Inferior Frontal Cortex Contributions to the Recognition of Spoken Words and Their Constituent Speech Sounds

Jack C. Rogers[1,2] and Matthew H. Davis[1]

Abstract

■ Speech perception and comprehension are often challenged by the need to recognize speech sounds that are degraded or ambiguous. Here, we explore the cognitive and neural mechanisms involved in resolving ambiguity in the identity of speech sounds using syllables that contain ambiguous phonetic segments (e.g., intermediate sounds between / b/ and /g/ as in "blade" and "glade"). We used an audio-morphing procedure to create a large set of natural sounding minimal pairs that contain phonetically ambiguous onset or offset consonants (differing in place, manner, or voicing). These ambiguous segments occurred in different lexical contexts (i.e., in words or pseudowords, such as blade–glade or blem–glem) and in different phonological environments (i.e., with neighboring syllables that differed in lexical status, such as blouse–glouse). These stimuli allowed us to explore the impact of phonetic ambiguity on the speed and accuracy of lexical decision responses (Experiment 1), semantic categorization responses (Experiment 2), and the magnitude of BOLD fMRI responses during attentive comprehension (Experiment 3). For both behavioral and neural measures, observed effects of phonetic ambiguity were influenced by lexical context leading to slower responses and increased activity in the left inferior fron-tal gyrus for high-ambiguity syllables that distinguish pairs of words, but not for equivalent pseudowords. These findings suggest lexical involvement in the resolution of phonetic ambigu-ity. Implications for speech perception and the role of inferior frontal regions are discussed. ■

INTRODUCTION

Speech perception and comprehension are often challenged by there being many different interpretations of a single stretch of speech. These multiple interpretations are often particularly apparent when we converse with people who speak with unfamiliar or foreign accents. For example, a Japanese speaker of English producing words like "right" and "light" may neutralize the third for-mant cues that ordinarily distinguish these word pairs (Ingvalson, McClelland, & Holt, 2011; Iverson et al., 2003). Similarly, a British English listener encountering American English speech for the first time may be initially surprised to hear them say something that sounds like /wɔ:də/ when requesting "water." In this case, however, the ambiguity in the identity of the second consonant (a flap rather than a stop) is more readily resolved be-cause this does not create an alternative word (unlike in "latter" and "ladder"). Nonetheless, it has been shown that hearing speech in a native or nonnative accent makes comprehension more challenging leading to slower and more error-prone responses in laboratory tasks (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009; Floccia, Goslin, Girard, & Konopczynski, 2006). Here, we explore the cog-nitive and neural processes that achieve accurate identification for speech tokens that are made deliberately ambiguous using an audio-morphing procedure. In particular, we consider the role that lexical knowledge plays in resolving phonetic ambiguity.

The impressive speed and accuracy of human speech comprehension when challenged by perceptual ambigu-ity belie the many processing stages involved. Even the recognition of single spoken words involves several hierarchically organized processing stages in which lower-level acoustic and phonetic features are identified and (potentially) categorized into larger units (phonemes or syllables) to recognize familiar words and then access syntactic and semantic properties (e.g., McClelland & Elman, 1986). This functional hierarchy has been proposed to map onto a neural hierarchy of temporal and frontal regions with multiple processing pathways that project from superior and lateral regions of the temporal lobe and map onto topographically organized regions of the inferior parietal and frontal cortex (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003).

Evidence for hierarchical neural organization of the processing stages involved in speech perception has come from functional neuroimaging data collected during the comprehension of ambiguous or degraded speech stimuli (see Peelle, Johnsrude, & Davis, 2010, for a sum-mary). For example, regions of the superior temporal gyrus (STG) close to the primary auditory cortex respond

[1]MRC Cognition & Brain Sciences Unit, Cambridge, UK,
[2]University of Birmingham

Journal of Cognitive Neuroscience 29:5, pp. 1–18
doi:10.1162/jocn_a_01096

to intelligible speech but are also sensitive to changes to the acoustic form of that speech (e.g., adding back-ground noise or interruptions). In an fMRI study, Davis and Johnsrude (2003) revealed a response profile within the STG such that BOLD responses differed for three types of degraded sentence that differed in their acoustic form but were matched for intelligibility. Similarly, in fMRI studies using multivoxel pattern analysis, STG response patterns differed for syllables spoken by different individ-uals (Evans & Davis, 2015; Formisano et al., 2008) or syllables presented with different forms of degradation (noise vocoded or sine-wave synthesized; Evans & Davis, 2015). These findings are consistent with the STG contrib-uting to processing stages that operate at relatively low levels of the functional hierarchy for speech perception (Evans & Davis, 2015; Evans et al., 2014; Okada et al., 2010; Davis & Johnsrude, 2003). In contrast, more distant regions of the lateral temporal lobe, adjacent inferior pa-rietal regions, and left inferior frontal gyrus (LIFG) and precentral gyrus (LPCG) also respond to intelligible speech but in a manner that is largely independent of the acoustic form of speech (Evans & Davis, 2015; Evans et al., 2014; Okada et al., 2010; Davis & Johnsrude, 2003; Scott, Blank, Rosen, & Wise, 2000). A response profile that is independent of the acoustic form of speech suggests a contribution to higher levels of the processing hierarchy (such as those processes involved in lexical and semantic access), although where and how these different brain regions contribute to word recognition and meaning ac-cess remain unclear (Lee, Turkeltaub, Granger, & Raizada, 2012; Binder, Desai, Graves, & Conant, 2009; Myers, Blumstein, Walsh, & Eliassen, 2009; Rauschecker & Scott, 2009; Lau, Phillips, & Poeppel, 2008; Hickok & Poeppel, 2007).

One question that continues to interest the field concerns the role of inferior frontal and motor regions (LIFG and LPCG) in speech perception and word recognition (see Lotto, Hickok, & Holt, 2009; Scott, McGettigan, & Eisner,2009, for discussion of motor contributions; Mirman, Yee, Blumstein, & Magnuson, 2011; Vaden, Piquado, & Hickok, 2011; Zhuang, Randall, Stamatakis, Marslen-Wilson, & Tyler, 2011, for inferior frontal regions). Some studies have shown that patients with inferior frontal (Broca's area) lesions were unimpaired on simple word-to-picture matching tests of speech comprehension—even if close phonological neighbors were used as foils (e.g., partic-ipants were required to distinguish the auditory stimulus "coat" from pictures of goat and boat; Rogalsky, Love, Driscoll, Anderson, & Hickok, 2011). These findings have been used to argue that inferior frontal regions play no role in speech comprehension at the level of single words (e.g., Hickok, Costanzo, Capasso, & Miceli, 2011). How-ever, other studies have shown perturbation of speech comprehension for Broca's aphasics presented with degraded speech (Moineau, Dronkers, & Bates, 2005). Aydelott Utman, Blumstein, and Sullivan (2001) have used cross-modal priming to show that Broca's aphasics are more severely affected by subphonetic variation. For example, a token of the prime word "king" that sounds more like "ging" reduced the magnitude of semantic priming for "queen" in Broca's aphasics more than was seen for healthy adults (see Andruski, Blumstein, & Burton, 1994, for data from healthy volunteers).

A number of functional imaging studies have demonstrated that inferior frontal and precentral gyrus regions are active during speech perception and comprehension (Pulvermüller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004), particularly when participants listen attentively to speech signals that are noisy or degraded (Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012; Wild et al., 2012; Osnes, Hugdahl, & Specht, 2011; Adank & Devlin, 2009). Furthermore, multivoxel patterns within LIFG and LPCG encode the perceptual identity of speech sounds (Arsenault & Buchsbaum, 2015; Correia, 2015; Lee et al., 2012), particularly for degraded speech sounds (Evans & Davis, 2015; Du, Buchsbaum, Grady, & Alain, 2014). These findings suggest a significant role for inferior frontal and precentral gyrus regions in speech perception, especially for stimuli that are difficult to perceive (see Guediche, Blumstein, Fiez, & Holt, 2014, for a review). This proposal is consistent with a number of TMS demonstrations showing that the stimulation of motor regions disrupts perceptual judgments on speech sounds (Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2015; Rogers, Möttönen, Boyles, & Watkins, 2014; D'Ausilio et al., 2009; Möttönen & Watkins, 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007).

What remains unclear, however, is whether these fron-tal and motor areas are contributing to perceptual identification and comprehension of speech or rather post-perceptual decision-making and executive functions. Evidence for LIFG contributions to perceptual decision-making has come from functional imaging findings of increased LIFG activity when participants are instructed to make segment-level rather than holistic judgments on the content of speech sounds (Burton, Small, & Blumstein, 2000) or during challenging listening situations when the observation that LIFG activity is correlated with RTs for the identification of syllables in noise (Binder, Liebenthal, Possing, Medler, & Ward, 2004).

One method for exploring this issue concerns neural responses to perceptual uncertainty in the identity of speech sounds. In several fMRI studies, Blumstein, Myers, and colleagues have repeatedly demonstrated additional LIFG activity during the perception of speech segments that are acoustically and phonetically ambiguous due to containing acoustic features that are at the boundary between two phonological categories. For instance, Blumstein, Myers, and Rissman (2005) showed additional activation for intermediate voice-onset time (VOT) values compared with more natural "end point" VOT values for a /da/-/ta/ continuum. Myers (2007) showed similar results for the comparison of boundary and extreme VOT values, and Myers and Blumstein (2008)

once more demonstrated additional LIFG activation for segments with ambiguous VOT values despite variation in the specific VOT values that are ambiguous due to lex-ical context. That is, different VOT values are perceptually ambiguous for a segment midway between /g/ and /k/ depending on whether this makes a familiar word "gift" or "kiss," due to the Ganong effect (Ganong, 1980). Yet, LIFG activity was greatest for the most ambiguous VOT value specific to that syllable.

Importantly, however, these studies explored neural activity observed while participants made overt judg-ments on the identity of the ambiguous segments. Thus, although these results serve to rule out the possibility of working memory or rehearsal-based processes (which are likely absent during simple listening tasks on single syllables), they do not speak to whether LIFG and LPCG contribute in listening situations focused on word recog-nition rather than phonological decisions. Similar chal-lenges have been raised for studies of TMS-induced disruption to speech perception; perceptual impairments after LPCG stimulation have been demonstrated in pho-nological judgment tasks (Möttönen & Watkins, 2009; Meister et al., 2007) that may be absent for semantic decisions (e.g., Krieger-Redwood, Gaskell, Lindsay, & Jefferies, 2013; see Schomers et al., 2015, for a counter-example). One fMRI study reported an interaction in LIFG such that subcategorical variation in VOT increased activity for words with a competing lexical prime (e.g., a modified token of "cap" that resembled "gap") relative to equivalent items that do not have a lexical neighbor (e.g., "coin"; Minicucci, Guediche, & Blumstein, 2013). Inter-estingly, this study used a semantic priming task in which listeners did not make overt judgments on the prime words but rather made lexical decisions on semantically related target words (e.g., "hat" or "penny").

In the present work, we return to the functional imag-ing contrast between more and less ambiguous speech sounds. Here, we use audio-morphed syllables created using STRAIGHT software (Kawahara & Morise, 2011; Kawahara, Masuda-Katsuse, & de Cheveigne, 1999). This is a form of perceptual challenge that has not previously been explored in the psycholinguistic and neuroscientific literatures. By using an audio-morphing procedure, we can create large sets of natural sounding syllables con-taining a mixture of different phonetically ambiguous segments (varying place, manner, and voicing features) rather than the more limited sets of syllables used previ-ously. This high degree of variation in our stimulus mate-rials allows us to explore the impact of ambiguity on perceptual performance and neural activity in more nat-ural listening situations and during tasks in which lis-teners are focused on recognizing words and accessing meaning rather than identifying single speech sounds. We anticipate that these more natural listening situations will minimize the need for making overt phonological judgments, along with the additional executive or meta-linguistic processes that these might entail.

In this work, we also explore the impact of lexical sta-tus (i.e., whether the ambiguous segments are heard in a real word or a pseudoword) on behavioral and neural re-sponses. We do this by comparing response latencies and neural activity for audio-morphed syllables that are syn-thesized from pairs of words (e.g., "blade"–"glade"), words and pseudowords (e.g., "bone"–"ghone", "bown"–"gown"), or pairs of pseudowords ("blem"– "glem"). As we will explain below, if lexical context can be shown to influence behavioral and neural responses to syllables containing ambiguous segments, then this suggests a model of speech perception in which lexical information is used to resolve segment level ambiguity in speech sounds rather than through purely pre-lexical processes.

Existing work has shown differential behavioral costs of another form of phonetic ambiguity (created by cross-splicing preclosure vowels and release bursts from two different syllables) in these different lexical contexts. For example, Marslen-Wilson and Warren (1994) showed that cross-spliced syllables made from two words ("jog" + "job") or pseudowords and words ("jod" + "job") led to slower responses when making auditory lexical deci-sions and phonological decisions relative to control spliced syllables ("job" + "job"). In contrast, there was no processing cost associated with segments cross-spliced between pseudowords ("smod" + "smob"). These results are interpreted as showing—perhaps contra to the TRACE model of speech perception (McClelland & Elman, 1986)—that the resolution of phonetic ambiguity is achieved through lexical rather than sublexical mecha-nisms. We note, however, that McQueen, Norris, and Cutler (1999) failed to replicate these findings for phono-logical decisions (although they did replicate for lexical decisions). In another study using the same cross-splicing manipulation, Dahan, Magnuson, Tanenhaus, and Hogan (2001) showed delayed recognition (measured using the timing of speech-contingent eye movements) only for syllables cross-spliced between words—a finding that could be simulated by the TRACE model.

Given these mixed results, additional behavioral data from phonetic ambiguity created with audio-morphed speech may help address long-standing issues concern-ing the role of lexical information in the resolution of phonetic ambiguity during spoken word recognition. We further collected fMRI data during recognition of these same stimuli in the context of a simple semantic lis-tening task (category monitoring) to also provide insights into the role of frontal and motor regions in the resolution of phonetic ambiguity. Inferior frontal contributions to sublexical stages of speech perception (e.g., during per-ceptual processing of speech sounds) would lead to a pre-diction of additional activation in these regions for phonetically ambiguous syllables irrespective of lexical status. However, an influence of lexical context on inferior frontal responses would instead suggest that activation increases in frontal regions may arise from higher-level

lexical selection or competition processes. Before we describe the specific behavioral and fMRI experiments reported, we will first describe some general methods (e.g., participants, stimulus preparation) that were used throughout these experiments.

GENERAL METHODS

Participants

Twenty participants (nine men) took part in the pilot experiment, 20 participants (10 men) completed the behavioral lexical decision task (LDT; Experiment 1), 23 participants (seven men) took part in the behavioral semantic decision task (SDT; Experiment 2), and 24 participants (10 men) took part in the fMRI experiment (Experiment 3). All were native British English speakers (aged 18–45 years) with nor-mal or corrected-to-normal vision and reporting no history of neurological disease, language impairment, or hearing loss. All participants self-reported as being right-handed. Participants were recruited from the MRC Cognition and Brain Sciences Unit volunteer panel with all experimental procedures approved by the Cambridge Psychology Research Ethics Committee and written informed consent obtained from all participants. None of the participants took part in more than one experiment.

Stimulus Preparation

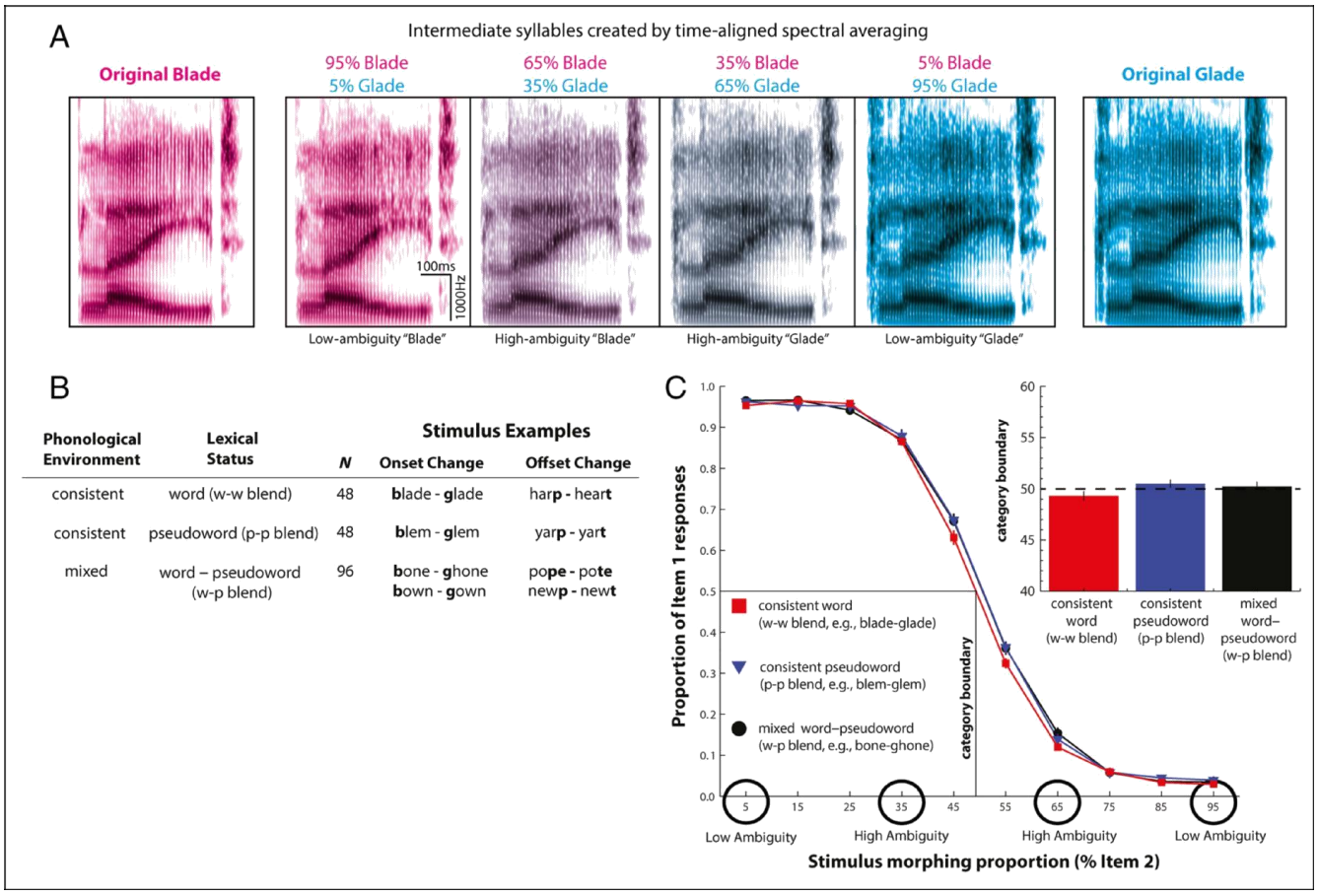
Three hundred twenty pairs of spoken syllables were chosen for use in the experiments described in this arti-cle. Each pair consisted of two syllables that were mini-mally phonetically different (i.e., differed in only voicing, manner, or place; cf. Ladefoged, 1975) at syllable onset ("porch"–"torch") or offset ("harp"–"heart"). Across the set of 320 pairs, changes were made to consonantal voicing (/b/-/p/, /d/-/t/, /g/-/k/, /θ/-/ð/, /s/-/z/, /tʃ/-/dʒ/, /ʃ/-/ʒ/, /f/-/v/), manner (/b/-/w/, /b/-/m/, /d/-/n/, /tʃ/-/ʃ/, /tʃ/-/t/, /dʒ/-/d/, /dʒ/-/ʒ/), or place (/p/-/t/-/k/, /b/-/d/-/g/, /m/-/n/-/ŋ/, /f/- /θ/-/s/-/ʃ/, /v/-/ð/-/z/-/ʒ/, /r/-/l/) of articulation. Pairs of syllables were divided into three categories that differed in terms of their phonological environment: 80 consistent word–word pairs (hereafter w–w blend, e.g., "blade"–"glade"), 80 consistent pseudoword–pseudoword pairs (hereafter p–p blend, e.g., "blem"–"glem"), and 160 mixed word– pseudoword pairs[1] (hereafter w–p blend, e.g., "gown"– "bown," "bone"–"ghone"; see Figure 1). These syllables were recorded by a single male, native English speaker (MHD) at a sampling rate of 44.1 kHz and edited into separate files using Adobe Audition 2.0.

We used time-aligned averaging of periodic, aperiodic, and F0 representations in the STRAIGHT channel vocoder (Kawahara & Morise, 2011; Kawahara et al., 1999) to gener-ate 10-step audio-morphed phonetic continua between all pairs of naturally recorded syllables. To ensure that equiva-lent positions in the pairs of syllables were averaged, we used dynamic time-warping code (www.ee.columbia.edu/~dpwe/resources/matlab/) implemented in MATLAB (The MathWorks, Inc., Natick, MA) to place anchor points at 50-msec intervals in the first syllable and maximally similar positions in the second syllable. This provides an auto-mated procedure for creating high-quality, natural sound-ing phonetic continua for the set of syllable pairs used in the experiments we describe. This further allows us to use the proportion of Sound token 1 compared with Sound token 2 as an independent measure when combining responses to different continua. For each syllable pair, we generated 10 intermediate syllables at 10% acoustic steps from 5% (highly similar to Syllable 1, e.g., "blade") to 95% (highly similar to Item 2, e.g., "glade"). Informal listening suggested increased perceptual ambiguity for syllables that were at intermediate steps (i.e., a 45% or 55% morph might sound like either "blade" or "glade" depending on listener and context).

Pilot Behavioral Experiment

To assess listeners' perception of these morphed sylla-bles, we conducted an initial pilot identification task. The 20 participants heard each of the 3200 tokens de-scribed above (10 tokens for each of the 320 syllable pairs). Five hundred milliseconds after syllable offset, they were provided with a visual prompt (the two written forms of the two possible syllables, with the critical segment underlined) and responded with a keypress to indicate which of the two source syllables they heard. A third-response alternative was offered for the possible occurrence that participants heard neither of the two syl-lable choices (only 0.7% of the trials). Proportions of re-sponses for each token were averaged over participants and transformed so that a logistic regression function could be fitted to the data for each syllable pair. From these resulting parameter estimates, we computed the position of the category boundary—that is, the estimated morphing percentage for which equal numbers of Syllable 1 and 2 responses would be expected. For most of the syllable pairs, this was close to 50%, although category boundaries varied between individual syllable pairs and were systematically shifted toward pseudo-words for w–p blend pairs (i.e., 50% stimuli were more often heard as words than as pseudowords). This is con-sistent with changes to phoneme category boundaries observed in the Ganong effect (cf. Ganong, 1980; see Rogers & Davis, 2009, for more details).

On the basis of the results of this listening test, we used MATCH software (Van Casteren & Davis, 2007) to select a subset of 192 syllable pairs to be used in the three experiments described in this article. This subset consisted entirely of syllable pairs for which the estimated category boundary was between 35% and 65% (mean = 53.65%, range = 35.34–64.79%). We selected pairs in each of the three phonological environments: 48 consistent word (w–w blend) pairs (e.g., "harp"–"heart"), 48 consis-tent pseudoword (p–p blend) pairs (e.g., "yarp"–"yart"), and



Phonological Environment	Lexical Status	N	Stimulus Examples: Onset Change	Stimulus Examples: Offset Change
consistent	word (w-w blend)	48	**b**lade - **g**lade	har**p** - hear**t**
consistent	pseudoword (p-p blend)	48	**b**lem - **g**lem	yar**p** - yar**t**
mixed	word – pseudoword (w-p blend)	96	**b**one - **g**hone **b**own - **g**own	po**pe** - po**te** new**p** - new**t**

Figure 1. (A) Spectrogram showing original stimuli and changes made during audio-morphing for a word–word minimal pair "blade" (pink) and "glade" (blue). Spectrograms show original tokens and low-ambiguity (5%/95%) and high-ambiguity (35%/65%) stimuli created using STRAIGHT software and time-aligned spectral averaging (see Rogers & Davis, 2009, for details). (B) Table showing example stimulus pairs and numbers of items illustrating changes made at syllable onset and offset in items with different types of change across lexical conditions (point of change highlighted in bold). (C) Proportion of responses matching Item 1 averaged over the phonetic continua in each of the phonological environment conditions; inset bar graph shows the mean position of the category boundary in each condition.

96 mixed word–pseudoword (w–p blend) pairs (48 items such as "newp"–"newt" and 48 items such as "pope"– "pote") such that specific segments appeared equally as words and pseudowords. By selecting a subset of items with a reduced range of category boundaries, we could ensure that lexical status (i.e., a word or a pseudoword) did not sys-tematically alter participants' perception of intermediate morph stimuli and that 35% and 65% morphed syllables were perceived as being exemplars of Syllables 1 and 2, respectively (see Figure 1C). In other words, lexical status and phonetic ambiguity were not confounded within this matched subset of 192 syllable pairs. This is confirmed by ANOVA with SPSS that revealed no significant difference between category boundary values as a function of lexical status ($F < 1$ across items; Figure 1). Analysis of response rates (proportion of Item 1 responses) recorded during the listening test for speech items from the 192 syllable pairs was also carried out using a 3 (phonological environment; w–w blend, p–p blend, w–p blend) × 4 (morph step; 5%, 35%, 65%, 95%) logistic mixed effects model with SPSS, appropriate for binomial data (Jaeger, 2008). Results revealed a highly significant effect of morph step, $F(3, 15{,}011) = 1{,}667.25$, $p < .0001$, but no significant effect of phonological environment, $F(2, 15{,}011) = 1.88$, $p > .1$, and no significant interaction, $F < 1$. Given the null effect of phonological environment for response proportions and category boundaries, this ensured that the subsequent experiments used high-ambiguity stimuli (35/65% morphs) that were equally ambiguous for the three different categories (w–w blend, p–p blend, and w–p blend) and differed only in terms of the outcome for word recognition (i.e., whether the syllables are recognized as a word or a pseudoword). We further ensured that approximately equal proportions of each type of phonetic change appeared in all three stimulus categories as well as equal proportions of changes at syllable onset and offset (see Figure 1B for examples).

Hence, the stimulus subset consisted of four tokens selected from each of the 192 stimulus pairs (total = 768 syllables). Of these syllables, 384 were high-ambiguity tokens (35% and 65% morphed syllables; Figure 1C); and 384, low-ambiguity tokens (5% and 95% morphed syllables; Figure 1C). These low-ambiguity stimuli are perceptually similar to the original recordings but have been

processed in the same way as the more ambiguous morphed stimuli. In this way, we can compare percep-tion of high- and low-ambiguity tokens so as to assess the effect of phonetic ambiguity on the identification of specific syllables. We can further look for lexical effects by comparing syllables perceived as words and pseudo-words and look for effects of phonological environment by comparing w–w blend or p–p blend syllable pairs for which both items have a consistent lexical status, with mixed w–p blend pairs in which changing the identity of the ambiguous segment turns a word into a pseudo-word or vice versa. Analysis of the characteristics of the speech tokens (duration and amplitude) was conducted with ANOVAs in SPSS and included the nonrepeated variables of Lexical status (word vs. pseudoword) and Phonological environment (w–w blend, p–p blend, w–p blend) and the repeated factor of Ambiguity (high vs. low ambiguity). Results revealed that the 768 syllables did not differ in duration (milliseconds) as a function of Lexical status, $F(1, 47) = 1.35$, $p > .1$, Ambiguity, $F < 1$, or Phonological environment, $F(1, 47) = 2.01$, $p > .1$. In addition, there was no significant difference in root mean square amplitude due to Lexical status, $F < 1$, Ambiguity, $F < 1$, or Phonological environment, $F(1, 47) = 1.16$, $p > .1$, for syllables from the 192 syllable pairs. There were no significant interactions between these factors for measures of stimulus duration or amplitude (all $ps > .1$).

Behavioral Effects of Speech Sound Ambiguity

Experiment 1 (Lexical Decision): Methods

Participants (n = 20) heard single audio-morphed sylla-bles in the context of a LDT (speeded word/pseudoword discrimination). Each participant heard a series of sylla-bles in a soundproof booth over high-quality headphones (Sennheiser HD 250) through a QED headphone ampli-fier using DMDX software (Forster & Forster, 2003) run-ning on a Windows personal computer (Dell Inc., Austin, TX). Participants made button press responses (word/ pseudoword) with both hands using a custom-made but-ton box. Equal numbers of participants pressed with their left and right hands to indicate whether a familiar word was heard. To avoid excessive stimulus repetition that may modify participant's responses to specific words and pseudowords, each participant heard half of the 768 syllables from the full item set (i.e., 384 syllables comprising one exemplar of each phonological form). These 384 syllables were divided into two experimental sessions presented with a short break between the two sessions. Each session included a single morphed syllable from each stimulus pair (48 syllables each from the w–w blend and p–p blend conditions, 96 syllables from the w–p blend condition). Syllable presentation was also rotated over two experimental versions to ensure that both high- and low-ambiguity tokens from each syllable pair were presented during the experiment but that no single phonological form was heard twice. For example, a partic-ipant might hear the low-ambiguity syllable "blade" (5% morph) during run 1 and the high-ambiguity syllable "glade" (65%) during run 2 or, alternatively, hear the high-ambiguity example of "blade" (35%) during run 1 and the clear, low-ambiguity example of "glade" (95%) in run 2. Each run contained an equal number of low-ambiguity (5/95%) and high-ambiguity (35/65%) morphed syllables. The order of stimulus presentation was also counter-balanced across participants (i.e., whether the low-ambiguity "porch" token was presented during run 1 or run 2). This resulted in four versions of the experiment, with participants pseudorandomly assigned to one of these four versions.

Results and Discussion

RTs faster than 300 msec and slower than 2000 msec (0.82% of the data) and incorrect responses (M = 9.31%; ranging from 7.43% to 12.38% across participants) were excluded from the RT analysis. RTs for the remain-ing trials are shown in Figure 2A. The erroneous re-sponses were analyzed separately and are shown in Figure 2B. Analysis of RTs included the variables of lexical status (word vs. pseudoword), ambiguity (high vs. low ambiguity), and phonological environment (w–w blend, p–p blend, w–p blend) using a linear mixed effects model with SPSS. Analysis revealed a robust main effect of pho-netic ambiguity with significantly slower responses to high- compared with low-ambiguity syllables, $F(1, 6661.55) = 40.4$, $p < .001$. Analysis also revealed a sig-nificant interaction between lexical status and ambiguity, $F(1, 6661.39) = 5.43$, $p < .05$, reflecting slower responses to high- compared with low-ambiguity words that were absent for pseudoword responses. Subsequent analysis of the simple effects revealed a reliable effect of ambigu-ity for words from w–w blend syllable pairs with a com-peting lexical neighbor (e.g., "blade"–"glade"), $t(19) = 3.82$, $p < .01$, and for words from w–p blend pairs with a pseudoword neighbor (e.g., "pope"–"pote"), $t(19) = 4.87$, $p < .001$ (see Figure 2A). Neither of the ambiguity effects for pseudoword responses were statistically reli-able; effects of ambiguity were absent both for p–p blend syllable pairs, $t(19) = 1.64$, $p > .1$, and for responses to w–p blend pairs, $t(19) = 1.48$, $p > .1$.

The analysis also revealed a main effect of lexical status with participants slower to respond to pseudowords than words, $F(1, 372.07) = 47.01$, $p < .001$. This is consistent with the findings from previous auditory LDT experi-ments (e.g., Rodd, Gaskell, & Marslen-Wilson, 2002; see also Goldinger, 1996). There was no main effect of phonological environment, $F(1, 372.14) = 1.35$, $p > .1$, with participants making lexical judgments to words and pseudowords from consistent w–w blend, p–p blend, and mixed w–p blend syllable pairs equally quickly. However, the analysis did reveal an interaction between lexical

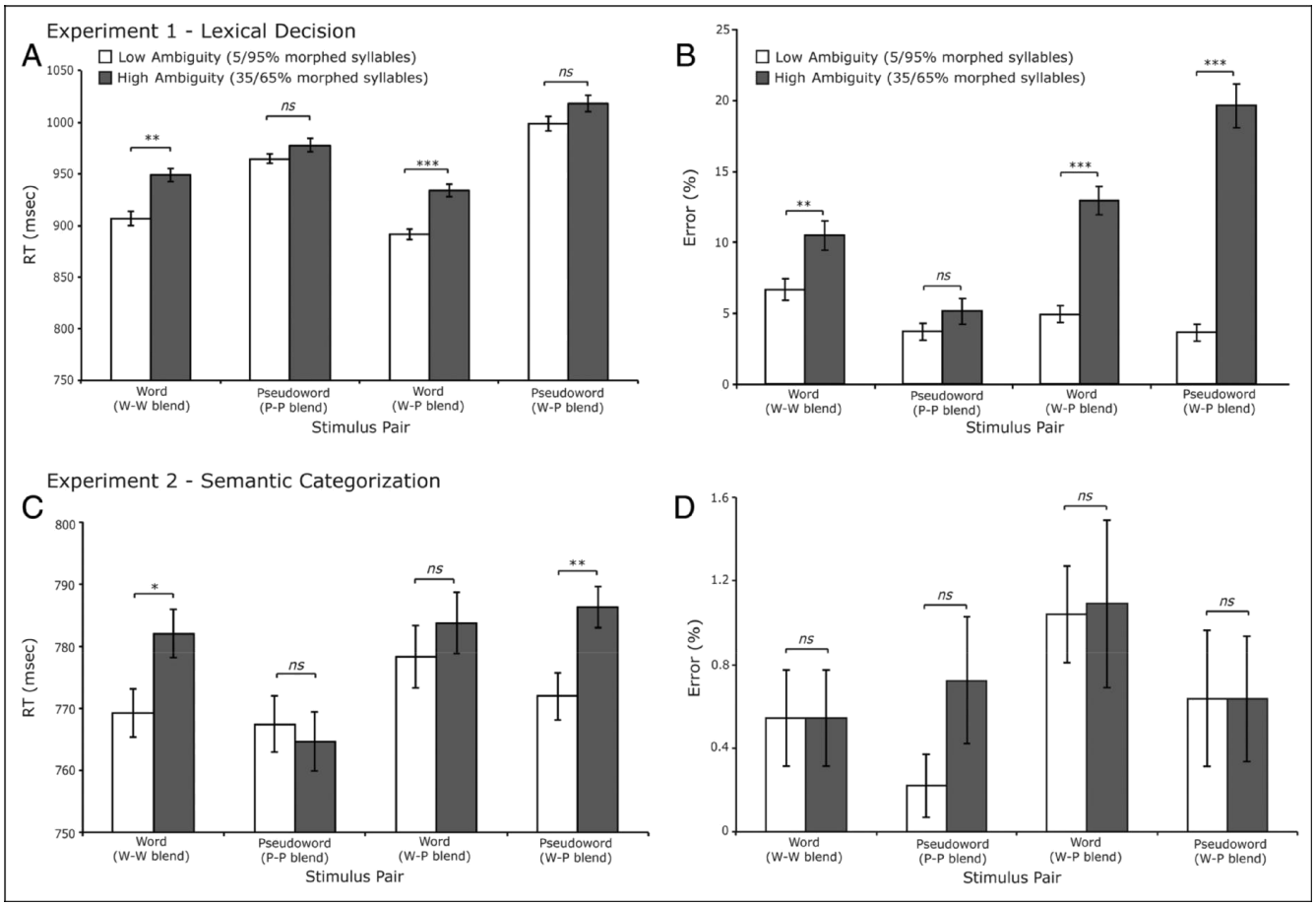


Figure 2. Graphs displaying RTs and error rates. (A, B) Results from Experiment 1 (LDT). (A) RTs in milliseconds. (B) Error rates (%). (C, D) Results from Experiment 2 (SCT). (C) RT (milliseconds). (D) Error rates (%). Error bars display the SEM with between-participant variability removed suitable for repeated-measures comparisons (cf. Loftus & Masson, 1994). *p < .05; **p < .01; ***p < .001. ns = nonsignificant.

status (words vs. pseudowords) and phonological environment, F(1, 371.79) = 7.55, p < .01. This reflects a greater slowing of lexical decisions to pseudowords from w–p blend stimulus pairs (e.g., slower responses to “bown” than “gown” from the pair “bown–gown”). This might suggest an additional source of slowing when par-ticipants make lexical decisions for pseudowords that are phonologically similar to real words, perhaps because of the conflicting (yes/no) responses required. No other two- or three-way interactions were significant.

Analysis of incorrect responses (M = 9.31%; 7.43–12.38% across participants) used a logistic mixed effects model with SPSS appropriate for binomial data and in-cluded the same factors of Lexical status, Ambiguity, and Phonological environment as used for the RT data. This revealed a highly significant effect of ambiguity on response accuracy, F(1, 7669) = 80.66, p < .001, reflect-ing increased error rates to high- versus low-ambiguity syllables. Although the interaction between lexical status and ambiguity was nonsignificant, F(1, 7669) = 2.15, p = .13, contra to the RT results, the error rates do follow the same trend as the RT results with more erroneous lexical decisions to ambiguous words from w–w blend syllable pairs, t(19) = 3.43, p < .01, that were absent for re-sponses to syllables ambiguous between two pseudo-words (p–p blend), t < 1 (Figure 2B). In contrast to effects seen in RT, no significant main effect of lexical status on error rates was observed (F < 1). This outcome suggests no overall bias in responding to words com-pared with pseudowords. However, we did see increased numbers of incorrect responses to words and pseudo-words from mixed phonological environments (i.e., w–p blend syllable pairs), and furthermore, these were increased for high- versus low-ambiguity words, t(19) = 7.69, p < .001, and pseudowords, t(19) = 10.29, p < .001 (Figure 2B). This profile is consistent with greater difficulty in generating an accurate lexical decision response to syllables that are confusable with items of op-posite lexical status (i.e., is it a word or a pseudoword?). This interpretation is supported by a significant three-way interaction of phonological environment, ambiguity, and lexical status F(1, 7670) = 4.92, p < .05 (Figure 2B). Phonetic ambiguity for items in mixed phonological environments (i.e., w–p blend syllable pairs) leads to re-sponse conflict for the LDT and hence more error-prone responses.

The presence of response conflict for phonetically ambiguous w–p blend items in the LDT and differences between the interactions seen in RTs and error rates make it hard to interpret the overall pattern of results in this experiment. Although the interaction of lexical status and ambiguity in RTs is consistent with lexical contributions to the resolution of phonetic ambiguity, response conflict is also apparent in responses to ambig-uous items from mixed w–p blend pairs. Yet, we do see reliable effects of ambiguity on RTs for w–w blend pairs for which response conflict is presumably absent. To remove response conflict as a potential explanation, we carried out a further behavioral experiment (Experiment 2) using a semantic categorization task (SCT). Because all critical trials in this experiment received a "no" response, response conflict was absent. However, for this task, word identification is still required, and ambiguity costs are again predicted for all items if phonetic ambiguity slows the early stages of speech perception but only in certain conditions if phonetic ambiguity is resolved during word recognition.

Experiment 2 (Semantic Categorization): Methods

In this experiment, (n = 23) participants performed a SCT. The 768 critical syllables from the 192 syllable pairs used in Experiment 1 were divided as before so as to en-sure that all syllable pairs were heard as high-ambiguity (35/65%) and low-ambiguity (5/95%) exemplars but that no specific phonological form was heard more than once by any participant. Rather than dividing the experiment into two sessions as in Experiment 1, we now divided the experiment into four sessions, each containing 96 test syllables. This allowed us to add 12 target words to each test session for which participants were expected to respond with a detection response while maintaining an acceptable target: nontarget ratio of 1:8 (12.5% targets). The four semantic categories used for the targets were monosyllabic color terms (e.g., "blue," "red"), weather terms (e.g., "wind," "frost"), girl's names (e.g., "Jane," "Sue"), and emotion terms (e.g., "fear," "love") selected from category norms (Battig & Montague, 1969). Spoken exemplars of these target items were recorded by the same speaker as the critical test items and analyzed/ resynthesized using STRAIGHT (Kawahara & Morise, 2011; Kawahara et al., 1999) to ensure that the sounds were matched for stimulus quality, but no morphing was applied. Participants were instructed to press a button after each item to indicate whether it was an exemplar be-longing to the current target category. Thus, all the critical high- and low-ambiguity test items (words and pseudo-words alike) should receive a nonexemplar, "no" re-sponse. The order of the four target categories was rotated across participants to control for order effects, and the hand used for "yes" and "no" responses was counterbalanced over participants as for Experiment 1.

Results and Discussion

RTs faster than 300 msec and slower than 2000 msec (0.16% of the data) and incorrect responses (M = 2.11%; ranging from 0.69% to 4.17% across participants) were excluded from the RT analysis as in Experiment 1. RTs for the remaining trials are shown in Figure 2C. As before, error rates are analyzed separately and shown in Figure 2D. Linear mixed effects analysis included the variables of lexical status (word vs. pseudoword), ambigu-ity (high vs. low ambiguity), and phonological environ-ment (w–w blend, p–p blend, w–p blend) as before. There was no significant main effect of lexical status, $F < 1$, suggesting no difference in responses to words compared with pseudowords. More importantly, this analysis revealed a significant main effect of ambiguity, $F(1, 8190.24) = 4.24$, $p < .05$, again reflecting slower RTs to high- versus low-ambiguity syllables consistent with the results from the LDT (Experiment 1). However, contra to the results from the LDT, the two-way inter-action between lexical status and ambiguity was not sig-nificant, $F < 1$. Instead, analysis revealed a significant three-way interaction between lexical status, ambiguity, and phonological environment, $F(1, 8190.26) = 3.92$, $p < .05$ (Figure 2C). This reflects an effect of phonetic ambi-guity on responses to words for which there is a compet-ing lexical neighbor (e.g., "blade"–"glade"; $t(22) = 2.31$, $p < .05$) that was absent for responses to ambiguous words paired with a pseudoword (e.g., "bone" from the pair "bone-ghone"; $t < 1$) and also absent for syllables that did not resemble real words (i.e., p–p blend pairs; $t < 1$). However, ambiguity did slow down responses to pseudo-words from w–p blend pairs in which the syllable pair in-cludes a real word competitor (e.g., "bown" from the pair "bown-gown"; $t(22) = 3.25$, $p < .01$; Figure 2C).

Incorrect responses were relatively rare in this SCT (2.11%; 0.69–4.17% across participants) and were ana-lyzed with logistic mixed effects analysis as before. This revealed no significant main effects or interactions and no difference among the different conditions tested (all ps > .1; Figure 2D).

In combination, these findings show that the effect of phonetic ambiguity on the speed and accuracy of speech perception and word recognition depends on the lexical status of the target item. In neither of these behavioral studies did we observe any effect of phonetic ambiguity on responses to pseudowords morphed with pseudo-word neighbors (e.g., "blem–glem" pairs). However, re-sponses to ambiguous syllables from w–w blend pairs (e.g., "blade–glade") were always slowed relative to un-ambiguous syllables. In line with results from previous studies of cross-spliced syllables (e.g., Marslen-Wilson & Warren, 1994, and follow-on studies cited in the Intro-duction), these findings might suggest that phonetic am-biguity is resolved by lexical rather than sublexical processes. However, we also see differential effects of ambiguity on LDT and SCT responses for syllables from

mixed phonological environments (e.g., "bown–gown" pairs). These findings are consistent with response con-flict or other task-specific influences on phonetic ambigu-ity resolution. We will explore the detailed implications of these findings in the General Discussion. However, to avoid task-specific effects, we will instead use fMRI to explore neural correlates of phonetic ambiguity reso-lution. By measuring BOLD responses during a category detection task in which all our critical items are non-targets, we can compare neural processes engaged dur-ing attentive comprehension of high- and low-ambiguity syllables without participants making any active response on critical trials.

Experiment 3 (Neural Effects of Speech Sound Ambiguity): Methods

As in our two previous behavioral experiments, each participant in the fMRI experiment (n = 24) heard half (384) of the full set of 768 syllables. Item selection was counterbalanced over two versions to ensure that partic-ipants heard only a single exemplar of each phonological form. The experiment was divided into two scanning ses-sions. During each session, participants heard one of the four syllables from a single morphed continuum (high-ambiguity 35/65% or low-ambiguity 5/95% stimuli). The order of presentation was counterbalanced so that the high-ambiguity stimulus from each syllable pair was pre-sented equally often in the first and second scanning sessions. Participants were asked to perform a semantic monitoring task, responding with a button press on an MR-compatible response box with the index finger of their left hand when they heard an exemplar of the intended cat-egory. They made no overt responses to the critical stimu-lus items (nontargets). This allowed us to assess the neural effects of phonetic ambiguity in the absence of activity due to task-induced decisions and button presses. There were 12 target stimuli in each of three possible semantic catego-ries (color terms, weather terms, and girl's names). All target stimuli were also analyzed/resynthesized using STRAIGHT (Kawahara & Morise, 2011; Kawahara et al., 1999) as described previously for Experiment 2.

Hence, each participant completed two runs,[2] each run containing 192 test syllables, 54 silent trials (20% null events) to provide a resting baseline, 12 run-relevant targets (e.g., "wind" if responding to weather targets), and 12 run-irrelevant target fillers (e.g., 12 word and pseudoword neighbors of weather terms such as "frosk" for "frost" or "wing" for "wind," ensuring that partial stim-ulus repetition could not be used to distinguish targets from nontargets). The ratio of targets to nontargets per run was 1:17 (5.88% of spoken words were targets). The order of presentation of events in each condition was pseudorandomized for each run and for each participant using MIX software (Van Casteren & Davis, 2006), ensur-ing that no more than four exemplars from one lexical condition (including null events) were heard in succes-sion, that no more than two targets were heard together, and that no more than 30 null events or nontarget items were heard between targets. Participants were notified which targets they should attend to (e.g., color terms) before the start of each run. All auditory stimuli were pre-sented at a comfortable listening volume through a pair of high-quality electrostatic headphones (Nordic Neuro Labs, Milwaukee, WI). Stimulus presentation and re-sponse measurement were controlled using custom soft-ware running on a Windows PC (Dell). Target responses and errors were recorded throughout and used to derive a signal detection measure of target detection accuracy (d^0). Participants responded correctly to nearly all run-relevant targets, M = 3.48, SD = 0.26.

Image acquisition. Imaging data were acquired from all 24 participants using a Siemens 3-T Tim Trio MR system (Siemens, Erlangen, Germany) with a 12-channel head coil. A total of 560 EPI volumes were acquired over two 13-min scanning runs (280 volumes per run, including five dummy scans at the start of each scanning run to al-low stabilization of longitudinal magnetization and five dummy scans at the end of each run to record the BOLD response to the final items). Each volume consisted of thirty-two 3-mm slices (slice order: descending, noninter-leaved; slice thickness = 3 mm, plus 0.75-mm interslice gap; in-plane resolution = 3 × 3 mm, field of view = 192 × 192 mm, matrix size = 64 × 64, echo time = 30 msec, acquisition time = 2000 msec, repetition time = 3000 msec, flip angle = 90°). Acquisition was transverse oblique, angled to avoid interference from the eyeballs and to cover the whole brain except for, in a few cases, the top of the parietal lobe. The temporal and frontal lobes were fully covered in all cases. To avoid interference from scanner noise, a rapid, fast sparse-imaging paradigm was employed (Peelle, 2014; Perrachione & Ghosh, 2013; Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999) in which stimuli were presented during the silent intervals between successive scans. A T1-weighted 3-D MPRAGE structural scan was also acquired for all participants for use during normalization (repeti-tion time = 2250 msec, echo time = 2.98 msec, flip angle = 9°, field of view = 256 mm × 240 mm × 160 mm, matrix size = 256 mm × 230 mm × 160 mm, spatial resolution = 1 × 1 × 1 mm).

Analysis of fMRI data. Data were processed and ana-lyzed using Statistical Parametric Mapping (SPM5; Well-come Department of Cognitive Neurology, London, United Kingdom) and the AA (Automatic Analysis) software package for the analysis of neuroimaging data (Cusack, 2015). Preprocessing steps included within-participant alignment of the BOLD time series to the first image of the first run, coregistration of the mean BOLD image with the structural image, and normalization of the structural image to the Montreal Neurological Institute

(MNI) average brain using the combined segmentation/ normalization procedure implemented within SPM5.

Data were analyzed using a participant-specific general linear model (GLM) with an event-related analysis procedure (Josephs & Henson, 1999) and spatially smoothed using a Gaussian kernel with a FWHM of 8 mm. The design matrix included eight test event types per run accounting for the effects of interest, notably lexical effects (words vs. pseudowords), ambiguity effects (high vs. low ambiguity), and phonological environment (w–w blend, p–p blend, w–p blend). Four additional event types also coded the target events and responses: correct target responses (hits), incorrectly identifying a nontarget as a target (false alarms), missed targets (misses), and correctly rejecting a run-irrelevant filler item (correct re-jections). Each of these 12 event types were convolved with the SPM canonical hemodynamic response function (HRF) and its temporal and dispersion derivatives (al-though contrasts were only computed using the canoni-cal response). Null events were left unmodeled and used as an implicit, silent baseline. Six additional parameters were included to account for movement-related artifacts estimated during realignment (i.e., three translation and three rotation parameters). A high-pass filter (cutoff = 128 sec) and AR(1) correction for serial autocorrelation were applied during the least mean square estimation of this GLM.

Contrasts of parameter estimates for the canonical HRF from single-participant models were entered into random effects analyses, one-sample t tests enabling in-ferences about significant effects of interest across par-ticipants. Results are reported significant at $p < .05$ whole-brain family-wise error (FWE) voxel-wise cor-rected, unless otherwise specified. We used MarsBar (MarsBar v0.41; Brett, Anton, Valabregue, & Jean-Baptiste, 2002) to analyze activation observed from the contrast of all test (nontarget) events compared with null events (implicit resting baseline). Reported lexical or ambiguity effects in this functional ROI do not constitute "double dipping" as the ROI was defined on the basis of an orthog-onal contrast in which all test items were included (see Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Friston, Rotshtein, Geng, Sterzer, & Henson, 2006).

Results and Discussion

Averaging across all speech test items compared with rest (null events) revealed large bilateral clusters in the primary auditory cortex (Heschl's gyrus) extending into the middle and superior temporal cortex, $p < .05$ (whole-brain FWE corrected; Figure 3A, Table 1). This contrast also revealed a significant cluster (47 contiguous voxels) of activation within an anterior region of LIFG (pars triangularis; peak voxel = −48, 24, 20; Figure 3A, Table 1). This is an area previously implicated in phonetic decision-making and the increased demands involved in identifying ambiguous speech tokens (e.g., Myers et al., 2009; Myers & Blumstein, 2008; Myers, 2007; Blumstein et al., 2005). However, the influence of lexical information on these phonetic categorizations and the role of LIFG in the competitive processes associated with phonetic ambiguity resolution and word recognition remain unclear.

To address this issue, the observed LIFG cluster was defined as an ROI (MarsBar v0.41; Brett et al., 2002), and we extracted the average parameter estimate for the canonical HRF for each of our eight conditions of in-terest (i.e., the magnitude of the BOLD response com-pared with rest). These parameter values were analyzed using a repeated-measures ANOVA with the factors of Lexical status (words vs. pseudowords), Ambiguity (high vs. low ambiguity), and Phonological environment (w–w blend, p–p blend, w–p blend) similar to the behavioral experiments described above. Results revealed a signifi-cant effect of Lexical status (words vs. pseudowords), $F(1, 23) = 7.36$, $p < .05$, reflecting increased activation within LIFG for words compared with pseudowords. Although the main effect of Ambiguity was not significant, $F < 1$, analysis revealed a significant interaction between Lexical status and Ambiguity, $F(1, 23) = 7.82$, $p < .05$, reflecting significantly increased activity within LIFG to high-compared with low-ambiguity words that was absent for pseudowords. We note that this significant two-way interaction is in line with the phonetic ambiguity effect on lexical decisions observed in Experiment 1. Post hoc pairwise comparisons revealed significantly in-creased activity within LIFG to ambiguous words from w–w blend stimulus pairs, $t(23) = 2.55$, $p < .05$, again in line with the behavioral results from Experiments 1 and 2 reported above, that was absent for ambiguous words paired with a pseudoword, w–p blend, $t < 1$, and for pseudowords from p–p blend, $t(23) = 1.48$, $p > .1$, and w–p blend stimulus pairs, $t < 1$ (see Figure 3B, inset graph). The three-way interaction between Lexical status, Ambiguity, and Phonological environment was not sig-nificant, $F < 1$, however, differing from the results of Experiment 2. Nonetheless, these fMRI results are in line with our behavioral results and provide compelling evi-dence that response time slowing and additional neural activity due to phonetic ambiguity resolution are observed during the recognition of words but have only a limited effect on pseudoword recognition.

For completeness, we also carried out an ROI analysis on the bilateral clusters observed in the temporal cortex for all speech test items compared with rest (null events; see Table 1) using a repeated-measures ANOVA with the factors of Lexical status (words vs. pseudowords), Ambi-guity (high vs. low ambiguity), and Phonological environ-ment (w–w blend, p–p blend, w–p blend) as before. Results revealed a significant effect of lexical status (words vs. pseudowords) in both the left, $F(1, 23) = 7.17$, $p < .05$, and right, $F(1, 23) = 7.13$, $p < .05$, hemi-sphere reflecting increased activation for pseudowords compared with words. No additional main effects or in-teractions were significant, with notably no significant

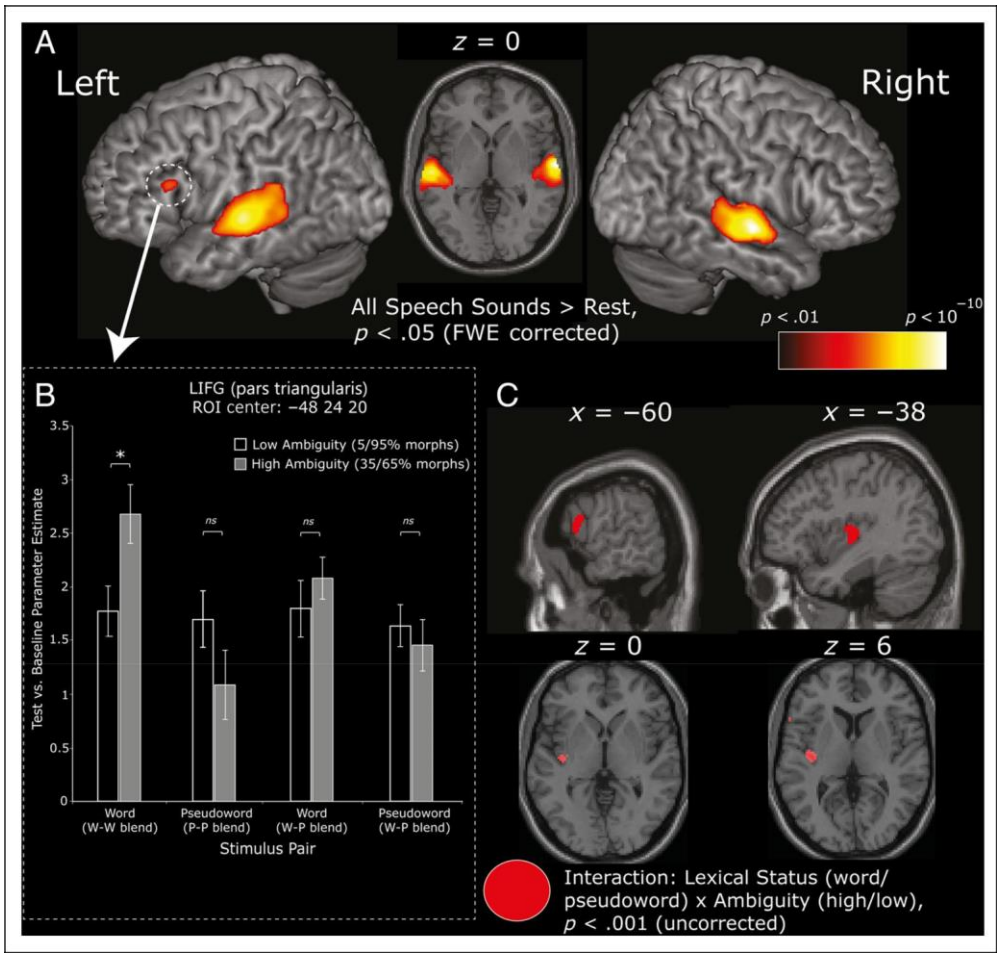


Figure 3. (A) Activation for the contrast all test items > rest (n = 24) averaged across conditions and rendered on an MNI-template canonical brain. Activation shown at p < .05, FWE corrected. Circled cluster highlights activation in LIFG (pars triangularis). (B) Bar graph shows parameter estimates for specific conditions compared with rest for the defined LIFG cluster highlighted in A. (C) Sagittal and axial slices show whole-brain results for the Lexical status (word vs. pseudoword) × Ambiguity (high vs. low) interaction, at p < .001 uncorrected. We note that the LIFG cluster in the

x = −60 slice fails to reach whole-brain-corrected significance (cluster p = .092), whereas the HG/insula cluster shown in the other slices reaches cluster-corrected significance (p = .025; Table 2).

main effect of Ambiguity (F < 1 in both the left and right hemispheres) or interaction between Lexical status and Ambiguity (F(1, 23) = 3.05, p = .09, in the left hemisphere; F(1, 23) = 2.87, p > .1, in the right hemisphere).

Additional contrasts of interest were also computed at the whole-brain level, comparing high- versus low-ambiguity items (collapsed across lexical status and phonological environment), which revealed no voxels even at an uncorrected voxel-wise threshold of p < .001. The contrast of all pseudowords compared with words did reveal reliable bilateral activation in the temporal cortex. More specifically, a large left-lateralized cluster (481 voxels) was

Table 1. All Test (Nontarget) Items Greater than Null Events (Resting Baseline)

					MNI Coordinates		
Anatomical Location[a]	Hemisphere	Voxels (n)	p Value[b]	Z Value	x	y	z
STG	Right	1523	.001	7.17	66	−10	−2
STG bordering			.001	6.00	46	−16	4
Heschl's gyrus							
Middle temporal gyrus	Left	2628	.001	6.72	−56	−20	2
Posterior STG			.001	6.37	−44	−34	10
Mid STG			.001	6.20	−64	−28	8
Inferior frontal gyrus (triangularis)	Left	47	.001	5.10	−48	24	20

[a]Areas shown in bold reflect the peak anatomical location, with the breakdown of local peaks within this cluster also shown. The table shows MNI coordinates and anatomical location of all peak voxels separated by more than 8 mm in clusters larger than 30 voxels.

[b]Column indicates voxel-wise corrected p values thresholded at p < .05, whole-brain peak level FWE correction.

Table 2. Lexical Status (Words/Pseudowords) × Ambiguity (High/Low) Interaction

Anatomical Location[a]	Hemisphere	Voxels (n)	p Value[b]	Z Value	MNI Coordinates		
					x	y	z
Insula bordering Heschl's gyrus	Left	122	.025	3.95	−38	−16	6
IFG (opercularis)	Left	63	.092	3.87	−60	14	20
IFG (triangularis)				3.49	−58	18	8

[a]Areas shown in bold reflect the peak anatomical location, with the breakdown of local peaks within this cluster also shown. The table shows MNI coordinates and anatomical location of all peak voxels separated by more than 8 mm in clusters larger than 30 voxels.

[b]Column indicates cluster-extent corrected p values thresholded at $p < .001$, uncorrected whole-brain.

observed in the left middle temporal gyrus extending to the left STG (peak voxel: $x = -62$, $y = -12$, $z = -6$, $Z = 4.04$, $p < .01$, cluster level corrected). A cluster of 244 voxels was also observed in the right STG (peak voxel: $x = 68$, $y = -26$, $z = 2$, $Z = 3.86$, $p < .05$, cluster-level corrected). This increased neural activity for pseudo-words compared with words is consistent with the main effect of lexical status observed for the bilateral temporal cortex ROIs defined using the contrast of all speech test items compared with rest (null events) reported above (see Figure 3A and Table 1). This finding is also consis-tent with a number of previous studies that have similarly demonstrated significant activation increases for pseudo-words in temporal regions (Davis & Gaskell, 2009). The reverse contrast (i.e., words greater than pseudowords) revealed two clusters in LIFG (pars orbitalis and triangu-laris), with the LIFG (pars triangularis) cluster close to the location of the LIFG ROI defined using the contrast between all speech test items and rest (null events) reported above (Table 1); however, neither of these clusters—or any others in this analysis—approached whole-brain or cluster-corrected significance.

To ensure that the LIFG ROI analysis (Figure 3A and B, Table 1) described above did not overlook other ambigu-ity effects, we also assessed the two-way interaction be-tween lexical status (words vs. pseudowords) and ambiguity (high vs. low ambiguity), collapsed across pho-nological environment, in a whole-brain analysis assessed at a voxel-wise threshold of $p < .001$, uncorrected. This interaction was observed in a region of posterior insula bordering the primary auditory cortex (Heschl's gyrus) in a cluster of 122 voxels that was significant using cluster-extent correction ($p = .025$; Table 2). Inspection of the neural response profile from the peak voxel in this cluster (−38, −16, 6) resembled that observed in the LIFG cluster described above with effects of ambiguity for syllables from w–w blend pairs. We also observed an interaction in LIFG (pars opercularis and pars triangu-laris; see Figure 3C and Table 2). This LIFG activation does not reach whole-brain-corrected significance at a cluster level ($p = .092$; Table 2). However, it is consistent with the interaction observed in an LIFG ROI defined on the basis of averaging over all test (nontarget) items (Figure 3A, circled), albeit in a slightly more posterior and lateral frontal location.

We also computed separate pairwise comparisons of high- versus low-ambiguity items for words (collapsed over w–w and w–p blends) and pseudowords (collapsed over p–p and w–p blends). Although none of these find-ings revealed clusters that reached corrected significance at the whole-brain level, they largely confirmed the loca-tions of the two-way interaction between lexical status and ambiguity. Furthermore, assessing the reverse inter-action (i.e., greater effects of ambiguity for pseudowords than for words) revealed no significant voxels even at $p < .001$ uncorrected. These nonsignificant findings suggest that our ROI and whole-brain analyses have not over-looked any significant effects of ambiguity for pseudo-words in other brain regions. We further computed the remaining two- and three-way interactions between lexi-cal status, phonological environment, and ambiguity revealing no neural effects of note or that approached a cluster-corrected threshold.

GENERAL DISCUSSION

Despite substantial variation in the sounds of speech, listeners typically perceive spoken words accurately; this is true even if the sensory input is ambiguous or de-graded. However, our ability to perceive syllables con-taining ambiguous speech sounds comes at a significant processing cost, as shown in the behavioral and neural data reported in this article. Both behavioral experiments revealed a significant ambiguity effect, reflecting slower and (for lexical decisions) more error-prone responses to high-compared with low-ambiguity morphed syllables. This processing cost is also seen in the fMRI data with ambiguous syllables increasing neural activity in LIFG regions. Strikingly, however, this processing cost is not seen for all syllables. In both our behavioral and imaging data, the effects of ambiguity interact with lexical context (i.e., whether ambiguous sounds are heard in words or pseudowords). Furthermore, the form of this interaction depends (somewhat) on the task and dependent mea-sure used. In the opening section of this discussion, we will summarize these interactions between ambiguity and

lexical context. We will then move on to discuss the implications of these observations for cognitive and neural accounts of speech perception and spoken word recognition. A particular focus will be to consider whether and how the identification of spoken words influences and is influenced by the identification of the individual speech sounds within those words.

Phonetic Ambiguity Resolution Depends on Lexical Context

In both of the behavioral experiments presented here, we report significant effects of phonetic ambiguity on measures of the ease of processing spoken words. Effects of ambiguity are confined to conditions in which participants hear stimuli that are identified as words or are audio-morphed from recordings of real spoken words. In neither of the two behavioral experiments (or indeed the fMRI data, which we will discuss subsequently) do we see evidence for significant phonetic ambiguity effects for audio-morphed stimuli created from pairs of pseudo-words (e.g., "blem–glem"). This is despite there being substantial and equivalent ambiguity in the phonetic form of audio-morphed p–p blend pairs as for the other conditions (see, for instance, the categorization function shown in Figure 1C).

To review these findings in detail, for the LDT (Experiment 1), we saw an interaction between lexical context and ambiguity such that ambiguity led to slower re-sponses for high-ambiguity syllables heard as words, but not for high-ambiguity pseudowords. For lexical de-cision errors, we saw a three-way interaction such that ambiguity had the largest numerical effect on error rates for mixed (w–p blend) pairs, irrespective of whether these were ultimately heard as words or pseudowords. This pattern is consistent with delays due to response uncertainty—the subtle acoustic differences that change a syllable from "bone" to "ghone" or from "gown" to "bown" impacts on lexical status (word to pseudoword) and hence on participants' responses. Yet, even for sylla-bles heard in consistent lexical contexts (i.e., w–w blend and p–p blend syllable pairs), increased ambiguity led to significantly slower and more error-prone responses for words with a competing lexical neighbor (e.g., "blade– glade") that was entirely absent for p–p blend pairs (e.g., "blem–glem"). This last finding cannot be explained by response uncertainty, which is equivalent for these two conditions.

A similar profile of ambiguity effects that depend on lexical context and that again cannot be explained by response uncertainty was also seen for our semantic categorization task (SCT, Experiment 2). For this study, we obtained a three-way interaction on RTs such that ambiguity effects depend on both lexical context and phonological environment. Ambiguity effects are reliable for both word (w–w blend) and w–p blend syllables heard as pseudowords, although (curiously) not for w–p blend syllables heard as words. Consistent with the LDT findings, there was no evidence for any effect of ambiguity on RTs for p–p blend pairs. Perhaps because of the universally low rates of errors for critical items in this study, there were no effects of ambiguity on participants' error rates irrespective of lexical context.

Our behavioral results for audio-morphed syllables are consistent with earlier findings of RT costs due to mismatching acoustic–phonetic information created by cross-splicing pairs of syllables before and after stop-consonant closure (i.e., subcategorical mismatches; Whalen, 1984). As described in the Introduction, seminal work from Marslen-Wilson and Warren (1994) revealed slower responses to cross-spliced syllables made from pairs of words that were absent for segments cross-spliced between two pseudowords when listeners made auditory lexical and phonological decisions. Although there has been some discussion of whether these findings rep-licate for phonological decisions (see Gaskell, Quinlan, Tamminen, & Cleland, 2008; McQueen et al., 1999), there have been several replications of the original Marslen-Wilson and Warren (1994) findings for lexical decision latencies (McQueen et al., 1999) and for the timing of speech-contingent eye movements (Dahan et al., 2001). These findings using cross-spliced syllables come from a rather limited number of segmental contrasts (typically place-of-articulation changes for word-final voiced stops like /b/, /d/, and /g/). Here, we report very similar results for a large set of audio-morphed syllables with more varied forms of phonetic ambiguity (changes to place, manner, or voicing) for consonants at syllable onset and at offset. We can therefore be more confident that our re-sults do not reflect idiosyncratic details of the acoustic form of specific tokens or segments (as might be possible for experiments in which large numbers of cross-spliced stimuli are presented).

Implications for Cognitive Models of Speech Perception and Comprehension

Our findings suggest lexical involvement in the resolution of phonetic ambiguity. When listening to spoken sylla-bles, the recognition of words is influenced by ambiguity in their constituent speech sounds. However, this effect is absent for lexical or semantic decisions on pseudo-words containing similarly ambiguous segments. One interpretation of this finding—as originally argued by Marslen-Wilson and Warren (1994)—is that categorical identification of individual speech sounds or phonemes does not occur at a pre-lexical stage during lexical identification of spoken words. Rather, listeners map the full details of the speech signal directly onto lexical representations, and phonetic ambiguity is resolved during spo-ken word recognition. Marslen-Wilson and Warren (1994) further argued that this interpretation is contra to models of speech perception in which categorical perception is achieved by competition processes at a

pre-lexical, phoneme level (as in the TRACE model; McClelland, Mirman, & Holt, 2006; McClelland & Elman, 1986). If ambiguity were resolved by recognizing phonemes before recognizing whole words, then responses to ambiguous syllables from p–p blend pairs (e.g., "blem–glem") should be disrupted similarly to those from w–w blend pairs (e.g., "blade–glade"). Simulations reported by Dahan and colleagues (2001) show that the TRACE model can simulate slowed identification of phonetically ambiguous words with a lexical competitor (i.e., additional slowing for w–w blend pairs due to increased top–down feedback to the phoneme level). Yet, in both behavioral experiments reported here, we also observed slower and/ or more error-prone responses for mixed word–pseudo-word conditions (i.e., ambiguity effects for w–p blend items). These findings have (thus far) proven difficult to simulate using the standard form of the TRACE model (see Dahan et al., 2001; Marslen-Wilson & Warren, 1994, for discussion); further simulations would be helpful in this regard.

These findings offer some support for alternative cognitive models in which speech perception is structured around two distinct processing goals: (1) recognizing familiar words and accessing their meaning (i.e., mapping heard speech onto lexical and semantic representations) and (2) identifying the phonological form of speech so that words and pseudowords can be repeated or so that phonological decisions can be made. Importantly, this dual-pathway account allows for phonetic ambiguity res-olution to operate differently during tasks in which the primary goal is to recognize words (such as in the pres-ent experiments) as for phonological tasks in which pho-netic ambiguity also leads to slower responses for pseudowords (as in more conventional categorical per-ception studies). In the context of these dual-process models, the effects seen here—with phonetic ambiguity influencing word but not pseudoword identification—are proposed to reflect processes in the lexical/semantic processing pathway.

Several dual-route models of this sort have been proposed in the literature, including the MERGE model of Norris, McQueen, and Cutler (2000) and the distributed cohort model (Gaskell & Marslen-Wilson, 1997, 1999). This latter model proposes that the task of generating a coherent phonological percept (suitable for verbal repetition or phonological decision tasks) is achieved in parallel and, to some degree, separately from the task of accessing lexical or semantic representations for familiar words. This model has been used to simulate how tasks that emphasize processing in lexical/semantic or phono-logical pathways can lead to differential influences of phonetic ambiguity on RTs and accuracy (e.g., simula-tions reported by Gaskell & Marslen-Wilson, 1997). In these views, categorical perception of speech segments is not a precursor to spoken word recognition but rather achieved in a separate processing pathway. A similar proposal has been made recently on the basis of dissociations of perceptual and lexical processing for incongru-ent audio-visual speech (Ostrand, Blumstein, Ferreira, & Morgan, 2016).

The parallel mappings proposed for accessing the phonological form and meaning of spoken words in these cognitive models, to some extent, resemble dorsal and ventral pathway accounts of the neural basis of speech perception and comprehension (see Davis, 2015; Rauschecker & Scott, 2009; Hickok & Poeppel, 2007, for discussion; see Ueno, Saito, Rogers, & Lambon Ralph, 2011, for illustrative simulations). These accounts similarly propose separate pathways for phonological and lexical/semantic processing during speech. In the final section, we will therefore consider the results of our fMRI study that localized a neural correlate of pho-netic ambiguity resolution processes in inferior frontal regions.

Inferior Frontal Contributions to Phonetic Ambiguity Resolution Depend on Task and Lexical Status

As introduced at the outset, a long-standing issue in the neural basis of speech perception and comprehension concerns the functional role of inferior frontal and pre-central gyrus regions. Demonstrations of prefrontal activation abound, particularly when listeners attentively process speech that is degraded or perceptually am-biguous (e.g., Evans & Davis, 2015; Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Hervais-Adelman et al., 2012; Lee et al., 2012; Wild et al., 2012; Adank & Devlin, 2009; Davis & Johnsrude, 2003) and if they are required to make overt decisions on the content of that speech (Du et al., 2014; Myers & Blumstein, 2008; Myers, 2007; Blumstein et al., 2005; Binder et al., 2004). How-ever, the question remains as to whether and how frontal processes contribute to speech perception per se or whether, instead, these frontal regions are associated with decision-making processes, executive functions, or other task demands. Some have argued for prefrontal contributions to attentive perceptual processing—for example, through contributions to top–down processes that are of particular importance for perception and learning of degraded speech (e.g., Sohoglu & Davis, 2016; Wild et al., 2012; Davis & Johnsrude, 2007). Others propose that prefrontal contributions are limited to tasks that require explicit segmentation and phonetic decision-making and so do not play an obligatory role in speech perception per se (for relevant imaging evidence, see Burton et al., 2000; Zatorre, Meyer, Gjedde, & Evans, 1996; see Lotto et al., 2009, for a strong form of these arguments).

In this context, then, our finding that LIFG activity is increased for phonetic ambiguity in spoken words but not pseudowords has much to contribute. First, in our fMRI experiment, overt responses were made only on semantic targets rather than critical items. We chose this

task to ensure that participants were required to listen attentively—previous work has shown significantly reduced frontal responses during inattentive listening that we wished to avoid (Wild et al., 2012). Yet, this task differs from previous studies that investigated neural responses to phonetic ambiguity because listeners were not required to make explicit phonetic judgments on ambiguous segments (e.g., Myers & Blumstein, 2008; Myers, 2007; Blumstein et al., 2005; although see, Minicucci et al., 2013; Myers et al., 2009). Hence, we explored the impact of phonetic ambiguity in a more natural listening situation in which listeners were focused on recognizing words and accessing meaning rather than identifying and responding to single speech sounds. Second, we used audio-morphing to manipulate the degree of ambiguity of a range of consonants at word onset and offset—not only changing the place or voicing of word-initial stop consonants. This variation again makes us more confi-dent that the results do not reflect idiosyncratic details of the acoustic form of specific tokens or segments or perceptual learning processes that are possible when specific, ambiguous segments are frequently repeated (cf. Norris, McQueen, & Cutler, 2003; see Kilian-Hütten, Vroomen, & Formisano, 2011, for findings linking per-ceptual learning to frontal activity).

We therefore propose that the LIFG activation we observed for phonetically ambiguous syllables from w–w blend pairs reflects the operation of prefrontal mecha-nisms that make a functional contribution to the identifi-cation and comprehension of spoken words. This conclusion goes beyond those that were possible from previous studies that employed explicit phonological judgment tasks. We note that, in these earlier studies, phonetic ambiguity often leads to increased frontal re-sponses irrespective of lexical status (e.g., in Blumstein et al., 2005, for simple nonwords like /da/ and /ta/; in Myers & Blumstein, 2008, for syllables from w–p blend pairs like "gift-kift"). Myers, Blumstein, and others have argued that greater demands on response selection in identifying segments in ambiguous syllables contribute to increased LIFG activity in these cases (Myers et al., 2008). However, in the context of our semantic monitor-ing task, response selection processes should only be engaged to the degree to which task-relevant semantic representations are activated (i.e., for exemplars of the categories that participants are monitoring for). Yet, we see increased LIFG activity for phonetically ambiguous w–w blend pairs for which neither of the words are semantic targets. We therefore propose that increased LIFG activity for these pairs can arise from increased demands on word identification processes (such as lexical competition or lexical selection) and not only from demands on nonlinguistic response selec-tion processes.

Interestingly, this conclusion supports a proposal previously made by Blumstein and colleagues from findings of impaired word recognition in patients with Broca's aphasia after lesions to left inferior frontal regions (summarized in Blumstein, 2009). In a series of semantic priming studies, they showed that these patients, like healthy controls, show reduced priming of semantically related targets for prime words with pseudoword neighbors (e.g., reduced but still significant priming of the target word "dog" from an acoustically modified token of "cat" that resembles the pseudoword "gat"; Misiurski, Blumstein, Rissman, & Berman, 2005; Utman, Blumstein, & Sullivan, 2001). However, for w–w blend pairs like "bear-pear," these patients show aberrant resolution of phonetic ambiguity, because a modified token of "pear" fails to prime the target word "fruit" unlike control participants (because of a failure to resolve competition cre-ated by the word neighbor "bear"; Utman et al., 2001). Similarly, the word "bear" will prime the related target word "wolf" (Misiurski et al., 2005), but an acoustically modified token (more similar to "pear") will not. Thus, patients with lesions to inferior frontal regions appear im-paired in resolving phonetic ambiguity in the same kind of w–w blend minimal pairs that gave rise to additional LIFG activity in our fMRI study.

These findings, along with our observations of slower and less accurate word recognition for phonetically ambiguous w–w blend pairs, suggest that phonetic ambiguity is resolved through lexical rather than pre-lexical processes and that these processes are associated with inferior frontal regions. This conclusion is consistent with the proposal made by Blumstein (2009), Thompson-Schill, D'Esposito, Aguirre, and Farah (1997), and others that the task of selecting appropriate semantic informa-tion from competing alternatives engages inferior frontal regions. However, we also note that, for tasks other than the comprehension of spoken words (e.g., in making phonetic category decisions to simple syllables like /da/ and /ta/), these inferior frontal regions also contribute to sublexical speech identification (cf. Blumstein et al., 2005). This therefore suggests that the functions sup-ported by inferior frontal regions arise from interactions between prefrontal regions and posterior superior and inferior temporal regions involved in identifying speech sounds and accessing word meanings (Davis, 2015). However, a more detailed specification of how these inferior frontal systems interact with posterior systems during different speech perception and comprehension tasks remains to be established.

Acknowledgments

We thank Hideki Kawahara for his support in using STRAIGHT, the CBU radiographers for their assistance with MRI data collection, and William Marslen-Wilson for guidance and support. This research was funded by the U.K. Medical Research Council funding (M. H. D.: MC-A060-5PQ80) and a PhD studentship awarded to J. C. R.

Reprint requests should be sent to Jack C. Rogers, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK, or via e-mail: j.rogers@bham.ac.uk.

Notes

1. In typical Ganong experiments using a limited set of phonetic contrasts, stimuli that are intermediate between a word and a pseudoword (e.g., "bone"–"ghone" and "bown"–"gown") would be assigned to different conditions—based on whether lexical context biases toward a / b/ or /g/ segment. However, because our stimulus set includes many different ambiguous segments, the division of items into these two item sets is arbi-trary. Instead, we group all the word–pseudoword minimal pairs into a single "mixed" condition and distinguish those tokens that are heard as words ("bone," "gown") as distinct from those heard as pseudowords ("ghone," "bown").
2. Three counterbalanced target semantic categories were used as participants completed three runs while in the scanner, two relevant to the fMRI experiment described in this article and one for a different fMRI experiment not described here.

REFERENCES

Adank, P., & Devlin, J. T. (2009). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. Neuroimage, 49, 1124–1132.

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. Journal of Experimental Psychology: Human Perception and Performance, 35, 520–529.

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

Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. Journal of Neuroscience, 35, 634–642.

Aydelott Utman, J., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. Brain and Language, 79, 444–472.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80, 1–46.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex, 19, 2767–2796.

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.

Blumstein, S. E. (2009). Auditory word recognition: Evidence from aphasia and functional neuroimaging. Language and Linguistics Compass, 3, 824–838.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice-onset time: An fMRI investigation of phonetic category structure. Journal of Cognitive Neuroscience, 17, 1353–1366.

Brett, M., Anton, J.-L., Valabregue, R., & Jean-Baptiste, P. (2002). Region of interest analysis using an SPM toolbox. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, 2–6 June 2002, Sendai, Japan. Available on CD-ROM in Neuroimage, 16.

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

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.

Correia, J. M. (2015). Neural coding of speech and language: fMRI and EEG studies. PhD thesis. Maastricht University.

Cusack, R. (2015). Automatic analysis (aa): Efficient neuroimaging workflows and parallel processing using Matlab and XML. Frontiers in Neuroinformatics, 8, 90.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes, 16, 507–534.

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

Davis, M. H. (2015). The neurobiology of lexical access. In G. Hickok & S. Small (Eds.), Neurobiology of language (pp. 541–555) San Diego, CA: Academic Press.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 364, 3773–3800.

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

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top–down influences on the interface between audition and speech perception. Hearing Research, 229, 132–147.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proceedings of the National Academy of Sciences, U.S.A., 111, 7126–7131.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisition. Human Brain Mapping, 7, 89–97.

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.

Evans, S., Kyong, J., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., et al. (2014). The pathways for intelligible speech: Multivariate and univariate perspectives. Cerebral Cortex, 24, 2350–2361.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? Journal of Experimental Psychology: Human Perception and Performance, 32, 1276–1293.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R., (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. Science, 322, 970–973.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behaviour Research Methods, Instruments, & Computers, 35, 116–124.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. A. (2006). A critique of functional localizers. Neuroimage, 30, 1077–1087.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance, 6, 110–125.

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.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition and blending in spoken word recognition. Cognitive Science, 23, 439–462.

Gaskell, M. G., Quinlan, P. T., Tamminen, J. T., & Cleland, A. A. (2008). The nature of phoneme representation in spoken word recognition. Journal of Experimental Psychology: General, 137, 282–302.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1166–1183.

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. Frontiers in Systems Neuroscience, 7, 1–16.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. P., et al. (1999). "Sparse" temporal sampling in auditory fMRI. Human Brain Mapping, 7, 213–223.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. Language and Cognitive Processes, 27, 1145–1166.

Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca's area in speech perception: Evidence from aphasia revisited. Brain and Language, 119, 214–220.

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

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. Journal of Phonetics, 39, 571–584.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition, 87, B47–B57.

Jaeger, T. F.(2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit mixed models. Journal of Memory and Language, 59, 434–446.

Josephs, O., & Henson, R. N. A. (1999). Event-related fMRI: Modelling, inference and optimisation. Philosophical Transactions of the Royal Society of London, 354, 1215–1228.

Kawahara, H., Masuda-Katsuse, I., & de Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. Speech Communication, 27, 187–207.

Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. SADHANA: Academy Proceedings in Engineering Sciences, 36, 713–722.

Kilian-Hütten, N., Vroomen, J., & Formisano, E. (2011). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. Neuroimage, 57, 1601–1607.

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. Journal of Cognitive Neuroscience, 25, 2179–2188.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience—The dangers of double dipping. Nature Neuroscience, 12, 535–540.

Ladefoged, P. (1975). A course in phonetics. Orlando, FL: Harcourt Brace. ISBN 0-15-507319-2.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. Nature Reviews Neuroscience, 9, 920–933.

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.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review, 1, 476–490.

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

Marslen-Wilson, W., & Warren, P. (1994). Levels of representation and process in lexical access. Psychological Review, 101, 653–675.

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

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? Trends in Cognitive Sciences, 10, 363–369.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. Journal of Experimental Psychology: Human Perception and Performance, 25, 1363–1389.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. Current Biology, 17, 1692–1696.

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.

Mirman, D., Yee, E., Blumstein, S., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. Brain and Language, 117, 53–68.

Misiurski, C., Blumstein, S. E., Rissman, J., & Berman, D. (2005). The role of lexical competition and acoustic-phonetic structure in lexical processing: Evidence from normal subjects and aphasic patients. Brain and Language, 93, 64–78.

Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. Journal of Speech, Language, and Hearing Research, 48, 884–896.

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. Journal of Neuroscience, 29, 9819–9825.

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

Myers, E. B., & Blumstein, S. E. (2008). The neural bases of the lexical effect: An fMRI investigation. Cerebral Cortex, 18, 278.

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., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23, 299–370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. Cognitive Psychology, 47, 204–238.

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.

Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage, 54, 2437–2445.

Ostrand, R., Blumstein, S. E., Ferreira, V. S., & Morgan, J. L. (2016). What you see isn't always what you get: Auditory word signals trump consciously perceived words in lexical access. Cognition, 151, 96–107.

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. Frontiers in Neuroscience, 8, 253.

Peelle, J. E., Johnsrude, I., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. Frontiers in Human Neuroscience, 4, 1–3.

Perrachione, T. K., & Ghosh, S. S. (2013). Optimized design and analysis of sparse-sampling fMRI experiments. Frontiers in Neuroscience, 7, 04.
Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. Proceedings of the National Academy of Sciences, U.S.A., 103, 7865–7870.
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.
Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. Journal of Memory and Language, 46, 245–266.
Rogalsky, C., Love, T., Driscoll, D., Anderson, S. W., & Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. Neurocase, 17, 178–187.
Rogers, J. C., & Davis, M. H. (2009). Categorical perception of speech without stimulus repetition. Paper presented at the Interspeech meeting, Brighton, UK.
Rogers, J. C., Möttönen, R., Boyles, R., & Watkins, K. E. (2014). Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. Frontiers in Psychology, 5, 754.
Schomers, M., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. Cerebral Cortex, 25, 3894–3902.
Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. Brain, 123, 2400–2406.
Scott, S. K., & Johnsrude, I. (2003). The neuroanatomical and functional organization of speech perception. Trends in Neurosciences, 26, 100–107.
Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action: Candidate roles for motor cortex in speech perception. Nature Reviews Neuroscience, 10, 295–302.
Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. Proceedings of the National Academy of Sciences, U.S.A., 12, E1747–E1756.
Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left prefrontal cortex in retrieval of semantic knowledge: A re-evaluation. Proceedings of the National Academy of Sciences, U.S.A., 94, 14792–14797. Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesising aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. Neuron, 72, 385–396.
Utman, J. A., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. Brain and Language, 79, 444–472.
Vaden, K. I., Piquado, T., & Hickok, G. S. (2011). Sublexical properties of spoken words modulate activity in Broca's area but not superior temporal cortex: Implications for models of speech recognition. Journal of Cognitive Neuroscience, 23, 2665–2674.
Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. Behavior Research Methods, 38, 584–589.
Van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. Behavior Research Methods, 39, 973–978.
Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. Perception & Psychophysics, 35, 49–64.
Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. Journal of Neuroscience, 32, 14010–14021.
Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. Nature Neuroscience, 7, 701–702.
Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: Review, replication, and reanalysis. Cerebral Cortex, 6, 21–30.
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.