Brain & Language 218 (2021) 104959

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Sentence predictability modulates cortical response to phonetic ambiguity

Hannah Mechtenberg [a,*], Xin Xie [b], Emily B. Myers [a,c]

[a] Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, Mansfield, CT 06269, USA
[b] Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA
[c] Department of Psychological Sciences, University of Connecticut, Storrs, Mansfield, CT 06269, USA

ARTICLE INFO

Keywords:
Semantic predictability
Speech perception
fMRI

ABSTRACT

Phonetic categories have undefined edges, such that individual tokens that belong to different speech sound categories may occupy the same region in acoustic space. In continuous speech, there are multiple sources of top-down information (e.g., lexical, semantic) that help to resolve the identity of an ambiguous phoneme. Of interest is how these top-down constraints interact with ambiguity at the phonetic level. In the current fMRI study, participants passively listened to sentences that varied in semantic predictability and in the amount of naturally-occurring phonetic competition. The left middle frontal gyrus, angular gyrus, and anterior inferior frontal gyrus were sensitive to both semantic predictability and the degree of phonetic competition. Notably, greater phonetic competition within non-predictive contexts resulted in a negatively-graded neural response. We suggest that uncertainty at the phonetic-acoustic level interacts with uncertainty at the semantic level—perhaps due to a failure of the network to construct a coherent meaning.

1. Introduction

Variability is an intrinsic property of perception—perceptual categories, be they visual objects (e.g., trees/bushes), facial expressions (e.g., anger/fear), or speech sounds (e.g., d/t) tend to partially overlap. The job of the perceiver is to balance variable or probabilistic information in the bottom-up signal with constraints imposed by context or expectation. In spoken language processing it is well established that no two productions of the same phoneme are identical (e.g., Chodroff & Wilson, 2017; Peterson & Barney, 1952), even in intelligible speech. An intriguing question is how sentence context influences a listener's ability to resolve phonetic uncertainty to choose the correct word. Highly-predictive contexts could help guide listeners towards likely lexical candidates and thus assist in resolving potential phonetic ambiguity. Consider the sentence "*he placed the saddle on the horse.*" Context leads listeners to the final word, "*horse,*" and away from phonetic competitors like "*hearse*." On the other hand, in a sentence like "*the bride needed to win the horse*," context insufficiently constrains these two possible interpretations of the final word. In the current study, we examine how ambiguity at the phonetic level interacts with sentence-level semantic predictability. Specifically, we ask whether sentence predictability modulates neural sensitivity to overlapping phonetic categories in continuous, naturally-produced speech.

Speech sound categories, even those belonging to a single talker, overlap substantially. In English, this is especially true for vowel categories by virtue of the language's relatively large vowel inventory (Bradlow, 1995). A single token (e.g., the vowel in "*kit*") might land in an acoustic space also occupied by vowels of other categories (such as those in "*cut,*" "*cot,*" "*cat,*" etc.). In contemporary models of language processing (Davis & Sohoglu, 2019; McClelland & Elman, 1986; Norris & McQueen, 2008), indeterminacy at the phonetic level cascades to the lexical level—meaning that the acoustics for one token (e.g., "*kit*") will also activate a set of partially overlapping alternative words that compete for selection. The prediction that phonetic ambiguity also leads to lexical ambiguity is well supported by behavioral data—even temporary ambiguity at the phonetic level slows access to the intended word, shows evidence of activating competing lexical alternatives, and introduces a processing cost that is observable in physiological measures such as pupil size (e.g., Kuchinsky et al., 2013; McMurray et al., 2002).

This processing cost is also observable in the neural systems that are sensitive to phonetic competition, with increasing activation associated with increasing phonetic ambiguity. These regions include those linked to phonetic processing (the superior temporal gyri, or STG) and those implicated generally in ambiguity or competition resolution (the left inferior frontal gyrus, or LIFG) (Adank, 2012; Davis et al., 2011; Rogers et al., 2017). Both the STG and LIFG show increasing activation when

* Corresponding author at: Department of Speech, Language, and Hearing Sciences, Box 1085, University of Connecticut, Storrs, Mansfield, CT 06269, USA.
E-mail addresses: hannah.mechtenberg@uconn.edu (H. Mechtenberg), xxie13@ur.rochester.edu (X. Xie), emily.myers@uconn.edu (E.B. Myers).

https://doi.org/10.1016/j.bandl.2021.104959
Received 13 July 2020; Received in revised form 2 March 2021; Accepted 9 April 2021
Available online 28 April 2021

listeners categorize digitally-manipulated ambiguous syllables (/da/ vs. /ta/), (Myers, 2007; Myers & Blumstein, 2008), or listen to words that are edited to have a partial phonological overlap with a visually-presented target option (Luthra et al., 2019). These areas also respond to naturally-occurring phonetic competition—prior work from our group found similar regions recruited to process ambiguous phonemes that emerged naturally in continuous speech. Sentences containing more vowel category overlap (sentences with vowels that fell in crowded regions of acoustic space) showed greater activation in LIFG compared to those that contained sounds with less overlap (Xie & Myers, 2018). These findings are consistent with the view that phonetic competition produces cascading effects within the neural systems for speech processing—we ascribe the role of resolving phonetic identity to posterior temporal regions, while mapping to a set of competing lexical items may be handled by the inferior frontal gyrus.

While there is widespread consensus that uncertainty at low levels percolates upwards to higher levels of processing, a longstanding debate concerns the mechanism of top-down feedback on lower levels of processing. Models like TRACE propose direct feedback between the lexical and phonetic levels while models like Shortlist B instantiate the use of top-down information through an offline integration process (McClelland & Elman, 1986; Norris et al., 2016; Norris & McQueen, 2008; Strauss et al., 2007). This debate has animated the field for many years, but what remains fairly uncontroversial is that lexical and semantic information does guide interpretation of the acoustic–phonetic signal, helping to resolve low-level ambiguities. Although the current study does not seek to adjudicate between models of speech perception, considering how feedback passes between levels of processing is relevant to how semantic and phonetic signals interact during receptive listening.

Indeed, phonetic ambiguity typically goes unnoticed by the listener precisely because sounds are embedded within lexical or message-level contexts that disambiguate the signal. For instance, in an eye tracking experiment, listeners heard target words with artificially altered initial phonemes (e.g., "*panda*" sounded more like "*banda*"). Listeners were quicker to access a target picture when distractors had no overlap with the spoken target word (e.g., "*wizard*") compared to when the distractor was a word that was momentarily consistent with the target (e.g., "*bandit*"), suggesting that lexical access is facilitated when the input is strongly consistent with only one possibility (Luthra et al., 2019). Similarly, Rogers et al. (2017) found increased activation in the left inferior frontal gyrus (LIFG) when phonetic ambiguity also led to lexical ambiguity (e.g., a blend between "*blade*" and "*glade*") but not when lexical information could resolve that ambiguity (e.g., in a blend between "*bone*" and "*ghone*", "*bone*" is the only likely resolution), indicating that lexical information acts quickly and efficiently to decrease the processing penalty for phonetic ambiguity.

Top-down effects are not limited to the lexicon. While computationally-instantiated interactive frameworks of receptive language (McClelland & Elman, 1986; Davis & Sohoglu, 2019) do not explicitly include a "message-level" node, there is a tacit assumption that other top-down sources of information help constrain the set of lexical options. Within this system, a coherent semantic context will activate of a group of likely lexical candidates which will in turn boost activation of their constituent phonemes, consistent with evidence that sentence context biases perception of ambiguous phonemes towards the more sensible alternative. For example, a word ambiguous between "*goat*" and "*coat*" will more likely be heard as "*goat*" when embedded in a sentence like "*he milked the*…" (Borsky et al., 1998), and the effect of this shifted phonetic category boundary can be seen early in the processing stream within the superior temporal lobe (Guediche et al., 2013). Beyond resolving lexical ambiguity, coherent sentence contexts rescue noise-obscured speech (Kalikow et al., 1977; Miller et al., 1951), suggesting that message-level information helps guide bottom-up perceptual processes. For instance, Obleser et al. (2007) systematically manipulated sentence predictability (high vs. low) as well as the degree of acoustic signal integrity while listeners processed vocoded speech. There was a substantial boost to identification accuracy for high-predictability sentences passed through an 8-band noise-vocoding routine, while low-predictability sentences masked with the same routine were perceived at chance. The effect of context not only helps in noisy environments, but also in perceiving reduced wordforms that commonly populate conversational speech (Ernestus et al., 2002).

Using semantic prediction to facilitate comprehension of noisy or degraded speech involves neural networks implicated in semantics as well as those associated with recruitment of domain-general resources (Obleser et al., 2007; Rysop et al., 2021; Vaden et al., 2017). Rysop and colleagues parametrically manipulated noise levels for semantically predictable and unpredictable sentences, calibrating the noise level to individual participant's speech reception threshold. They found that semantic predictability differentially drove activation in angular gyrus, supramarginal gyrus and posterior middle temporal gyrus. Notably, the activation difference between high and low semantic predictability was most evident at medium levels of noise—suggesting that the effects of predictability are not linear across the range of noise. This nonlinearity effect across noise levels is reflected in behavioral data; the effects of semantic constraint are greatest when noise levels are challenging, but not impossible (see also Obleser et al., 2007). The angular gyrus has been specifically implicated in the integration of semantic information with speech obscured by noise (Obleser et al., 2007; Obleser & Kotz, 2010; Rysop et al., 2021). Similarly, the left (and often right) STG respond to the intelligibility of the signal in noisy or degraded speech, a response which, especially in the posterior portions of the left STG, is modulated by semantic constraint. Although differing in their details, studies have shown that under conditions of higher constraint, the effect of signal degradation is dampened in the STG—suggesting that high semantic constraint lightens the burden on lower-level acoustic–phonetic processing.

Of note, prior studies examining the neural basis of semantic constraint on speech perception manipulated signal quality writ large, resulting in a global degradation of the acoustic signal and reduction of intelligibility. Of interest is whether the same networks emerge when listeners are confronted with natural phonetic variability that increases or decreases phonetic competition without impacting intelligibility. In the current study, we ask whether coherent semantic context diminishes the processing penalty for phonetic ambiguity, since less is known about the neural architecture underlying sentence context effects on phonetic competition resolution. As described above, prior work from our lab showed that when exposed to sentences varying naturally in the degree of phonetic competition, listeners showed increased recruitment of inferior frontal and posterior tempoparietal areas as phonetic ambiguity increased (Xie & Myers, 2018). However, by design, the stimuli were nonsensical (e.g., "*The trout is straight and also writes brass*"), and as such contained no semantic cues that could limit potential lexical alternatives. An open question is whether phonetic competition effects persist within semantically meaningful sentences, or whether sentence predictability constrains lexical, and therefore phonetic, interpretation to the extent that naturally-occurring phonetic variation poses no processing cost. To our knowledge, the question of whether naturally-occurring phonetic variation within semantically constrained sentences taxes the language processing system has yet to be explored. To probe this question, we varied the degree of word predictability based on surrounding sentence context. Theoretically, activation of multiple lexical alternatives due to bottom-up phonetic ambiguities might be observed in the brain regardless of whether lexical possibilities are constrained by semantic context (Forster, 1981). Conversely, interactive views predict that the degree of neural sensitivity to phonetic competition will ultimately depend on the amount of semantic constraint (McClelland et al., 2014; Mirman et al., 2005). Within the scope of this study, it may be that phonetic competition effects disappear within highly-predictive contexts, while persisting in non-predictive contexts.

To address this question, participants passively listened to sentences that unfolded either predictively or non-predictively. To control for lexical competition driven by competition at the acoustic-phonetic level (Luce & Pisoni, 1998), identical content words were used in both highly-predictive and non-predictive sentences (albeit in different orders and combinations). Phonetic competition was also equated across sentence types. At issue is whether sensitivity to phonetic competition persists when semantic context constrains the number of possible lexical alternatives (and thus also constrains the phonetic interpretations of ambiguous phonemes). As such, we anticipate a reduction in neural sensitivity to phonetic competition in highly predictive sentences. However, in non-predictive sentences, we expect to find similar neural regions to Xie and Myers (2018)—LIFG and left temporoparietal areas—that respond to phonetic competition and that the weak semantic coherence in non-predictive sentences will drive positively-graded activation. By minimizing differences between sentences to isolate semantic predictability, we can specifically investigate how phonetic competition is processed depending on the availability of top-down information to constrain lexical selection.

2. Methods

2.1. Stimuli

2.1.1. Norming: predictability

Sentences that varied in their semantic predictability were adapted from Kalikow et al. (1977) and Bradlow and Alexander (2007). In a series of studies, hosted through Amazon's Mechanical Turk crowd-sourcing platform, we normed the predictability of key words in each sentence. Using a Cloze procedure presented in Qualtrics, participants were instructed to fill in the blank with the first word that came to mind given the rest of the sentence text, e.g., "The soccer player _ a goal." The position of the omitted content word in the sentence was counter-balanced across participants, and no participant saw a sentence more than once. In spoken language processing, listeners only have the prior sentence context available to judge the predictability of the upcoming input (e.g., "The soccer __…"). We elected to measure the predictability of individual key words given the entire context (both before and after the key word) because the temporal resolution of fMRI prevents definitive separation of incremental context processing from the wrap-up effects of the predictability of the entire sentence. Our approach of using the full sentence context and one missing word gives us a more holistic measure of the predictability of each sentence.

Adults ($n = 225$) between the ages of 18 and 45 were recruited from Amazon's Mechanical Turk. All participants were located in the United States, indicated that they were native speakers of North American English, and had not completed the task in a previous session. Thirteen participants were excluded for not following instructions (filling in multi-word phrases or obvious nonwords) or for failing to complete the entire task. After exclusions, $n = 212$ participants contributed to all subsequent analyses (82 females, 130 males; mean age = 32, SD = 5.9).

A preliminary set of sentences (75 each highly-predictive and non-predictive) were normed with 18 participant responses for each Cloze position (range of 2–4 positions per sentence). Sentences in the highly-predictive category were culled if the final word was predicted less than 30% of the time (i.e., fewer than 5/18 participants guessed the intended word). After culling, 65 highly-predictive sentences remained. We created 65 non-predictive sentences to match the number of highly-predictive sentences. To equate lexical frequency and phonological neighborhood density across highly-predictive and non-predictive sentence sets, a subset of the non-predictive sentences was rearranged to maintain the content words present in the final highly-predictive set, such that the collection of content words was identical in highly-predictive and non-predictive sentences. The resulting non-predictive sentences were normed with 10 participant responses at each Cloze position. The percentage predictability was capped at ≤20% correct (2/10 participants) at the final position to be considered sufficiently non-predictive.

Predictability was analyzed at two levels: global (mean predictability of content words across the entire sentence) and only at the final word (see Fig. 1 for an example). A two-sample t-test confirmed a statistically significant difference between the two sentence sets at both levels. Globally, content words in highly-predictive contexts were guessed with greater frequency than those same content words in non-predictive contexts (average proportion correct highly vs. non: 0.54 vs. 0.02 respectively; $t(128) = 23.21$, $p < 0.001$; see Supplementary Materials). An identical pattern appeared when only assessing the predictability of the final word (0.72 for highly-predictive vs. 0.01 for non-predictive; $t(128) = 26.61$, $p < 0.001$).

This set of 65 highly-predictive and 65 non-predictive sentences were presented during the MRI session as critical trials while 14 additional sentences served as catch trials (seven each highly-predictive and non-predictive). The last author, a female native speaker of North American English, produced each sentence a total of six times. Recording occurred in a sound-isolated room with a microphone and digital recorder that sampled at 44.1 kHz. Final tokens were selected based on natural prosody and clarity of pronunciation. Stimuli were individually normalized to 70 dB root mean square amplitude. Acoustic analyses were conducted in Praat (Boersma & Weenink, 2018) for the 130 critical sentences.

2.1.2. Acoustic measures

Acoustic measures included the mean and standard deviation of pitch (F0) and duration. These measures were statistically equivalent across highly and non-predictive sentences (see Supplementary Materials). Non-predictive sentences were marginally longer than highly predictive sentences (high vs. non: 1812 vs. 1879 ms, $t(128) = -1.94$, $p = 0.05$), and the range in duration for all sentences was between 1406 and 2457 ms. There was no difference in mean pitch ($t(128) = 1.89$, $p = 0.06$) nor variation in the standard deviation of F0 ($t(128) = 1.71$, $p = 0.09$).

2.1.3. Vowel properties

To assess the degree of sentence-by-sentence phonetic competition, we followed procedures in Xie and Myers (2018) to analyze the acoustics of all stressed vowels. Vowel boundaries were identified in a first pass using the Penn Forced Aligner (Yuan & Liberman, 2008). The first author then manually adjusted the output boundaries to ensure full capture of each stressed vowel. The midpoints of F1 and F2 were extracted using GSU Praat Tools (Owren, 2008). We chose to use the midpoint values of F1 and F2 for monophthong as well as for diphthong vowels to fairly represent all vowel types present in the stimuli and to compare vowels along the same metric.

Notably, the mean and standard deviation of F1 and F2 of each vowel type did not differ across highly-predictive and non-predictive sentence sets (see Supplementary Materials). One sentence containing the vowel category /ɔɪ/ was omitted from all analyses (acoustic and fMRI), as it

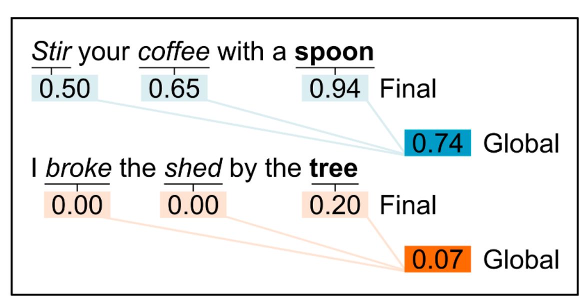

Fig. 1. Examples of the predictability norming Cloze test. Predictability was evaluated using proportion correct at the final word and across the whole sentence (global). For each sentence, only one key word was blanked for each participant. **Top**: highly-predictive sentence, **Bottom:** non-predictive sentence.

only appeared in a single instance ("lawyer" in the highly-predictive set).

We estimated the amount of phonetic competition for each vowel using procedures established in previous publications (Wright, 2004). Put simply, if a particular vowel token was only surrounded by tokens that belong to the same category, that vowel token would have a low phonetic competition value. Conversely, if a vowel token occupies an acoustic space crowded by vowels of different categories, that token would have a high value of phonetic competition. The average of the inverse squared distances from a given vowel token to every other vowel token belonging to a different phonetic category was calculated for each stressed vowel. The resulting values were then log-transformed. To visualize this more intuitively, Fig. 2B applies a blue-to-red gradient for low-to-high phonetic competition. A vowel token with a blue shading has a relatively lower degree of phonetic competition than a token shaded in red.

2.1.4. Norming: Intelligibility

To ensure high intelligibility across both sets of sentences, 10 native English speakers (females = 9, males = 1) transcribed all 144 sentences (130 critical and 14 catch sentences). These 10 participants did not participate in either the predictability norming or the main fMRI experiment. Assessment of transcription accuracy of content words between highly and non-predictive sets confirmed no difference in intelligibility (high versus non: 93.8% (SD = 0.24), 95.4% (SD = 0.21); $t(128) = -0.39$, $p = 0.7$).

2.2. Participants: fMRI

Twenty-four adults (21–36 years of age, females = 15, males = 9) were recruited from the University of Connecticut community. All indicated that they were right-handed, native monolingual speakers of North American English, and had no hearing or vision deficits. One female participant was excluded due to excessive motion in the scanner, resulting in $n = 23$ for all further analyses. All participants provided written consent per the guidelines by the University of Connecticut's Institutional Review Board. After obtaining written consent, all participants were screened for MRI safety (no ferromagnetic materials). Participants were paid for their time and debriefed after completion of the fMRI task.

2.3. fMRI design and procedure

Before entering the scanner, participants were told that they were going to listen to sentences through headphones, and that occasionally a word would appear on the screen. Participants held an MRI-compatible button box in one hand and were instructed to press the button under their index finger if they heard the word in the previous sentence, or the button under their middle finger if they did not hear the word in the previous sentence. To ensure that participants fully understood the task instructions, they completed four practice trials (taken from the set of 14 catch sentences) during acquisition of the anatomical scan. Participants were also told to remain still and to keep their eyes open. Accuracy on catch trials were analyzed post-scan to confirm that all participants responded appropriately. All stimuli were presented using OpenSesame v3.2.8 (Mathôt et al., 2012).

The fMRI experiment consisted of five runs presented in a fixed order across all participants. Trials within each run were presented in a fixed and pseudorandom order, which was determined using the OptSeq2 tool (https://surfer.nmr.mgh.harvard.edu/optseq/). Each run had 13 highly-predictive trials, 13 non-predictive trials, and two catch trials; for a total of 28 stimulus presentations per run (see Fig. 2D for schematic). No content words were repeated within a run. For the catch trials, the probe word was in the sentence 50% of the time. Catch trials were modeled into the participant-level regressions but not analyzed at the group-level.

Trials were presented at SOAs ranging from 4 to 16 s, in multiples of 4 s, with a total of 84 volumes per run. All auditory stimuli were delivered through MRI-compatible headphones (Avotech Silent Scan SS-3300, Stuart, FL) and responses for the catch trials were recorded with an MRI-compatible button box (Current Designs, 932, Philadelphia, PA).

2.4. fMRI acquisition

A 3-T Siemens Prisma scanner (Erlanger, Germany) collected anatomical and functional MRI data. A multiecho magnetization prepared rapid gradient echo sequence (MPRAGE: repetition time [TR] = 2400 ms, echo time = 2.98 ms, inversion time = 1000 ms, 0.8-mm^3

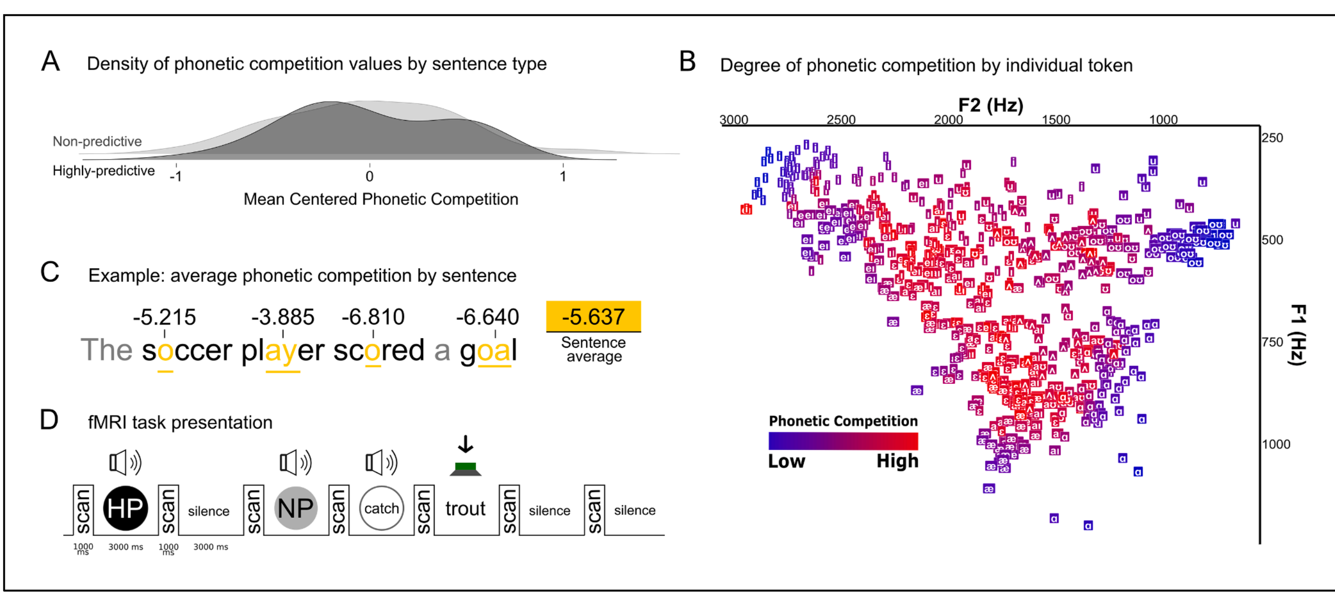

Fig. 2. Methods. (A) Density plot for mean-centered phonetic competition values, plotted by sentence type. Grey is non-predictive and black is highly-predictive. (B) Vowel space graph for stressed vowels in the stimuli set. Color scaling indicates degree of phonetic competition for each token. Blue is low while red is high. (C) Extraction of sentence-level values of phonetic competition. (D) fMRI paradigm schematic. Depiction of sentence presentation in between EPI scans. Only catch trials required a button press.

isotropic voxels, 300 × 320 matrix) was used to acquire the high resolution 3-D T1-weighted anatomical images, reconstructed into 208 slices. The functional EPIs were collected in a rapid, sparse sampling design, with the functional volumes acquired in 1000 ms and followed by 3000 ms of silence where the auditory stimuli were played (effective TR = 4000 ms). All auditory stimuli began 254 ms into the silent gap between scans. Functional EPIs were collected in an ascending, interleaved order with an accelerated multiband sequence (52 slices, 2.5-mm thick, 2 mm^2 axial in-plane resolution, 110 × 110 matrix, 220 mm^3 field of view, flip angle = 62).

2.5. fMRI data analysis

Functional and anatomical fMRI images were analyzed using AFNI (Cox, 1996). Preprocessing consisted of transforming the images from an oblique to cardinal orientation, then correcting for motion using a six-parameter rigid body transform that were then aligned to each participant's reconstructed anatomical images. This was followed by normalization to Talairach space (Talairach & Tournoux, 1988) and spatial smoothing with a 4-mm Gaussian kernel. All motion and signal fluctuation outliers were eliminated following standard procedures. Individual participant masks were created using their respective anatomical data to restrict functional data to voxels located within the brain. Those individual masks were then combined to create a group-level mask with overlapping voxels for at least 21 out of 23 participants.

We created three time series vectors for each participant: highly-predictive trials, non-predictive trials, and catch trials. Two participants had a single accidental button press during a non-catch trial; these instances were moved to the catch trial time series file. The vectors were aligned to stimulus onset for each sentence type (highly-predictive, non-predictive, and catch trials). Each vector was convolved with a canonical gamma HRF. The three condition vectors were regressed with six nuisance movement parameters (generated during preprocessing) which resulted in by-voxel fit coefficients for each condition, for each participant.

A second analysis investigated the interaction between phonetic competition and semantic predictability (for distribution, see Fig. 2A). The participant-level regression was completed with an amplitude-modulated approach in AFNI (using the -stim_times_AM2 flag in *3dDeconvolve*). Sentence-level measures of phonetic competition (see Fig. 2C) were added as a second regressor to the time series vectors as described above (excluding catch trials). Additionally, sentence-level values for lexical frequency ($SUBTLWF_{US}$, Balota et al., 2007) and phonological neighborhood density from the IPhOD (Vaden et al., 2009) were included as third and fourth regressors to model out their influence. All continuous regressors were log-transformed, values for content words were averaged across each sentence, then mean-centered. Convolution was done with the stereotypical gamma HRF, and the same six nuisance regressors as described above were also included. We generated amplitude-modulated by-voxel fit coefficients for each participant for both conditions of interest.

For each analysis, we performed group-level comparisons with an ANOVA (using *3dANOVA2*, AFNI). The first was an estimation of the main effect of predictability (highly-predictive versus non-predictive). Highly-predictive and non-predictive beta coefficients were also compared to an implicit baseline. The second group-level analysis searched for the hypothesized interaction between sentence predictability and phonetic competition. The outputs of both group-level analyses were convolved with a small-volume corrected group mask that was constrained with the following bilateral anatomically-defined language regions: angular gyrus, superior parietal lobule, inferior parietal lobule, supramarginal gyrus, middle temporal gyrus, Heschl's gyrus, superior temporal gyrus, insula, middle frontal gyrus, superior frontal gyrus, and inferior frontal gyrus (see Fig. 3C). Outputs were also subject to cluster thresholding, determined by running 10,000 Monte Carlo iterations on the small-volume corrected group mask. The -acf flag in *3dFWHMx* and *3dClustSim* in AFNI estimated spatial smoothness and generated the voxel- and cluster-level thresholds to minimize instances of false-positives in the fMRI data. Data thresholds were set at a corrected threshold of $p < 0.05$ (voxelwise threshold of $p < 0.05$, 2-sided thresholding, 274 contiguous voxels).

3. Results

3.1. Effects of predictability

The comparison of non-predictive trials with highly-predictive trials revealed a large cluster in the left superior temporal gyrus (STG) extending from anterior regions to the temporoparietal junction (see Fig. 3A, Table 1). All regions showed greater activity during non-predictive trials than in highly-predictive trials.

3.2. Interactions between phonetic competition and predictability

To test our primary question about the interaction between phonetic competition and semantic predictability, we first looked at the effect of phonetic competition in non-predictive sentences compared to in highly-predictive sentences. Lexical frequency and phonological neighborhood density were also included as regressors to control for their potential influence. Three regions showed differences in their response to phonetic competition as a function of semantic predictability: left angular gyrus (AG) extending into the superior portion of the left posterior middle temporal gyrus (MTG), the left inferior middle frontal gyrus (LMFG), and the pars orbitalis region of the left inferior frontal

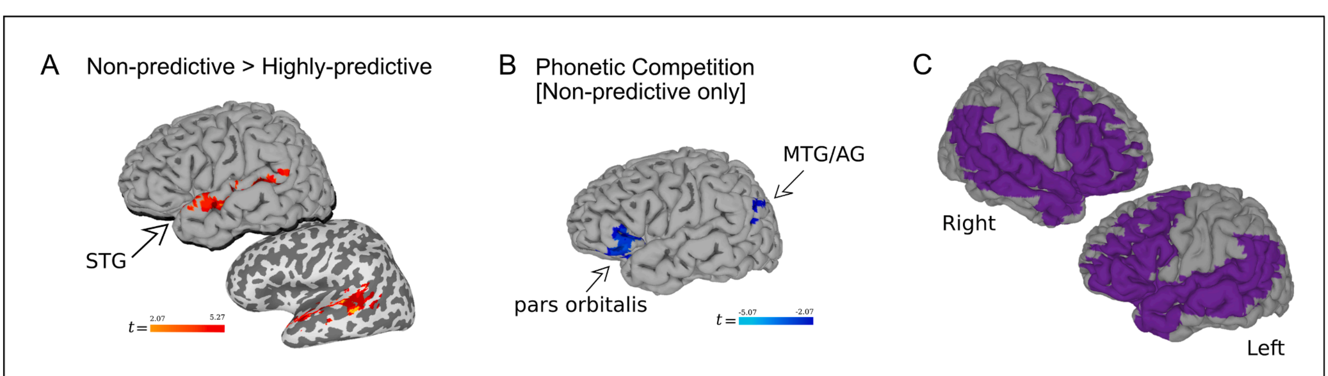

Fig. 3. (A) Non-predictive versus highly-predictive trials. Clusters corrected at $p < 0.05$ (voxelwise $p < 0.05$, 274 contiguous voxels). (B) Results of amplitude-modulated phonetic competition in non-predictive trials compared to implicit baseline. All regions show a negative correlation with variability in phonetic competition. (C) Language regions used for the small-volume group mask.

Table 1
Results of fMRI ANOVAs.

Area	*Cluster Size in Voxels*	*Maximum Intensity Coordinates*			*Maximum t Value*
		x	y	z	
Non-predictive > highly predictive[a]					
Left STG	784	−53	−5	−4	5.27
Interaction of phonetic competition with sentence predictability.[b]					
Left angular gyrus (AG), left posterior MTG	498	−47	−67	32	2.35
Left inferior MFG	289	−47	41	14	−2.66
Left IFG, pars orbitalis	278	−43	25	−10	2.57
Effect of phonetic competition: non-predictive sentences[b]					
Left posterior MTG, left angular gyrus (AG)	393	−39	−65	24	−5.07
Left IFG, pars orbitalis	401	−47	−25	0	−3.93

[a] Clusters corrected at $p < 0.05$ (voxel level $p < 0.05$, 274 contiguous voxels).
[b] Results of the amplitude-modulated analysis. The above clusters correlated significantly with by-trial phonetic competition beyond what could be attributed to event timing. Variability attributed to lexical frequency and phonological neighborhood density were also regressed out. Cluster size of 274 contiguous voxels at a voxelwise threshold of $p < 0.05$ yields a cluster-level threshold of $p < 0.05$.

gyrus (LIFG) (see Fig. 4A, Table 1). Since we found an interaction between phonetic competition and semantic predictability, we then looked at the effect of by-trial variability in phonetic competition within non-predictive and highly-predictive trials separately. We found that there were no significant clusters correlating with phonetic competition in highly-predictive contexts. However, for phonetic competition in non-predictive sentences we saw similar regions as found in the initial interaction (Fig. 3B), but the left middle frontal gyrus cluster no longer emerged. Notably, by-trial phonetic competition showed a negative correlation with activity in the left AG/MTG and anterior LIFG clusters.

To characterize the interaction between phonetic competition and sentence predictability discovered in the amplitude modulated regression, we extracted by-trial beta weights for every participant within the three ROIs that emerged in the omnibus analysis. By-trial beta estimates were then averaged across participants and plotted by condition (see Fig. 4B). We analyzed the correlation between by-item phonetic competition and beta estimates separately for each cluster using the *cor* function within R Studio (R Core Team, 2019). Two general patterns emerged: in left AG/MTG and left IFG, increasing phonetic competition was associated with increased activation in high-predictive sentences but decreasing activation in non-predictive sentences. The opposite relationship was seen in the left MFG, with a non-significant relationship within highly predictive sentences, but a positive relationship to phonetic competition within non-predictive sentences.

4. Discussion

Spoken language is rife with phonetic ambiguity. Despite this, listeners deftly juggle multiple sources of probabilistic information (phonetic, lexical, semantic) to arrive at a coherent message. Although the process of resolving competition between speech sounds is seemingly automatic during passive listening, there are different processing demands if those ambiguous sounds are heard in semantically-predictive compared to unpredictive contexts. If listeners use predictable semantic context to constrain the possible set of lexical alternatives, this in turn reduces uncertainty at the phonetic level. Consistent with this assertion, in a whole-brain analysis considering only sentences with high semantic constraint, no regions showed sensitivity to phonetic competition, suggesting that top-down influence on lexical selection reduces the processing cost incurred by phonetic category overlap. When the input is highly predictive, listeners need less support from the incoming signal, and ambiguities in the signal can be resolved quickly. Strikingly, when sentences were not predictive, increased phonetic competition reduced activation levels in left IFG (pars orbitalis) and left AG/MTG, showing less activation when phonetic competition effects were high. At the same time, one cluster in left MFG showed the reverse pattern, with activation increasing in proportion to phonetic competition for low-predictability sentences, a finding that replicates prior work (Xie & Myers, 2018). We discuss the implications of these findings below in the context of other work on interactions between bottom-up and top-down aspects of language processing.

4.1. *Left superior temporal gyrus responds to semantic predictability*

Even during passive listening and in highly intelligible speech, sentence predictability modulates activation. Specifically, we found increased activity in left STG for semantically non-predictive contexts compared to highly-predictive contexts. Other work routinely reports increased recruitment of left STG for processing semantically unrelated cue-target words in semantic priming paradigms and for auditory sentence processing after presentation of a nonword cue (Minicucci et al., 2013; Rissman et al., 2003). In both studies, participants performed a lexical decision task on the target word of single word prime-target pairs. While our task did not require explicit decision making, we propose that there are similar underlying neural processes for assessing semantic relatedness across word pairs as for whole sentences. Indeed, there is evidence that suggests that listening to semantically coherent vs. anomalous sentences drives activity in left STG, with greater activation for sentences with semantic anomalies (Friederici et al., 2003). Although our sentences were all semantically coherent, the content of non-predictive sentences was far more unusual than the highly-predictive sentences—it is perhaps unsurprising to see similar neural patterns for perceiving semantically disjointed stimuli.

Left STG is also associated with processing novel auditory stimuli. Recruitment of STG is observed for listening to novel auditory tones compared to nontarget stimuli (Kiehl et al., 2001; Müller et al., 2002). This phenomenon is paralleled in the speech domain, such that exposure to novel nonwords correlates with increased activity in left STG (Rauschecker et al., 2008). In our study, sensitivity of the STG to auditory novelty could extend to processing non-predictive sentences. For example, the sentence "*we sat in the snow by the pool*" is likely a more novel message compared to "*swimming is common at the pool.*" Predictability in semantically coherent stimuli also tends to recruit prefrontal regions (Badre et al., 2005; Copland et al., 2003; Zhu et al., 2012), which is notably absent in the analysis contrasting predictable with non-predictable sentences. Prefrontal cortex is thought to direct goal or task-related behaviors (see Alvarez & Emory, 2006; Cabeza & Nyberg, 2000, for reviews) and its recruitment is most commonly seen in speech processing tasks that require metalinguistic lexical or semantic judgements, which may not reflect the processing demands of typical phonological comprehension (Hickok & Poeppel, 2004). Our task did not require listeners to make explicit semantic or metalinguistic judgements, thus minimizing task-induced involvement of frontal regions. As we detail below, LIFG activation only emerged when considering the joint effects of semantic predictability and phonetic competition, leaving it more likely that this region integrates across multiple levels of uncertainty in the signal.

4.2. *Highly predictive contexts reduce processing cost of phonetic ambiguity*

While Xie and Myers (2018) showed that a network of regions responded to phonetic competition in nonsense sentences, of interest is whether those effects persist when listeners can engage normal comprehension processes once words are embedded within semantically meaningful sentence contexts. One region, the left MFG, showed an

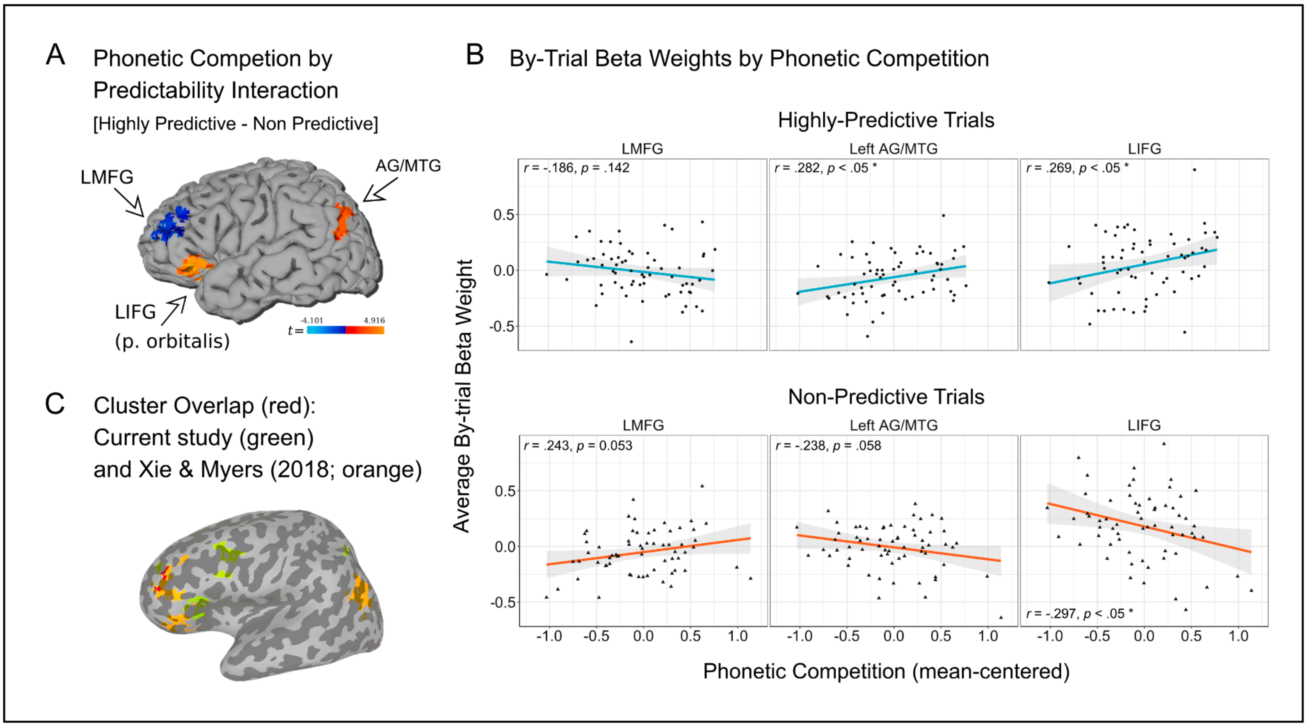

Fig. 4. (A) Interaction of amplitude-modulated regression of PC in highly-predictive versus non-predictive trials. Cluster corrected at p < 0.05, 274 contiguous voxels). (B) Graph showing correlation of average beta-weights for each trial by mean-centered phonetic competition. Faceted by cluster (defined by the amplitude modulated interaction analysis), and by stimulus predictability (blue for highly-predictive, red for non-predictive). Pearson correlation coefficients and significance reported for each panel. (C) Cluster overlap comparison for Xie and Myers (2018) and the current study. Overlap between clusters indicated in red—60 voxels in the MFG. No other clusters overlapped.

interaction between semantic constraint and phonetic ambiguity resulting in increasing activation as a function of phonetic competition—a pattern that replicates findings from our prior work (Xie & Myers, 2018). Extending the interpretation from that study, we suggest that this response reflects the cascade of phonetic ambiguity to the lexical level—recruiting regions involved in lexical selection.

The MFG cluster was an exception from the general trend, namely that sensitivity to phonetic competition was greatly diminished in highly-predictive sentences. When we considered sentence-by-sentence variability of phonetic competition in highly-predictive (and necessarily highly semantically coherent) sentences, we found no clusters that significantly responded to phonetic competition. We cannot simply attribute these results to a lack of attention to the stimuli—participants' high degree of accuracy on catch trials suggests continued engagement in the task. However, if strong message-level context sufficiently activates lexical items, this may diminish attention to the acoustic signal or decrease reliance on phonetic-level competition resolution (Gaston & Marantz, 2018). For instance, when listening to the sentence, "*For your birthday, I baked a* ___" the final word is strongly predicted by the preceding context (i.e., "*cake*"). If phonetic competition in the bottom-up signal leads listeners to some confusion about the final word (e.g., "*cake*" or "*kick*"), this competition will quickly be resolved by context. Beyond accurate anticipation of a single word at the end of a phrase, the entire message is also highly internally coherent[1]. Thus, small phonetic ambiguities likely do not affect the rapid and accurate lexical selection that occurs during continuous speech perception.

The assertion that strong semantic cues can overcome degraded or ambiguous acoustic–phonetic signals is a highly consistent effect across neuroimaging and behavioral studies. Speech-in-noise is more intelligible when embedded in highly-predictive contexts compared to semantically unpredictive or anomalous contexts (Boothroyd & Nittrouer, 1988; Bradlow & Alexander, 2007; Giolas et al., 1970; Kalikow et al., 1977; Miller et al., 1951). In addition to the benefit of semantic information for speech-in-noise, there are also reports of increased intelligibility for synthesized or accented speech when listeners are provided with broader conversational discourse (Drager & Reichle, 2001) as well as from context gleaned from a single sentence (Behrman & Akhund, 2013; Bradlow & Alexander, 2007; Clopper, 2012). While the current study only included highly-intelligible sentences presented in silence, it is clear that semantic context helps to disambiguate speech sounds at multiple levels.

4.3. *Phonetic ambiguity weakens top-down influence in non-predictive contexts*

There were robust effects of phonetic competition within non-predictive sentences. Surprisingly, unlike prior findings from our lab (Xie & Myers, 2018), these results show a *negatively*-graded relationship between by-sentence measures of phonetic competition and activity in the pars orbitalis of the LIFG and left AG/MTG. While at odds with results from Xie and Myers, these data suggest that a flexible cortical network rapidly integrates sentence-level context with the incoming acoustic–phonetic signal to select lexical items.

Broadly, these results are consistent with spreading activation models of language processing (McClelland & Elman, 1986), where competition at the phonetic level is thought to cascade to lexical levels of processing (thus activating multiple lexical targets). Within a hierarchical processing structure, weak activation at one level will necessarily lead to weaker activation in all other levels. When semantic predictability is low, increased phonetic competition could lead to more

[1] For the current stimuli, global (full sentence) and final word Cloze predictability values were highly correlated ($r = 0.91$, $p < 0.01$), thus preventing clean separation to explore potentially distinct interactions of top-down context with acoustic–phonetic representations.

widespread activation of lexical competitors such that more lexical items are weakly activated, resulting in overall weaker lexical activation. Alternatively, low semantic predictability coupled with high phonetic competition may elicit greater uncertainty about each lexical candidate, leading to weaker activation for each candidate without changing the *number* of competing representations. Our findings do not distinguish between these two possibilities, as both predict weaker activation for each lexical competitor (regardless of the number of competitors), ultimately leading to weaker semantic activation. This interpretation is consistent with the regions that are modulated in the current study, left IFG (pars orbitalis) and left angular gyrus, which, as we review below, both play a role in semantic structure-building (Badre et al., 2005; Sharp et al., 2010).

A paradoxical finding in the current study is that greater phonetic competition leads to *less* activation in LIFG (pars orbitalis) and left AG/MTG, whereas in a prior study, greater phonetic competition led to *greater* activation in a similar, but distinct set of areas. Notably, in Xie and Myers (2018), all sentences were nonsensical (e.g., "*Relax the idea of the thin graceful code*"). We hypothesize that participants in the prior study may have engaged in a word-by-word rather than structure-building processing strategy, and that they did not recruit mechanisms for semantic assembly. If listeners are not attempting to assemble a coherent meaning, effects of phonetic competition would be free to emerge without modulation from the context of the containing sentence. As such, the observed patterns of activation in LIPL and LIFG (pars triangularis/opercularis) in Xie and Myers (2018) were interpreted as resolving competition between multiple phonological and lexical alternatives. Critically, in a post-hoc comparison of regions activated in Xie and Myers (2018) and the current study, overlap was only found for the MFG clusters and not for LIFG and angular gyrus clusters. Dynamically selecting the appropriate contextual constraint, be that lexical or sentential, is consistent with proposed theoretical frameworks that consider how uncertainty unfolds during speech perception (Heald & Nusbaum, 2014). This further supports our suggestion that different speech processing strategies were chosen based on the availability of broader sentence context.

This pattern is broadly consistent with work from Blank et al. (2016), who showed, using a multivariate analysis approach, that when top-down expectations were strong, decreased sensory detail (in this case, a 4-channel filtered signal) led to a sharpened representation of the speech signal. The opposite pattern was found when there were no top-down expectations. This result is congruous with a prediction error account, where the degree of mismatch between the predicted input (top-down cues from context or expectation) and the actual input (quality and content of the bottom-up signal) drives activation patterns. Interpreting the current findings within this framework, we suggest that when top-down expectation is weak (i.e., non-predictive contexts), enhanced signal clarity (i.e., less phonetic competition) sharpens or increases the precision of the neural signal. However, we cannot draw a direct parallel between these studies (and thus an explicit test of the prediction error account) due to differences in the imaging methods—a multivariate approach in Blank et al. (2016) and a univariate approach in the current study.

4.3.1. LIFG and left AG/MTG sensitivity to the strength of the semantic message

Both LIFG and left AG/MTG have been linked to semantic processing. The LIFG is anatomically connected to the left AG/MTG via the third branch of the superior longitudinal fasciculus (Makris et al., 2005) as well as functionally connected (for review, see Hagoort, 2014). LIFG is implicated in a range of language-related cognitive functions ranging from phonetic categorization (Myers, 2007; Rogers et al., 2017) to resolving semantic competition (Grindrod et al., 2008; Hirshorn & Thompson-Schill, 2006). The majority of studies investigating the role of LIFG in linguistic processing involve active decision making, raising the question if the prefrontal cortex is necessary for receptive language processing (Hickok & Poeppel, 2004). Our results, as well as those in Xie and Myers (2018), found LIFG activity driven by the degree of by-trial phonetic competition with a simple passive listening task, thus overcoming the active-task confound.

LIFG may serve as an integration site between uncertainty at multiple levels of the language hierarchy. A posterior to anterior functional gradient within the LIFG (Hagoort, 2013; Poldrack et al., 1999) ascribes phonological processing to posterior regions (pars opercularis) while semantic and lexical processing functions tend to cluster towards anterior regions (pars orbitalis). For instance, Badre et al. (2005) found that activation of anterior portions of the LIFG (BA 47) depended on the strength of semantic association between cue and target word pairs. Other work associates anterior LIFG with resolving competition between semantic alternatives (Thompson-Schill et al., 1997) or with switching between semantic categories (Hirshorn & Thompson-Schill, 2006). In the lens of our proposed framework, anterior LIFG activation could reflect the ease of constructing a semantic message while balancing phonetic ambiguity and message predictability.

A large body of previous work implicates left angular gyrus in processing semantic information as well as in predictive processing more broadly (see Seghier, 2013, for review). This region is reliably modulated by lexico-semantic information within tasks that require participants to make decisions along semantic dimensions (Démonet et al., 1992; Petersen et al., 1988; Sharp et al., 2010; Thompson-Schill et al., 1997). Studies also suggest that left AG/MTG is important at the sentence level for processing syntactic information (Ni et al., 2000) as well as overall sentence meaning (Humphries et al., 2007). While we did not find a main effect of semantic predictability in this region (perhaps due to the passivity of the task), we did find that left AG/MTG shows a different activation pattern depending on the strength of the sentence-wide message in relation to the speech signal. Notably, studies tend to find that left AG/MTG activation increases when semantic processing demands also increase (Hagoort, 2013; Mashal et al., 2009; Ye & Zhou, 2009). We found the opposite effect, as there was a decrease in activity in this region while processing non-predictive, and thus more semantically demanding, sentences as phonetic competition increased. There is some evidence to suggest that activation of the inferior parietal cortex (which includes the angular gyrus) is weaker when a highly degraded speech signal is paired with disrupted access to semantic knowledge (Obleser & Kotz, 2010). It is not unreasonable to view increasing phonetic competition as a type of "degradation", though to a much lesser degree than occurs in a noise-vocoded mask. Our data indicate that left AG/MTG is sensitive to both the quality of the incoming acoustic–phonetic information as well as the overarching semantics of the message. So, as the bottom-up speech signal becomes increasingly convoluted, this region fails to construct a message-level representation.

5. Conclusion

The current study, in combination with prior literature, highlights the interactive nature of spoken language processing. Sensitivity to temporary ambiguities in the bottom-up signal (that is, the effects of phonetic competition) depend on sentence context itself. We found that the strength of the semantic message influences the importance of bottom-up phonetic signals in the LIFG and left temporo-parietal junction. Listeners are adept at dynamically allocating neural processing resources that reflect the relative reliability of the signal at multiple levels of processing—consistent with our finding that phonetic competition effects disappear in highly-predictive sentences. A potential avenue for future experimentation could be in characterizing the role of attention when there is fluctuating semantic constraint. There is evidence to suggest that listener attention modulates perception of lexicality during lexically guided perceptual learning studies (Mirman et al., 2008; Pitt & Szostak, 2012). A theoretically interesting question is whether listeners' attention can be flexibly routed to either processing message-level information or on a word-by-word basis during continuous speech perception. Taken together, these data suggest that natural variability in the speech signal interacts with semantic processing, such that the strength of semantic coherence moderates the influence of phonetic ambiguity during receptive listening.

Author contributions

HM and EM designed the study. HM created the stimuli and collected the data. HM analyzed the data in consultation with XX and EM. All authors were involved in preparing the manuscript for submission.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by NIH NIDCD grant R01 CD013064 to E.B. M. The authors thank the members of the Language and Brain Lab for help with data collection and feedback about the project and Elisa Medieros for assisting with fMRI data collection.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2021.104959.

References

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language, 122*(1), 42–54.

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16*(1), 17–42.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron, 47*(6), 907–918.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods, 39*(3), 445–459.

Behrman, A., & Akhund, A. (2013). The influence of semantic context on the perception of Spanish-accented American English. *Journal of Speech, Language, and Hearing Research, 56*(5), 1567–1578.

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLOS Biology, 14*(11).

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America, 84*(1), 101–114.

Boersma, & Weenink. (2018). Praat: Doing phonetics by computer. *Version, 6*, 37.

Borsky, S., Tuller, B., & Shapiro, L. P. (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.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*(3), 1916–1924.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*(1), 1–47.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47.

Clopper, C. G. (2012). Effects of dialect variation on the semantic predictability benefit. *Language and Cognitive Processes, 27*(7-8), 1002–1020.

Copland, D. A., de Zubicaray, G. I., McMahon, K., Wilson, S. J., Eastburn, M., & Chenery, H. J. (2003). Brain activity during automatic semantic priming revealed by event-related functional magnetic resonance imaging. *NeuroImage, 20*(1), 302–310.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*(3), 162–173.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience, 23*(12), 3914–3932.

Davis, M. H., & Sohoglu, E. (2019). Three functions of prediction error for Bayesian inference in speech perception. *The Cognitive Neurosciences*.

Démonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., … Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115*(6), 1753–1768.

Drager, K. D. R., & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *Journal of Speech, Language & Hearing Research, 44*(5), 1052–1057.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*(1-3), 162–173.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology Section A., 33*(4), 465–495.

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170–177.

Gaston, P., & Marantz, A. (2018). The time course of contextual cohort effects in auditory processing of category-ambiguous words: MEG evidence for a single "clash" as noun or verb. *Language, Cognition and Neuroscience, 33*(4), 402–423.

Giolas, T. G., Cooker, H. S., & Duffy, J. R. (1970). The predictability of words in sentences. *The Journal of Auditory Research, 10*, 328–334.

Grindrod, C. M., Bilenko, N. Y., Myers, E. B., & Blumstein, S. E. (2008). The role of the left inferior frontal Gyrus in implicit semantic competition and selection: An event-related fMRI study. *Brain Research, 1229*, 167–178.

Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience, 25*(5), 706–718.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4*.

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

Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*.

Hickok, G., & Poeppel, D. (2004). Dorsal and Ventral Streams: A Framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67–99.

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia, 44*(12), 2547–2557.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage, 36*(3), 924–932.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337–1351.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., & Liddle, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: An event-related fMRI study. *Psychophysiology, 38*(1), 133–142.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology, 50*(1), 23–34.

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

Luthra, S., Guediche, S., Blumstein, S. E., & Myers, E. B. (2019). Neural substrates of subphonemic variation and lexical competition in spoken word recognition. *Language, Cognition and Neuroscience, 34*(2), 151–169.

Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., & Pandya, D. N. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cerebral Cortex, 15*(6), 854–869.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality, 14*(1), 30–54.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.

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

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science, 38*(6), 1139–1189.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*(5), 329–335.

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*(10), 1980–1988.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language, 52*(3), 416–435.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science, 32*(2), 398–417.

Müller, B. W., Jüptner, M., Jentzen, W., & Müller, S. P. (2002). Cortical activation to auditory mismatch elicited by frequency deviant and complex novel sounds: A PET study. *NeuroImage, 17*(1), 231–239.

Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia, 45*(7), 1463–1473.
Myers, E. B., & Blumstein, S. E. (2008). The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex, 18*(2), 278–288.
Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., … Shankweiler, D. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12*(1), 120–133.
Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395.
Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition, and Neuroscience, 31*(1), 4–18.
Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex, 20*(3), 633–640.
Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience, 27*(9), 2283–2289.
Owren, M. J. (2008). GSU Praat Tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods, 40*(3), 822–829.
Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(6157), 585–589.
Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*(2), 175–184.
Pitt, M. A., & Szostak, C. M. (2012). A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes, 27*(7-8), 1225–1239.
Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage, 10*(1), 15–35.
R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.
Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping, 29*(11), 1231–1242.
Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience, 15*(8), 1160–1175.
Rogers, J. C., Davis, D., & Matthew, H. (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of Cognitive Neuroscience, 29*(5), 919–936.
Rysop, A. U., Schmitt, L.-M., Obleser, J., & Hartwigsen, G. (2021). Neural modelling of the semantic predictability gain under challenging listening conditions. *Human Brain Mapping, 42*(1), 110–127.
Seghier, M. L. (2013). The angular Gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist, 19*(1), 43–61.
Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J. S., Vigliocco, G., & Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Human Brain Mapping, 31*(3), 365–377.
Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39*(1), 19–30.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.* Stuttgart, New York: Thieme Medical Publishers.
Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United State of America, 94*(26), 14792–14797.
Vaden, K.I., Halpin, H.R., & Hickok, G.S. (2009). Irvine phonotactic online dictionary, Version 2.0. [Data file]. Available from http://www.iphod.com.
Vaden, K. I., Teubner-Rhodes, S., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2017). Cingulo-opercular activity affects incidental memory encoding for speech in noise. *NeuroImage, 157*, 381–387.
Wright, R. (2004). Factors of lexical competition in vowel articulation. Phonetic Interpretation: Papers in Laboratory Phonology VI, 75–86.
Xie, X., & Myers, E. (2018). Left inferior frontal gyrus sensitivity to phonetic competition in receptive language processing: A comparison of clear and conversational speech. *Journal of Cognitive Neuroscience, 30*(3), 267–280.
Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: FMRI evidence. *NeuroImage, 48*(1), 280–290.
Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *The Journal of the Acoustical Society of America, 123*(5), 3878.
Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H.-C., Bastiaansen, M., & Wang, S. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *NeuroImage, 60*(4), 2230–2237.