsky & Hickok, 2011; Buchsbaum & D'Esposito, 2008; Smith, Jonides, Marshuetz, & Koeppe, 1998). However, because the same set of sentences was 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 versus conversational sentences. Another concern is that working memory load may increase with the amount of PC 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 PC, for example, by maintaining the acoustic–phonetic information until the category membership is resolved. For now, whether this is true is inconsequential to our interpretation that left frontal and parietal regions are involved in processing PC. Working memory may be one of the core cognitive processes on which the resolution of PC 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 PC.

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., Lash, Rogers, Zoller, & Wingfield, 2013). Although semantically anomalous sentences are rare in the listening environment, challenging listening conditions—that is, hearing fragments of sentences and 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, 2007), there is no inherent reason to suspect that the patterns seen here are specific to anomalous sentences.

Interestingly, in comparison with the activation patterns in frontal–parietal regions, the left STG and HG exhibited a greater response for clear speech than for conversational speech. With respect to the perception of specific speech sounds, studies have shown graded activation in bilateral STG as a function of token typicality as members of a particular sound category (Myers et al., 2009; Myers, 2007). To the extent that, overall, carefully articulated speech tokens are further away from category boundaries and better exemplars (see Figure 1) compared with 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 PC in the right IFG (RIFG). This finding is consistent with previous reports on the effects of PC 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 that are within a category (lesser competition). Beyond speech and language processing, bilateral IFGs are implicated in tasks that broadly engage cognitive control resources (e.g., Aron, Robbins, & Poldrack, 2004, 2014; Levy & Wagner, 2011; Badre & Wagner, 2007; Novick, Trueswell, & Thompson-Schill, 2005). 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).

Although 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 PC 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 PC itself. We suggest that such investigations might further elucidate the nature of LIFG's role in processing PC.

In summary, 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 is 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 PC 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 signals (Evans & Davis, 2015; Davis & Johnsrude, 2003) or to guide perceptual adaptation (e.g., Sohoglu, Peelle, Carlyon, & Davis, 2012). Importantly, although 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 LIFG lesions. Taken together, these findings support the notion that resolution of PC is inherent to receptive language processing and is not limited to unusual or exceptional cases.

Notes

1. We 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 readjusted 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 postexperiment survey questions in a pilot study (see Stimulus norming section).

2.   In an exploratory analysis, we also tested the possibility that by-participant variability in difficulty drove activation differences across participants. To this end, we fitted a mixed effects model (@3dLME, AFNI): Fixed effects included Condition (clear vs. conversational) as a within-participant factor and by-participant, by-condition RT in the behavioral task as a continuous covariate; random effects included by-participant 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.

3.   To test the replicability of these effects, a jackknifing procedure was used in which separate analyses were conducted, leaving one participant out in succession. All of the clusters reported here except one are robust to this test, emerging in all combinations of 14 participants. The exception is the STG cluster reported for the clear versus conversational contrast, which emerged in 5 of 15 simulations, an indicant that this difference is weaker than the other findings. Notably, at a slightly reduced threshold ($p < .01$, 59 contiguous voxels), the STG cluster emerged in every simulation, which rules out the possibility that one outlier participant drives this result.

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

REFERENCES

Adank, P., Nuttall, H. E., Banks, B., & Kennedy-Higgins, D. (2015). Neural bases of accented speech perception. *Frontiers in Human Neuroscience, 9,* 558.

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.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effects of subphonetic differences on lexical access. *Cognition, 52,* 163–187.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right interior frontal cortex. *Trends in Cognitive Sciences, 8,* 170–172.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right interior frontal cortex one decade on. *Trends in Cognitive Sciences, 18,* 177–185.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45,* 2883–2901.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience, 7,* 295–301.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.57). Retrieved from http://www.praat.org/.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory-an aggregate analysis of lesion and fMRI data. *Brain and Language, 119,* 119–128.

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience, 20,* 762–778.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13,* 1428–1432.

Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience, 33,* 5208–5215.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

D'Ausilio, A., Craighero, L., & Fadiga, L. (2012). The contribution of the frontal lobe to the perception of speech. *Journal of Neurolinguistics, 25,* 328–335.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience, 23,* 3914–3932.

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

Dial, H., & Martin, R. (2017). Evaluating the relationship between sublexical and lexical processing in speech perception: Evidence from aphasia. *Neuropsychologia, 96,* 192–212.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience, 30,* 7179–7186.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences, U.S.A., 113,* 7900–7905.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron, 88,* 33–46.

Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex, 25,* 4772–4788.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50,* 1241–1255.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16,* 262–268.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66,* 789–806.

Gaskell, M. G., & Marslen Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gow, D. W., Jr. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language, 121,* 273–288.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology, 28,* 136–141.

Herman, R., & Pisoni, D. B. (2003). Perception of "elliptical speech" following cochlear implantation: Use of broad phonetic categories in speech perception. *Volta Review, 102,* 321–347.

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,* 393–402.

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.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099–3111.

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. *Cerebral Cortex, 17,* 2084–2093.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st Session of the 10th International Symposium,* Tokyo, Japan.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69,* 505–528.

Kocagoncu, E., Clarke, A., Devereux, B. J., & Tyler, L. K. (2017). Decoding the cortical dynamics of sound-meaning mapping. *Journal of Neuroscience, 37,* 1312–1319.

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and entropy in spoken word recognition: Effects of age and hearing acuity. *Experimental Aging Research, 39,* 235–253.

Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience, 32,* 3942–3948.

Leonard, M. K., & Chang, E. F. (2014). Dynamic speech representations in the human temporal lobe. *Trends in Cognitive Sciences, 18,* 472–479.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences, 1224,* 40–62.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America, 117,* 3879–3889.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36.

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin and Review, 21,* 748–754.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59,* 203–243.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language, 65,* 145–160.

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

McCloy, D. R., Wright, R. A., & Souza, P. E. (2015). Talker versus dialect effects on speech intelligibility. A symmetrical study. *Language and Speech, 58,* 371–386.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1609–1631.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352.

Minicucci, D., Guediche, S., & Blumstein, S. E. (2013). An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical–semantic network. *Neuropsychologia, 51,* 1980–1988.

Myers, E. B. (2007). Dissociable effects of phonetic competition and category typically in a phonetic categorization task: An fMRI investigation. *Neuropsychologia, 45,* 1463–1473.

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science, 20,* 895–903.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 263–281.

Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience, 27,* 2283–2289.

Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebral Cortex, 17,* 2251–2257.

Okada, K., & Hickok, G. (2006). Identification of lexical–phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *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.

Owren, M. J. (2008). GSU Praat tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods, 40,* 822–829.

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2011). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience, 23,* 593–603.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America, 24,* 175.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing: I. Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96–103.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., et al. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience, 13,* 687–697.

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.

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.

Righi, G., Blumstein, S. E., Mertus, J., & Worden, M. S. (2010). Neural systems underlying lexical competition: An eye tracking and fMRI study. *Journal of Cognitive Neuroscience, 22,* 213–224.

Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience, 23,* 1664–1680.

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,* 167–169.

Rogers, J. C., & Davis, M. H. (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of Cognitive Neuroscience, 29,* 919–936.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400–2406.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *Journal of the Acoustical Society of America, 120,* 1075–1083.

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,* 182–186.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America, 118,* 1677–1688.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5–42.

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences, 95,* 12061–12068.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences, 95,* 876–882.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top–down integration of prior knowledge during speech perception. *Journal of Neuroscience, 32,* 8443–8453.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.* Stuttgart, Germany: Thieme.

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology, 15,* 219–224.

Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language, 114,* 1–15.

Venezia, J. H., Saberi, K., Chubb, C., & Hickock, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology, 3,* 157.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41,* 262–275.

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage, 60,* 1490–1502.

Wright, R. A. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation* (pp. 75–87). Cambridge University Press.

Zhuang, J., Randall, B., Stamatakis, E. A., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The interaction of lexical semantics and cohort competition in spoken word recognition: An fMRI study. *Journal of Cognitive Neuroscience, 23,* 3778–3790.

Zhuang, J., Tyler, L. K., Randall, B., Stamatakis, E. A., & Marslen-Wilson, W. D. (2014). Optimally efficient neural systems for processing spoken language. *Cerebral Cortex, 24,* 908–918.