Neural substrates of subphonemic variation and lexical competition in spoken word recognition

Sahil Luthra[a], **Sara Guediche**[b], **Sheila E. Blumstein**[c,d], and **Emily B. Myers**[a,e,f]

[a]Department of Psychological Sciences, University of Connecticut 406 Babbidge Road, Unit 1020, Storrs, CT, USA 06269

[b]BCBL. Basque Center on Cognition, Brain and Language Mikeletegi Pasealekua, 69, 20009 Donostia, Gipuzkoa, Spain

[c]Department of Cognitive, Linguistic & Psychological Sciences, Brown University 190 Thayer Street, Providence, RI, USA 02912

[d]Brown Institute for Brain Science, Brown University 2 Stimson Ave, Providence, RI, USA 02912

[e]Department of Speech, Language & Hearing Sciences, University of Connecticut 850 Bolton Road, Unit 1085, Storrs, CT, USA 06269

[f]Haskins Laboratories 300 George Street, Suite 900, New Haven, CT, USA 06511

Abstract

In spoken word recognition, subphonemic variation influences lexical activation, with sounds near a category boundary increasing phonetic competition as well as lexical competition. The current study investigated the interplay of these factors using a visual world task in which participants were instructed to look at a picture of an auditory target (e.g., *peacock*). Eyetracking data indicated that participants were slowed when a voiced onset competitor (e.g., *beaker*) was also displayed, and this effect was amplified when acoustic-phonetic competition was increased. Simultaneously-collected fMRI data showed that several brain regions were sensitive to the presence of the onset competitor, including the supramarginal, middle temporal, and inferior frontal gyri, and functional connectivity analyses revealed that the coordinated activity of left frontal regions depends on both acoustic-phonetic and lexical factors. Taken together, results suggest a role for frontal brain structures in resolving lexical competition, particularly as atypical acoustic-phonetic information maps on to the lexicon.

Keywords

sahil.luthra@uconn.edu (S. Luthra, corresponding author).

Introduction

Spoken language comprehension is inherently a process of ambiguity resolution. At the acoustic-phonetic level, speech sounds map probabilistically onto the candidate set of phonemes for a language — for instance, a production of a /t/ sound with a somewhat short voice-onset time (VOT) might match the /t/ category with a probability of 0.75 but also match the /d/ category with 0.25 probability, whereas a /t/ with a very long VOT value might be assigned to the /t/ category with a probability of 0.99. Natural variability in the production of speech sounds (across phonological contexts, talkers, regional dialects, and accents) leads to greater or lesser perceptual confusion among individual speech sounds (Hillenbrand, Getty, Clark, & Wheeler, 1995; Miller & Nicely, 1955).

A separable but related source of uncertainty arises at the lexical level. Most current models of lexical access propose that words in the lexicon are activated according to the degree to which they match the acoustic-phonetic input and that these words compete for ultimate selection, leading to slower access to words with, for instance, many phonologically similar neighbours (Luce & Pisoni, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007). In this way, the phonological structure of words in the lexicon influences which candidate items are activated; for example, lexical competitors *beaker* and *beetle* are activated given the input /bi/ (Allopenna, Magnuson, & Tanenhaus, 1998; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). For present purposes, we take resolution of competition to reflect all of the processes that enable the listener to select a target stimulus amongst multiple lexical candidates in order to ultimately guide behaviour.

It can be methodologically difficult to distinguish between the influence of acoustic-phonetic variability on *phoneme-level* processes and its influence on *lexical-level* processes. In current models of auditory word recognition, an exemplar of the word *pear* with the initial /p/ close to the /p/-/b/ phonetic boundary would result in reduced activation of /p/ and increased activation of /b/ at the phonemic level as the acoustic-phonetic information is mapped on to a phonemic representation. This in turn would result in reduced activation of *pear* and increased activation of its competitor *bear* as the phonemic level information is mapped onto its lexical representation. Indeed, results of several studies suggest that subphonemic variation cascades through the phoneme level and ultimately affects the activation of lexical representations (e.g., Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, & Aslin, 2002; 2009). Using a visual world eye tracking paradigm, McMurray et al. (2002) showed that fixations to a named picture (e.g., *pear*) were affected by word-initial VOT manipulations of the auditory target in the presence of a picture of a voiced competitor (e.g., *bear*). The closer the VOT was to the boundary, the more participants looked at the voiced competitor, consistent with within-category phonetic detail influencing the activation of word representations. Of interest in the present investigation is how these two sources of indeterminacy, phonetic variation and lexical competition, are resolved by the brain as listeners map speech to meaning.

The bulk of studies that have investigated the neural substrates of phonetic variability as it contacts the lexicon have discussed their results in terms of phonetic competition; however, subphonemic variation does not always result in more ambiguous speech. A token that is

distant from the category boundary (e.g., a /t/ sound with an uncharacteristically long VOT) will be less ambiguous than a token with more typical VOT values, but it will still be an atypical or poor exemplar. Indeed, listeners appear to track two distinct but interrelated aspects of phonetic variability: the *goodness of fit* between a token and a category exemplar (Miller, 1994) as well as the degree of *phonetic competition* between phonetically similar categories (i.e., proximity to the category boundary; Pisoni & Tash, 1974). When a speech token approaches a phonetic category boundary, it is likely to induce more phonetic competition with the contrasting category, but it is also less typical as a member of the phonetic category. Tokens that are far from the category centre but also not near a competing category will be atypical but not subject to increased competition.

Neuroimaging studies have suggested these two aspects of subphonemic variability, namely competition and goodness of fit, are also neuroanatomically separable (Myers, 2007). In general, a role for the LIFG has been proposed in resolving *phonetic competition* (Binder et al., 2004; Myers, 2007), with the LIFG showing greater activation to more phonetically ambiguous stimuli as participants perform a phonetic identification task (e.g. "*d* or *t*?"). In contrast, the bilateral superior temporal gyri (STG), especially in areas lateral and posterior to Heschl's gyrus, are sensitive to the *goodness of fit* between a token and the phonetic categories in one's language (Blumstein, Myers, & Rissman, 2005; Chang et al., 2010; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Myers, 2007) and respond in a graded fashion to degradations in the speech signal (Obleser, Wise, Alex Dresner, & Scott, 2007; Scott, Rosen, Lang, & Wise, 2006).

Here, we consider how each dimension of phonetic variability – *phonetic competition* and *goodness of fit* – affects lexical access and how they impact brain activity. A network of regions in temporal, parietal, and frontal areas respond to lexical competition in a variety of contexts, including activation of phonologically and semantically-related words (Gow, 2012; Hickock & Poeppel, 2007; Prabhakaran et al., 2006; Righi, Blumstein, Mertus & Worden, 2010). Among these regions, frontal brain regions including the left inferior frontal gyrus (LIFG) are strong candidates for showing integration of phonetic competition with lexical selection because they have been implicated in resolving competition, both in domain-general contexts (Badre & Wagner, 2004; Chrysikou, Weber, & Thompson-Schill, 2014) and in language processing (Thompson-Schill, D'Esposito, Aguirre, & Farrah, 1997). In a visual world study examining the neural correlates of lexical competition, Righi et al. (2010) found that the visual presence of an onset competitor (e.g., *beaker* as a competitor for *beetle*) led to increased activity in left supramarginal gyrus (LSMG) and left inferior frontal gyrus (LIFG). Similarly, a study by Minicucci, Guediche, and Blumstein (2013) found that the LIFG was modulated by an interaction between lexical competition and phonetic competition during a lexical decision task.

It is less clear how the goodness of fit dimension of acoustic-phonetic variation might interact with lexical competition. Theoretically, the increased activation of the STG for less-degraded versions of the stimulus could be linked to stronger access to higher levels of processing (lexical, etc.) that a more intelligible signal affords. However, studies using decoding techniques show that even when individuals only listen to syllables, regions in the STG respond in a graded fashion to degradations in the input (Evans et al., 2013; Pasley et

al., 2012), leading to the hypothesis that the STG should show preferential tuning (i.e., increased activation) for "good" versions of the phonetic category. Supporting this hypothesis, Myers (2007) found that bilateral STG showed graded activation in line with the goodness of fit of syllable tokens, whether or not they were close to the category boundary.

In assessing the interactions between different aspects of phonetic variability and lexical competition, we use a paradigm that better approximates ecologically valid listening conditions. Notably, the vast majority of studies suggesting that frontal areas may be sensitive to within-category phonetic detail have employed metalinguistic tasks (e.g., lexical decision; Aydelott Utman, Blumstein & Sullivan, 2001; Minicucci et al., 2013) or else have used nonsense syllables devoid of meaning (Myers, 2007; Myers, Blumstein, Walsh, & Eliassen, 2009). Hickok and Poeppel (2004) have argued that such paradigms may unduly tax the processing system by engaging metalinguistic strategies or speech segmentation processes that may or may not emerge when listeners map speech to conceptual representations under natural listening conditions. This may be a particular concern for assessing the role of frontal areas (e.g., LIFG) in processing phonetic variation, as frontal areas are generally linked to domain-general cognitive control processes; that is, the previously observed recruitment of frontal areas may be a consequence of task-related decision-level processing rather than an aspect of naturalistic language processing.

Though any laboratory task is in some sense artificial, we suggest that ecologically valid tasks (1) should entail mapping sound to meaning, (2) should require tasks that de-emphasise metalinguistic judgments, and (3) should not impose unnatural burdens on phonetic processing. A couple of recent studies examining the role of frontal areas in processing phonetic variation have satisfied some of these criteria, though perhaps not all of them. For instance, Xie and Myers (2018) observed LIFG involvement for naturally-occurring phonetic variation in the comparison of conversational speech compared to clear speech registers. The hypoarticulation of conversational speech leads to a more densely-packed vowel space, thereby increasing phonetic competition. The authors found that trial-by-trial variation in phonetic competition (as estimated from measures of vowel density) correlated with activation in the LIFG as listeners engaged in a simple lexical probe verification task. These findings suggest that LIFG is recruited under conditions of phonetic competition even in the absence of an explicit sub-lexical task. A recent semantic monitoring study by Rogers and Davis (2017) further suggests that the LIFG may be involved in resolving phonetic competition, at least as such competition cascades to the lexical level. In that study, the authors morphed minimal pairs to create phonetically ambiguous blends and asked participants to press a button when they heard an exemplar of a target semantic category. They found that LIFG activation was sensitive to the degree of phonetic competition when both ends of the morphed continuum were real words (e.g. *blade-glade*) but not when one or both ends of the continuum were non-words (e.g. *gleam-*bleam*). Taken together, these studies provide additional evidence that the LIFG is recruited for processing phonetic category detail, particularly in situations when lexical competition can arise.

In real language comprehension, however, a rich mosaic of linguistic and visual cues limit the set of likely utterances. For instance, semantic context may lead listeners to prefer one

member of a minimal pair over another (*The farmer milked the GOAT/COAT*; Borsky, Tuller, & Shapiro, 1998), or items in a visual array may re-weight the probability of a particular referential target (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Previous work by Guediche, Salvata, and Blumstein (2013) examined the interaction between sentence context and acoustic phonetic properties of speech and found that frontal responses were modulated by the semantic bias of the sentential context and by the phonetic properties of the stimuli; however, in that study, the interaction between the two factors only emerged in temporal areas. Such context was not a factor in either the Rogers and Davis (2017) or Xie and Myers (2018) studies. In particular, listeners in Rogers and Davis (2017) heard single words produced in isolation, whereas listeners in Xie & Myers (2018) heard semantically anomalous sentences (e.g., "Trout is straight and also writes brass") where semantic context could not be used to guide recognition of the target word. As the field moves to increasingly naturalistic paradigms, it is important to also consider the influence of contextual information. Notably, constraining contexts can come from the sentence itself (e.g. "I put on my winter __" strongly biases the listener towards the word "coat"), or can come from factors external to language, such as the presence of a picture of a likely lexical target in a visual array.

The current study used the visual world paradigm to ascertain whether subphonemic detail influences the activation of lexical representations, with eye tracking and fMRI measures collected simultaneously. Participants were asked to look at a named target (e.g., *peacock*) for which the initial VOT varied within the voiceless category (i.e., the initial sound was never voiced), with some VOTs *shortened* such that they were closer to the voiced category (i.e., increased phonetic competition, decreased fit to the category) and others *lengthened* so that they were even further from the category boundary (decreased phonetic competition, decreased fit to the category). By manipulating VOT in this way, effects of phonetic competition can be dissociated from the influence of goodness of fit. Lexical competition was manipulated by displaying either a voiced onset competitor (e.g., *beaker*) or an unrelated distractor (e.g., *sausage*) with the target picture. A strength of this paradigm is its ecological validity, as the paradigm requires mapping acoustic-phonetic input to lexical semantics, does not require participants to make metalinguistic decisions, and does not place explicit demands on phonetic processing because the word referent is always present in the visual array. Furthermore, the use of only two pictures places strong constraints on lexical activation, effectively biasing lexical activation towards two candidates. In theory, limiting the scope of likely referents to only two pictured items might attenuate (or even eliminate) any effect of phonetic competition when those possible referents are phonologically distinct, but might amplify competition effects when the two possible referents are phonologically similar. The use of only two visual referents allows us to determine whether phonetic competition automatically triggers frontal activation, or whether frontal structures are only recruited when some uncertainty remains at the lexical level. In our study, we take a two-fold approach to identifying potential interactions between phonetic variation and lexical competition. First, we employ a univariate approach to assess whether activation of any individual region is modulated by both factors. Second, we analyse functional connectivity to examine whether an interaction emerges in the pattern of activation across multiple brain regions.

Materials and Methods

Participants

Participants were recruited from the Brown University community and were paid for their participation. Eight participants completed a pretest to collect normative data on the stimuli, 15 completed a behavioural pilot of the fMRI experiment, and 22 completed the fMRI experiment; individuals participated in only one of the three experiments. Data from one fMRI participant were lost due to an eye tracker battery failure, and an additional three participants were excluded from analyses due to excessive motion, resulting in a total of 18 fMRI participants (age: mean = 22.1, SD = 3.3; 8 female) included in the analyses reported here.

Stimuli

Sixty word triads (see Appendix) were created, each triad comprising three items:

1. a voiceless *target* (e.g., *peacock*) with a word-initial voiceless stop consonant and no voicing minimal pair (i.e. *beacock* is not a real word)

2. a voiced *onset competitor* (e.g., *beaker*) that differed from the corresponding target item by the voicing of the initial phoneme but otherwise shared the first syllable, or at least the first vowel, with the target token

3. a phonologically *unrelated* item (e.g., *sausage*) with a word-initial non-plosive consonant.

Each item was a polysyllabic noun with initial stress. Separate one-way ANOVAs revealed no significant difference between word types (voiceless, voiced, unrelated) in written frequency (Kucera & Francis, 1967) [$F_{(2,177)} = 0.330$, $p = 0.719$], imageability (Coltheart, 1981) [$F_{(2,130)} = 0.977$, $p = 0.379$] or verbal frequency (Brysbaert & New, 2009) [$F_{(2,93)} = 1.334$, $p = 0.268$]. For each item, a greyscale-filtered line drawing was selected, with images either drawn by the first author or taken from clip art and public domain databases.

An auditory version of each item was recorded in a soundproof room using a microphone (Sony ECMMS907) and digital recorder (Roland Edirol R-09HR). A female native speaker of American English produced each item three times, and the first author selected the best token. Stimuli were scaled to a mean intensity of 70.4 dB SPL (SD: 2.5). A one-way ANOVA revealed a significant difference in duration between word type [$F_{(2,177} = 3.184$, $p = 0.044$; see Table 1]. Post-hoc t-tests revealed that this effect was due to a significant difference in the duration of unrelated items and voiced items [$t(118) = 2.43$, $p = 0.016$]. There was no significant difference between voiceless targets and voiced items [$t(118) = 1.68$, $p = 0.096$], nor between voiceless targets and unrelated items [$t(118) = 0.85$, $p = 0.396$].

Since voice-onset time is the primary cue that distinguishes between word-initial voiced and voiceless stop consonants (Abramson & Lisker, 1985; Francis, Kaganovich, & Driscoll-Huber, 2008; Lisker & Abramson, 1964), we manipulated this parameter to examine the effects of subphonemic variation on lexical activation. To do so, the VOT of the voiceless *unaltered* tokens were measured in BLISS. For each item, two additional voiceless tokens

were created by manipulating the voice-onset time. *Shortened* tokens were created by removing the middle two-thirds of the VOT, yielding tokens with an initial consonant closer to the voiced-voiceless category boundary. *Lengthened* tokens were created by duplicating the middle two-thirds of the VOT and inserting this segment at the midpoint of the VOT, resulting in tokens that were farther from the phonetic category boundary and poorer exemplars of the voiceless category. For the sake of comparison, the VOT of voiced items was also measured. VOT measurements are reported in Table 1, and sample waveforms and spectrograms are shown in Figure 1A.

Pretesting and piloting

Stimuli were pretested to ensure that participants were sensitive to the VOT manipulation and that all three variants (*shortened*, *unaltered*, *lengthened*) were perceived as voiceless. Eight participants heard only the first syllable of each word (to avoid lexical biases) and were asked to press one button if the stimulus began with /b/, /d/ or /g/ and a different button if it began with /p/, /t/ or /k/. Each subject heard each variant of the 60 voiceless tokens once (180 trials) and each voiced item three times (an additional 180 trials). To avoid ordering effects of the three VOT types (*lengthened*, *unaltered*, *shortened*), subjects received an equal number of items in each of the six possible orders. This same pseudo-randomisation was used for all subjects with one exception due to experimenter error. Subjects were above 98 percent accuracy for all three types of voiceless items and on voiced trials, demonstrating that altered VOT tokens were perceived as voiceless. A within-subjects ANOVA on reaction time data (Figure 1B) demonstrated that participants were sensitive to the acoustic manipulation [$F(2,14) = 4.380$, $p = 0.033$]. Newman-Keuls post-hoc tests showed a significant RT difference between *shortened* (mean: 764 ms) and *unaltered* (mean: 742 ms) tokens as well as between *lengthened* (mean: 770 ms) and *unaltered* tokens. There was no significant difference between *shortened* and *lengthened* tokens ($p > 0.05$).

fMRI/eye tracking experiment

Immediately prior to scanning, participants completed a self-paced familiarisation task to learn the intended image names (e.g. *pistol* instead of *gun*). Participants saw each of the 180 items individually, under which was written the name of the item. Stimulus order was randomised once and held constant for all subjects. Participants then viewed the images without the names and were asked to name each image; a new randomisation was used for this test, and all subjects received the same randomisation at test. The experimenter checked each participant's response and provided the correct response when the participant made an error.

The eye tracking task was completed while participants lay supine in a 3T Siemens PRISMA scanner using a 64-channel head coil. During the task, participants saw visual displays with two images on a white background and were instructed to look at the image that corresponded to the auditory target. Each participant saw 240 unique visual displays, half corresponding to critical trials (voiceless item as a target) and half corresponding to non-critical trials (voiced *onset competitors* or *unrelated* words as a target). Each participant received two critical trials in which a given voiceless item (e.g., *peacock*) was the target; in one, the *onset competitor* (*beaker*) served as the distractor image, and in the other, the

unrelated item (*sausage*) was the distractor image. For each voiceless item, the *Acoustic Modification* (*shortened, unaltered, lengthened*) was counterbalanced such that each subject received equal exposure to the three phonetic conditions across the experiment.

Each item was presented in all three acoustic conditions across subjects, and no subject heard a particular token (e.g. the lengthened version of *peacock*) more than once during the experiment. On non-critical trials, either the *onset competitor* or *unrelated* item served as the auditory target. For these trials, the distractor picture was of a voiceless item from a separate word triad selected to match the onset of the voiced target as much as possible (e.g., *peanut*); in this way, subjects did not receive the same visual display more than once. The visual display was sized to subtend 20° of visual angle. Target location was counterbalanced across displays, and displays were also balanced such that each word type (e.g. voiced) appeared equally often on the left as on the right. A sample array of displays is shown in Figure 1C.

The experiment was divided into four blocks, each consisting of 60 trials. Each block included only one critical or non-critical trial for each word triad (i.e. each trial in Figure 1C appeared in a separate block). Trial order was pseudorandomised once such that the same picture did not appear in consecutive trials, and this order was used for all participants. Each trial began with a black fixation cross at the centre of the screen. The black fixation was replaced by a green fixation 500 ms before the onset of the visual display. Once the visual display appeared, subjects were given 250 ms to preview the pictures before the onset of the auditory stimulus. The pictures remained on screen for 1500 ms following the onset of the audio, after which the black fixation reappeared on the screen. This design is shown in Figure 1D.[1]

Eye tracking was completed using a long-range eye tracker (EyeLink 1000 Plus, SR Research, Ontario, Canada) with monocular tracking and an MRI-compatible tracker. Standard five-point calibration and validation procedures were used to ensure that the tracker was functioning appropriately, and these procedures were repeated as necessary between blocks. Eye data were collected at a frequency of 250 Hz.

Anatomical images were acquired using a T1-weighted magnetisation-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 1900 ms, TE = 3.02 ms, FOV = 256 mm, flip angle = 9 degrees) with 1 mm sagittal slices. Functional images were acquired using a T2*-weighted multi-slice, interleaved ascending EPI sequence and a rapid event-related design (TR = 3.6 s [effective TR of 2.0 s with a 1.6 s delay], TE = 25 ms, FOV = 192 mm, flip angle = 90 degrees, slice thickness = 2.5 mm, in-plane resolution: 2 mm x 2 mm). A sparse sampling design was employed so that auditory stimuli fell in silent gaps between scans (Figure 1D). Auditory stimulus onset occurred 100 ms into the silent phase. Because the longest auditory stimulus was 812 ms, there was a guarantee of at least 88 ms silence before the next scan phase. To appropriately model the hemodynamic response, three jitter times (3.6 s, 7.2 s, and 10.8 s) were distributed randomly across the experimental conditions, and each jitter was used an equal number of times in each run. A total of 181 volumes were

---

[1]Prior to conducting the fMRI experiment, we conducted a behavioural pilot experiment (n = 15) outside of the scanner. This pilot employed an analogous design to the one used in the fMRI experiment. While analyses of the pilot data are not reported here, the results were similar to those observed in the fMRI experiment.

acquired for each functional run. In total, there were four runs, each lasting approximately 11 minutes. Due to a programming error, neural data were not collected for the final trial of each run.

Preprocessing of eye tracking data

Growth curve analysis (GCA) was used to examine eye tracking data. Traditional approaches for analysing visual world data (e.g., t-tests, ANOVAs) can yield different results depending on time-bin selection and inappropriate assumptions about the independence of different time bins. By contrast, GCA approaches analysis hierarchically; a first-order model is used to capture the time course of looks across all conditions, and a second-order model describes how the fixed factors of interest affect the first-order time course. In the present study, the fixed factors of interest were *Lexical Competition* (*onset competitor*, *unrelated distractor*) and *Acoustic Modification* (*shortened*, *unaltered*, *lengthened*). An advantage to this approach is that it isolates which time component (e.g., intercept, linear term) is affected by each fixed factor, providing a precise understanding of how eye looks change over time. (For introductions to GCA, see Mirman, 2014; Mirman, Dixon, & Magnuson, 2008.)

Following McMurray et al. (2002), we defined an area of interest for each picture by taking each cartoon and expanding the area by 100 pixels in each direction. A fixation report for each subject was compiled using the EyeLink DataViewer, with saccades within an area of interest considered as part of the fixation in that area. Fixation data were then downsampled to a rate of 20 Hz to reduce the likelihood of false positives (Mirman, 2014). Mean proportion of fixations to the distractor at each time point were calculated by averaging across trials in each condition; the mean proportion of fixations to the target item was similarly computed. Growth curve analyses were then performed in R using the *lmer* function of the lme4 package.

Eye tracking analyses were performed using a time window of 300 ms to 1000 ms after audio onset. Analyses were limited to critical trials (that is, limited to trials where the target word began with a voiceless stop) in which participants fixated on the target picture for at least 100 ms, effectively excluding "incorrect" trials (in which the subject never looked at the target picture) and trials in which the eye tracker did not track eye movements (e.g., due to a battery failure). This resulted in the exclusion of 9.2% trials (4.2% due to incorrect responses, 5.0% due to technical failures). For brevity, we report only analyses of fixations to distractors; a similar pattern of fixed effects also emerged in analyses of target trials.

A cubic orthogonal polynomial (first-order model) was used to capture the overall time course of fixations to the distractor. A second-order model examined how fixed effects of interest interacted with each time term in the first-order model. The fixed effect of *Lexical Competition* (*onset competitor*, *unrelated*) was backward difference-coded with a (1/2, −1/2) contrast, while *Acoustic Modification* (*shortened*, *unaltered*, *lengthened*) was backward difference-coded with contrasts of (−2/3, 1/3, 1/3) and (−1/3, −1/3, 2/3). By using these two contrasts for *Acoustic Modification*, the model is able to separately analyse two pairwise differences: *unaltered-shortened* and *lengthened-unaltered*. To also examine the *lengthened-shortened* pairwise difference, a second analysis used contrasts of (−2/3, 1/3, 1/3) and (−1/3, 2/3, −1/3) for *Acoustic Modification.* The second-order model also included random subject

and subject-by-condition effects for the intercept, linear and quadratic terms; the cubic term tends to only be informative about variation in the tails of the distribution, so estimating random effects for this term is generally not worth the amount of data needed to do so (Mirman, Dixon, & Magnuson, 2008). The normal approximation (i.e. that t values approach z values as degrees of freedom increase) was used to estimate p values for significance testing.

Preprocessing of fMRI data

Preprocessing of functional data was completed in AFNI (Cox, 1996). Functional images were transformed to a cardinal orientation using the de-oblique command, and the remaining preprocessing was done separately on each run using an afni_proc.py pipeline. Functional data were aligned to the anatomy, registered to the third volume of each run to correct for motion, and warped to Talairach and Tournoux (1988) space with an affine transformation; these transformations were applied simultaneously to minimise interpolation. Data were then smoothed using a 4-mm full-width half-maximum Gaussian kernel and scaled such that the mean of each run was 100. To estimate the hemodynamic response, a gamma function was convolved with stimulus onset times for each of the six critical (*onset competitor* / *unrelated* distractor $\times$ *shortened* / *unaltered* / *lengthened* token) and non-critical (*voiced*, *unrelated*) conditions. Preprocessed functional data were submitted to a regression analysis that included the idealised hemodynamic response functions for each condition and also included the six rigid-body parameters from the volume registration as nuisance regressors.

Beta coefficients were submitted to a mixed-factor ANOVA with *Lexical Competition* (*onset competitor* / *unrelated*) and *Acoustic Modification* (*shortened* / *unaltered* / *lengthened*) as fixed factors and *Subject* as a random factor. A group mask containing only voxels imaged in all 18 participants was applied at the ANOVA step. A subsequent small volume correction was applied, limiting analyses to anatomically defined regions broadly associated with speech processing: bilateral transverse temporal gyri (TTG), superior temporal gyri (STG), middle temporal gyri (MTG), middle frontal gyri (MFG), inferior frontal gyri (IFG), angular gyri (AG) and supramarginal gyri (SMG); these regions were defined automatically in AFNI using the Talairach and Tournoux (1988) atlas (Figure 3A). Monte Carlo simulations performed on the small-volume corrected mask (10,000 iterations) indicated that a cluster threshold of 164 voxels was necessary to reach significance at the $p < 0.05$ level (corrected for multiple comparisons). For these simulations, a Gaussian filter was applied, with the filter width in each direction set as the average noise smoothness values from the residual time series.

To separately examine effects of phonetic competition and goodness of fit, two planned comparisons were employed. A planned comparison to assess *phonetic competition* assumed that competition increases linearly as VOT values approach the phonetic boundary and thus used a contrast (*shortened* = 1, *unaltered* = 0, *lengthened* = −1). A *goodness of fit* contrast assumed that goodness of fit would be optimal for *unaltered* tokens and decrease for both types of modified tokens and therefore used the contrast (*shortened* = −0.5, *unaltered* = 1, *lengthened* = −0.5).

Finally, a psychophysiological interaction (gPPI; McLaren, Ries, Xu, & Johnson, 2012) analysis was performed to examine whether regions sensitive to *Lexical Competition* change in their connectivity to any other region as a function of *Acoustic Modification*. To this end, regions identified in the univariate analysis as sensitive to *Lexical Competition* were used as seed regions, and stimulus onset vectors were generated for each *Acoustic Modification* (*shortened*, *unaltered*, *lengthened*). For each seed, a deconvolution analysis was run for each subject that included: the de-trended time course of the seed region; timing vectors for the acoustic conditions (*shortened*, *unaltered*, *lengthened*, *voiced*, *unrelated*) convolved with a gamma function; two-way interactions between the seed and three convolved conditions of interest; and nuisance regressors for the six rigid-body motion parameters. As in the univariate analysis, planned comparisons for phonetic competition and goodness of fit were used. Monte Carlo simulations were performed to determine a minimum significant cluster size of 265 voxels ($p < 0.05$, corrected).

Results

Eye tracking results

The time course of distractor fixations in the behavioural experiment is shown in Figure 2, with model fits and data points shown for each condition. Significant effects from the growth curve analyses are summarised in Table 2.

The significant effect of *Lexical Competition* on the intercept indicates that across all time terms, participants looked more at the distractor if it was a lexical competitor (e.g. *beaker* when *peacock* is the target) than if it was an unrelated distractor (e.g. *sausage*). An interaction between *Lexical Competition* and *Acoustic Modification* emerged on the intercept term for the *lengthened-shortened* pairwise comparison. Follow-up models[2] indicated that this interaction was driven by a significant effect of *Lexical Competition* on the intercept time term for the *shortened* tokens but no such effect for *unaltered* or *lengthened* tokens. Finally, the interaction between *Lexical Competition* and *Acoustic Modification* also emerged on the cubic time terms in the *lengthened-shortened* and *lengthened-unaltered* pairwise comparisons. Neither appears to be theoretically meaningful; the former captures a small difference in the tails of the curves, whereas the latter reflects a difference in the growth curve trajectories outside the time range being analysed.

Univariate fMRI analysis

Results from the univariate analysis are summarised in Table 3. An effect of *Lexical Competition* emerged in bilateral middle/inferior frontal gyri as well as in bilateral temporal/supramarginal gyri (Table 3, Figure 3B). Within these clusters, activation was greater when there was an onset competitor shown (e.g., *beaker*) on the screen than an unrelated distractor

---

[2] These additional models used treatment-coded factors instead of the backward difference coding scheme described above. These follow-up models do not differ in their fit to the data; the only difference lies in what is captured by the beta values. In treatment coding, one level (e.g., the *shortened* level of *Acoustic Modification*) is set as a reference level; the beta value for the other factor (e.g., *Lexical Competition*) then reflects a simple effect within that reference level. (In this example, the *Lexical Competition* beta terms would reflect the simple effect of *Lexical Competition* for *shortened* tokens on the intercept, linear, quadratic and cubic terms.) Constructing models for each level of *Acoustic Modification* allowed us to examine the effect of *Lexical Competition* separately for each level.

(e.g., *sausage*). A main effect of *Acoustic Modification* emerged in two clusters in Heschl's gyrus bilaterally (Table 3, Figure 3C). These clusters were nearly identical to those revealed in the planned comparison designed to select regions that were sensitive to phonetic competition. Though phonetic competition is greatest when VOTs are *shortened* (that is, when the initial segment is closest to the voiced competitor), the clusters in Heschl's gyrus showed greatest activation for *lengthened* tokens, with less activation for both *unaltered* and *shortened* tokens[3]

No clusters reflecting a *Lexical Competition* (*onset competitor* / *unrelated*) × *Acoustic Modification* (*shortened* / *unaltered* / *lengthened*) interaction survived correction for multiple comparisons. This is striking given the evidence for an interaction between these two factors in the eye tracking data.[4] However, as noted previously, it is possible that the interaction between lexical competition and subphonemic variation is not tied to activation in one particular region but rather to the pattern of activity (i.e., functional connectivity) across several regions.

gPPI analysis

Potential interactions between *Acoustic Modification* and *Lexical Competition* might emerge in the pattern of brain activity across regions. As such, a gPPI analysis was conducted to examine the influence of *Acoustic Modification* on functional connections with seed regions that were sensitive to *Lexical Competition* (LSTG, RIFG, LIFG [*pars opercularis*/*pars triangularis*], LIFG [*pars orbitalis*] and RSTG). Results are visualised in Figures 4 and 5 and summarised in Table 4. We observed significant effects for our planned comparison of phonetic competition and our planned comparison of goodness of fit. Each set of results is briefly described in turn.

The phonetic competition comparison (Figure 4) reflects functional connections that are strongest for *shortened* tokens relative to *unaltered* and *lengthened* ones. We observed this pattern in the functional connections between the left STG seed and the right TTG. The same pattern emerged in functional connections with right hemisphere seeds – namely, in connections between the right IFG seed and the left MFG; between the right IFG seed and bilateral cingulate cortex; the right IFG seed and the left/right thalamus; and the right STG seed and the right inferior parietal lobule.

---

[3]Because of concerns that the effects of *Acoustic Modification* might be driven by differences in overall stimulus length and not by the VOT manipulation, a control analysis was conducted that also included post-consonant stimulus length as a nuisance regressor. The VOT manipulation only affected the duration of the word-initial consonant, so examining post-consonant stimulus length affords us an orthogonal measure of stimulus length and gives us more confidence that the effects of *Acoustic Modification* reflect our VOT manipulation and not overall differences in stimulus length. All clusters reported in Table 3 also emerged in this follow-up analysis.
[4]To account for individual differences in behaviour on functional activation, a control analysis was conducted that included behavioural effect sizes as a continuous covariate in the group-level fMRI analysis. To estimate each subject's effect size, we extracted subject-by-condition random effects from the second-order model in the growth curve analysis. In particular, we measured for each subject how much larger their competitor effect was in the *shortened* condition than in the *lengthened* condition. (Recall that the difference in the lexical competition effect between the *shortened* and *lengthened* conditions was the only significant interaction in the eyetracking analysis, and note also that these conditions differ in *phonetic competition* but not in *goodness of fit*.) A region in right superior / transverse temporal gyrus [(61, −13, 6), 219 voxels, F = 18.2] was sensitive to the size of this behavioural covariate. For the fixed effects of interest, the same clusters emerged in this control analysis as in the main analysis, albeit at a slightly reduced voxel-level threshold ($p < 0.06$; 195 voxels required for a cluster-level alpha of 0.05).

The planned comparison for goodness of fit (Figure 5) reflects connectivity that changes as a function of the degree of match between the token and its phonetic category (i.e. for *unaltered* tokens relative to the two altered (*shortened* and *lengthened*) forms). We observed this pattern with the two left IFG seeds – in particular, in the connectivity between a left IFG seed (*pars opercularis / triangularis*) and the left MFG, as well as in the connectivity between the other left IFG seed (*pars orbitalis*) and anterior midline structures. Finally, this same pattern of connectivity was observed in functional connections between the left STG seed and the medial frontal gyrus/anterior cingulate, as well as in functional connections between the left STG seed and the cerebellum.

Discussion

Cascading effects of phonetic variation

Lexical access is subject to competition between acoustically similar items, and ambiguity in the speech signal can influence the degree to which various candidates are activated. The present study used a visual world paradigm in which the target voiceless stimuli were either shortened, lengthened or unaltered. Consistent with previous findings (Andruski et al., 1994; Clayards, Tanenhaus, Aslin, & Jacobs, 2008; McMurray et al., 2002), eye tracking data indicated that listeners entertained multiple visual targets as possible lexical candidates and that the closer the initial VOT was to the voiced/voiceless category boundary, the more a voiced competitor was considered, as indexed by increased looks to the distractor picture. Of note, this interaction emerged in the context of minimal acoustic-phonetic overlap between the target and competitor; that is, the critical onset competitors used here differed in initial voicing of the onset consonant and overlapped at most through the first syllable (though in many cases, only through the first vowel). These behavioural data add to a body of evidence showing that within-category phonetic detail cascades to the level of lexical processing, persisting despite disambiguating phonological information as the word unfolds.

Frontal and temporal processing of lexical competition

Of interest in the present study is whether phonetic variation and lexical competition are neurally dissociable and whether specific brain regions are sensitive to their interaction. Functional neuroimaging and neuropsychological studies have routinely implicated frontal brain regions — most notably, the LIFG — in speech perception, particularly under perceptually challenging conditions or when laboratory tasks require explicit sublexical decisions (Burton, Small, & Blumstein, 2000; Poldrack et al., 2001; Zatorre et al., 1996). However, explicit phonological tasks (e.g. identification and discrimination tasks) may not reflect the higher-level processes involved in mapping the speech signal to conceptual representations (e.g., Hickok & Poeppel, 2004) or may also require additional processing demands apart from mapping the acoustic signal to a linguistically meaningful unit. As such, a number of recent studies (e.g., Myers & Xie, 2018; Rogers & Davis, 2017) have eschewed metalinguistic tasks in favour of more ecologically valid paradigms (i.e., those that require mapping to a conceptual representation in the absence of making any metalinguistic judgment and without undue pressure on the phonetic processing system) when probing the role of frontal regions in processing phonetic variation. Importantly, because mapping the speech signal to conceptual representations necessarily involves activating lexical

representations, these studies have queried the neural basis of phonological competition by investigating how such subphonemic variation affects lexical competition. Results from those experiments have indicated that even with increasingly ecological tasks, the LIFG is implicated in resolving phonologically-mediated lexical competition under conditions of phonetic competition. However, these studies entailed presenting stimuli to participants in the absence of any linguistic or visual context. As the field moves to using increasingly naturalistic tasks, it is worth probing the impact of context-based expectations, which listeners routinely use in natural instances of spoken word recognition. Thus, the present study provided visual cues to limit the set of lexical candidates to an extreme degree – only the two candidates shown on any given trial – while still employing an ecologically valid task.

Results indicated that frontal (bilateral IFG, MFG) and temporo-parietal regions (STG, MTG, and SMG) were recruited to a greater degree when a lexical competitor was visually present in the display. This set of regions is largely consistent with the findings of Righi et al. (2010), who manipulated the presence of a cohort competitor (e.g., *beaker* when the target was *beetle*) in a visual world experiment, though notably the degree of lexical competition in the current study was reduced as the target and competitor (e.g., *peacock* and *beaker*, respectively) also differed in word-initial voicing. Importantly, activity in these regions has been shown to be modulated by a number of phonological and lexical properties (e.g., Peramunage et al., 2011; Prabhakaran et al., 2006; Raettig & Kotz, 2008). Studies have suggested that anatomically distinct regions within the IFG are sensitive to different linguistic properties, with *pars opercularis* (BA 44) and *pars triangularis* (BA 45) playing a role in phonological processing and *pars orbitalis* (BA 47) being sensitive to lexical-semantic details (Fiez, 1997; Nixon et al., 2004; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). Gold and Buckner (2002) found that these frontal regions coactivated with the SMG during phonological processing and with the MTG in mapping the speech signal to semantic information, consistent with the functional roles ascribed to SMG and MTG elsewhere in the literature (Gow, 2012; Hickok & Poeppel, 2007). Activation in the left middle frontal gyrus, on the other hand, has been specifically tied to lexical retrieval (Gabrieli, Poldrack, & Desmond, 1998; Grabowski, Damasio, & Damasio, 1998).

Notably, the current results implicate not only left frontal areas in the processing of lexical competition but also right frontal regions, with the presence of an onset competitor eliciting activation of the right IFG/MFG. Typically, neurobiological models of language processing hold that spoken word recognition recruits bilateral temporal regions, but is left-lateralised in the anterior regions that underlie higher-level language processing (e.g., Hickok & Poeppel, 2004). Based on our findings, additional work is needed to clarify the extent to which right frontal regions are also engaged in higher-level aspects of naturalistic language processing.

While the bilateral posterior STG are most prominently associated with early acoustic-phonetic processing (Hickok & Poeppel, 2004; 2007), neuroimaging studies suggest that activation of the STG may also be modulated by top-down feedback from higher-order cortices (Gow & Olson, 2016; Gow, Segawa, Ahlfors, & Lin, 2008; Guediche et al., 2013; Leonard, Baud, Sjerps, & Chang, 2016; Myers & Blumstein, 2008). However, given that the temporal resolution of fMRI precludes an inference about the timecourse of activation, it is

difficult to determine whether the pattern in the STG in the present data reflects bottom-up sensitivity to lexical competition (i.e. two phonologically-similar targets are activated and those lexical candidates implicitly compete for access) or whether it reflects post-phonetic-analysis feedback from higher-level areas (e.g. the LIFG) that are implicated in competition resolution. Overall, it is striking that even under conditions of relatively little lexical competition, the set of implicated regions is consistent with results from previous studies (e.g., Righi et al., 2010).

In addition to the modulation of the LIFG by lexical competition, we predicted that this region would show sensitivity to the cascading influence of subphonemic variation on lexical competition. However, the univariate analysis of activation data showed no significant interaction between acoustic modification and lexical competition, in IFG or in any other brain areas. Once contact with the lexicon has been made, phonetic competition may not be heavily weighted by the LIFG, in contrast to tasks that don't involve lexical processing such as phoneme identification tasks (e.g., Myers, 2007). Furthermore, as noted previously, the degree of phonological overlap between the target and the competitor was smaller than in previous studies; for instance, the present study used lexical competitors *kettle* and *gecko*, which mismatch in voicing in the initial phoneme and overlap only in the following vowel. While this minimal phonological overlap was sufficient to show a main effect of lexical competition in a wide network of brain areas, it may not have been enough to yield a neural interaction between acoustic modification and lexical competition, particularly given the strong top-down support in the current task (as pictures related to only two potential lexical targets). Future investigations that probe the interaction of less subtle manipulations with lexical competition may successfully evoke a neural interaction.

Although we did not observe the predicted interaction between lexical competition and acoustic modification, we did find that functional connections between LIFG and other frontal regions were modulated by acoustic-phonetic information in the target, specifically by the "goodness of fit" manipulation. In particular, functional connectivity between LIFG and LMFG decreased when the target was acoustically modified (shortened or lengthened). Considered with evidence that these regions are implicated in executive control (e.g., Badre & Wagner, 2004; Braver et al., 2001), the present connectivity results may reflect an influence of phonetic variation on lexical competition at a selection level. The LMFG specifically has been associated with mapping atypical acoustic-phonetic information to internal categories (e.g., Desai, Liebenthal, Waldron, & Binder, 2008; Myers & Blumstein, 2008, Myers & Swan, 2012). Thus, it may be that when listeners encounter *typical* acoustic information (i.e., tokens with high goodness of fit to a phonetic category), strong activation for the phoneme target cascades to the lexical level and results in efficient selection among the set of possible lexical/semantic items. Under this interpretation, greater connectivity between left frontal regions and an adjacent domain-general executive processing network may reflect the increased ease with which listeners are able to select the appropriate target in precisely the situations when acoustic information is unaltered.

It is somewhat surprising that connectivity results implicated frontal regions in processing the *goodness of fit* between a token and its phonetic category, given that previous findings have linked this dimension of phonetic variation to temporal regions (e.g., Myers, 2007) and

that the studies that have looked at how acoustic-phonetic information influences lexical competition have shown an increase in frontal sensitivity to lexical competition precisely when phonetic competition is high (Miniccuci, et al., 2013; Rogers & Davis 2017). However, such studies did not test for interactions between lexical competition and goodness of fit specifically, and indeed we interpret the changes of functional connectivity as goodness of fit interacts with lexical competition to be primarily driven by activation of multiple lexical candidates.

We have focused our discussion principally on functional connections with left frontal seed regions, as a primary theoretical goal of the current work is to identify the extent to which frontal regions are sensitive to acoustic-phonetic detail and lexical competition. Nonetheless, the functional connectivity analysis demonstrated that subphonemic variability also impacted functional connections with other seed regions that were sensitive to lexical competition. For instance, we observed task-based modulation of connectivity between a left STG seed and the cerebellum as well as between the left STG seed and anterior cingulate / middle frontal gyrus, with increased connectivity for unaltered tokens compared to modified (shortened or lengthened) ones. The cerebellum has been proposed to be involved in encoding auditory events with high temporal precision (Schwartze & Kotz, 2016). Within this framework, enhanced cerebellar recruitment and connectivity with temporal and frontal regions may better guide attention for more efficient perceptual integration of the speech signal. In the present study, the pattern of connectivity with the STG seed might reflect the relative ease of integration when acoustic-phonetic properties are typical of the phonetic category; that is, unaltered tokens may facilitate access to phonemic representations and enhance the adaptive mechanisms involved in resolving ambiguities among competing lexical items. Additionally, we observed that connectivity between left and right temporal regions depended on the voice-onset time of the stimuli, with increased connectivity observed as VOT was reduced. Such a finding is consistent with the notion that activation of the bilateral temporal lobes is tied to early acoustic analysis, with increased right temporal activation in tasks that do not require explicit phonetic analysis (Turkletaub & Branch Coslett, 2010). Finally, we found increased connectivity between a right IFG seed and regions implicated in domain-general executive functioning (left MFG, bilateral cingulate gyrus); as discussed above, such a finding supports the idea that acoustic detail affects the resolution of lexical competition at a selection stage.

Temporal lobe sensitivity to subphonemic variation

Behavioural sensitivity to subphonemic variation was relatively subtle. In-scanner eye tracking data revealed a main effect of the acoustic manipulation such that participants' looks to the target were fastest when the VOT was shortened, and looks were slowest when the VOT was lengthened. This finding stands in contrast to previous work (e.g., Myers, 2007) showing that altered tokens slowed processing in a phonetic decision task (e.g. "/d/ or /t/?"). The pattern in the eye tracking data here seems to instead mirror the fact that manipulation of the VOT also manipulates the timecourse of the rest of the word — that is, for words with shorter VOTs, listeners also hear the subsequent vowel (as well as the rest of the word) earlier. Thus, it appears that earlier looks to the target for shortened tokens reflect the fact that the identity of the word was accessible earlier than for unaltered and lengthened

tokens, rather than that shortened VOTs somehow facilitate access to the voiceless initial stop. This view is supported by our behavioural pretest. When confronted with syllables with shortened VOTs, listeners were slower to identify these as voiceless stops than they were for unaltered tokens.

A parallel effect emerges in the bilateral STG, which showed greatest activation for *lengthened* tokens, with diminished activation for *unaltered*, and then *shortened* tokens. This pattern is inconsistent with what might be expected from previous studies (e.g., Myers, 2007) that show sensitivity to *goodness of fit* in the temporal lobes – that is, increased activation for altered tokens compared to unaltered. Further, this region also shows no increased activation for tokens approaching the category boundary, unlike studies using explicit categorisation tasks (Blumstein, et al., 2005, Myers, 2007, Myers & Blumstein, 2008). At least two candidate explanations might account for these differences. First, the use of nonwords in the aforementioned studies may focus participant attention on phonetic level processes, whereas the current study engaged lexical processing. As such, the increased activation for *lengthened* tokens in the current study may reflect increased strength of activation of the target word, with greatest strength when the initial segment is entirely unambiguous (*lengthened*) providing an unambiguous match to the target picture. This possibility is supported by findings that the STG shows stronger responses to unambiguous compared to degraded speech signals (e.g., Davis & Johnsrude, 2003; Erb, Henry, Eisner, & Obleser, 2013; Wild, Davis, & Johnsrude, 2012). A second possibility is that these clusters are responding to the overall length of stimuli (which co-varies with the VOT manipulation) rather than to the subphonemic details of the stimuli (e.g. Ranaweera et al., 2016). To evaluate this possibility, a secondary analysis considered the length of the entire stimulus as a regressor of no interest in order to account for the variance associated with responses to stimulus length (Footnote 3). STG clusters sensitive to the VOT manipulation persisted after controlling for the overall length of the stimuli, suggesting that the observed effects of VOT on STG activation are above and beyond those due to stimulus length.

Conclusions

The results of the present investigation support a long-standing view in the spoken word recognition literature that variability at the phonological level cascades to the lexical level and influences lexical dynamics as listeners access meaning (McClelland & Elman, 1986; McMurray et al., 2002). The present neural findings implicate a network of frontal regions in the processing of subphonemic variation as it cascades to the lexical level, even in a paradigm that does not require explicit sublexical decisions. In the current study, the strength of functional connections between frontal regions changes as a function of the quality of acoustic-phonetic information, showing greater connections for typical compared to less typical acoustic-phonetic. Future investigations will be needed to further dissociate effects of phonetic competition from effects of phonetic goodness of fit and to determine how the informativeness of top-down context shapes sensitivity to acoustic-phonetic detail, both in behaviour and in the brain.

Acknowledgements

Appendix

Appendix:

Stimuli

	Triad			Distractor for Non-Critical Trial
	Voiceless (Critical Target)	Voiced	Unrelated	
1	carpet	garbage	finger	carton
2	cargo	gargoyle	magnet	cardboard
3	padlock	badger	faucet	paddle
4	pepper	berry	radio	petal
5	toilet	doily	sandwich	toaster
6	palette	balcony	river	poodle
7	carton	garden	hammer	carpet
8	tiger	diver	newspaper	tunnel
9	kettle	gecko	rabbit	kayak
10	puppet	bucket	circus	puddle
11	cobra	gopher	firework	coconut
12	people	beaver	chimney	pistol
13	parchment	barber	flashlight	parsley
14	coffee	goblin	racket	comet
15	coconut	goalie	hanger	cobra
16	pudding	butcher	rainbow	pulley
17	poodle	boomerang	fairy	palette
18	puddle	butter	leopard	puppet
19	teapot	demon	monkey	tentacle
20	puppy	button	muscle	puzzle
21	paddle	battery	necklace	padlock
22	palace	ballot	money	package
23	table	daisy	wagon	taxi
24	pancake	bandage	vacuum	panda
25	parrot	barrel	genie	passport
26	comet	goggles	ruler	coffee
27	petal	belly	mushroom	pepper
28	passport	basket	chocolate	parrot
29	puzzle	bubble	waffle	puppy
30	cauldron	golfer	microphone	cobweb

	Triad			Distractor for Non-Critical Trial
	Voiceless (Critical Target)	Voiced	Unrelated	
31	tortoise	doorknob	window	tissue
32	parsley	barstool	cereal	parchment
33	pastry	baseball	sandal	paper
34	package	banner	laundry	palace
35	tentacle	dentist	rattle	teapot
36	panther	banjo	lightning	parsnip
37	tunnel	dungeon	robot	tiger
38	pocket	bottle	hammock	poncho
39	camel	gambling	ninja	candle
40	candle	gander	scissors	camel
41	pavement	bagel	shovel	patient
42	cobweb	goblet	motorcycle	cauldron
43	poncho	bonnet	rocket	pocket
44	paper	baby	medal	pastry
45	tissue	discus	needle	tortoise
46	taxi	dagger	lobster	table
47	patient	bacon	ladder	pavement
48	cabin	gavel	rooster	cabbage
49	toaster	doughnut	lawnmower	toilet
50	pulley	bullet	lemon	pudding
51	timer	dinosaur	ginger	typewriter
52	parsnip	barley	honey	panther
53	peacock	beaker	sausage	peanut
54	typewriter	diamond	ladle	timer
55	panda	bandit	wizard	pancake
56	pistol	biscuit	lollipop	people
57	cardboard	garlic	farmer	cargo
58	peanut	beetle	locker	peacock
59	cabbage	gallery	monocle	cabin
60	kayak	geyser	feather	kettle

References

Abramson AS, & Lisker L (1985). Relative power of cues: F0 shift versus voice timing. Phonetic linguistics: Essays in honor of Peter Ladefoged, 25–33.

Allopenna PD, Magnuson JS, & Tanenhaus MK (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language, 38(4), 419–439.

Andruski J, Blumstein SE, & Burton M (1994). The effects of subphonetic differences on lexical access. Cognition, 52, 163–187. [PubMed: 7956004]

Aydelott Utman J, Blumstein SE, & Sullivan K (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. Brain and Language, 79(3), 444–472. [PubMed: 11781053]

Badre D, & Wagner AD (2004). Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. Neuron, 41(3), 473–487. [PubMed: 14766185]

Blumstein SE, Myers EB, & Rissman J (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. Journal of Cognitive Neuroscience, 17(9), 1353–1366. [PubMed: 16197689]

Borsky S, Tuller B, & Shapiro LP (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. The Journal of the Acoustical Society of America, 103(5), 2670–2676. [PubMed: 9604360]

Braver TS, Barch DM, Gray JR, Molfese DL, & Snyder A (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. Cerebral Cortex, 11(9), 825–836. [PubMed: 11532888]

Brysbaert M & New B (2009) Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. Behavior Research Methods, 41(4), 977–990. [PubMed: 19897807]

Burton MW, Small SL, & Blumstein SE (2000). The role of segmentation in phonological processing: An fMRI investigation. Journal of Cognitive Neuroscience, 12(4), 679–690. [PubMed: 10936919]

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, & Knight RT (2010). Categorical speech representation in human superior temporal gyrus. Nature Neuroscience, 13(11), 1428–1432. [PubMed: 20890293]

Chrysikou EG, Weber MJ, & Thompson-Schill SL (2014). A matched filter hypothesis for cognitive control. Neuropsychologia, 62, 341–355. [PubMed: 24200920]

Clayards M, Tanenhaus MK, Aslin RN, & Jacobs RA (2008). Perception of speech reflects optimal use of probabilistic speech cues. Cognition, 108(3), 804–809. [PubMed: 18582855]

Coltheart M (1981). The MRC Psycholinguistic Database. Quarterly Journal of Experimental Psychology, 33A, 497–505.

Cox RW (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, 29, 162–173. [PubMed: 8812068]

Davis MH, & Johnsrude IS (2003). Hierarchical processing in spoken language comprehension. Journal of Neuroscience, 23(8), 3423–3431. [PubMed: 12716950]

Desai R, Liebenthal E, Waldron E, & Binder JR (2008). Left posterior temporal regions are sensitive to auditory categorization. Journal of Cognitive Neuroscience, 20(7), 1174–1188. [PubMed: 18284339]

Erb J, Henry MJ, Eisner F, & Obleser J (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. Journal of Neuroscience, 33(26), 10688–10697. [PubMed: 23804092]

Evans S, Kyong JS, Rosen S, Golestani N, Warren JE, McGettigan C, Mourão-Miranda J, Wise RJS, & Scott SK (2013). The pathways for intelligible speech: multivariate and univariate perspectives. Cerebral Cortex, 24(9), 2350–2361. [PubMed: 23585519]

Fischl B (2012). FreeSurfer. NeuroImage, 62(2), 774–781. [PubMed: 22248573]

Fiez JA (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. Human Brain Mapping, 5(2), 79–83. [PubMed: 10096412]

Francis AL, Kaganovich N, & Driscoll-Huber C (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. Journal of the Acoustical Society of America, 124(2), 1234–1251. [PubMed: 18681610]

Gabrieli JD, Poldrack RA, & Desmond JE (1998). The role of left prefrontal cortex in language and memory. Proceedings of the National Academy of Sciences, 95(3), 906–913.

Grabowski TJ, Damasio H, & Damasio AR (1998). Premotor and prefrontal correlates of category-related lexical retrieval. Neuroimage, 7(3), 232–243. [PubMed: 9597664]

Gold BT, & Buckner RL (2002). Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. Neuron, 35(4), 803–812. [PubMed: 12194878]

Gow DW (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. Brain and Language, 121(3), 273–288. [PubMed: 22498237]

Gow DW, & Olson BB (2016). Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. Language, Cognition and Neuroscience, 31(7), 841–855.

Gow DW, Segawa JA, Ahlfors SP, & Lin FH (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. Neuroimage, 43(3), 614–623. [PubMed: 18703146]

Guediche S, Salvata C, & Blumstein SE (2013). Temporal cortex reflects effects of sentence context on phonetic processing. Journal of Cognitive Neuroscience, 25(5), 706–718. [PubMed: 23281778]

Hickok G, & Poeppel D (2000). Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences, 4(4), 131–138. [PubMed: 10740277]

Hickok G, & Poeppel D (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition, 92(1), 67–99. [PubMed: 15037127]

Hickok G, & Poeppel D (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8(5), 393–402. [PubMed: 17431404]

Hillenbrand J, Getty LA, Clark MJ, & Wheeler K (1995). Acoustic characteristics of American English vowels. The Journal of the Acoustical Society of America, 97(5), 3099–3111 [PubMed: 7759650]

Kucera H, & Francis WN (1967). Computational analysis of present-day American English. Providence, RI, Brown University Press.

Leonard MK, Baud MO, Sjerps MJ, & Chang EF (2016). Perceptual restoration of masked speech in human cortex. Nature Communications, 7.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, & Medler DA (2005). Neural substrates of phonemic perception. Cerebral Cortex, 15(10), 1621–1631. [PubMed: 15703256]

Lisker L, & Abramson AS (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20(3), 384–422.

Luce PA, & Pisoni DB (1998). Recognizing spoken words: the neighborhood activation model. Ear & Hearing, 19(1), 1–36. [PubMed: 9504270]

Magnuson JS, Dixon JA, Tanenhaus MK, & Aslin RN (2007). The dynamics of lexical competition during spoken word recognition. Cognitive Science, 31(1), 133–156. [PubMed: 21635290]

Marslen-Wilson W, & Welsh A (1978). Processing interactions during word-recognition in *continuous speech*. Cognitive Psychology, 10, 29–63.

McClelland JL, & Elman JL (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1–86. [PubMed: 3753912]

McLaren DG, Ries ML, Xu G, & Johnson SC (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. Neuroimage, 61(4), 1277–1286. [PubMed: 22484411]

McMurray B, Tanenhaus MK, & Aslin RN (2002). Gradient effects of within-category phonetic variation on lexical access. Cognition, 86(2), B33–B42. [PubMed: 12435537]

McMurray B, Tanenhaus MK, & Aslin RN (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. Journal of Memory and Language, 60(1), 65–91. [PubMed: 20046217]

Miller GA, & Nicely PE (1955). An analysis of perceptual confusions among some English consonants. The Journal of the Acoustical Society of America, 27(2), 338–352.

Miller JL (1994). On the internal structure of phonetic categories: a progress report. Cognition, 50(1), 271–285. [PubMed: 8039364]

Minicucci D, Guediche S, & Blumstein SE (2013). An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. Neuropsychologia, 51(10), 1980–1988. [PubMed: 23816958]

Mirman D, Dixon JA, & Magnuson JS (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. Journal of Memory and Language, 59(4), 475–494. [PubMed: 19060958]

Mirman D (2014). Growth curve analysis and visualization using R. Boca Raton, FL: CRC Press.

Misiurski C, Blumstein SE, 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(1), 64–78. [PubMed: 15766769]

Myers EB (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. Neuropsychologia, 45(7), 1463–1473. [PubMed: 17178420]

Myers EB, & Blumstein SE (2008). The neural bases of the lexical effect: an fMRI investigation. Cerebral Cortex, 18(2), 278–288. [PubMed: 17504782]

Myers EB, Blumstein SE, Walsh E, & Eliassen J (2009). Inferior frontal regions underlie the perception of phonetic category invariance. Psychological Science, 20(7), 895–903. [PubMed: 19515116]

Myers EB, & Swan K (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. Journal of Cognitive Neuroscience, 24(8), 1695–1708. [PubMed: 22621261]

Nixon P, Lazarova J, Hodinott-Hill I, Gough P, & Passingham R (2004). The inferior frontal gyrus and phonological processing: An investigation using rTMS. Journal of Cognitive Neuroscience, 16(2), 289–300. [PubMed: 15068598]

Obleser J, Wise RJS, Alex Dresner M, & Scott SK (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. Journal of Neuroscience, 27(9), 2283–2289. [PubMed: 17329425]

Oldfield RC (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9(1), 97–113. [PubMed: 5146491]

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, & Chang EF (2012). Reconstructing speech from human auditory cortex. PLoS Biology, 10(1), e1001251. [PubMed: 22303281]

Peramunage D, Blumstein SE, Myers EB, Goldrick M, & Baese-Berk M (2011). Phonological neighborhood effects in spoken word production: An fMRI study. Journal of Cognitive Neuroscience, 23(3), 593–603. [PubMed: 20350185]

Pisoni DB, & Tash J (1974). Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics, 15(2), 285–290 [PubMed: 23226881]

Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, & Gabrieli JD (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. Journal of Cognitive Neuroscience, 13(5), 687–697. [PubMed: 11506664]

Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, & Britton B (2006). An event-related fMRI investigation of phonological–lexical competition. Neuropsychologia, 44(12), 2209–2221. [PubMed: 16842827]

Raettig T, & Kotz SA (2008). Auditory processing of different types of pseudo-words: An event-related fMRI study. Neuroimage, 39(3), 1420–1428. [PubMed: 17981055]

Ranaweera RD, Kwon M, Hu S, Tamer GG, Luh W-M, & Talavage TM (2016). Temporal pattern of acoustic imaging noise asymmetrically modulates activation in the auditory cortex. Hearing Research, 331, 57–68. [PubMed: 26519093]

Righi G, Blumstein SE, Mertus J, & Worden MS (2010). Neural systems underlying lexical competition: An eye tracking and fMRI study. Journal of Cognitive Neuroscience, 22(2), 213–224. [PubMed: 19301991]

Rogers JC, & Davis MH (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. Journal of Cognitive Neuroscience.

Saad ZS, & Reynolds RC (2012). Suma. Neuroimage, 62(2), 768–773. [PubMed: 21945692]

Schwartze M, & Kotz SA (2016). Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. Brain and Language, 161, 28–32. [PubMed: 26362972]

Scott SK, Rosen S, Lang H, & Wise RJ (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. The Journal of the Acoustical Society of America, 120(2), 1075–1083. [PubMed: 16938993]

Talairach J, & Tournoux P (1988). Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging.

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, & Sedivy JC (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 1632–1634.

Thompson-Schill SL, D'Esposito M, Aguirre GK, & Farah MJ (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. Proceedings of the National Academy of Sciences, 94(26), 14792–14797.

Turkeltaub PE, & Branch Coslett H (2010). Localization of sublexical speech perception components. Brain and Language, 114(1), 1–15. [PubMed: 20413149]

Wagner AD, Paré-Blagoev EJ, Clark J, & Poldrack RA (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron, 31(2), 329–338. [PubMed: 11502262]

Wild CJ, Davis MH, & Johnsrude IS (2012). Human auditory cortex is sensitive to the perceived clarity of speech. NeuroImage, 60(2), 1490–1502. [PubMed: 22248574]

Xie X, & Myers EB (2018). LIFG sensitivity to phonetic competition in receptive language processing: A comparison of clear and conversational speech. Journal of Cognitive Neuroscience, 30(3), 267–280. [PubMed: 29160743]

Zatorre RJ, Meyer E, Gjedde A, & Evans AC (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. Cerebral Cortex, 6(1), 21–30. [PubMed: 8670635]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
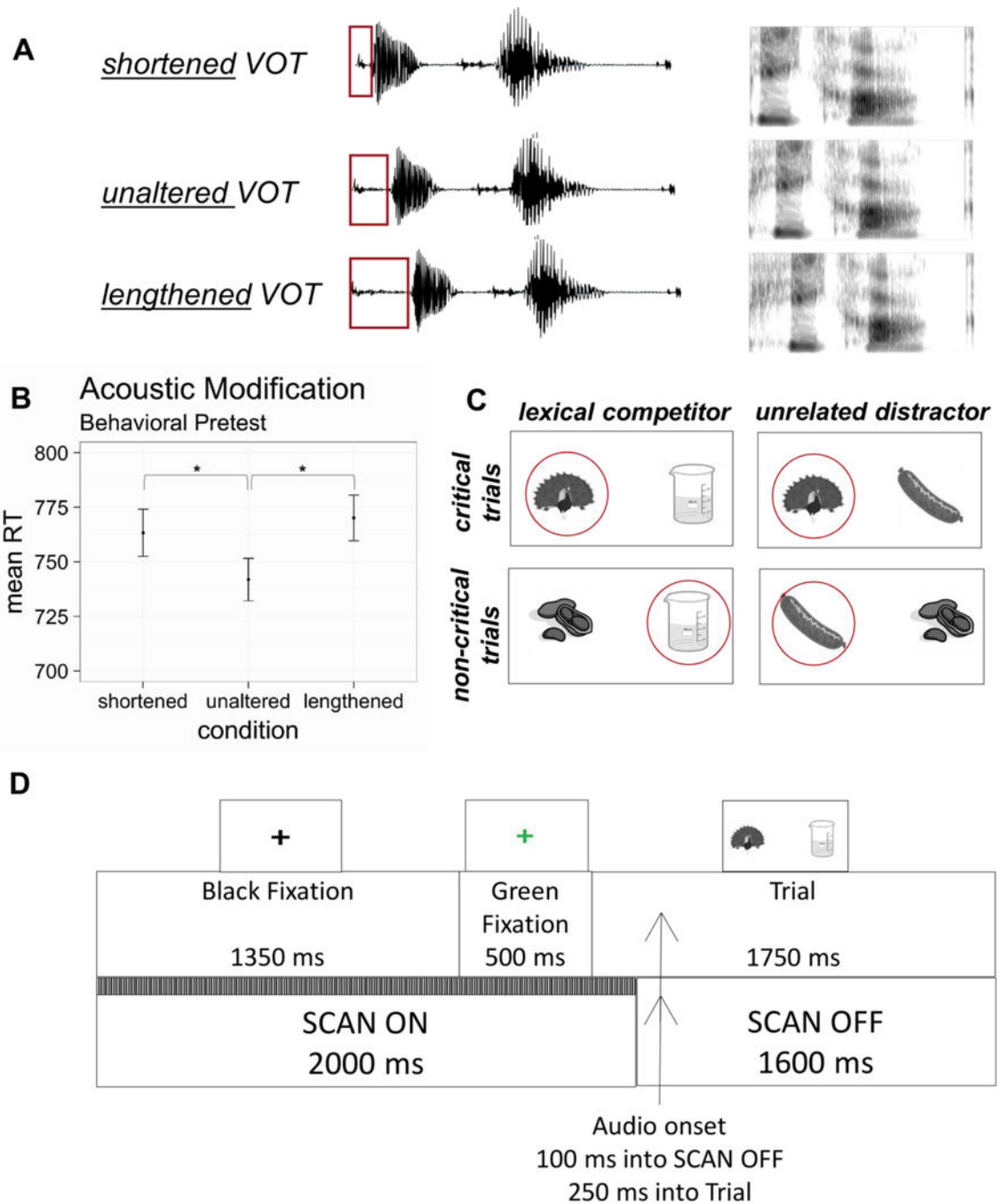
Figure 1.

(A) Three different versions of each critical auditory target (in this example, the stimulus *peacock*) were used in this experiment. In *shortened* tokens, the middle 2/3 of the VOT was excised, and in *lengthened* tokens, the middle 2/3 of the VOT was reduplicated. *(B)* A behavioral pretest indicated that while participants correctly recognized all critical tokens as voiceless, they were significantly slower to categorize modified tokens as voiceless relative to *unaltered* tokens. Error bars indicate standard error of the mean, and the asterisk (*) indicates a significant difference based on Newman-Keuls post-hoc tests. *(C)* A sample set

of displays corresponding to the peacock-beaker-sausage triad, with auditory targets circled in red. On critical trials, the voiceless item served as the auditory target. On non-critical trials, a voiceless item from a different triad (e.g., peanut) served as the distractor picture to avoid repetition of visual displays. *(D)* Schematic of the timing for the fMRI eye tracking experiment.
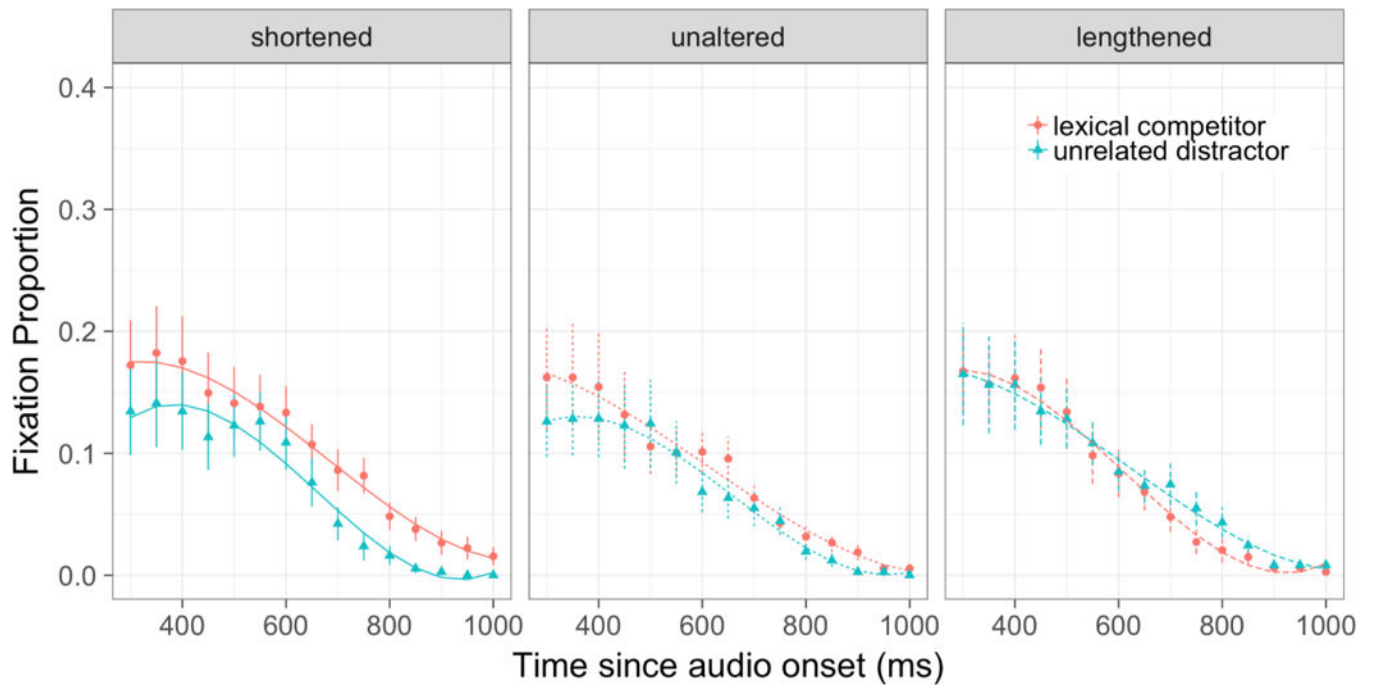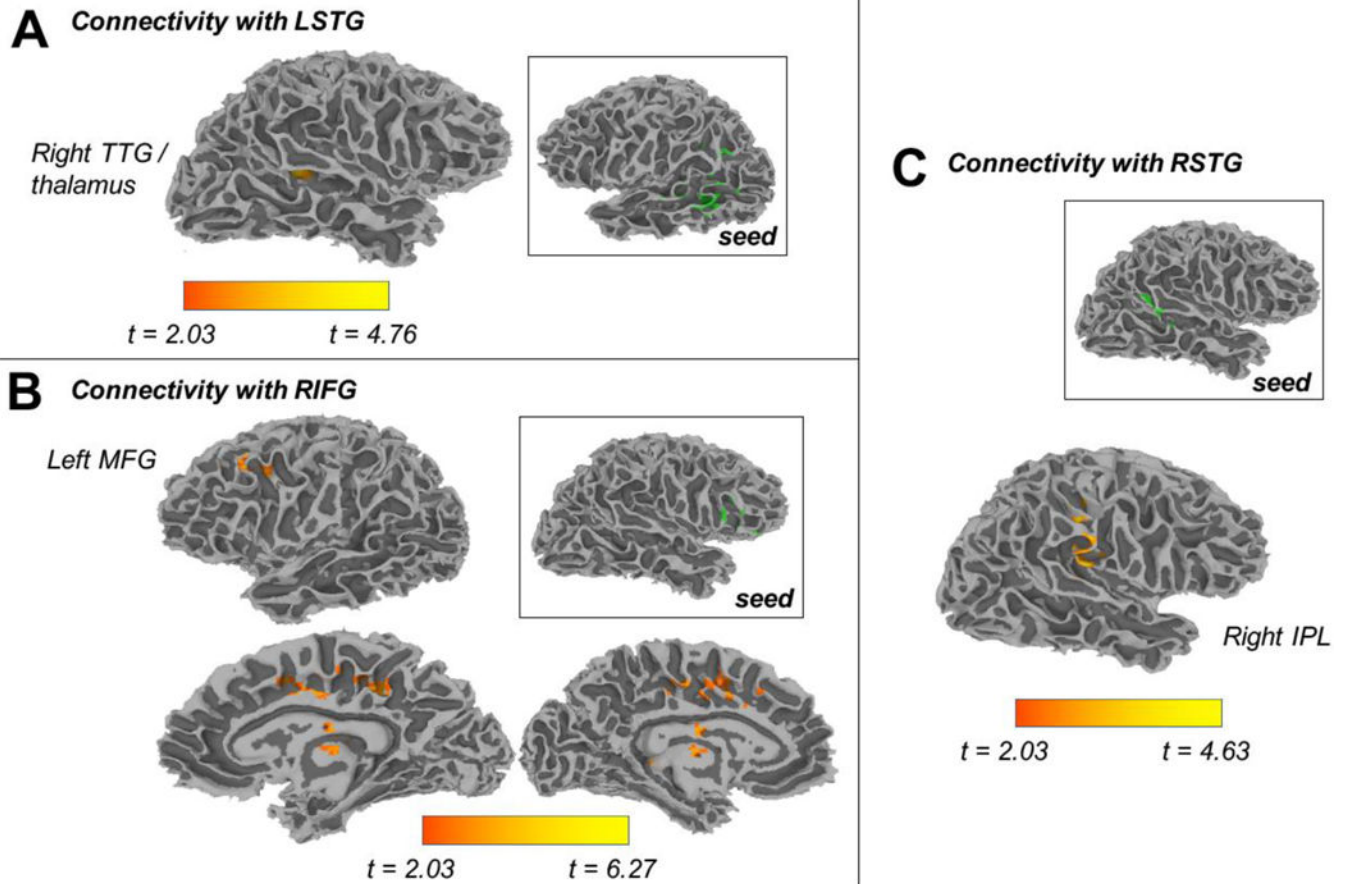
Figure 2.
Time course of looks to distractors for critical trials in fMRI experiment. Lines indicate model fits.
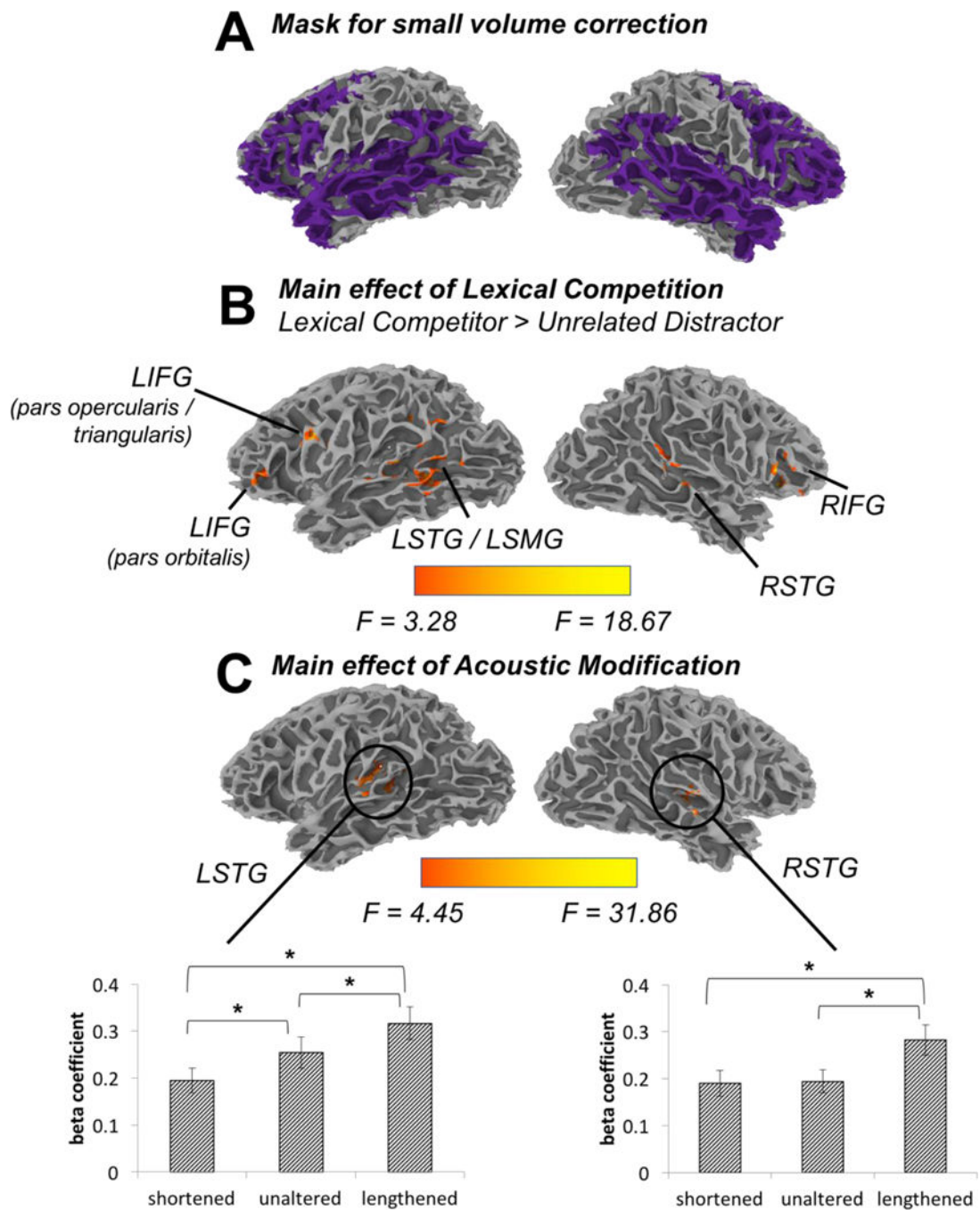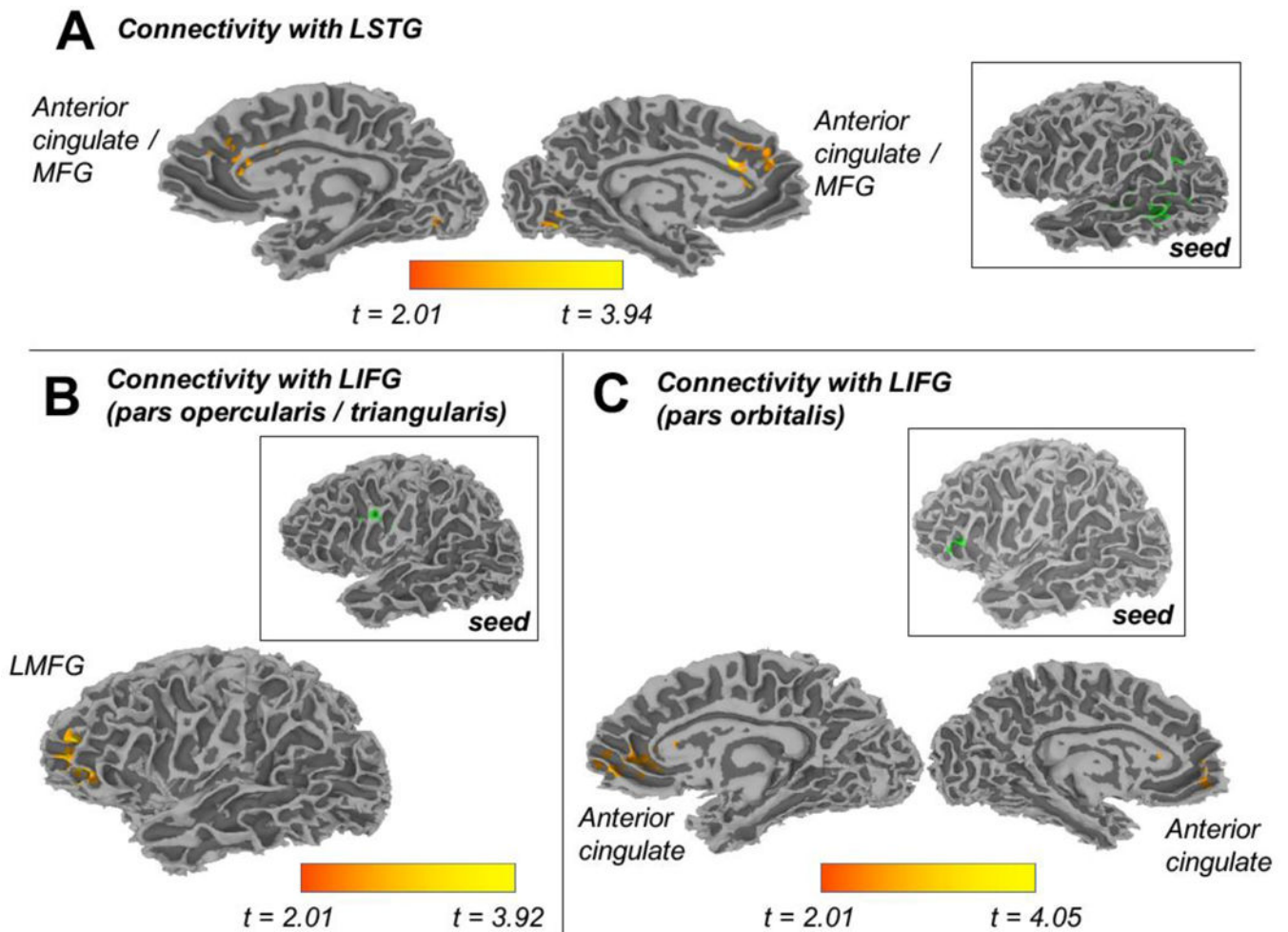
Figure 3.
The univariate analysis was limited to (A) a set of regions typically thought to be involved in language processing. We observed effects of *(B) Lexical Competition* and *(C) Acoustic Modification*. Asterisks indicate significant (Bonferroni-corrected) post-hoc pairwise comparisons. For purposes of visualization (e.g., particularly of clusters extending into sulci), volume-based statistics were registered to a single subject's reconstructed anatomy. Reconstruction was achieved using FreeSurfer, and alignment of the volumetric statistics done with SUMA (Saad & Reynolds, 2012).

Figure 4.
Regions that were sensitive to the *Lexical Competition* manipulation were used as seed regions (shown in green) in a generalized PPI analysis of fMRI data. Seeds and target regions (shown in orange) differed in functional connectivity as a function of phonetic competition.

Figure 5.
Regions that were sensitive to the *Lexical Competition* manipulation were used as seed regions (shown in green) in a generalized PPI analysis of fMRI data. Seeds and target regions (shown in orange) differed in functional connectivity as a function of acoustic goodness of fit. Though not displayed, we also observed increased connectivity between the LSTG seed and two clusters in the cerebellum for unaltered tokens compared to modified (shortened or lengthened) ones.

Table 1.

Details of auditory stimuli. (Mean ± SE)

	Voiceless			Voiced	Unrelated
	Shortened	Unaltered	Lengthened		
Duration	482 ± 11 ms	583 ± 12 ms	603 ± 12 ms	515 ± 12 ms	558 ± 13 ms
VOT	31 ± 1 ms	92 ± 3 ms	153 ± 5 ms	14 ± 1 ms	*n/a*

Table 2.

Significant effects from growth curve analysis of eye fixations to distractor

Fixed effect (Second order model)	Time term (First order model)	β	SE	t	p
Lexical Competition	Intercept	0.013	0.007	2.041	0.041
Lexical Competition × Acoustic Modification (lengthened-unaltered)	Cubic	0.044	0.019	2.363	0.018
Lexical Competition × Acoustic Modification (lengthened-shortened)	Intercept	−0.037	0.016	−2.296	0.022
	Cubic	0.048	0.019	2.559	0.010

Table 3.

Significant clusters in univariate fMRI analysis

Anatomical region	Maximum intensity coordinates			Number of activated voxels	t-value / F-value
	x	y	z		
Main Effect of Lexical Competition *(Onset-Unrelated)*					
1. Left superior temporal gyrus / Left middle temporal gyrus / Left supramarginal gyrus	−53	−47	6	634	4.76
2. Right inferior frontal gyrus / Right middle frontal gyrus	49	21	8	213	5.64
3. Left inferior frontal gyrus (pars opercularis & triangularis) / Left middle frontal gyrus	−43	9	28	179	4.52
4. Right superior temporal gyrus / Right middle temporal gyrus	55	−31	4	178	3.54
5. Left inferior frontal gyrus (pars orbitalis) / Left middle frontal gyrus	−45	35	8	168	3.66
Main Effect of Acoustic Modification					
1. Left insula / Left superior temporal gyrus / Left transverse temporal gyrus	−39	−29	12	335	18.67
2. Right superior temporal gyrus / Right insula / Right transverse temporal gyrus	45	−23	12	243	11.01
Planned Comparison: *Phonetic Competition*					
1. Left insula / Left superior temporal gyrus / Left transverse temporal gyrus	−41	−15	6	346	−5.19
2. Right insula / Right superior temporal gyrus / Right transverse temporal gyrus	43	−17	10	301	−6.99
Planned Comparison: *Goodness of Fit*					
No significant clusters					

Table 4.

Results from the generalized psychophysiological interaction (gPPI) analysis

Seed	Connectivity modulated by phonetic competition					Connectivity modulated by goodness of fit				
	Anatomical region	Maximum intensity coordinates			Number of activated voxels	t–value	Anatomical region	Maximum intensity coordinates			Number of activated voxels	t–value
		x	y	z				x	y	z		
Left superior temporal gyrus (−53, −47, 6) 634 voxels	1. Right thalamus/ Right transverse temporal gyrus	17	−25	12	341	4.52	1. Cerebellum	−11	−67	−10	385	3.65
							2. Bilateral anterior cingulate / Medial frontal gyrus	−1	25	22	379	3.47
							3. Right cerebellum	37	−47	−32	265	3.94
Right inferior frontal gyrus (49, 21, 8) 213 voxels	1. Right thalamus/ Left thalamus	11	−21	12	543	6.27	_No significant clusters_					
	2. Bilateral cingulate gyrus	15	−33	38	462	3.77						
	3. Left middle frontal gyrus (BA 9)	−47	31	32	275	3.75						
Left inferior frontal gyrus (pars opercularis / triangularis) (−43, 9, 28) 179 voxels	_No significant clusters_						1. Left middle frontal gyrus (BA 10) / Left superior frontal gyrus	−27	53	12	288	3.67
Right superior temporal gyrus (55, −31, 4) 178 voxels	1. Right inferior parietal lobule/ Right postcentral gyrus	41	−23	44	539	4.19	_No significant clusters_					
Left inferior frontal gyrus (pars orbitalis) (−45, 35, 8) 168 voxels	_No significant clusters_						1. Bilateral anterior cingulate / Medial frontal gyrus	3	49	0	362	3.66