WILEY

Listening under difficult conditions: An activation likelihood estimation meta-analysis

Claude Alain[1,2] 🄳 | Yi Du[3] | Lori J. Bernstein[4,5] | Thijs Barten[1] | Karen Banai[6]

[1]Rotman Research Institute, Baycrest Health Centre, Toronto, Ontario, Canada

[2]Department of Psychology, University of Toronto, Toronto, Ontario, Canada

[3]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[4]Department of Supportive Care, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

[5]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

[6]Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel

Correspondence
to Claude Alain, Rotman Research Institute at Baycrest, 3560 Bathurst Street, Toronto, ON M6A 2E1, Canada.
Email: calain@research.baycrest.org

Abstract

The brain networks supporting speech identification and comprehension under difficult listening conditions are not well specified. The networks hypothesized to underlie effortful listening include regions responsible for executive control. We conducted meta-analyses of auditory neuroimaging studies to determine whether a common activation pattern of the frontal lobe supports effortful listening under different speech manipulations. Fifty-three functional neuroimaging studies investigating speech perception were divided into three independent Activation Likelihood Estimate analyses based on the type of speech manipulation paradigm used: Speech-in-noise (SIN, 16 studies, involving 224 participants); spectrally degraded speech using filtering techniques (15 studies involving 270 participants); and linguistic complexity (i.e., levels of syntactic, lexical and semantic intricacy/density, 22 studies, involving 348 participants). Meta-analysis of the SIN studies revealed higher effort was associated with activation in left inferior frontal gyrus (IFG), left inferior parietal lobule, and right insula. Studies using spectrally degraded speech demonstrated increased activation of the insula bilaterally and the left superior temporal gyrus (STG). Studies manipulating linguistic complexity showed activation in the left IFG, right middle frontal gyrus, left middle temporal gyrus and bilateral STG. Planned contrasts revealed left IFG activation in linguistic complexity studies, which differed from activation patterns observed in SIN or spectral degradation studies. Although there were no significant overlap in prefrontal activation across these three speech manipulation paradigms, SIN and spectral degradation showed overlapping regions in left and right insula. These findings provide evidence that there is regional specialization within the left IFG and differential executive networks underlie effortful listening.

KEYWORDS
auditory, listening effort, prefrontal cortex, speech

1 | INTRODUCTION

The difficulty of speech identification and comprehension can be manipulated using a variety of methods (for a review, see Mattys, Davis, Bradlow, & Scott, 2012). The three most common methods are amplifying background noise, degrading the spectral quality of the auditory signal using filtering techniques, and presenting unfamiliar or complex linguistic materials. For instance, accuracy in identification of speech sounds decreases as background noise increases (Anderson Gosselin & Gagne, 2011; Gosselin & Gagne, 2011; Zekveld, Kramer, & Festen, 2011). Similarly, in studies manipulating the spectral quality of speech sounds (as in those using noise-vocoded speech stimuli), speech comprehension decreases with decreasing amounts of spectral information (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Word intelligibility is also reduced when high-frequency speech information is attenuated with a low-pass filter (Bhargava & Başkent, 2012; Eckert et al., 2008; Vaden et al., 2011). In addition, speech comprehension is more effortful with increased linguistic complexity—operationalized as the use of semantically ambiguous words or syntactically complex sentences (Bilenko, Grindrod, Myers, & Blumstein, 2009). These manipulations in perceptual and cognitive difficulty are reflected in behavioral performance (e.g., lower accuracy in speech identification or comprehension, increased response time), and physiological measures in pupil dilation (Koelewijn, Zekveld, Festen, & Kramer, 2012; Kuchinsky et al., 2016; Zekveld, Kramer, & Festen, 2010). Increased pupil diameter and decreased behavioral performance are considered manifestations of

mental effort (Kahneman, 1973; McGarrigle et al., 2014; Westbrook & Braver, 2015), and these indices accompany effortful listening when speech stimuli are manipulated using either of the three manipulations described above.

Listening under difficult conditions has been associated with enhanced activity in prefrontal regions implicated in cognitive control, attention, and working memory processes (Du, Buchsbaum, Grady, & Alain, 2016; Erb & Obleser, 2013; Love, Haist, Nicol, & Swinney, 2006; Peelle, Troiani, Wingfield, & Grossman, 2010b; Rodd, Johnsrude, & Davis, 2010a; Vaden, Kuchinsky, Ahlstrom, Dubno, & Eckert, 2015; Vaden et al., 2013; Wong et al., 2009a; Zekveld, Rudner, Johnsrude, Heslenfeld, & Ronnberg, 2012). For instance, evidence from functional magnetic resonance imaging (fMRI) suggests that accurate speech-in-noise (SIN) processing depends on a widely distributed neural network, including the left ventral premotor cortex (PMv) and Broca's area (i.e., the speech motor system), bilateral dorsal prefrontal cortex, superior temporal gyrus (STG), and parietal cortices (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Du, Buchsbaum, Grady, & Alain, 2014; Du et al., 2016; Hervais-Adelman, Pefkou, & Golestani, 2014; Wong et al., 2009a). Similarly, positron emission tomography (Scott, Rosen, Lang, & Wise, 2006) and other fMRI studies (Davis & Johnsrude, 2003; Evans & Davis, 2015; Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004; Vaden et al., 2011; Wild et al., 2012b) have shown increased activation in left inferior frontal gyrus (IFG), bilateral STG, and middle temporal gyrus (MTG) when hearing noise-vocoded and low-pass filtered speech stimuli. Increased levels of syntactic, lexical, and semantic complexity have also been associated with activation in the left IFG (Bilenko et al., 2009; Friederici, Fiebach, Schlesewsky, Bornkessel, & von Cramon, 2006), left middle frontal gyrus (MFG) (Caplan, Alpert, & Waters, 1999; Meltzer, McArdle, Schafer, & Braun, 2010), and left STG (Grindrod, Bilenko, Myers, & Blumstein, 2008).

The above studies suggest that the prefrontal cortex is ubiquitously involved in supporting effortful listening when speech sounds are masked by background noise, or spectrally degraded, or when linguistic information is unfamiliar or complex. That is, when the auditory cortex in the STG cannot effectively process speech sounds, executive functions such as attentional, working memory, and speech-motoric processes in the prefrontal regions may be recruited to compensate for speech comprehension (Du et al., 2014, 2016; Pichora-Fuller et al., 2016; Ronnberg et al., 2016; Skipper, Devlin, & Lametti, 2017; Van Engen & Peelle, 2014). According to sensorimotor integration theories of speech perception, predictions generated from the frontal articulatory network, including Broca's area and PMv, impose phonological constraints to auditory representations in sensorimotor interface areas, for example, the Sylvian-parietal-temporal area (Spt) in the posterior planum temporale (Hickok, 2009; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). This kind of sensorimotor integration is proposed to facilitate speech perception, especially in adverse listening environments. For instance, Du et al. (2014) found greater specificity of phoneme representations in the left PMv and Broca's area than in bilateral auditory cortices during syllable identification in high background noise. Older adults also showed greater specificity of phoneme representations in frontal articulatory regions than auditory regions (Du

et al., 2016). Thus, upregulated activity in prefrontal speech motor regions may provide a means of compensation for decoding impoverished speech representations.

Together, this indicates that effortful listening under increased background noise, poor spectral details or increased linguistic complexity likely requires greater recruitment of the brain networks underlying speech comprehension (described above). What remains unclear, however, is the extent to which increased prefrontal activations during effortful speech comprehension reflects: (a) the recruitment of the speech motor system; (b) a common anterior attention network needed to successfully identify speech stimuli; and/or (c) distinct patterns of prefrontal activations as a function of the listening challenges. For instance, increasing linguistic complexity may engage different brain regions within the left IFG from those recruited when the signal is degraded or masked by noise because linguistic complexity presents higher level cognitive and semantic challenge in deciphering the meaning of speech sounds, rather than lower level sensory and perceptual challenge. Similarly, effortful listening when the speech signal is spectrally degraded versus when masked by background noise might engage different neural networks.

There is some evidence suggesting that increasing background noise and spectrally degrading the signal yield similar changes in brain activity (Davis & Johnsrude, 2003), but that finding should be interpreted cautiously because it is based on a small sample of only 12 participants. Different patterns of activations may have been missed or undetected even if the net effect of both manipulations on the speech intelligibility was similar. SIN tasks require individuals to suppress competing auditory inputs, whereas no such suppression is needed while processing spectrally degraded speech. Consequently, SIN tasks may require greater executive control needed to focus attention on task-relevant speech sounds. Moreover, SIN tasks require segregation of concurrent sound objects which are presented in their entirety. By contrast, in spectrally degraded speech paradigms, parts of the speech signal are missing, so listeners must "fill in" the missing parts. Therefore, even though SIN, spectrally degraded speech, and linguistic complexity all require effortful listening, these three speech manipulation paradigms may recruit different networks and elicit different subpatterns of activation.

Meta-analyses of published neuroimaging studies provide a means of identifying brain regions that are reliably recruited across multiple studies and/or in different laboratories. Several such meta-analyses have been conducted on speech perception and production (Adank, 2012; Adank, Nuttall, Banks, & Kennedy-Higgins, 2015; DeWitt & Rauschecker, 2012; Rodd, Vitello, Woollams, & Adank, 2015; Samson et al., 2001), but none have directly compared activation patterns elicited between distinct methods of speech manipulation. For instance, Adank's (2012) meta-analysis grouped studies using different types of speech distortions (e.g., accented speech, background noise, compressed speech, and noise-vocoded speech). In another meta-analysis, Rodd et al. (2015) found differences in patterns of activation for semantic versus syntactic processing of spoken and written words in medial prefrontal regions. However, they also observed substantial overlap between semantic and syntax processing in the left IFG, even

with more than 50 studies included in the meta-analysis. This suggests that semantic and syntactic processing in prefrontal cortex may share common resources, yet it remains unknown whether the patterns of activations observed for increasing linguistic complexity would differ from those reported in SIN studies and studies manipulating the spectral quality of speech sounds.

The aims of this study are to (a) determine patterns of brain activation that support effortful listening when there is background noise, when the auditory signal is spectrally degraded, and when linguistic information is unfamiliar or complex and (b) identify regions of overlap (if any) among these three different effortful listening situations. We focus on three particular paradigms of speech manipulation: SIN comprehension, spectrally distorted speech, and linguistic complexity—because those are the most commonly used paradigms and thus have a sufficiently large number of neuroimaging studies to assess patterns of brain activation (Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012; Eickhoff, Laird, Fox, Lancaster, & Fox, 2017; Eickhoff et al., 2016). Of course, there are many other methods of speech manipulation techniques used to increase listening effort (e.g., increase speech rate or use accented speech); however, the number of studies published to date using any of these other methods is insufficient to obtain reliable estimates. In this study, the linguistic complexity category grouped together auditory studies that manipulated syntactic, lexical, and semantic ambiguity because the number of studies within these subgroups was too small to stand alone. Importantly for our study objectives, none of the studies included in linguistic complexity involved manipulation of the signal-to-noise ratio, providing a valid comparison of challenging listening situations without overlapping manipulation methods.

We sought to identify which (if any) brain areas are uniquely recruited during listening in SIN, spectral degradation, and linguistic complexity scenarios. To accomplish this, we used Activation Likelihood Estimation software (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012) to analyze activation results from relevant fMRI and PET studies. We anticipated that all three manipulations would yield patterns of prefrontal and temporal activations. We also performed contrast analyses to compare the patterns of activations within the prefrontal cortex and within the STG across different manipulations. The regions that are consistently recruited with increasing listening effort can be identified from the conjunction analyses between the three manipulations and from the omnibus analysis that comprises all studies regardless of the manipulation used to increase listening effort. If the analyses reveal significant overlapping prefrontal regions, then this will indicate that effortful listening recruits nonspecific neural networks regardless of the particular conditions that make it effortful. Conversely, if different manipulation paradigms elicit different patterns within the prefrontal cortex, this will indicate that listening effort involves brain regions which are differentiated as a function of listening conditions.

2 | MATERIALS AND METHODS

2.1 | Search strategy

The PubMed (www.pubmed.org) and Psychlib (hosted by the Department of Psychology at University of Toronto) database were searched using the following terms: "speech intelligibility," "speech in noise," "SIN," "vocoded speech," "distorted speech," "degraded speech," "listening effort," "semantic ambiguity," "linguistic complexity," crossed with "fMRI," "PET," "positron emission tomography," and "neuroimaging." All relevant review articles retrieved in the primary search were then hand-searched for articles not previously identified. The search included studies published in peer-reviewed journals and in English as of October 2017.

2.2 | Screening process

Each record was screened based on its title and abstract against the eligibility criteria outlined in Figure 1, which depicts the complete screening process. Full texts of potentially eligible articles were retrieved and screened, and any disagreements were settled by consensus among all authors.

Articles were included if they met the following criteria: (a) speech materials were presented aurally; (b) the study included a behavioral task performed prior to, during, or after scanning; (c) the study included a control experimental condition; (d) whole-brain analysis from fMRI or PET on 3D coordinates in either Talairach (Talairach & Tournoux, 1988) or Montreal Neurological Institute (MNI) standardized space were reported; (e) the study reported higher activity associated with increased listening difficulty; and (f) participants were young adults, without any hearing problems, psychiatric or neurological disorders, or brain abnormalities. Studies which combined data from healthy and patient participants were excluded. Studies meeting these selection criteria use a variety of stimuli and tasks, such as identification of isolated phonemes and comprehension of continuous speech. Studies were included whether or not there was behavioral difference between experimental conditions because comparable task performance does not always mean that listening effort is comparable (Alain, McDonald, Ostroff, & Schneider, 2004). That is, in more difficult listening conditions, participants may invest extra attention to maintain a comparable level of behavioral performance to that of easier listening conditions. Studies without behavioral data were included if participants were instructed to perform a task during scanning such as mental addition of numbers presented in noise or repeating aloud the last word of a sentence in noise.

2.3 | Activation likelihood estimate (ALE)

Coordinate-based quantitative meta-analyses of neuroimaging results were performed using the GingerALE software (version 2.3.6) available on the BrainMap website (http://brainmap.org/ale/index.html). The MNI coordinates were converted to Talairach space using the Lancaster transformation (Lancaster et al., 2007) before being entered into the analysis. This software generates a brain activation map based on coordinates provided in the included articles and uses a permutation test to determine whether the group mean activation is statistically reliable or not. It calculates above-chance clustering maps between experiments by modeling a three-dimensional Gaussian probability distribution centered on each focus reported in a study (weighted by
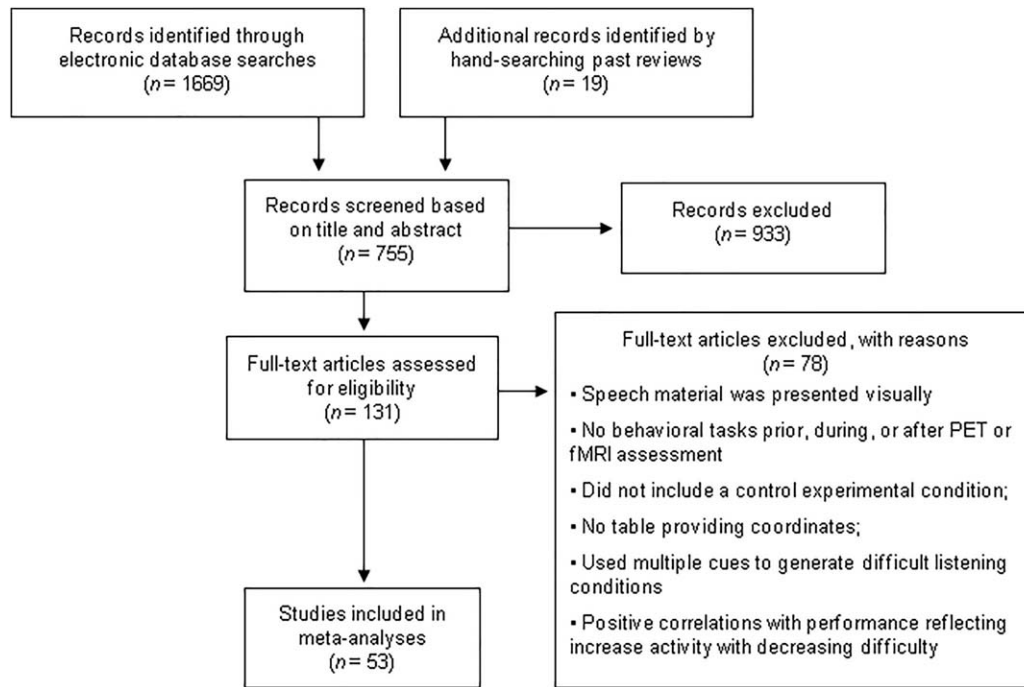
FIGURE 1 Screening process for studies included in the meta-analysis

the number of participants) and combining the probabilities of activation for each voxel. It then calculates voxel-wise scores representing convergence in similar brain locations across experiments. We used the smaller mask size and the random effect Turkeltaub nonadditive method, which minimizes both within-experiment and within-group effects by limiting probability values of neighboring foci from the same experiment (Turkeltaub et al., 2012). Cluster-level inference was used to identify brain areas consistently recruited during SIN, spectrally degraded speech, and linguistic complexity. We used 1,000 permutations to test for statistical significance. We also used an uncorrected p value of .001 for "cluster-forming threshold" and .05 for cluster-level inference as being statistically significant (Eickhoff et al., 2012, 2016, 2017).

Coordinates were selected for inclusion if they reflected: (a) activations from a direct comparison of the task of interest with a control task (e.g., speech-in-noise versus speech without noise) or (b) a correlation between brain activation and signal-to-noise ratio (SNR) when studies manipulated SNR for speech sounds. Separate analyses were performed for SIN, spectrally degraded speech, and linguistic complexity.

Three sets of follow-up contrast analyses were performed to determine whether different effortful listening situations yielded a different pattern of brain activation. These contrasts revealed the similarity (conjunction image) between data sets and contrast images created by directly subtracting one condition from the other. Pairwise contrast and conjunction analyses were performed between (a) SIN and spectrally degraded speech, (b) SIN and linguistic complexity, and (c) spectrally degraded speech and linguistic complexity. In GingerALE, contrasts are calculated using a voxel-wise minimum statistic (Nichols, Brett, Andersson, Wager, & Poline, 2005), which ascertains the

intersection between the individually thresholded meta-analysis results and produces a new thresholded ($p < .05$) ALE image and cluster analysis. For contrast, we used uncorrected p value of .05, with 10,000 permutation and minimum volume of 100 mm^3.

For visualization, BrainNet software was used to display foci (Xia, Wang, & He, 2013). The ALE-statistic maps were projected onto a cortical inflated surface template using SUrface MApping (SUMA) with Analysis of Functional Neuroimages software (Cox, 1996, 2012).

3 | RESULTS

Fifty-three studies met inclusion criteria (SIN: 16 studies, involving 224 participants; spectrally degraded speech: 15 studies, involving 270 participants; linguistic complexity: 22 studies, involving 368 participants; Table 1).

3.1 | Neural substrates of SIN, degraded speech, and linguistic complexity

The coordinates of the cluster-level brain areas consistently activated in SIN, spectrally degraded speech, or linguistic complexity studies are shown in Table 2. Figure 2 shows the individual foci used in the meta-analyses in each listening condition. Figure 3 displays the ALE-statistic maps for regions of statistically significant concordance for each of the three different listening conditions and the results from the omnibus analysis that included all studies.

3.1.1 | Speech in noise

The total number of foci from SIN manipulations was 129, with 61 of them located in the left hemisphere. The left and right foci were

TABLE 1 Neuroimaging studies included in the meta-analyses

Paper	Stimuli and contrast revealing increasing activation	Task and behavioral findings: during, no task, outside scanner	Number of participants (sex and age)	Foci	Coordinate space	Source in paper
Speech in noise						
Adank, 2012	Sentences in speech-shaped noise > sentences in quiet	Indicate if a sentence was true or false, RT for speech in noise > clear speech	26 (20F; 18–28 years)	8	MNI	Table 2
Bee, 2014	Arithmetic addition in noise > addition in quiet	Perform arithmetic addition, accuracy not reported	18 (0F; 20–28 years)	3	MNI	Table 1
Binder, 2004	Phonemes in noise: negative correlation with SNR	Two-interval forced-choice procedure, decrease accuracy and increase RT with increasing noise	18 (10F; 24–47 years)	3	Talairach	Table 1
Dos Santos Sequiera, 2010	Consonant-vowel syllables in babble noise > babble noise	Dichotic listening task, accuracy in quiet > babble conditions	21 (8F; 19–35 years)	19	MNI	Table 4
Du, 2014	Phonemes in broadband noise: negative correlation with SNR-modulated accuracy	Phoneme identification task, decrease accuracy and increase RT with decreasing SNR	16 (8F; 21–34 years)	8	Talairach	SI, Table 2
Evans, 2016	Two talker setting vs single talker	*Participants were asked to listen to a target speaker, hit rate for clear speech > masked speech*	20 (10F; 19–36 years)	8	MNI	Table 1
Golestani, 2013	Word in noise; negative correlation with SNR-modulated accuracy	Choose between two visual probes, increase RT with decreasing SNR	9 (6F, not provided)	6	MIN	Table 2
Guerreiro, 2016	Words in low SNR > high SNR	One-back task, accuracy for high SNR > low SNR	9 (7F; 19–56 years)	9	Talairach	Table 3
Hervais-Adelman, 2014	Word in noise, negative correlations with noise	Choose between two visual probes, increase RT with decreasing SNR	9 (6F, not provided)	6	MNI	SI, Table 1
Peelle, 2010	Sentences in continuous scanning EPI > quiet EPI	Respond to a probe word on a screen, no difference in behavioral performance	6 (3F; 20–26 years)	7	MNI	Table 4
Salvi, 2002	Sentences in noise > sentences in quiet	Repeat aloud the last sentence word, not reported	10 (5F; 23–34 years)	7	Talairach	Table 1
Scott, 2004	Sentences in noise masker and speech masker, negative correlation with SNR	*Participants were asked to listen to the female speaker for meaning, accuracy for higher SNRs > lower SNRs*	7 (0F; 35–52 years)	2	MNI	Figure 4
Sekiyama, 2003	Syllables at low SNR minus visual control	McGurk effect for lower SNRs > higher SNRs	8 (1F; 22–46 years)	9	Talairach	Table 1
Vaden, 2013	CVC syllables in multitalker babble: +3 dB > +10 dB SNR	Repeat each word aloud, hit rate for +10 dB > +3 dB SNR	18 (10F; 20–38 years)	5	MNI	Table 1
Wong, 2008	Words in multitalker babble: −5 dB SNR > quiet	Word identification, longer RT and lower accuracy for SNR −5 dB than SNR 20 dB	11 (7F; 20–34 years)	20	Talairach	Table 1
Zekveld, 2012	Sentences in long-term average speech spectrum > speech in quiet	Report sentence aloud, accuracy for speech in quiet > in noise	18 (9F; 20–29 years)	12	MNI	Table 2
Spectrally degraded speech						
Eckert, 2009	Filtered words, parametric increased in activity with decreasing intelligibility	Repeat word aloud, word recognition (accuracy) decrease with decreasing high-frequency information	11 (5F; mean age 32.8 ± 10.9 years)	7	MNI	SI, Table 1
Erb, 2013	Noise-vocoded sentence > clear sentence	Repeat sentence aloud, accuracy for clear > noise-vocoded speech	30 (15F; 21–31 years)	5	MNI	Table 1
Evans, 2015	Noise-vocoded phoneme > clear phoneme	One back task, d' scores for clear > noise-vocoded speech	17 (12F; 18–40 years)	3	MNI	Table 1
Hervais-Adelman, 2012	Noise-vocoded words > clear words	Respond to target sounds, accuracy for clear > noise-vocoded speech	15 (10F; 18–35 years)	4	MNI	Table 2

(Continues)

TABLE 1 (Continued)

Paper	Stimuli and contrast revealing increasing activation	Task and behavioral findings: during, no task, outside scanner	Number of participants (sex and age)	Foci	Coordinate space	Source in paper
Hesling, 2004	Expressive filtered sentences > normal sentences	Shadowing of normal speech > expressive filtered speech	12 (0F; 24–38 years)	4	Talairach	Table 5
Kyong, 2014	Noise-vocoded sentences: negative correlates of sentence intelligibility	Participants were asked to try to understand what was being said. Clear speech > noise vocoded speech	19 (18–40 years)	6	MNI	Table 5
Lee, 2016	Noise-vocoded sentences > clear sentences	Indicate gender of the character performing the action, accuracy for clear speech marginally higher than noise vocoded speech	26 (12F; 20–34 years)	7	MNI	Table 4
Meyer, 2002	Prosodic filtered (PURR-filtering) sentences > normal sentences	Indicate if the sentence was normal or has prosody, accuracy for clear > filtered speech	14 (6F; mean age 25.2 ± 6 years)	8	Talairach	Table 3
Meyer, 2004	Distorted sentences > normal sentences	Prosody comparison task, no behavioral data provided.	14 (8F; 18–27 years)	10	Talairach	Table 1
Sharp, 2010	Noise-vocoded words > clear words	Indicate whether two words were related or not, RT for noise-vocoded speech > clear speech	12 (5F; 35–61 years)	2	MNI	Table 2
Takeichi, 2010	Modulated sentences > nonmodulated sentences	Participants asked to listen to a narrative, accuracy higher for nonmodulated than modulated speech	23 (12F; 20–38 years)	2	MNI	Table 1
Tuennerhoff, 2016	Filtered sentences > normal sentences	No difference in target detection between filtered and normal speech, unprimed filtered speech less intelligible than normal speech	20 (10F; median age 24.05 years)	12	MNI	Table 1
Wild, 2012a	Noise-vocoded sentences vs clear sentences, main effect of intelligibility, noise-elevated response	Repeat the sentence aloud, accuracy for clear > noise-vocoded speech	21 (13F; 19–27 years)	4	MNI	Table 1
Wild, 2012b	Noise-vocoded sentences > clear speech and rotated noise-vocoded sentences	Participants were asked to listen attentively	19 (11F; 18–26 years)	1	MNI	Table 1
Zekveld, 2014	Noise-vocoded sentences > clear sentences	Indicate whether a visual probe were part of the sentence, accuracy for clear > noise-vocoded speech	17 (9F; 19–33 years)	4	MNI	Table 4
Linguistic complexity						
Bekinschtein, 2011	Semantically ambiguous > unambiguous sentence	Rate sentences as funny or not, no significant difference in performance between ambiguous and unambiguous sentences	12 (not specified)	2	MNI	Table 1
Bilenko, 2009	Semantically ambiguous > unambiguous words	Speeded lexical decision task, slower RT for ambiguous than unambiguous words	16 (9F; 18–21 years)	2	Talairach	Table 3
Buchweitz, 2014	Unfamiliar > familiar sentences	True–false question that probe comprehension, RT for unfamiliar > familiar passages	9 (3F; 18–25 years)	6	MNI	Table 3
Caplan, 1999	Semantically implausible > plausible sentences	Plausibility judgment task, RT for implausible > plausible sentences	16 (8F; 22–34 years)	3	Talairach	Table 2
Davis, 2007	High ambiguity > low ambiguity sentences	Participants listened passively	12 (3F; 29–42 years)	7	MNI	Supporting Information, Table 3

(Continues)

TABLE 1 (Continued)

Paper	Stimuli and contrast revealing increasing activation	Task and behavioral findings: during, *no task, outside scanner*	Number of participants (sex and age)	Foci	Coordinate space	Source in paper
Friederici, 2006	Increase activity with increasing grammatical complexity of sentences	Rate acceptability of the sentence, RT increase with increasing complexity	13 (7F; mean age 23.1 years)	2	Talairach	Figure 4
Friederici, 2010	Syntactically incorrect > correct sentences	*Participants were asked to listen attentively*	17 (9F; 20–30 years)	5	MNI	Table 1
Grindrod, 2008	Sequences of semantically disconcordant > neutral words	Lexical decision task, RT for discordant > concordant or neutral words	15 (8F; 19–29 years)	1	Talairach	Table 3
Guediche, 2013	Semantically ambiguous > unambiguous sentences	Press a button to target word, longer RT, and lower accuracy in ambiguous than unambiguous target	17 (8F; 24.5 ± 3.6 years)	3	Talairach	Table 3
Herrmann, 2012	Syntactically incorrect > correct sentences	Grammatically judgment task, accuracy for correct > incorrect grammar	25 (12F; 22–32 years)	5	MNI	Table 2
Lopes, 2016	Semantically complex > easy semantic decision task	Semantic decision task, accuracy for easy- > complex semantic decision	24 (15F; 20–31 years)	6	MNI	Table 2
Meltzer, 2010	Main effect of syntactic sentence complexity	Match heard spoken sentence with visual probes, RT for reversible > irreversible sentence	24 (12F; 22–37 years)	1	Talairach	Table 3
Obleser, 2011	Positive correlation with syntactic complexity	*Participants were asked to listen attentively,* no difference in accuracy	14 (8F; 23.4 ± 2.1 years)	1	MNI	Table 2
Peelle, 2004	Object-relative > subject-relative sentences	Indicate the gender of the character performing the action, longer RT, and lower accuracy for object related than subject-related sentences	8 (4F; 19–27 years)	2	Talairach	Table 1
Rissman, 2003	Semantically unrelated > related words	Lexical decision task, longer RT, and lower accuracy for unrelated than related words	15 (8F; 18–44 years)	5	Talairach	Table 2
Rodd, 2005	High semantic ambiguity > low ambiguity sentences	Relatedness judgment task, no difference in performance between high and low ambiguity	15 (10F; 18–40 years)	4	MNI	Table 4
Rodd, 2010	High > low semantic ambiguity words	Indicate whether the incoming word was related or not to the sentence meaning	14 (19–37 years)	1	MNI	Table 3
Ruff, 2008	Semantic unrelated > related words	Relatedness judgment task, RT for unrelated > related words	15 (8F; 19–31 years)	9	Talairach	Table 2
Tyler, 2005	Regular vs irregular verb	Same or different judgment task, RT regular > irregular verb, no difference in error rate	18 (8F, mean age 24 ± 7)	9	MNI	Table 2
Tyler, 2011	Control group: syntactically ambiguous > unambiguous sentences	*Listen to spoken sentence.* Longer RT and lower accuracy for ambiguous than unambiguous sentences	15 (8F; 46–74 years)	3	MNI	Table 3
Vitello, 2014	Semantic ambiguous > unambiguous semantic sentences	Participant indicated whether a probe was related or unrelated to the sentence they just heard	20 (11F; 18–35 years)	3	MNI	Table 4
Wright, 2010	Complex > simple words, main effect of complexity	Lexical decision task, RT for non-words > real words, RT for complex- > simple words	14 (sex not reported; 19–34 years)	1	MNI	Table 3

TABLE 2 Brain areas consistently activated as a function of listening conditions using cluster-level inference with a thresholding value = 0.05

Studies	Brain region, BA	Talairach. coordinates (x, y, z)	Cluster size (mm³)	#Studies/cluster
Speech in noise				
	Left inferior frontal gyrus, 45	−36, 19, 8	1,616	6
	Right insula, 13	31, 18, 12	1,248	5
	Left inferior parietal lobule, 40	−35, −50, 36	800	3
Spectrally degraded speech				
	Right insula, 13	28, 18, 7	1,704	6
	Left insula, 13	−34, 16, 5	1,384	5
	Left superior temporal gyrus, 13	−37, −29, 10	1,144	3
Linguistic complexity				
	Left inferior frontal gyrus, 44	−49, 13, 19	5,544	15
	Right middle frontal gyrus, 46	43, 22, 23	1,896	7
	Left middle temporal gyrus, 37	−45, −48, −11	1,088	4
	Right superior temporal gyrus	51, −25, 4	952	3
	Left superior temporal gyrus, 22	−55, −16, 3	928	5
All studies combined				
	Left insula, 13	−43, 13, 16	9,400	29
	Right insula, 13	35, 20, 14	5,392	19
	Left superior temporal gyrus	−42, −29, 9	1,376	5
	Right superior temporal gyrus	52, −22, 4	920	4

primarily distributed in prefrontal and posterior temporal and parietal regions (Figure 2). Processing SIN was associated with peak activations in the left IFG, right insula, and left inferior parietal lobule (Table 2 and Figure 3).

3.1.2 | Spectrally degraded speech

Studies that manipulated the spectral quality of the speech sounds were associated with 79 foci, 40 of which were located in the left hemisphere. The analysis of foci from studies using spectrally degraded speech yielded peak activity in the right and left insula and left STG (Table 2 and Figure 3). There was no common area of activity in prefrontal regions, which could be accounted for by the spread of foci within prefrontal cortex (Table 2 and Figure 2).

3.1.3 | Linguistic complexity

Linguistic complexity manipulations yielded 81 foci, with 60 of them in the left hemisphere (Figure 2). These were primarily distributed within the prefrontal cortex. Studies that focused on linguistic complexity consistently showed increased activation of the left IFG and right MFG, left MTG, and bilateral STG (Table 2 and Figure 3).

3.1.4 | Networks involved in effortful listening independent of manipulation

This meta-analysis aimed to identify brain areas that are consistently recruited during effortful listening by including all studies regardless of

manipulation (i.e., all 53 studies) into a single analysis. Figure 3 shows the brain areas that were consistently activated across studies. These include the insula and STG bilaterally (Table 2). It also included the anterior and dorsal portion of the IFG bilaterally.

3.2 | Contrast between patterns of activations

In three separate contrast analyses, we tested whether the spatial patterns of activations observed in one task differed with those from another task. The brain areas that were significantly different between the pairwise subtractions are listed in Table 3. The contrast between SIN and spectrally degraded speech revealed greater left STG activation in SIN than in spectral degradation. Moreover, SIN studies yielded greater activation in the left IFG, right insula, and left IPL activation than studies that manipulated linguistic complexity. Studies that manipulated spectral quality of the speech signal showed greater activation in the insula bilaterally and in the left transverse temporal gyrus than studies that manipulated linguistic complexity. Last, linguistic complexity studies generated greater left IFG activation than studies manipulating SIN or spectral quality.

3.3 | Overlap between patterns of activations

The three types of speech manipulation paradigms included in this meta-analysis are associated with changes in accuracy and/or response
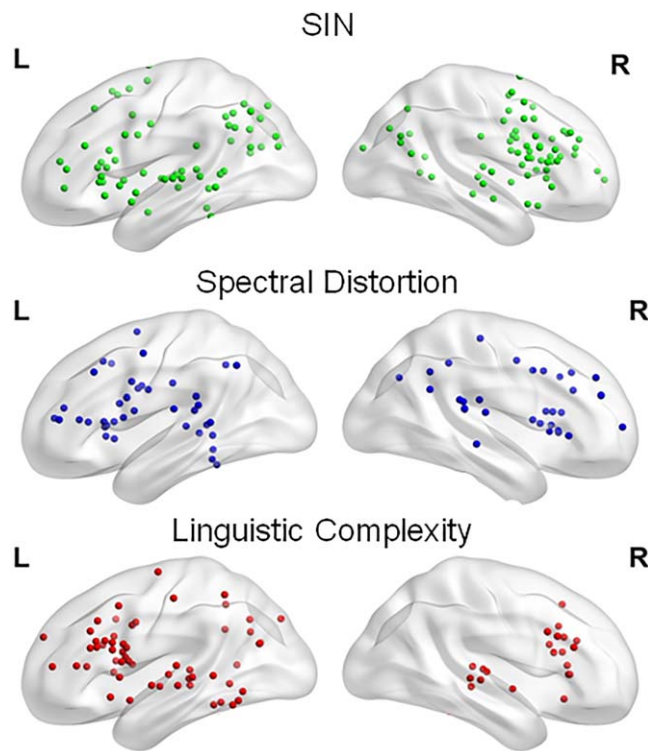
FIGURE 2 Foci from the speech in noise (SIN), spectrally degraded speech, and linguistic complexity studies

time, reflecting changes in the difficulty of speech identification and comprehension. In three separate conjunction analyses, we tested whether the spatial patterns of activation observed in one task overlap with those from another task. The conjunction analyses between SIN and spectrally degraded speech revealed significant bilateral overlap in the insula. There was no significant overlap in prefrontal regions. Neither the conjunction analysis between SIN and linguistic complexity nor the conjunction analysis between spectrally degraded speech and linguistic complexity yielded any significant overlapping regions in either the STG or in prefrontal cortices.

4 | DISCUSSION

This study aimed to identify unique and overlapping activation patterns across three types of effortful listening conditions.

4.1 | Speech in noise

Comprehending speech in noise is challenging. Several speech-motor networks have been proposed to compensate for challenging listening conditions. Our results revealed a left fronto-parietal network that includes the IFG (Broadmann area 45) and IPL. Brodmann area 45, known as the Pars Triangularis of Broca's area, is believed to be involved in language perception and production (Hickok, 2009; Rauschecker & Scott, 2009), and has a role in the cognitive control of memory (Badre, 2008; Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Grady, Yu, & Alain, 2008). The observed activation in left IFG is
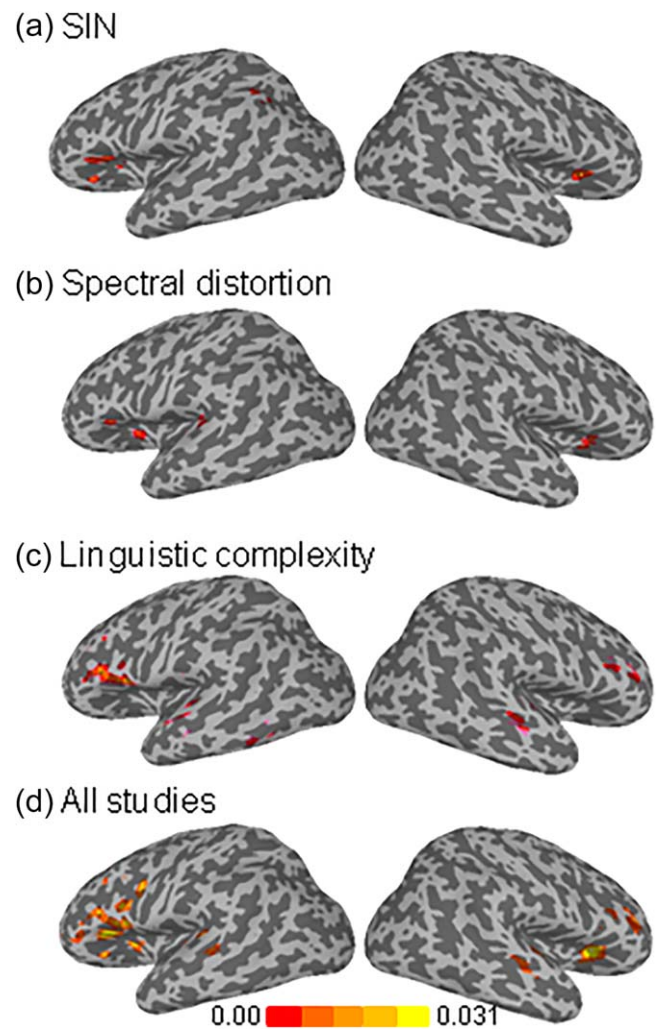

FIGURE 3 ALE-statistic maps for regions of significant concordance in neuroimaging studies: (a) speech in noise (SIN), (b) spectral distortion, (c) linguistic complexity studies, and (d) all studies together

consistent with findings from a prior meta-analysis (Adank, 2012) and provides converging evidence supporting sensorimotor accounts of speech processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). This account posits that the speech motor system generates internal models which predict sensory consequences of articulatory
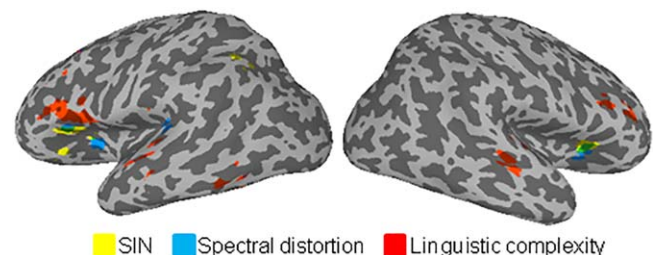

FIGURE 4 Overlay of ALE-statistic maps for all three tasks. Blue = SIN; yellow = spectral distortion; red = linguistic complexity; and green = significant overlap between spectral distortion and SIN

TABLE 3 Pairwise contrasts across SIN, spectrally degraded speech, and linguistic complexity studies

Contrast analysis	Brain region, BA	Talairach coordinates (x, y, z)	Cluster size (mm³)
SIN vs spectrally degraded speech	Left superior temporal gyrus, 39	−36, −51, 34	440
Spectrally degraded speech vs SIN	Right lentiform nucleus	26, 17, 2	144
SIN vs linguistic complexity	Left inferior frontal gyrus, 13	−35, 19, 8	1,392
	Right insula, 13	30, 17, 13	904
	Left inferior parietal lobule, 40	−35, −50, 36	800
Linguistic complexity vs SIN	Left inferior frontal gyrus, 44	−50, 12, 18	4,032
Spectrally degraded speech vs linguistic complexity	Right insula, 13	28, 17, 7	1,328
	Left insula, 13	−31, 17, 2	592
	Left transverse temporal gyrus, 41	−34, −32, 15	288
	Right insula, 13	55, −32, 19	272
Linguistic complexity vs spectrally degraded speech	Left inferior frontal gyrus, 44	−50, 17, 20	3,100
	Right middle frontal gyrus, 46	45, 18, 22	400

gestures under consideration, and such forward predictions are matched with acoustic representations in sensorimotor interface areas located in the left pSTG or IPL to constrain perception. Prefrontal speech motor-based predictive coding is believed to be especially useful for disambiguating phonological information under adverse listening conditions thus increasing the probability of extracting target signals from noise (Du et al., 2014, 2016).

The observed activity in the left IFG may also reflect compensatory mechanisms or increased attentional demands needed to segregate and identify speech sounds. The decline-compensation hypothesis posits that deficits in sensory processing regions can be mitigated by recruitment of more general executive control areas in the prefrontal cortex (Cabeza, Anderson, Locantore, & McIntosh, 2002; Wong, Perrachione, & Margulis, 2009b). That is, when peripheral and central auditory systems cannot effectively process speech sounds, the relative contributions of sensory, prefrontal, and parietal cortices may change, and the nature of that change depends on the task. Prefrontal activity may reflect engagement of predictive processes based on lexical and grammatical knowledges to offset the impoverished encoding and representations of speech sounds in auditory short-term memory. However, such knowledge-based predictive processes may not be of much

use for identifying isolated syllables or phonemes in noise. In that context, participants may rely on articulatory representations of speech stimuli held in working memory, which could act as templates against which an incoming sound of syllable or phoneme could be compared.

4.2 | Spectrally degraded speech

Studies that manipulate the spectral quality of incoming speech stimuli using noise-vocoded speech (e.g., Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Shannon et al., 1995) or low-pass filtered speech (Bhargava & Başkent, 2012; Eckert et al., 2008; Vaden et al., 2011) have shown reduced speech intelligibility with decreased speech spectral details. The present meta-analysis results showed that decreasing speech intelligibility with spectral-filtering techniques consistently yields increased activation in the insula bilaterally and left STG.

These areas are part of the auditory ventral stream, which is thought to be particularly important for auditory object formation and sound identification. The observed insula activation is also consistent with the idea that, when presented with a degraded signal, acoustical details need to be analyzed to a greater extent than in normal conditions—because the level of available acoustic information might not be

TABLE 4 Brain areas that overlap across SIN, spectrally degraded speech and linguistic complexity studies

Conjunction analysis	Brain region, BA	Talairach coordinates (x, y, z)	Cluster size (mm³)
SIN ∧ spectrally degraded speech			
	Right insula, 13	29, 19, 11	592
	Left insula, 13	−36, 18, 8	456
SIN ∧ linguistic complexity			
	No significant cluster		
Spectrally degraded speech ∧ linguistic complexity			
	No significant cluster		

sufficient for lexical access. Although 13 out of 15 studies reported foci in the frontal lobes, the analysis indicated that there was no significant overlap in prefrontal activation across studies. That is, no prefrontal cortex areas were consistently recruited in studies using spectrally degraded stimuli. This was unexpected considering that most studies were fairly similar in methodology and stimuli. This is also surprising given the evidence from transcranial direct current stimulation (tDCS) studies suggesting that stimulation of the left IFG can improve identification of degraded speech (Sehm et al., 2013). The lack of a common pattern of activation in prefrontal cortex suggests that activation within the prefrontal cortex is widely distributed during tasks involving spectrally degraded stimuli. This spread of activation may be related to differences in task instructions, with a third of the studies included having participants listen to the stimuli without having to generate a response during scanning. In fact, there was higher proportion of spectral degradation studies without behavioral measures during scanning compared to SIN or linguistic complexity studies. The absence of behavioral measurements during scanning makes it difficult to determine participants' attention and engagement in processing the speech stimuli and could likely contribute to variability in prefrontal activation. Further research is needed to better understand the source of this variability in prefrontal activation within studies using spectrally degraded speech stimuli.

4.3 | Linguistic complexity

Linguistic complexity is operationalized as situations in which the acoustic quality of the incoming signal is kept constant while the semantic, lexical and grammatical aspects of the speech materials are manipulated using more semantically or syntactically ambiguous words or sentences. Results showed that there were consistent activations of the left IFG (Broadmann area 44), right MFG (Broadmann area 46), and bilateral STG when processing more ambiguous or complex sentences. The observed common pattern of activations in the left IFG is consistent with that of an earlier meta-analysis showing consistent left IFG activity during semantic and syntactic processing (Rodd et al., 2015). Brodmann area 44 in the left hemisphere, also known as the Pars Opercularis of Broca's area, is recognized as one of the main language areas. It is important for inner speech (Kuhn et al., 2013), perception and expression of prosodic and emotional information (Merrill et al., 2012), and auditory working memory (Alain, He, & Grady, 2008; Buchsbaum et al., 2005; Grady et al., 2008). Brodmann area 46, the MFG, which is a part of the dorsolateral prefrontal cortex, is involved in motor planning, organization, regulation, and working memory. The pattern of prefrontal activation may reflect enhanced verbal working memory needed to process meaning with increasing syntactic, lexical and semantic intricacy. It may also reflect attention to internal memory representations, that is, "listening" back in time to what a person just said (Backer & Alain, 2012, 2014; Backer, Binns, & Alain, 2015). Indeed, to string spoken phonemes into words and words into sentences, the acoustic representations must be maintained in auditory short-term memory, retrospectively processed, and bound with incoming acoustic signals

(Alain & Bernstein, 2008). For effective comprehension, the semantics of these sentences must also be grouped together in a meaningful way. Hence, reflective attention to internal sound representations is crucial for understanding what another person is saying, especially in the midst of other concurrent conversations. In this study, different types of linguistic complexity tasks were grouped together under one single paradigm and provided a comparison group against the two other manipulations affecting the acoustic quality of the speech signals. Further research is needed to determine whether the effort exerted into different linguistic complexity tasks also recruit distinct areas (for further discussion of this issue, see Rodd et al., 2015).

4.4 | Patterns of activation during difficult listening conditions

This meta-analysis aimed to determine if there are overlapping activation patterns across speech identification and comprehension studies, which varied either in background noise level, spectral quality of the incoming speech sounds, or in linguistic complexity. The results revealed that brain activation patterns associated with increasing listening difficulty differed across the manipulation paradigms. Both SIN and linguistic complexity studies yielded consistent activations in left the IFG, but spectrally degraded speech sounds did not. Importantly, contrast analyses revealed differential involvement: activation of the Broca's areas (i.e., Brodmann area (BA) 45) was more associated with SIN whereas linguistic complexity recruited more dorsal prefrontal areas (BA 44 and 46). This distinction along the dorsal–ventral axis of prefrontal cortex could reflect the differences in the stimuli used—SIN studies more often used phoneme and single word stimuli, whereas linguistic complexity studies tended to use sentences. In SIN studies, effortful listening may recruit speech motor representations and inhibitory control processes within the left IFG, whereas linguistic complexity may place more demand on verbal working memory processes by recruiting distinct prefrontal regions. Evidence indeed suggests that working memory capacity is correlated with comprehension of complex sentences (Andrews, Birney, & Halford, 2006). Taken together, these results suggest that listening during difficult conditions is not a unitary concept, but rather depends on the specific acoustic and semantic characteristics of the speech signals. Hence, in effortful listening situations, different tasks and stimuli lead to differential recruitment of processes relying on the left IFG.

It is not surprising that right and left insula were consistently activated in both SIN and spectrally degraded speech studies as they directly manipulate acoustic quality of speech. However, the meta-analysis also revealed specific differences in the pattern of activation in the STG across manipulation paradigms, and this suggests that the recruitment of distinct processes is needed for successful identification of speech sounds in different effortful conditions. For instance, SIN requires participants to segregate and identify phonemes, words, or sentences against either broadband noise (energetic masking) or multi-talker babble (informational masking). Unlike spectrally degraded speech, speech stimuli embedded in noise are not directly distorted;

instead, the main challenge is to separate task-relevant speech signals from irrelevant background noise (figure-ground perception). This may explain the difference observed in the STG, with spectrally degraded speech consistently showing increased activation in the anterior portion of STG, that is, the planum polare, which is important for speech and sound object identification (Alain, Arnott, Hevenor, Graham, & Grady, 2001; Alain, Reinke, He, Wang, & Lobaugh, 2005; DeWitt & Rauschecker, 2012; Rauschecker & Scott, 2009).

5 | LIMITATIONS

Meta-analyses of neuroimaging studies enable identification of core brain regions that support task performance independent of differences in methodology (e.g., sample size, materials used, and scanner type). However, one needs to consider the number of studies, the type of studies, and the methodology and statistical thresholds used because these variables may influence the interpretation. We followed recent guidelines for meta-analyses using ALE software. The number of studies included in each meta-analysis was sufficient to yield reliable patterns of activation. Nevertheless, it is possible that these patterns of activation may change as additional studies are published and can be included in each condition.

6 | CONCLUDING REMARKS

Processing requirements in speech comprehension are higher when the physical quality of the incoming speech sounds declines or when the speech materials are more linguistically complex. These challenges usually imply an increase in listening effort, which places greater demands on cognitive resources to resolve perceptual or semantic ambiguity. This study indicates that different neural networks are used to resolve this ambiguity depending on the specific type of perceptual or semantic difficulty introduced. These results reveal that functional specialization within the left IFG supports effective listening under effortful conditions. The concept of listening effort is broad, and we recommend that future research studies continue to investigate subtypes within the three speech manipulation paradigms examined here and others that are less commonly explored, with an emphasis on listening resources rather than effort. Ultimately, listening under difficult conditions recruits different pools of resources to support perceptual and cognitive processes during verbal communication.

ADDITIONAL INFORMATION

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

CA, YD, LJB, and KB designed the study. TB and CA were involved in data collection. CA analyzed the data. CA, YD, LJB, and KB interpreted data and wrote the manuscript.

ORCID

Claude Alain http://orcid.org/0000-0003-4459-1538

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.

Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, *50*(1), 77–84.

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

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12301–12306.

Alain, C., & Bernstein, L. J. (2008). From sounds to meaning: The role of attention during auditory scene analysis. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *16*(5), 485–489.

Alain, C., He, Y., & Grady, C. (2008). The contribution of the inferior parietal lobe to auditory spatial working memory. *Journal of Cognitive Neuroscience*, *20*(2), 285–295.

Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2004). Aging: A switch from automatic to controlled processing of sounds? *Psychology and Aging*, *19*(1), 125–133.

Alain, C., Reinke, K., He, Y., Wang, C., & Lobaugh, N. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. *Journal of Cognitive Neuroscience*, *17*(5), 811–818.

Anderson Gosselin, P., & Gagne, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, *54*(3), 944–958.

Andrews, G., Birney, D., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory &Amp; Cognition*, *34*(6), 1325–1340.

Backer, K. C., & Alain, C. (2012). Orienting attention to sound object representations attenuates change deafness. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1554–1566.

Backer, K. C., & Alain, C. (2014). Attention to memory: Orienting attention to sound object representations. *Psychological Research*, *78*(3), 439–452.

Backer, K. C., Binns, M. A., & Alain, C. (2015). Neural dynamics underlying attentional orienting to auditory representations in short-term memory. *Journal of Neuroscience*, *35*(3), 1307–1318.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200.

Badre, D., Poldrack, R. A., Pare-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.

Bee, N. S., Yusoff, A. N., Ling, T. X., & Abd Hamid, A. I. (2014). Investigating brain activation and neural efficacy during simple arithmetic addition task in quiet and in noise: An fMRI study. *Jurnal Sains Kesihatan Malaysia, 12*(1), 23–33.

Bekinschtein, T. A., Davis, M. H., Rodd, J. M., & Owen, A. M. (2011). Why clowns taste funny: The relationship between humor and semantic ambiguity. *Journal of Neuroscience, 31*(26), 9665–9671.

Bhargava, P., & Başkent, D. (2012). Effects of low-pass filtering on intelligibility of periodically interrupted speech. *Journal of the Acoustical Society of America, 131*(2), EL87–EL92.

Bilenko, N. Y., Grindrod, C. M., Myers, E. B., & Blumstein, S. E. (2009). Neural correlates of semantic competition during processing of ambiguous words. *Journal of Cognitive Neuroscience, 21*(5), 960–975.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience, 7*(3), 295–301.

Buchsbaum, B., Pickell, B., Love, T., Hatrak, M., Bellugi, U., & Hickok, G. (2005). Neural substrates for verbal working memory in deaf signers: fMRI study and lesion case report. *Brain and Language, 95*(2), 265–272.

Buchweitz, A., Mason, R. A., Meschyan, G., Keller, T. A., & Just, M. A. (2014). Modulation of cortical activity during comprehension of familiar and unfamiliar text topics in speed reading and speed listening. *Brain and Language, 139*, 49–57.

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging gracefully: Compensatory brain activity in high-performing older adults. *NeuroImage, 17*(3), 1394–1402.

Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *NeuroImage, 9*(3), 343–351.

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

Cox, R. W. (2012). AFNI: What a long strange trip it's been. *NeuroImage, 62*(2), 743–747.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., ... Menon, D. K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences of the United States of America, 104*(41), 16032–16037.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience, 23*(8), 3423–3431.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology. General, 134*(2), 222–241.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences of the United States of America, 109*(8), E505–E514.

Dos Santos Sequeira, S., Specht, K., Moosmann, M., Westerhausen, R., & Hugdahl, K. (2010). The effects of background noise on dichotic listening to consonant-vowel syllables: An fMRI study. *Laterality, 15*(6), 577–596.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of National Academy of Sciences of the United States of America, 111*(19), 7126–7131.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nature Communications, 7*, 12241.

Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J. R. (2009). At the heart of the ventral attention system: The right anterior insula. *Human Brain Mapping, 30*(8), 2530–2541.

Eckert, M. A., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J. R. (2008). Age-related effects on word recognition: Reliance on cognitive control systems with structural declines in speech-responsive cortex. *Journal of the Association for Research in Otolaryngology, 9*(2), 252–259.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage, 59*(3), 2349–2361.

Eickhoff, S. B., Laird, A. R., Fox, P. M., Lancaster, J. L., & Fox, P. T. (2017). Implementation errors in the GingerALE Software: Description and recommendations. *Human Brain Mapping, 38*(1), 7–11.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping, 30*(9), 2907–2926.

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., ... Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage, 137*, 70–85.

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience, 33*(26), 10688–10697.

Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience, 7*, 116.

Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex (New York, N.Y. : 1991), 25*(12), 4772–4788.

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the cocktail party started: Masking effects in speech perception. *Journal of Cognitive Neuroscience, 28*(3), 483–500.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex (New York, N.Y. : 1991), 16*(12), 1709–1717.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping, 31*(3), 448–457.

Gosselin, P. A., & Gagne, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology, 50*(11), 786–792.

Grady, C. L., Yu, H., & Alain, C. (2008). Age-related differences in brain activity underlying working memory for spatial and nonspatial auditory information. *Cerebral Cortex (New York, N.Y. : 1991), 18*(1), 189–199.

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.

Guerreiro, M. J., Putzar, L., & Roder, B. (2016). The effect of early visual deprivation on the neural bases of auditory processing. *Journal of Neuroscience, 36*(5), 1620–1630.

Herrmann, B., Obleser, J., Kalberlah, C., Haynes, J. D., & Friederici, A. D. (2012). Dissociable neural imprints of perception and grammar in auditory functional imaging. *Human Brain Mapping*, 33(3), 584–595.

Hervais-Adelman, A., Pefkou, M., & Golestani, N. (2014). Bilingual speech-in-noise: Neural bases of semantic context use in the native language. *Brain and Language*, 132, 1–6.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, 27(7–8), 1145–1166.

Hesling, I., Clement, S., Bordessoules, M., & Allard, M. (2005). Cerebral mechanisms of prosodic integration: Evidence from connected speech. *NeuroImage*, 24(4), 937–947.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews*, 6(3), 121–143.

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

Kahneman, D. (1973). *Attention and effort*. New Jersey, NJ: Prentice-Hall.

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, 33(2), 291–300.

Kuchinsky, S. E., Vaden, K. I., Jr., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2016). Task-related vigilance during word recognition in noise for older adults with hearing loss. *Experimental Aging Research*, 42(1), 50–66.

Kuhn, S., Schmiedek, F., Brose, A., Schott, B. H., Lindenberger, U., & Lovden, M. (2013). The neural representation of intrusive thoughts. *Social Cognitive and Affective Neuroscience*, 8(6), 688–693.

Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., & McGettigan, C. (2014). Exploring the roles of spectral detail and intonation contour in speech intelligibility: An FMRI study. *Journal of Cognitive Neuroscience*, 26(8), 1748–1763.

Lancaster, J. L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., … Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, 28(11), 1194–1205.

Lee, Y. S., Min, N. E., Wingfield, A., Grossman, M., & Peelle, J. E. (2016). Acoustic richness modulates the neural networks supporting intelligible speech processing. *Hearing Research*, 333, 108–117.

Lopes, T. M., Yasuda, C. L., de Campos, B. M., Balthazar, M. L., Binder, J. R., & Cendes, F. (2016). Effects of task complexity on activation of language areas in a semantic decision fMRI protocol. *Neuropsychologia*, 81, 140–148.

Love, T., Haist, F., Nicol, J., & Swinney, D. (2006). A functional neuroimaging investigation of the roles of structural complexity and task-demand during auditory sentence processing. *Cortex*, 42(4), 577–590.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, 53(7), 433–440.

Meltzer, J. A., McArdle, J. J., Schafer, R. J., & Braun, A. R. (2010). Neural aspects of sentence comprehension: Syntactic complexity, reversibility, and reanalysis. *Cerebral Cortex (New York, N.Y. : 1991)*, 20(8), 1853–1864.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., & Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, 3, 76.

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17(2), 73–88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, 89(2), 277–289.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, 25(3), 653–660.

Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *NeuroImage*, 56(4), 2310–2320.

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010a). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, 52(4), 1410–1419.

Peelle, J. E., McMillan, C., Moore, P., Grossman, M., & Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain and Language*, 91(3), 315–325.

Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2010b). Neural processing during older adults' comprehension of spoken sentences: Age differences in resource allocation and connectivity. *Cerebral Cortex (New York, N.Y. : 1991)*, 20(4), 773–782.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., … Wingfield, A. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37 Suppl 1, 5S–27S.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.

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.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex (New York, N.Y. : 1991)*, 15(8), 1261–1269.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010a). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, 115(3), 182–188.

Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010b). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, 48(5), 1324–1335.

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. *Brain and Language*, 141, 89–102.

Ronnberg, J., Lunner, T., Ng, E. H., Lidestam, B., Zekveld, A. A., Sorqvist, P., … Stenfelt, S. (2016). Hearing impairment, cognition and speech understanding: Exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *International Journal of Audiology*, 55(11), 623–642.

Ruff, I., Blumstein, S. E., Myers, E. B., & Hutchison, E. (2008). Recruitment of anterior and posterior structures in lexical-semantic processing: An fMRI study comparing implicit and explicit tasks. *Brain and Language*, 105(1), 41–49.

Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., & Frisina, D. R. (2002). PET imaging of the normal human auditory

system: Responses to speech in quiet and in background noise. *Hearing Research*, 170(1–2), 96–106.

Samson, Y., Belin, P., Thivard, L., Boddaert, N., Crozier, S., & Zilbovicius, M. (2001). Auditory perception and language: Functional imaging of speech sensitive auditory cortex. *Revue Neurologique (Paris)*, 157(8–9 Pt 1), 837–846.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech–a positron emission tomography study. *Journal of the Acoustical Society of America*, 120(2), 1075–1083.

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America*, 115(2), 813–821.

Sehm, B., Schnitzler, T., Obleser, J., Groba, A., Ragert, P., Villringer, A., & Obrig, H. (2013). Facilitation of inferior frontal cortex by transcranial direct current stimulation induces perceptual learning of severely degraded speech. *Journal of Neuroscience*, 33(40), 15868–15878.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3), 277–287.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.)*, 270(5234), 303–304.

Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J., 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.

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105.

Takeichi, H., Koyama, S., Terao, A., Takeuchi, F., Toyosawa, Y., & Murohashi, H. (2010). Comprehension of degraded speech sounds with m-sequence modulation: An fMRI study. *NeuroImage*, 49(3), 2697–2706.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York, NY: Thieme Medical Publishers.

Tuennerhoff, J., & Noppeney, U. (2016). When sentences live up to your expectations. *NeuroImage*, 124(Pt A), 641–653.

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, 33(1), 1–13.

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., ... Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: Function, structure and behaviour in patients with left hemisphere damage. *Brain*, 134(2), 415–431.

Vaden, K. I., Jr., Kuchinsky, S. E., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2015). Cortical activity predicts which older adults recognize speech in noise and when. *Journal of Neuroscience*, 35(9), 3929–3937.

Vaden, K. I., Jr., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *Journal of Neuroscience*, 33(48), 18979–18986.

Vaden, K. I., Jr., Kuchinsky, S. E., Keren, N. I., Harris, K. C., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2011). Inferior frontal sensitivity to

common speech sounds is amplified by increasing word intelligibility. *Neuropsychologia*, 49(13), 3563–3572.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 577.

Vitello, S., Warren, J. E., Devlin, J. T., & Rodd, J. M. (2014). Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Frontiers in Human Neuroscience*, 8, 530.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective &Amp; Behavioral Neuroscience*, 15(2), 395–415.

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012a). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, 60(2), 1490–1502.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012b). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, 32 (40), 14010–14021.

Wong, P. C., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009a). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, 47(3), 693–703.

Wong, P. C., Perrachione, T. K., & Margulis, E. H. (2009b). Effects of asymmetric cultural experiences on the auditory pathway: Evidence from music. *Annals of the New York Academy of Sciences*, 1169, 157–163.

Wong, P. C., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *Journal of Speech, Language, and Hearing Research*, 51(4), 1026–1041.

Wright, B. A., Wilson, R. M., & Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task. *Journal of Neuroscience*, 30(35), 11635–11639.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS One*, 8(7), e68910.

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, 101, 76–86.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, 31(4), 480–490.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, 32(4), 498–510.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Ronnberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language*, 122(2), 103–113.