Neuropsychologia 50 (2012) 77–84

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Neural dissociation in processing noise and accent in spoken language comprehension

Patti Adank[a,b,*], Matthew H. Davis[c], Peter Hagoort[b,d]

[a] School of Psychological Sciences, University of Manchester, Manchester, United Kingdom
[b] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands
[c] Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom
[d] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 7 January 2011
Received in revised form 9 October 2011
Accepted 31 October 2011
Available online 4 November 2011

Keywords:
Speech
Accent
Noise
Auditory cortex
Prefrontal cortex
fMRI

ABSTRACT

We investigated how two distortions of the speech signal – added background noise and speech in an unfamiliar accent – affect comprehension of speech using functional Magnetic Resonance Imaging (fMRI). Listeners performed a speeded sentence verification task for speech in quiet in Standard Dutch, in Standard Dutch with added background noise and for speech in an unfamiliar accent of Dutch. The behavioural results showed slower responses for both types of distortion compared to clear speech, and no difference between the two distortions. The neuroimaging results showed that, compared to clear speech, processing noise resulted in more activity bilaterally in Inferior Frontal Gyrus, Frontal Operculum, while processing accented speech recruited an area in left Superior Temporal Gyrus/Sulcus. It is concluded that the neural bases for processing different distortions of the speech signal dissociate. It is suggested that current models of the cortical organisation of speech are updated to specifically associate bilateral inferior frontal areas with processing external distortions (e.g., background noise) and left temporal areas with speaker-related distortions (e.g., accents).

1. Introduction

Humans can generally understand each other despite a range of naturally occurring distortions to the speech signal (Mattys, Brooks, & Cooke, 2009). Some distortions are related to properties of the channel between conversation partners (*external distortions*), such as background noise, interruptions or signal degradation due to speaking over a telephone (Davis & Johnsrude, 2003; Wong, Uppanda, Parrish, & Dhar, 2008). Other distortions are related to the way in which the speaker produces utterances (*speaker-related distortions*), e.g., those due to anatomical and physiological differences between speakers, differences in speech rate, accent differences related to speakers' regional background, and differences related to speech style such as mumbling, reading aloud versus spontaneous speech (Adank, Evans, Stuart-Smith, & Scott, 2009; Dupoux & Green, 1997; Floccia, Goslin, Girard, & Konopczynski, 2006; Peterson & Barney, 1952). Both external and speaker-related distortions negatively affect speech processing: comprehension is generally slower, more effortful, and listeners make more errors for distorted input (e.g., Adank et al., 2009; Dupoux & Green, 1997; Munro & Derwing, 1995; Plomp & Mimpen, 1979).

Recent behavioural studies have addressed the cognitive mechanisms underlying the way listeners process variations in the speech signal. For added background noise, it was shown that listeners rely on fine acoustic cues in the speech signal (Mattys, White, & Melhorn, 2005). Mattys et al. demonstrated that listeners rely more on coarticulatory word-boundary cues in background noise than in quiet. Studies on how listeners process accent variation show that listeners perceptually compensate for speaker idiosyncrasies, such as ambiguous speech segments (Norris, McQueen, & Cutler, 2003) and accent variation (Evans & Iverson, 2003), by retuning their internal (phonetic) category boundaries to fit those of the speaker when confronted with this type of variation.

In the last decade, neuroimaging studies have studied neural underpinnings of the ability to successfully comprehend distorted speech. Studies investigating the effect of external distortions on speech processing generally do so use added background noise (Davis & Johnsrude, 2003; Scott, Rosen, Wickham, & Wise, 2004; Wong et al., 2008). Wong et al. presented words in background noise (multi-talker babble) in various signal-to-noise (SNRs) in decibel (dB) ratios (quiet, +20 dB, −5 dB) and reported an increase in the Blood-Oxygenated Level Dependent (BOLD) response for decreasing SNRs across a wide network of cortical regions, involving Superior Temporal Gyrus (STG) including Heschl's Gyrus (HG)

* Corresponding author at: School of Psychological Sciences, University of Manchester, Zochonis Building, Brunswick Street, Manchester M139PL, United Kingdom. Tel.: +44 161 2752693.
E-mail address: patti.adank@manchester.ac.uk (P. Adank).

0028-3932/$ – see front matter.
doi:10.1016/j.neuropsychologia.2011.10.024

bilaterally, left Inferior Frontal Gyrus (IFG) and the Frontal Operculum (FO). Davis and Johnsrude presented sentences in various types of distortion including added background noise and speech interrupted with bursts of noise (segmented speech). A number of left lateralised regions (STG, MTG, IFG, Precentral Gyrus and Thalamus) showed an elevated response to distorted speech compared to clear speech and unintelligible noise. They also found that left FO and posterior areas of Middle Temporal Gyrus (MTG) and STG responded differently to the three distortions. It thus appears that *temporal* areas associated with lower level auditory processing (such as HG) as well as *frontal* areas commonly found to be associated with higher level processing of intelligible speech (including FO and IFG) show sensitivity to background noise.

In addition, several studies evaluating speaker-related distortions investigated the neural bases of processing the effect of variations in speech rate using artificially time-compressed stimuli (Peelle, McMillan, Moore, Grossman, & Wingfield, 2004; Poldrack et al., 2001). Peelle et al. and Poldrack et al. report an increase in BOLD activity in temporal regions (including posterior STG) for processing time-compressed speech compared to normal-speed speech. Others investigated the effect of listening to speech in an unfamiliar accent (Adank, Noordzij, & Hagoort, 2011). Adank et al. found that processing speech in an unfamiliar accent leads to greater activity in posterior STG bilaterally compared to processing speech in a familiar accent. Finally, one study assessed perception of mispronunciations that can be heard as real words (Kotz, Cappa, Von Cramon, & Friederici, 2002; Raettig & Kotz, 2008) and found increased activation in STG bilaterally. Speaker-related distortions thus appear to be processed predominantly in *temporal* areas such as STG.

In the present study we will investigate whether the neural bases for processing external and speaker-related distortions dissociate into frontal areas for external distortions and temporal areas for speaker-related distortions for two types of distortions used in previous neuroimaging studies. We selected one type of external distortion, i.e., added background noise, and a specific type of speaker-related distortion, i.e., speech in an unfamiliar accent.

A challenge in studying differences between added background noise and accent is the possibility of confounds due to differences in speech intelligibility. For instance, Adank et al. (2011) presented sentences in Standard Dutch and in unfamiliar accent of Dutch. Sentences in an unfamiliar accent are generally less intelligible than sentences in a familiar accent (Adank et al., 2009). We therefore aimed to assess the neural responses to speech in an unfamiliar accent and speech with added background noise while equating intelligibility. We predict that added background noise will be processed predominantly in regions in inferior frontal areas and temporal areas. In addition, we predict that processing accented speech will lead to increased activity in areas found to be active for time-compressed speech and an unfamiliar accent, i.e., posterior STG. It has been suggested that areas in posterior STG serve as a computational hub for processing spectrotemporal variation in the speech signal (Griffiths & Warren, 2002). On this basis, we might expect that spectrotemporal variation due to processing accent variation in the speech signal will recruit STG.

We used sparse functional Magnetic Resonance Imaging (fMRI) (Hall et al., 1999) to compare BOLD responses to three types of sentences: (1) sentences in the listeners' native accent in quiet, (2) sentences in a familiar accent with added background noise and (3) sentences in an unfamiliar accent in quiet.

2. Methods

2.1. Participants

We tested twenty-six participants (20 F and 6 M, mean 21.2 years, range 18–28 years). All participants were right-handed, native monolingual speakers of Dutch, with no history of oral or written language impairment, or neurological or psychiatric disease. All gave written informed consent and were paid for their participation or received course credit. The study was approved by the local ethics committee.

Table 1
Intended vowel conversions for obtaining the novel accent. The left column shows the altered orthography in Standard Dutch, and the right column shows the intended change in pronunciation of the vowel in broad phonetic transcription, using the International Phonetic Alphabet (IPA, 1999).

Orthography	Phonetic (IPA)
a→aa	/ɑ/→/a:/
aa→a	/a:/→/ɑ/
e→ee	/ɛ/→/e:/
ee→e	/e:/→/ɛ/
i→ie	/I/→/i:/
ie→i	/i:/→/I/
o→oo	**/ɔ/** →/o:/
oo→o	/o:/→**/ɔ/**
uu→u	/y:/→/Y/
u→uu	/Y/→/y:/
oe→u	/u/→/Y/
eu→u	**/ø/** →/Y/
au→oe	**/ɔ/** u/→/u/
ei→ee	/ɛi/→/e:/
ui→uu	**/œy/** y/→/y:/

2.2. Materials

The stimulus set consisted of 204 sentences recorded in Standard Dutch and in an unfamiliar (novel) accent. This novel accent, as used previously (Adank, Hagoort, & Bekkering, 2010; Adank et al., 2011), was created by instructing the speaker to read sentences with an adapted orthography. The orthography was systematically altered to achieve the following changes in all 15 Dutch vowels as listed in Table 1. All sentences are listed in Appendix A. Only vowels bearing primary or secondary stress were included in the orthography conversion. An example of a sentence in Standard Dutch and a converted version is given below, including a broad phonetic transcription using the International Phonetic Alphabet (IPA, 1999):

Standard Dutch: "De bal vloog over de schutting"
/də bɑl flox o:fər də sxʏtɪŋ/
[The ball flew over the fence]
After conversion: "De baal flog offer de schuuttieng"
/də ba:l flo:x ɔfər də sxy:tiŋ/

The stimulus materials used in the scanner were created as follows. First, the speech rate differences across the tokens of a specific sentence (matched pairs of Standard Dutch and artificial accent tokens) were equalised using PSOLA (Moulines & Charpentier, 1990), as implemented in the Praat software package, version 4.501 (Boersma & Weenink, 2003), so that every token for a given sentence had the same length for *clear* and *accent*. The sentences in background noise (*noise*) were created by adding continuous speech-shaped noise to the Standard Dutch sentences in quiet. Noise was added using Matlab (Mathworks) so that the resulting sentence was played at a signal-to-noise ratio (SNR) of +2 dB. Sentences in all conditions were subsequently peak-normalised using Praat. This SNR was determined in a behavioural pilot test with 11 Dutch listeners. Participants in this test were presented with 200 sentences, 100 true, 100 false, in five conditions of 40 sentences each (sentences 1–100 and 101–200 in Appendix A). The Standard Dutch sentences were presented in quiet and at three signal-to-noise levels (+2 dB, 0 dB, −2 dB), while the sentences in the novel accent were presented in quiet only. Participants were instructed to decide as quickly as possible whether the sentence was true (e.g., "Makrelen ademen door kieuwen", *Mackerels breathe through gills*) or false ("Giraffes zijn fruit", *Giraffes are fruit*). Responses were measured from the offset of the sound file, allowing for negative response times. Responses were made using a computer keyboard by pressing 'p' with the index finger of the right hand (true responses) and by pressing 'q' with the index finger of the left hand (false response). The results for the response times (RTs) and percent error are shown in Fig. 1. A one-way ANOVA was run on the RTs with *condition* (Standard Dutch in quiet, at +2 dB, at 0 dB and at −2 dB, and novel accent in quiet) as the independent variable. Only correct responses were included. A main effect of *condition* on the RTs was found, ($F(2.73, 27.30) = 23.77$, $p < 0.05$, Huyhn–Feldt-corrected for non-sphericity), and post hoc analyses showed that the sentences in Standard Dutch in quiet differed from all other conditions ($p < 0.01$, Bonferroni-corrected for multiple comparisons) and that the sentences in the novel accent differed only from the sentences in Standard Dutch, but not from the three noise levels +2 dB, 0 dB and −2 dB. We then calculated the % error per condition (cf. Fig. 1) and repeated the one-way ANOVA on the error percentages per participant per condition with condition as a factor. The results showed a main effect of *condition* ($F(2.87, 28.67) = 23.77$, $p < 0.05$, Huyhn–Feldt-corrected for non-sphericity), and post hoc analyses showed that the sentences in Standard Dutch in quiet differed from the

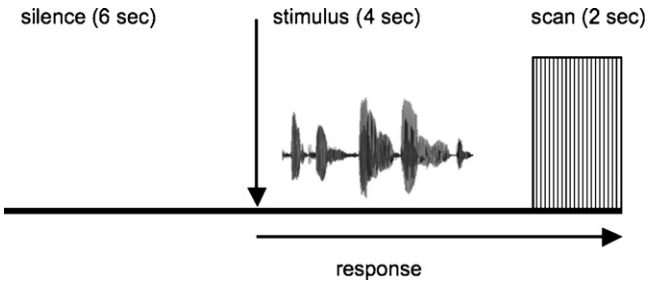

Fig. 1. Timeline of stimulus presentation and EPI acquisition during sparse scanning sequence. The repetition time (TR) is 12 s.

three conditions with added noise, but not from the sentences spoken in the novel accent. Second, the sentences in the novel accent in quiet differed significantly from the sentences presented at 0 dB and −2 dB ($p < 0.01$, Bonferroni-corrected for multiple comparisons). Based on these results, we decided to present all sentences in the speaker-external condition at a signal-to-noise ratio of +2 dB.

2.3. Design and procedure

One functional scan was obtained over 12 s (sparse temporal sampling, Hall et al., 1999), with the sentence onset at 4 s before each scan (cf. Fig. 2). A single trial began with a pause of 5000 ms, a tone with a duration of 200 ms, a pause of 300 ms and finally a single sentence was presented, followed by acquisition of a single volume 10 000 ms after the beginning of the trial (duration of single volume acquisition was 2000 ms). The procedure was similar to the pilot experiment. Participants were instructed to decide as quickly as possible whether the sentence was true or false and responses were measured from the offset of the sound file. Responses were made using a button box with the index finger (true responses) and middle (false response) finger of the right hand.

The study presented 204 sentences in three conditions: *clear* (Standard Dutch sentences in quiet), *noise* (Standard Dutch sentences with added background noise at an SNR of +2 dB) and *accent* (accented sentences in quiet). This three-way design permits comparison of effortless comprehension (*clear*) with either of two forms of effortful comprehension (*noise/accent*), which can also be directly compared to each other. Every unique sentence was presented only once in the experiment and all sentences were presented in a semi-randomised order and counterbalanced across conditions. True and false sentences were counterbalanced across conditions: each condition contained 34 true and 34 false sentences.

All participants confirmed their ability to hear and understand the sentences and the task during a familiarisation session in which six sentences in Standard Dutch in quiet (not included in the main experiment) were presented. Stimulus presentation was performed using Presentation (Neurobehavioral Systems, Albany, CA), running on a Pentium 4 with 2 GB RAM, and a 2.8 GHz processor. The presentation of all 204 trials (cf. Appendix A) lasted approximately 40 min.

2.4. Functional MRI data acquisition

Whole-brain imaging was performed at the Donders Institute for Brain, Cognition, and Behaviour, Centre for Cognitive Neuroimaging, using a 3T MR scanner (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany). The sentences were presented over headphones (MRConFon, Magdeburg, Germany) during sparse sampling acquisition (GE-EPI, repetition time = 12 s; TA, acquisition time = 2 s, echo time = 35 ms; 32 axial slices; slice thickness = 3 mm; voxel size = 3.5 × 3.5 × 3.5 mm; field of view = 224 mm; flip angle = 70°). All functional images were acquired in a single run. Listeners watched a fixation cross that was presented on a screen and viewed through a mirror attached to the head coil. After the acquisition of functional images, a high-resolution structural scan was acquired (T1-weighted MP-RAGE, 192 slices, repetition time = 2282 ms; echo time = 3.93 ms; field of view = 256 mm, slice thickness = 1 mm) was obtained. Total scanning time was 50 min.

2.5. Data analysis

The neuroimaging data were pre-processed and analysed using SPM8 (Wellcome Imaging Department, University College London, London, UK). The first volume of every functional run from each participant was excluded to minimise T1-saturation effects. Next, the time series were spatially realigned using a least-squares approach estimating six rigid-body transformation parameters (Friston et al., 1995) by minimising head movements between each image and a reference image, i.e., the first image in the time series. Subsequently, images were normalised onto a custom Montreal Neurological Institute (MNI)-aligned EPI template using both linear and nonlinear transformations and resampled at an isotropic voxel size of 2 mm. All participants' functional images were smoothed using an 8m6m FWHM Gaussian filter. Each participant's structural image was spatially co-registered to the mean of the functional images (Ashburner & Friston, 1997) and spatially normalised with the same transformational matrix applied to the functional images. A high-pass filter was applied with a 0.0039 Hz (256 s) cut-off to remove low-frequency components from the data, such as scanner drifts.

The fMRI time series were analysed within the context of the General Linear Model (GLM) using an event-related approach. One GLM was estimated per participant, consisting of three regressors: (1) Standard Dutch sentences spoken in quiet (*clear*), (2) Standard Dutch sentences in +2 dB SNR (*noise*) and (3) Dutch sentences in the unfamiliar accent (*accent*). We decided to exclude volumes associated with incorrect responses from the analysis as to assess successful speech comprehension only, and previous studies using similar sentence verification tasks also included only correct responses (e.g., Adank & Devlin, 2010). All regressors were estimated with a finite impulse response basis function (order 1 and window length 1) such that the response to each condition is estimated based on the single scan that followed each sentence. An additional six covariates were added to the GLM to capture head-movement effects estimated from the realignment stage of preprocessing. The least mean square parameter estimates were calculated for each voxel in each participant and contrasts of parameter estimates taken forward to a second-level analysis.

Linear weighted contrasts were used to identify six contrasts of interest. First, we identified regions that showed greater activation for *noise* than for either *clear* or *accent* (*noise* > *clear* and *noise* > *accent*). Second, we identified regions that showed increased BOLD-activity for the *accent* than the *clear* and *noise* conditions (*accent* > *clear* and *accent* > *noise*). Finally, we identified regions that showed increased BOLD-activity for processing both types of distortions versus no distortion (*clear* < [*noise* + *accent*]) and vice versa (*clear* > [*noise* + *accent*]).

The statistical thresholding of the second-level activation maps associated with these contrasts was an uncorrected threshold of $p < 0.001$ in combination with a minimal cluster extent of 43 voxels. This yields a whole-brain alpha of $p < 0.05$, determined using a Monte-Carlo Simulation with 1000 iterations, using a function implemented in Matlab (Slotnick, Moo, Segal, & Hart, 2003).

3. Results

3.1. Behavioural results

Fig. 3 shows the average response times and average error percentages for both groups per speech type. First, a repeated-measures ANOVA was run with the response times as the dependent variable and with *speech type* as an independent variable. Only correct responses were analysed. The results showed a main effect of *speech type* ($F(2, 54) = 41.51$, $p < 0.05$) on the response times, and post hoc analyses showed again that *noise* and *accent* differed from *clear* ($p < 0.017$, Bonferroni-corrected for multiple comparisons), while there was no difference between *noise* and *accent*. Second, a repeated-measures ANOVA was run with the average number of errors (missing responses were also coded as errors) per participant as the dependent variable and with *speech type* as an independent variable. The results showed a main effect of *speech type* ($F(1.56, 38.90) = 39.70$, $p < 0.05$, Huyhn–Feldt-corrected for non-sphericity) on the error rates. Post hoc analyses ($p < 0.017$, Bonferroni-corrected) indicated that both *noise* and *accent* differed from *clear*, while there was a significant difference between the error rates for *noise* and *accent* ($t(25) = 2.63$, $p < 0.017$). The behavioural results show that the stimuli in the *accent* and *noise* conditions were less intelligible than the sentences in *clear*, and that there was no difference in intelligibility between *accent* and *noise* in processing speed. A small difference was found in the error scores, with participants making slightly more errors for the accent condition, yet the relevance of this difference should not be overrated, as significantly more errors were made for both *accent* and *noise* than for *clear*, and there was no difference between the response times of *accent* and *noise*.

3.2. Neuroimaging results

For the contrast *noise* > *clear*, increases in BOLD-activity were seen bilaterally in IFG extending into FO (Fig. 4 and Table 2), and extending into left insula. Furthermore, increased activity was found in left caudate and right Cingulate Gyrus, Paracingulate/Cingulate Gyrus and right Frontal Pole. The results for *clear* > *noise* showed increased activity in left HG extending into left insula, in left Precentral Gyrus, left Inferior Frontal Gyrus (Pars

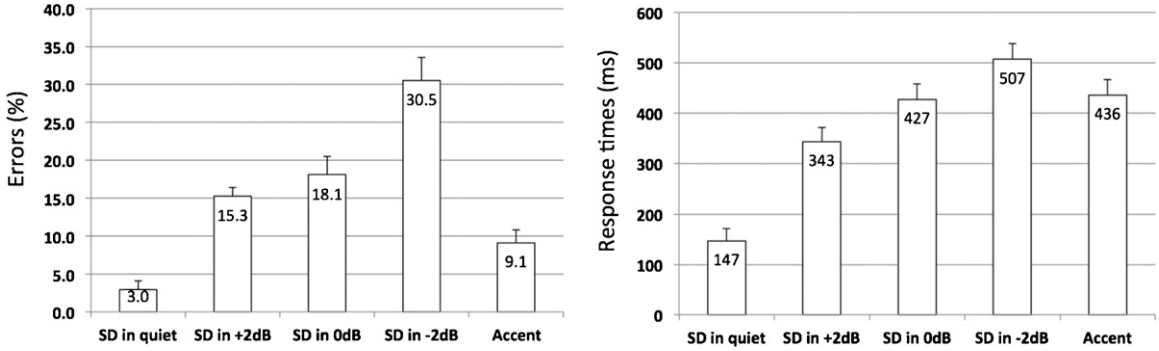

Fig. 2. Average response times (ms) and average percent error per condition (*SD in quiet*, *SD in +2 dB*, *SD in +2 dB*, *SD in −2 dB* and *Accent*) for the pilot study, SD: Standard Dutch. Error bars represent one standard error of the mean.

Triangularis), right insula, right frontal pole and Middle Temporal Gyrus (MTG) bilaterally.

The contrast *accent* > *clear* showed an increase in left STG/STS. *Clear* > *accent* showed an increase in left Angular Gyrus extending into left Supramarginal Gyrus, an increase in right Paracingulate/Cingulate Gyrus and finally in left Frontal Pole.

The contrast *noise* > *accent* resulted in increases in BOLD activity in right IFG and extending into FO and MFG, left insula, and in right FO extending to the insula, and Paracingulate/Cingulate Gyrus bilaterally. The results for the contrast *accent* > *noise* showed increases in STG/HG bilaterally, with activation extending into PT on the left and PT and the insula on the right.

Finally, the contrast *clear* < [*noise* + *accent*] showed increased BOLD-activity in right Frontal Pole (FP) extending into right insula, while *clear* > [*noise* + *accent*] showed activation peaks in left Parietal Opercular Cortex extending into HG, in left Precentral Gyrus, extending into FP and Superior Frontal Gyrus, bilateral MTG, Lateral Occipital Cortex (LOC).

4. Discussion

The aim of the study was to identify the neural bases associated with processing external (*noise*) and speaker-related (*accent*) distortions during speech comprehension. Our results indicate that despite being relatively well matched on difficulty, different sets of brain regions show additional activation when processing these two types of naturally occurring distortions of the speech signal relative to processing speech in a familiar accent in quiet listening conditions (*clear*). Comprehension of speech with added background noise results in increased BOLD-activation in bilateral inferior frontal areas, including IFG and FO, compared to processing clear speech. Processing accented speech recruits a left temporal area on the border between STG and STS more than processing clear speech. Processing clear speech versus the two distortions combined resulted in more activity in a network of frontal and temporal areas, including left frontal (e.g., Precentral Gyrus, FO), bilateral temporal (e.g., MTG, HG) and right insula, areas commonly found for processing intelligible speech (e.g., Davis & Johnsrude, 2003).

4.1. Noise versus clear speech

Processing speech in background noise recruits bilateral FO, IFG and the insula to a greater extent than processing clear speech. The results for processing background noise differ from Wong et al. (2008) and Davis and Johnsrude (2003), as we found increased BOLD activity only in frontal areas compared to processing speaker-related distortions and clear speech, while these previous studies found increases in both frontal and temporal areas. We may not have found additional activity for temporal areas because of our efforts to equate the relative intelligibility of both distortions, which was not done in Wong et al. (2008). Alternatively, this difference between Wong et al.'s results and our results may be attributable to stimulus differences, as the present study used speech-shaped noise in the noise condition, while Wong et al. embedded their sentences in multi-speaker babble. Davis and Johnsrude (2003), on the other hand, showed that superior temporal activation was particularly elevated for a distortion created by interrupting speech with bursts of noise, whereas inferior frontal regions showed an elevated response to distorted speech compared to clear speech that was common to three different forms of external distortion. It might be that speech perceived in fluctuating or changing background sounds creates additional activity in regions of Superior Temporal Gyrus, but that processing speech in background noise (any form of energetic masking) relies to a greater extent on activation of inferior frontal regions. Note that, right IFG showed more activation for processing background noise than for processing clear speech in our results, while Wong et al. and Davis and Johnsrude report increases for processing distorted speech in left IFG. Finally, a recent

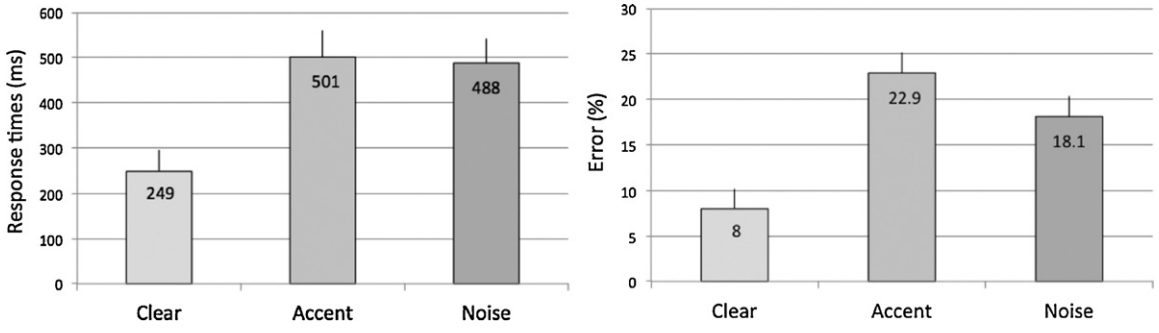

Fig. 3. Average response times (ms) and average percent error per condition (*clear*, *noise* and *accent*). Error bars represent one standard error.

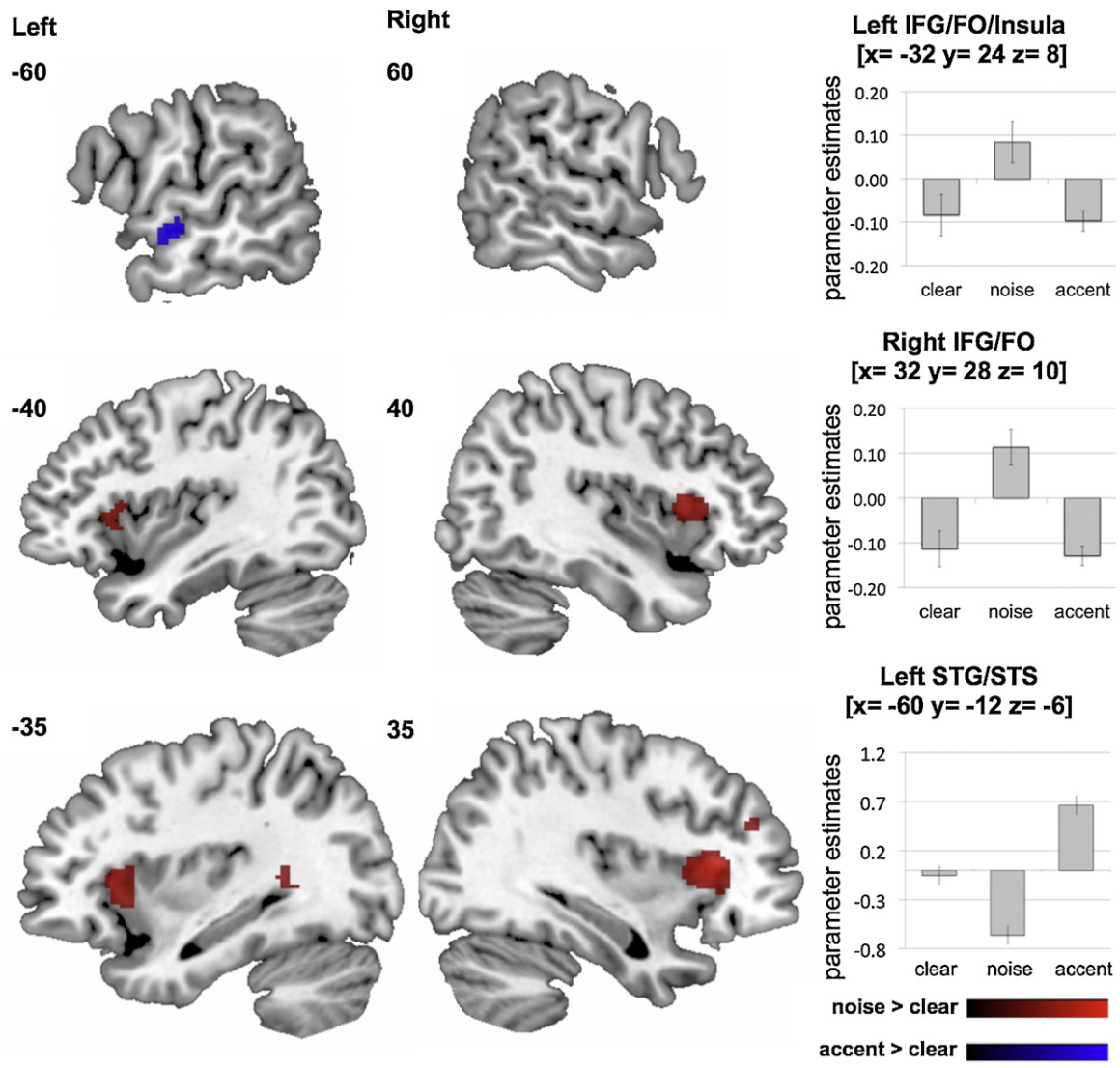

Fig. 4. Red: increases for *noise* > *clear*, red: increases for *accent* > *clear*, *t*-values indicated in the legend. FO: Frontal Operculum, IFG: Inferior Frontal Gyrus, STG: Superior Temporal Gyrus, STS: Superior Temporal Sulcus.

meta-analysis (Vigneau et al., 2011) suggests that activation in IFG during sentences processing may not be specific to the linguistic content of spoken sentences, but instead reflects increased involvement of working memory and attentional resources. Our results seem consistent with this suggestion, as we compared successful sentence processing in clear speech with background noise.

4.2. Accent versus clear speech

Processing accent variation recruited a small area in left STG/STS to a greater extent than processing clear speech. This finding is in line with previous studies (Adank et al., 2011; Raettig & Kotz, 2008); Raettig and Kotz report increased activations in anterior and posterior STG for processing pseudowords that can – with effort – be perceived as instances of familiar words (a common occurrence for foreign accented speech). Adank et al. (2011) report increases in left STG/STS for processing unfamiliar accented speech versus familiarly accented speech. Our results for accent variation are moreover largely in line with those of previous studies that also show activity in posterior temporal regions for processing acoustic-phonetic variation in speech (Adank et al., 2011; Peelle et al., 2004; Poldrack et al., 2001).

4.3. Noise and accent

Scott et al. (2004) contrasted different types of added distortions, namely added speech-shaped noise ("speech-in-noise"), and added speech from a competing speaker ("speech-in-speech") in a Positron Emission Tomography (PET) study. Scott et al. report increased activation in left prefrontal areas and right parietal areas for the contrast speech-in-noise > speech-in-speech, and bilateral activations for the contrast speech-in-speech > speech-in-noise in anterior STG extending posteriorly to HG. Our results for the contrasts *noise* > *accent* and *accent* > *noise* show a pattern similar to Scott et al.; when the speech is masked by background noise, activations can be observed in (pre)frontal areas. When listeners were listening to speech in an unfamiliar accent, temporal regions including STG bilaterally were more active than for processing added background noise. In our study, processing the variation in an unfamiliar accent recruits regions that were also activated in for processing the speech masker in Scott et al. Scott et al. propose several explanations for the neural dissociation between both speech maskers in posterior STG/HG bilaterally. First, they suggest that the increased activation for speech-in-speech may reflect the involvement of posterior STG/HG in grouping of auditory objects (i.e., separating the voices of the two speakers in the stimulus in

Table 2
Activation for peaks separated by at least 8 mm for the contrasts *noise* > *clear*, *clear* > *noise*, *clear* > *accent*, *accent* > *clear*, *noise* > *accent*, *accent* > *noise*, *clear* > (*noise* + *accent*) and (*noise* + *accent*) > *clear*. Coordinates in MNI standard space. FO: Frontal Operculum, FP: Frontal Pole, FOC: Frontal Orbital Cortex, HG: Heschl's Gyrus, IFG: Inferior Frontal Gyrus, LOC: Lateral Occipital Cortex, MFG: Middle Frontal Gyrus, MTG: Middle Temporal Gyrus, PT: Planum Temporale; SFG: Superior Frontal Gyrus, SMG: Supramarginal Gyrus, STG: Superior Temporal Gyrus.

Structure	Hemisphere	Size (voxels)	*x*, *y*, *z*	*T*	*Z*
Noise > *clear*					
IFG/FO	Right	925	32, 28, 10	6.49	4.92
FO/IFG/Insula	Left	265	−32, 24, 8	5.20	4.24
Cingulate Gyrus	Right	59	6, −14, 28	4.91	4.07
Parahippocampal Gyrus	Left	144	−24, −40, −2	4.63	3.90
Caudate	Left	267	−12, 10, −2	4.46	3.79
Paracingulate/Cingulate Gyrus	Right	88	12, 20, 36	4.31	3.69
Frontal Pole	Right	74	30, 40, 24	4.20	3.62
Cingulate Gyrus	Right	53	8, 22, 18	4.19	3.61
Clear > *noise*					
HG/Insula	Left	1118	−40, −22, 12	7.36	5.32
Precentral Gyrus	Left	1022	−40, −20, 54	5.61	4.47
Insula	Right	1128	34, −22, 10	5.46	4.39
IFG/PT	Left	68	−36, 22, −26	5.39	4.35
FP/FOC	Right	69	34, 42, −14	4.59	3.88
MTG	Left	192	−54, −44, −6	4.27	3.66
Subcortical	Right	50	14, −50, −20	4.15	3.59
FP/SFG	Left	65	−8, 50, 40	4.12	3.57
MTG	Right	55	66, −42, −8	4.10	3.55
MTG	Left	192	−54, −44, −6	4.27	3.66
Accent > *clear*					
STG/STS	Left	56	−60, −12, −6	4.70	3.94
Clear > *accent*					
FP	Left	71	−14, 56, 40	4.43	3.77
Paracingulate/Cingulate Gyrus	Right	57	14, 46, 16	4.08	3.54
Angular Gyrus/SMG	Left	222	−52, −56, 46	4.07	3.53
Noise > *accent*					
FO/IFG	Right	970	32, 28, 8	6.38	4.87
FP/MFG	Right	133	30, 42, 28	5.31	4.31
Insula	Left	309	−34, 18, 0	5.30	4.30
Paracingulate/Cingulate Gyrus	Right	197	10, 22, 40	5.25	4.27
Paracingulate/Cingulate Gyrus	Left	224	−8, 34, 24	4.95	4.09
Accent > *noise*					
STG/HG/PT	Left	2247	−40, −20, 8	8.25	5.68
STG/HG/Insula	Right	2003	46, −14, 4	7.86	5.53
Intracalcarine cortex/Lingual		113	12, −72, 6	3.94	3.44
Clear < [*noise* + *accent*]					
FP/Insula	Right	190	32, 36, 10	5.27	4.22
Subcortical	Left	47	−12, 10, −2	3.82	3.36
[*Noise* + *accent*] < *clear*					
Parietal Opercular Cortex/HG	Left	366	−42, −24, 16	5.62	4.48
Precentral Gyrus	Left	734	−40, −20, 54	5.39	4.35
FOC/FP	Left	112	−36, 22, −24	5.30	4.30
FP/SFG	Left	214	−10, 52, 42	4.99	4.12
MTG	Left	163	−62, −50, −6	4.41	3.76
LOC	Left	310	−34, −68, 36	4.28	3.67
MTG	Right	92	62, −42, −8	4.15	3.58
MFG	Left	46	−44, 16, 46	4.03	3.50
FP	Right	78	16, 50, 50	3.90	3.42

the speech-in-speech condition). A second interpretation is that the activations in posterior STG/HG reflect "glimpsing" processes (Cooke, 2003), due to larger portions of the speech signal being available for processing in the speech-in-speech than in the speech-in-noise conditions. It seems possible that our results are partially explained by the greater availability of the speech in the accent condition than in the noise condition, as no masker was used in the accent condition. However, it does not seem plausible that our results for *accent* > *noise* may be explained by the involvement of auditory objects segregation processes, as the distortion in our *accent* condition was not constructed by adding a masker signal as in Scott et al. Instead, the distortion was contained within the sentence spoken by the speaker, or intrinsic to the stimulus. Instead, we propose that the activations in posterior STG/HG, extending to PT may reflect additional phonetic/phonological processing for the accent versus the noise condition (Griffiths & Warren, 2002).

5. Conclusion

A variety of studies have sought to uncover the neural bases of processing intelligible speech (Obleser, Wise, Dresner, & Scott, 2007; Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006) by assessing comparing stimulus conditions in which speech in relatively quiet listening conditions was compared with unintelligible speech, or unintelligible speech. However, speech comprehension in daily life rarely occurs in quiet listening conditions; people often have to converse in background noise and tolerate variability in the form of speech due to accents and other forms of speaker-related variation. The dual-stream model proposed in Hickok and Poeppel (2007) suggests several routes to successful speech comprehension, and proposes a key role for areas in the temporal lobes bilaterally and frontal left-lateralised areas. They suggest that the activations of these areas vary depending on the ambient listening

conditions. Our results provide support for this model by demonstrating that speech comprehension under challenging conditions recruits different cortical areas to a greater or lesser extent based on the specific type of distortion. However, our results differ from the proposed model as we found activation in right instead of left frontal areas as well as in bilateral temporal areas. Finally, the present study represents a next step in identifying and differentiating between different neural mechanisms involved in different strategies for effortful listening that play a critical role in natural comprehension.

In sum, our results show dissociation in the neural bases for processing two types of speech distortions. Compared to clear speech, processing added background noise recruits cortical areas also involved in higher order processing (such as semantics and syntax) of spoken language, such as the insula and IFG (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Rodd, Davis, & Johnsrude, 2005). Processing accented speech seems to rely to a greater extent on cortical areas involved mainly in lower order, auditory, processing, such as STS/STG. This neural dissociation may support listeners' remarkable ability to understand speech under a variety of adverse listening conditions.

Our results have implications for theories on processing of distortions in speech. Most neural models for processing intelligible speech do not explicitly address this issue (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003). However, others have suggested that an increased cognitive load as a result from processing distortions/variations in the speech signal lead to the recruitment of additional neural resources (Skipper, Nusbaum, & Small, 2006), thought to be located in areas associated with speech production, such as left IFG. Our results suggest that current models for neural processing of speech are to be updated to incorporate hypotheses about the comprehension of speech under adverse listening conditions. Furthermore, our results suggest that the neural systems associated with processing these distortions are not generic, but may differ as function of the type of distortion applied. Our results imply that distortions that are related to the speaker or other paralinguistic variations including regional accents are processed differently from distortions that do not affect the acoustic signal as produced by the speaker, but that somehow affect the signal in the environment. We therefore suggest that existing models are modified to reflect cognitive mechanisms involved in processing speech under challenging listening conditions. Specifically, we propose that current models link processing of external distortions (e.g., speech distorted by added noise) with bilateral inferior frontal areas, and processing speaker-related distortions (e.g., phonetic/phonological variation related to the speaker's accent or style of speech) with left temporal areas.

Finally, from a cognitive point of view, it may be the case that these distortions are best compensated by processes operating at different levels in the linguistic hierarchy. For example, speaking with a regional accent may introduce variation at phonetic and phonological levels that can be adapted to[1] (Adank et al., 2010; Maye, Aslin, & Tanenhaus, 2008). In contrast, background noise affects the signal at mostly acoustic levels and mechanisms of perceptual compensation may be less effective (Peelle & Wingfield, 2005) and comprehension is best achieved by using higher level cues. However, more studies are required to explicitly address the effect of different distortions operating at acoustic, phonetic/phonological, and higher levels to establish the cortical networks associated in effectively processing such distortions.

[1] We made sure that no perceptual adaptation took place by using a design in which accented sentences were not presented in blocks (cf. Adank & Devlin, 2010). During the analysis, we have verified whether perceptual adaptation occurred for the accented speech (also for the other two conditions), by setting up first-level models contrasting the first with the second half of the stimuli per condition and testing for an interaction between experimental half and accent. No areas showed such an interaction at uncorrected levels ($p < 0.001$). In addition, we also checked for a similar interaction in the behavioural stimuli, and also found no effect.

Acknowledgments

We wish to thank Paul Gaalman for technical assistance, Benedikt Poser for his help in the experimental design, and Shirley-Ann Rueschemeyer for useful comments. This research was supported by the Netherlands Organization for Research (NWO) under project number 275-75-003.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuropsychologia.2011.10.024.

References

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, *49*(1), 1124–1132.
Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology Human Perception and Performance*, *35*(2), 520–529.
Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903–1909.
Adank, P., Noordzij, M. L., & Hagoort, P. (2011). The role of Planum Temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping*,
Ashburner, J., & Friston, K. (1997). Multimodal image coregistration and partitioning – A unified framework. *NeuroImage*, *6*, 209–217.
Boersma, P., & Weenink, D. (2003). *Praat: Doing phonetics by computer*. http://www.fon.hum.uva.nl/praat (retrieved 11.8.08)
Cooke, M. (2003). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, *119*(3), 1562–1573.
Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neurscience*, *23*(8), 3423–3431.
Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914–927.
Evans, B. G., & Iverson, P. (2003). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, *115*(1), 352–361.
Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1276–1293.
Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *2*, 165–189.
Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neuroscience*, *25*(7), 348–353.
Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.
Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliot, M. R., et al. (1999). Sparse temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*, 213–223.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
IPA. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
Kotz, S. A., Cappa, S. F., Von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the lexical-semantic network by auditory semantic priming: An event-related functional MRI study. *NeuroImage*, *17*, 1761–1772.
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203–243.
Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*, 477–500.
Maye, J., Aslin, R. N., & Tanenhaus, M. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*, 543–562.
Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5–6), 453–467.
Munro, M. J., & Derwing, T. M. (1995). Foreign accent comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

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

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*, 315–325.

Peelle, J. E, & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1315–1330.

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*, 175–184.

Plomp, R., & Mimpen, A. M. (1979). Speech reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America*, *66*(5), 1333–1342.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., et al. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *13*(5), 687–697.

Raettig, T., & Kotz, S. A. (2008). Auditory processing of different types of pseudowords: An event-related fMRI study. *NeuroImage*, *39*, 1420–1428.

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

Scott, S. K., Blank, S. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*, 100–107.

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

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (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.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 250–285). Cambridge, MA: Cambridge University Press.

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, *17*, 75–82.

Vigneau, M., Beaucousin, V. V., Herveĭ, P., Jobard, G., Petit, L., Crivello, F., et al. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*, 577–593.

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

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–1836.