Human Brain Mapping

The role of Planum Temporale in processing accent variation in spoken language comprehension

Journal:	*Human Brain Mapping*
Manuscript ID:	HBM-10-0238.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Adank, Patti; University of Manchester, School of Psychological Sciences Noordzij, Matthijs; University of Twente, Department of Cognitive Psychology and Ergonomics Hagoort, Peter; Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour
Keywords:	spoken language comprehension, fMRI, auditory, repetition suppression, planum temporale

THE ROLE OF PLANUM TEMPORALE IN PROCESSING ACCENT VARIATION IN SPOKEN LANGUAGE COMPREHENSION

Patti Adank[1,2], Matthijs L. Noordzij[2,3], and Peter Hagoort[2, 4]

[1] School of Psychological Sciences, University of Manchester, Manchester, United Kingdom

[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, the Netherlands

[3] Department of Cognitive Psychology and Ergonomics, University of Twente, Enschede, The Netherlands

[4] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Short title: Neural activity associated with processing variation in speech

Keywords: repetition-suppression, speech, speaker, auditory cortex

Correspondence:

Patti Adank

School of Psychological Sciences

University of Manchester

Zochonis Building

Brunswick Street

M20 6HL

Manchester, UK

Email: patti.adank@manchester.ac.uk

Phone: 0044-161-2752693

Abstract

A repetition-suppression fMRI paradigm was employed to explore the neuroanatomical substrates of processing two types of acoustic variation - speaker and accent - during spoken sentence comprehension. Recordings were made for two speakers and two accents: Standard Dutch and a novel accent of Dutch. Each speaker produced sentences in both accents. Participants listened to two sentences presented in quick succession while their haemodynamic responses were recorded in an MR scanner. The first sentence was spoken in Standard Dutch; the second was spoken by the same or a different speaker and produced in Standard Dutch or in the artificial accent. This design made it possible to identify neural responses to a switch in speaker and accent independently. A switch in accent was associated with activations in predominantly left-lateralised areas including posterior temporal regions, including Superior Temporal Gyrus (STG), Planum Temporale (PT), and Supramarginal Gyrus, as well as in frontal regions, including left Pars Opercularis of the Inferior Frontal Gyrus (IFG). A switch in speaker recruited a predominantly right-lateralised network, including Middle Frontal Gyrus and Prenuneus. It is concluded that posterior temporal areas, including PT, and frontal areas, including IFG, are involved in processing accent variation in spoken sentence comprehension.

Introduction

The human speech comprehension system seems to effortlessly extract the linguistic message from the acoustic signal. This is a remarkable feat, given the variability inherent to this signal, for instance as a result from speaker differences (Peterson and Barney 1952). These differences are not only anatomical/physiological in nature, but also emerge from social factors such as a speaker's geographical background and socioeconomic status. These social factors result in different spoken varieties of the standard language, which can be exemplified by phonetic and phonological variation in the sounds of a language (Wells 1982). For instance, the Dutch word *bed* (bed) is pronounced with the vowel /ɛ/ in the western part of the Netherlands, but with a vowel close to /a/ as in *bad* (bath) in the south-eastern part (Adank, et al. 2007). Listeners are continuously confronted with ambiguities in speech that they have to resolve perceptually (or, *normalise*) to extract the linguistic message (Nearey 1989). This disambiguating process requires cognitive effort; reflected in longer response times for comprehension of sentences spoken with an unfamiliar regional or foreign accent compared to listeners' native accent (Adank, et al. 2009; Floccia, et al. 2006; Rogers, et al. 2004; Van Wijngaarden 2001).

Behaviourally, listeners process accented speech by shifting their phonetic boundaries to match those of the speaker, when confronted with a speaker whose speech displays accent or specific idiosyncrasies (Evans and Iverson 2003; Norris, et al. 2003). It has finally been suggested that the adaptation process involves pattern matching mechanisms (Hillenbrand and Houde 2003; Nearey 1997) that are based on statistical learning (Nearey and Assmann 2007).

The neural bases underlying processing of accent-related variation are largely unknown. It has been hypothesized that the Planum Temporale (PT) is involved in processing complex spectrotemporal variation in speech (Griffiths and Warren, 2002; Warren, Wise, Warren, 2005). PT is a large region in the temporal lobe, posterior to Heschl's Gyrus in the posterior Superior Temporal Gyrus (STG) and represents the auditory association cortex. PT is involved in elementary acoustic pattern perception (Binder, et al. 2000; Giraud, et al. 2000; Hall, et al. 2002; Penhune, et al. 1998), spatial processing (Warren, et al. 2005) auditory scene analysis (Bregman 1990), musical perception (Zatorre, et al. 1994), and more specifically, speech perception, (Binder, et al. 1996; Giraud and Price 2001; Shtyrov, et al. 2000). PT is hypothesized to be involved in continuous updating of incoming traces required for phonological working memory and speech production (Binder, et al. 2000). Griffiths and Warren (Griffiths and Warren 2002) propose a functional model for the processing in PT of spectrotemporally complex sounds that change over time. PT continuously analyses these incoming signals and compares them with those previously experienced using pattern matching. Griffiths and Warren furthermore suggest that PT is associated with "…constructing a transient representation of the spectrotemporal structures embodied in spoken words, regardless of whether these are heard or retrieved from lexical memory (i.e. a phonological template.)".

The present study aimed to provide insights into the neural locus of processing accent and speaker variation using functional Magnetic Resonance Imaging (fMRI). We investigated whether PT is involved in disambiguation processes required for understanding accented speech using a repetition-suppression fMRI design. Repetition suppression is based on the finding that the repeated presentation of a stimulus induces a decrease in brain activity. This decrease can be detected using fMRI (Grill-Spector, et al. 1999; Grill-Spector and Malach 2001). This technique can be used to identify brain areas involved in processing specific stimulus characteristics.

By varying the property that is repeated, the neural bases involved in processing that specific property are uncovered. For example, repetition suppression paradigms have been employed to locate the neural substrates of speaker processing (Belin and Zatorre 2003), spoken syllables (Zevin and McCandliss 2005), spoken words (Orfanidou, et al. 2006), and spoken sentences (Dehaene-Lambertz, et al. 2006).

In the experiment, listeners heard two sentences presented in quick succession. The first sentence was spoken in Standard Dutch; the second sentence was spoken by the same or a different speaker, in Standard Dutch or in a novel accent of Dutch. This design allowed us to identify neural responses to a switch in speaker, in accent, or both. Recordings were made for a male and a female speaker of Dutch, to maximise the amount of variation related to anatomical/physiological differences between speakers. *Accent* and *speaker* were implemented in a factorial design with both factors crossed, allowing us to determine the neural bases associated with processing both variation types independently. Phonological/phonetic variation was introduced into the speech signal by creating an artificial, non-existing, accent. Using a non-existing accent has two advantages: first, speaker and accent were not confounded as both factors were manipulated independently. Second, the use of a novel accent ensures that all listeners are equally unfamiliar with the accent. This is necessary as familiarity with an accent affects language comprehension: spoken language comprehension slows when listeners are unfamiliar with the accent (Floccia, et al. 2006), especially in noisy conditions (Adank, et al. 2009).

Materials & Methods

Participants

Twenty participants (14F and 6M, mean 21.2 years, range 18-26 years) took part in the study, although the data from two (2F) were subsequently excluded due to i) to excessive head movement (>3mm) and ii) an unexpected brain anomaly. The remaining 18 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. The study was approved by the local ethics committee.

Experiment and Design

The present repetition-suppression fMRI experiment used a mini-block design, with continuous scanning. The choice of continuous rather than sparse sampling was based on a trade-off between the ability to reliably detect suppression in the blood oxygen level dependent (BOLD) signal and the length of the experiment. Continuous sampling results in both acoustic masking of the auditory sentences (Shah, et al. 1999) and contamination of the BOLD signal response in auditory regions (Bandettini, et al. 1998; Hall, et al. 1999; Talavage, et al. 1999). The former, however, was not a problem as a relatively quiet acquisition sequence (~80dB) coupled with sound attenuating headphones (~30dB attenuation) ensured the sentences were easily heard. Indeed, all participants confirmed their ability to hear and understand the sentences during a familiarisation session in which only sentences in Standard Dutch (not included in the main experiment) were presented. Contamination of the BOLD signal was potentially more problematic because scanner noise elevates BOLD responses in auditory areas (Gaab, et al. 2006; Hall, et al. 1999), and these effects need not be identical across regions (Tamer, et al. 2009; Zaehle, et al. 2007). In the current experiment, however, we were specifically interested in relative *reductions* in BOLD signal. As a result, elevated BOLD responses may not be problematic; only responses driven to saturation levels by the scanner noise would reduce sensitivity and previous studies have clearly shown that typical EPI sequences reduce, but do not eliminate, the dynamic range of the BOLD response (Gaab, et al. 2006; Zaehle, et al. 2007). Moreover, to avoid scanner-noise contamination and ensure an adequate sampling of the evoked haemodynamic response function (HRF) requires silent periods between volume acquisitions lasting between 16 and 32 seconds (Eden, et al. 1999; Edmister, et al. 1999; Hall, et al. 1999; Hickok, et al. 1997; Tamer, et al. 2009). A sparse design would therefore result in the experiment lasting up to twice as long as using a continuous design, which was deemed likely to reduce participants' ability to attend to the sentences. Consequently, we chose to use a continuous sampling paradigm.

Listeners were presented with two sentences in quick succession in four conditions as in Table I. The first sentence was always spoken in Standard Dutch, followed by the same sentence spoken by the same speaker in the same accent (condition SS, same speaker, same accent), spoken by a different speaker in the same accent (DS, different speaker, same accent, representing a switch of speaker), by the same speaker in a different accent (DS, same speaker, different accent, representing a switch of accent), or finally by a different speaker in a different accent (DSDA, different speaker, different accent, representing a switch of speaker and accent). Thirty-two sentences were presented per condition, in eight mini-blocks of four stimuli. Participants were required to listen to the sentences and to pay close attention. There was no additional task.

Insert Table I about here

Stimulus materials

The total stimulus set consisted of 256 sentences. The sentences were taken from the speech reception threshold corpus, or SRT (Plomp and Mimpen 1979). This corpus has been widely used for assessing intelligibility of different types of stimuli, e.g., for speech in noise (Zekveld, et al. 2006), or foreign-accented speech (van Wijngaarden, et al. 2002). The SRT consists of 130 sentences designed to resemble short samples of conversational speech. All consist of maximally eight or nine syllables, and do not include words longer than three syllables. Two versions of 128 of the SRT-sentences were recorded, in Standard Dutch and in the novel accent. The novel accent was designed to merely sound different from Standard Dutch, and was not intended to replicate an existing accent.

The novel accent, also used in (Adank and Janse 2010), 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: the switching of all tense-lax vowel pairs (e.g., /e:/ was pronounced as /ɛ/ and vice versa), /u/ (not having a lax counterpart in Dutch) was pronounced as /ʏ/, and all diphthongal vowels were realised as monophthongal vowels (e.g., /ɛi/ was pronounced as /e:/). All changes are listed in Table II and all sentences are listed in Appendix I. Only vowels bearing primary or secondary stress were included in the conversion of the orthography. 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 baːl floːx ɔfər də sxyːtiŋ/

Insert Table II about here

These sentences were recorded in both accents by a female and a male speaker of Dutch. The recordings were made in a sound-attenuated booth. Sentences were presented on the screen of a desktop computer. The speakers were instructed to read the sentences aloud as a declarative statement and with primary sentence stress on the first noun, as to keep the intonation pattern relatively constant across all sentences. First, two tokens were recorded of each Standard Dutch version followed by one token of the artificial accent version. Every sentence in the artificial accent was repeated until it was pronounced as instructed and sounded as fluent as the Standard Dutch sentences. The speakers were monitored from sentence to sentence during recording by the first author (a trained phonetician). After recording, the sentences were checked by the first author, and all sentences with mistakes were re-recorded, using the same procedure. Finally, 14 additional sentences were recorded in Standard Dutch for the control task in the fMRI experiment. All sentences were recorded to hard disk directly via an Imix DSP chip plugged into the USB port of an Apple Macbook.

Next, all sentences were saved into separate sound files with begin and end trimmed at zero crossings and re-sampled at 16 kHz. Subsequently, the speech rate differences across all six tokens of a specific sentence (two Standard Dutch tokens and one artificial accent token, for two speakers) were equalized, so that every token for a given sentence had the same length. This ensured that both sentences in each repetition-suppression stimulus pair were equally long. First, for each of the 128 sentences (4 experimental conditions × 32 sentences) the average duration across all six tokens for that sentence was calculated. Second, each token was digitally shortened or lengthened to fit the average length for the sentence, using PSOLA (Moulines and Charpentier 1990), as implemented in the Praat software package, version 4.501 (Boersma and Weenink 2003). Second, every sentence was peak-normalized at 99% of its maximum amplitude and then saved at 70 dB (SPL).

Procedure

The participants listened to the stimuli and were instructed to pay close attention and told that they would be tested after the experiment. A single trial (see Fig. 1) began with a tone signal of 200msec, followed by a pause of 300msec, the presentation of the first sentence of the pair (always in Standard Dutch), a pause of 300msec, and the second sentence of the pair. The inter-stimulus-interval was effectively jittered by adding a waiting period that was randomly varied between 4000-6000msec, to the offset of the second sentence. The average sentence duration was 2495msec (range 2074-3064msec).

Insert Figure 1 about here

To improve statistical power, trials occurred in short blocks of four sentences of one experimental condition, followed by a silent baseline trial (duration randomly varied from 4000-6000msec). The identity of the speaker did not vary across the first sentences of a pair in a mini-block. Every unique sentence was presented only once in the experiment and all were presented in a semi-randomised order and counterbalanced across conditions, so that the 128 sentences were presented in all four conditions across participants. The presentation of the 128 sentence trials and the 32 silent trials lasted 23 minutes. Before the main experiment, listeners were presented with six sentences (not included in the main experiment) spoken by the same speaker as in the main experiment to ensure that the sentences were audible over the scanner noise and presented at a comfortable loudness level (adapted for each individual participant).

Participants were informed that the sentences were presented in pairs and that the first sentence of a pair was always spoken in Standard Dutch and that the second one often varied in speaker, accent, or both. They were also informed that the sentence itself did not vary, i.e., that they would be presented with two tokens of the same sentence within a stimulus pair. Presenting participants with a sentence in Standard Dutch ensured that they would be able to understand the linguistic message and second, that they would not have to make additional cognitive efforts to understand the linguistic content of the second sentence in a pair.

After the main experiment, participants heard 28 single sentences in Standard Dutch. Half of these sentences had not been presented before (these sentences were not part of the SRT corpus, but had been constructed to resemble the SRT sentences as much as possible) - and the other half had been presented in the main experiment. After the main experiment had finished, participants responded through a button-press with their right index finger when they had heard the sentence in the main experiment. Stimulus presentation was performed using Presentation (Neurobehavioral Systems, Albany, CA), running on a Pentium 4 with 2 GB RAM, and a 2.8GHz processor.

Functional MRI data acquisition

Whole-brain imaging was performed at the Donders Centre for Brain, Cognition, and Behaviour, Centre for Cognitive Neuroimaging, at a 3T MR scanner (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany). The sentences were presented over electro-static headphones (MRConFon, Magdeburg, Germany) during continuous scanner acquisition (GE-EPI, repetition time = 2282msec; echo time = 35msec; 32

axial slices; slice thickness = 3 mm; voxel size = 3.5 x 3.5 x 3.5 mm; field of view = 224 mm; flip angle = 70º) – in other words, over the noise of the scanner. All participants confirmed their ability to hear and understand the sentences during a short practice session when the scanner was on. 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= 2282ms; echo time = 3.93ms; field of view = 256mm, slice thickness = 1mm). Total scanning time was 40 minutes.

Analyses

The neuroimaging data were pre-processed and analysed using SPM5 (Wellcome Imaging Department, University College London, London, UK). The first two volumes of every functional run from each participant were excluded from the analysis to minimize T1-saturation effects. Next, the image time series were spatially realigned using a least-squares approach that estimates six rigid-body transformation parameters (Friston, et al. 1995) by minimizing head movements between each image and the reference image, i.e., the first image in the time series. Next, the time series for each voxel was temporally realigned to acquisition of the middle slice. Subsequently, images were normalized onto a custom Montreal Neurological Institute (MNI)-aligned EPI template (based on 28 male brains acquired on the Siemens Trio at the Donders Institute for Brain, Cognition and Behaviour, Centre for Neuroimaging) using both linear and nonlinear transformations and resampled at an isotropic voxel size of 2mm. All participants' functional images were smoothed using an 8 mm

FWHM Gaussian filter. Each participant's structural image was spatially co-registered to the mean of the functional images (Ashburner and Friston 1997) and spatially normalised with the same transformational matrix applied to the functional images. A high-pass filter was applied with a 0.0078 Hz (128s) 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 using an event-related approach. Repetition suppression was operationally defined as the difference between stimulus pairs for each of the four condition (SS, DS, DA, and DSDA), following Noppeney & Penny's (2006) *categorical* approach for analysing repetition-suppression designs (see also (Chee 2009; Chee and Tan 2007; Henson, et al. 2004)). The rationale behind this approach is as follows: as the first sentence in each stimulus pair was always spoken in Standard Dutch, and all 128 sentences were randomised and counterbalanced across in all four conditions across all participants, it may be assumed that the first sentences did not differ systematically across conditions. Activation differences between conditions could therefore only be caused by the different patterns of neural suppression after presentation of the second sentence per condition, i.e., be due to an interaction between an overall suppression effect and the speaker or accent variation present in the second sentences.

Four events of interest (EV) were identified and entered into a subject-specific General Linear Model, consisting of the 32 stimulus pairs per condition [SS, DS, DA, DSDA]. All onsets within these events were modelled with a length equalling the duration of the both sentences presented and started at the onset of the first sentence in a stimulus pair. Parameter estimates were calculated for each voxel and contrast maps were constructed for each participant. Finally, the statistical model also considered six separate covariates describing the head-related movements (as estimated by the spatial realignment procedure).

Linear weighted contrasts were used to specify four contrasts. The conditions SS, DS, DA, and DSDA were analysed in an 2 × 2 factorial design with accent and speaker as factors. A switch of accent occurred in DA and DSDA, a switch of speaker in DS and DSDA, while SS was associated with neither a switch of accent or speaker. We determined main effects of each factor and the interaction term. A main effect of processing a switch of accent was assessed by (*DA* + *DSDA*) - (*SS* + *DS*), a main effect of processing a switch of speaker was assessed by (*DS* + *DA*) - (*SS* + *DA*), and the interaction term by (*SS* + *DSDA*) - (*DS* + *DA*).

The statistical thresholding of the second-level activation maps associated with these three contrasts was an uncorrected threshold of $p < 0.001$ in combination with a minimal cluster extent of 80 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, et al. 2003).

Results

Behavioural results

For each participant, the proportion of correct responses was calculated for the after-task. A response was correct whenever the participant had pressed the button and the sentence had been present in the main experiment, or whenever the participant had not pressed the button and the sentence had not been present in the main experiment. Participants correctly detected whether a sentence was present (or not) on average for 79.2% (SD 10.1%, range 60.7-96.4%) of the sentences, which is significantly higher than chance level (50%), $t(17)= 12.142$, $p < 0.05$. All individual participants' scores were significantly higher than chance level ($p < 0.05$). Given that all participants could judge whether a sentence had been present in the main experiment above chance level, it seems plausible that participants paid attention to the sentences played in the scanner in the main experiment.

Accent

We assessed which cortical regions showed an effect when a switch of accent was present, vs. no switch, (*DA* + *DSDA*) - (*SS* + *DS*). The coordinates of the peak voxels for these effects are listed in Table III and displayed in Figure 2. The peak of the relative increase in BOLD signal for a switch of accent was located in posterior STG bilaterally extending to the ventral supramarginal gyrus (SMG) and left PT. The activation appears more widespread on the left. A second cluster is found in left Inferior Frontal Gyrus (IFG), in Pars Opercularis (POp), extending into Pars Triangularis.

We ensured that the activations in left posterior STG/PT for the contrasts (*DA* + *DSDA*) - (*SS* + *DS*) was located in PT using the probability map in Westbury, Zatorre, and Evans (1999). The group peak activation results in Figure 2 for (*DA* + *DSDA*) - (*SS* + *DS*) in left posterior STG/PT (-60, -34, 8) is inside the 25-45% probability area, after conversion from MNI to Talairach coordinates ((Talairach and Tournoux 1988), which was necessary to use the Westbury et al. probability map.

Speaker

We assessed which cortical regions showed an effect when a switch of speaker was present, vs. no switch, (*DS* + *DSDA*) - (*SS* + *DA*) (cf. Table III and Figure 2). Peaks for a relative increase in BOLD for a switch of speaker were located in Lateral

Occipital Cortex bilaterally, right Precuneus, and right Middle Frontal Gyrus extending into Frontal Pole. Activations appeared to be more right-lateralised. Finally, no activated clusters were found at the selected significance level for the interaction term (*SS* + *DSDA*) - (*DS* + *DA*).

Insert Table III and Figure 2 about here

Discussion

The present study aimed to establish the neural bases of processing variation in spoken sentences related to speaker and accent in general, and to investigate the role of PT in processing accent variation. Specifically, it was hypothesized that BOLD-activity in PT would vary as a result of an increase in accent-related phonetic/phonological variation in the speech signal.

Accent

Several areas in the temporal lobes, including anterior and posterior STG, PT and SMG, and in the frontal lobes, including POp, showed a relative increase whenever there was a switch of accent present: (*DA* + *DSDA*) - (*SS* + *DS*).

Bilateral STG has been associated with different cognitive functions. Left pSTG (including PT) is generally regarded as part of a pathway for processing comprehensible speech (Davis and Johnsrude 2003; Poldrack, et al. 2001), and it has been suggested that it serves as an interface between the perception and long-term representation in mental lexicon of familiar words (Wise, et al. 2001), and is implicated in resolving semantic ambiguities in spoken sentence comprehension (Rodd, et al. 2005).

Earlier studies demonstrate that (left) POp (Blumstein, et al. 2005; Burton 2001; Burton, et al. 2000; Golestani, et al. 2002; Myers 2007; Wilson and Iacoboni 2006; Zatorre, et al. 1996) as well as posterior STG/PT (Callan, et al. 2004; Warren, et al. 2005; Wilson and Iacoboni 2006) are associated with phonetic-analytic listening to speech sounds or syllables. In addition, POp has previously been associated with disambiguation tasks at a syntactic processing level (Fiebach, et al. 2004) and has been named as a key structure in models for processing phonetic/phonological variation in speech comprehension (Callan, et al. 2004; Skipper, et al. 2006). Furthermore, left POp has been associated with implicit phonemic analysis processes in speech comprehension (Burton, et al. 2000) and it may be expected that processing accented speech may rely in part on increased (low-level) auditory analysis of the speech signal.

We also found activations in left SMG for a switch in accent. SMG shows sensitivity to phonological changes in speech (Dehaene-Lambertz, et al. 2005). It has been suggested (Obleser and Eisner 2009) that SMG has access to abstract (i.e., normalised) phonological units. Activations in SMG in speech perception tasks are thus in many cases interpreted as reflecting involvement in phonological working memory (e.g., (Jacquemot, et al. 2003).

Speaker

We found several areas that showed a relative increase when a switch of accent was present for the contrast (*DA* + *DSDA*) - (*SS* + *DS*). These activations included a relative increase in areas in Lateral Occipital Cortex bilaterally, right Precuneus, and right Middle Frontal Gyrus extending into frontal pole. These results are generally in line with earlier studies investigating neural activation related to the speaker's voice.

(Stevens 2004) and Belin et al. (2000) found increases in BOLD activation in MFG for processing voice variation versus non-vocal stimuli. Belin et al. (2000), (Kriegstein and Giraud 2004) both report activations in the Precuneus for processing voices.

Nevertheless, most studies on processing speaker variation report more extensive activations in predominantly right-lateralised temporal areas. Studies on processing speaker-related information show a wide variety in the neural locus of this process; some report activation in an area close to the left temporal pole in left anterior STS (Belin and Zatorre 2003), while others report that perceptual normalisation for the speaker occurs in the superior/middle temporal gyri bilaterally and the superior parietal lobule bilaterally (Wong, et al. 2004). Furthermore, a recent repetition-suppression study on spoken sentence processing also does not report whole-brain effects for switching between speakers (Dehaene-Lambertz, et al. 2006). Dehaene-Lambertz et al. report a small normalisation effect for speaker differences in left STG, after applying a more sensitive analysis. It seems plausible that the differences between our results and the aforementioned studies arise both from differences in task. More over, previous studies did not explicitly control for regional accent differences between speakers. The present study shows predominantly activations outside the temporal regions when accent differences between speakers are accounted for. Therefore, it could be the case that accent differences between speakers used in previous studies could have affected results, especially in temporal areas.

Speaker and accent normalisation

The question arises whether phonological and phonetic variation in speech is processed in the same way as speech stimuli that have been distorted or degraded, for instance by presenting sentences at a lower signal-to-noise ratio (Zekveld, et al. 2006), or by noise-vocoding (Obleser, et al. 2007). Relatively few studies investigate the specific effect of different types of distortion of the speech signal and identified areas that are involved more when the intelligibility decreases. Only one study addresses this question in-depth (Davis and Johnsrude 2003). Davis and Johnsruhe evaluated the effect of three types of distortions (speech in noise, noise-vocoded speech, and segmented speech) on speech processing. They found that left STG and left IFG became more active for distorted versus intelligible speech. However, activity in left posterior STG varied dependent on the type of distortion, while left anterior STG's responses were form-independent (i.e., showed elevated activation independent from the type of distortion used). It is at this point not feasible to determine whether accented speech and distorted speech are processed differently in left posterior STG, as Davis and Johnsruhe did not include accented speech and our study did not include distorted speech.

Conclusion

We conclude that bilateral posterior STG (including PT) and POp are involved in processing various types of distortions in the speech signal. However, further study is required to establish whether these areas differentiate between various types of speech-*intrinsic* (such as accent, speech rate, or clarity of speech) and speech-*extrinsic* variation (such as added noise).

Finally, our results provide further evidence for the hypothesis that PT is associated with processing accent and speaker variation during spoken language comprehension, and thus support the theory (Griffiths and Warren 2002; Warren, et al. 2005) that PT serves as a computational hub for processing spectrotemporal variation in auditory perception.

Acknowledgments

We wish to thank Paul Gaalman for technical assistance, Esther Aarts for lending her voice, and Joseph T. Devlin and Penny Lewis for useful comments. This research was supported by the Netherlands Organization for Research (NWO) under project number 275-75-003.

References

Adank P, Evans BG, Stuart-Smith J, Scott SK. 2009. Familiarity with a regional accent facilitates comprehension of that accent in noise. Journal of Experimental Psychology: Human Perception and Performance 35(2):520-529.

Adank P, Janse E. 2010. Comprehension of a novel accent by younger and older listeners. Psychology and Aging 25(3):736-740.

Adank P, van Hout R, Van de Velde H. 2007. An acoustic description of the vowels of Northern and Southern Standard Dutch II: Regional Varieties. Journal of the Acoustical Society of America 121:1130-1141.

Ashburner J, Friston K. 1997. Multimodal image coregistration and partitioning -- a unified framework. NeuroImage 6:209-217.

Bandettini PA, Jesmanowicz A, van Kylen J, Birn RM, Hyde JS. 1998. Functional MRI of brain activation induced by scanner acoustic noise. Magnetic resonance in medicine 39(3):410-416.

Belin P, Zatorre RJ. 2003. Adaptation to speaker's voice in right anterior temporal lobe NeuroReport 14(16):2105-2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. Nature 403(6767):309-12.

Binder J, Frost J, Hammeke T, Rao S, Cox R. 1996. Function of the left planum temporale in auditory and linguistic processing. Brain 119:1229-1247.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10( ):512-528.

Blumstein SE, Myers EB, Rissman J. 2005. The perception of voice onset time: an fMRI investigation of phonetic category structure. Journal of Cognitive Neuroscience 17(9):1353-1366.

Boersma P, Weenink D. 2003. Praat: doing phonetics by computer.

Bregman AS. 1990. Auditory Scene Analysis. Cambridge: MIT Press.

Burton MW. 2001. The role of inferior frontal cortex in phonological processing. Cognitive Science 25:695-709.

Burton MW, Small SL, Blumstein SE. 2000. The role of segmentation in phonological processing: an fMRI investigation. Journal of Cognitive Neuroscience 12(4):679-690.

Callan DE, Jones JA, Callan AM, Akahane-Yamada R. 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/or- osensory internal models. NeuroImage 22:1182-1194.

Chee MWL. 2009. fMR-Adaptation and the bilingual brain. Brain and Language 109(2-3):75-79.

Chee MWL, Tan CC. 2007. Inter-relationships between attention, activation, fMR adaptation and long-term memory. NeuroImage 37:1487-1495.

Davis MH, Johnsrude IS. 2003. Hierarchical processing in spoken language comprehension. J Neurosci 23(8):3423-31.

Dehaene-Lambertz G, Dehaene S, L. AJ, Campagne L, Ciuciu P, Dehaene G, Denghien I, Jobert A, LeBihan D, Pallier C and others. 2006. Functional segregation of cortical language areas by sentence repetition. Human Brain Mapping 27:360-371.

Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S. 2005. Neural correlates of switching from auditory to speech perception. Neuroimage 24:21-33.

Eden GF, Joseph JE, Brown HE, Brown CP, Zeffiro TA. 1999. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. Magnetic Resonance in Medicine 41:13-20.

Edmister WB, Talavage TM, Ledden TJ, Weisskoff RM. 1999. Improved auditory cortex imaging using clustered volume acquisitions. Human Brain Mapping 7(2):89-97.

Evans BG, 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 KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ. 1995. Spatial registration and normalization of images. Human Brain Mapping 2:165-189.

Gaab N, Gabrieli JD, Glover GH. 2006. Assessing the influence of scanner background noise on auditory processing. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Human Brain Mapping 28(8):703-720.

Giraud A-L, Lorenz C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. 2000. Representation of the temporal envelope of sounds in the human brain. Journal of Neurophysiology 84:1588-1598.

Giraud AL, Price CJ. 2001. The constraints functional neuroimaging places on classical models of auditory word processing. J Cogn Neurosci 13(6):754-65.

Golestani N, Paus T, Zatorre RJ. 2002. Anatomical correlates of learning novel speech sounds. Neuron 35(5):997-1010.

Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. Trends Neurosci 25(7):348-53.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital cortex. Neuron 24:187-203.

Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologia 107:293-321.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliot MR, Gurney EM, Bowtell RW. 1999. "Sparse" temporal sampling in auditory fMRI. Human Brain Mapping 7:213-223.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ. 2002. Spectral and temporal processing in human auditory cortex. Cerebral Cortex 12:140-149.

Henson RN, Rylands A, Ross E, Vuilleumeir P, Rugg MD. 2004. The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming NeuroImage 21:1674-1689.

Hickok G, Love T, Swinney D, Wong EC, Buxton RB. 1997. Functional MR imaging during auditory word perception: A single trial presentation paradigm. Brain and Language 58:197-201.

Hillenbrand JM, Houde RA. 2003. A narrow band pattern-matching model of vowel perception. Journal of the Acoustical Society of America 113(2):1044-1055.

IPA. 1999. Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge: Cambridge University Press.

Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E. 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. Journal of Neuroscience 23(29):9541-9546.

Kriegstein K, Giraud A. 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NeuroImage 22:948-955.

Moulines E, Charpentier F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Communication 9(5-6):453-467.

Myers EB. 2007. Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. Neuropsychologia 45(7):1463-1473.

Nearey TM. 1989. Static, dynamic, and relational properties in speech perception. Journal of the Acoustical Society of America 85:2088-2113.

Nearey TM. 1997. Speech perception as pattern recognition. Journal of the Acoustical Society of America 101(6):3241-3254.

Nearey TM, Assmann PF. 2007. Probabilistic 'sliding-template' models for indirect vowel normalization. In: Solé MJ, Beddor PS, Ohala M, editors. Experimental Approaches to Phonology. Oxford: Oxford University Press. p 246-269.

Noppeney U, Penny W. 2006. Two approaches to repetition suppression. Human Brain Mapping 27:411-416.

Norris D, McQueen JM, Cutler A. 2003. Perceptual learning in speech. Cognitive Psychology 47:204-238.

Obleser J, Eisner F. 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends in Cognitive Sciences 13(1):14-19.

Obleser J, Wise RJS, Dresner MA, Scott SK. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions Journal of Neuroscience 27:2283-2289.

Orfanidou E, Marslen-Wilson WD, Davis MH. 2006. Neural response suppression predicts repetition priming of spoken words and pseudowords. Journal of Cognitive Neuroscience 18:1237-1252.

Penhune VB, Zatorre RJ, Evans AE. 1998. Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. Journal of Cognitive Neuroscience 10(6):752-765.

Peterson GE, Barney HL. 1952. Control methods used in a study of the vowels. The Journal of the Acoustical Society of America 24:175-184.

Plomp R, Mimpen AM. 1979. Improving the reliability of testing the speech reception threshold for sentences in quiet for sentences. Audiology 18:42-53.

Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, Gabrieli JDE. 2001. Relations between the Neural Bases of Dynamic Auditory Processing and Phonological Processing: Evidence from fMRI. Journal of Cognitive Neuroscience 13(5):687-697.

Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15(8):1261-9.

Rogers CL, Dalby J, Nishi K. 2004. Effects of noise and proficiency level on intelligibility of Chinese-accented English. Language and Speech 47:139-154.

Shah NJ, Jäncke L, Grosse-Ruyken ML, Muller-Gartner HW. 1999. Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. Journal of Magnetic Resonance Imaging 9(1):19-25.

Shtyrov Y, Kujala T, Palva S, Ilmoniemi RJ, Naatanen R. 2000. Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. Neuroimage 12(6):657-63.

Slotnick SD, Moo LR, Segal JB, Hart JJ. 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cognitive Brain Research 17:75-82.

Stevens AA. 2004. Dissociating the neural bases for voices, tones, and words. Cognitive Brain Research 18:162-171.

Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.

Talavage TM, Edmister WB, Ledden TJ, Weisskoff RM. 1999. Quantitative assessment of auditory cortex responses induced by imager acoustic noise. Human Brain Mapping 7(2):79-88.

Tamer G, Talavage T, Wen-Ming L. 2009. Characterizing response to acoustic imaging noise for auditory event- related fMRI. IEEE Transactions in Biomedical Engineering 56(7):1919-1928.

Van Wijngaarden SJ. 2001. Intelligibility of native and non-native Dutch speech. Speech Communication 35(103-113).

van Wijngaarden SJ, Steeneken HJ, Houtgast T. 2002. Quantifying the intelligibility of speech in noise for non-native talkers. Journal of the Acoustical Society of America 112(6):3004-3013.

Warren JE, Wise RJS, Warren JD. 2005. Sounds do-able: auditory–motor transformations and the posterior temporal plane Trends in Neurosciences 28(12):636-643.

Wells JC. 1982. Accents of English. Three volumes and cassette. Cambridge: Cambridge University Press.

Westbury CF, Zatorre RJ, Evans AC. 1999. Quantifying variability in the planum temporale: A probability map. Cerebral Cortex 9:392-405.

Wilson SM, Iacoboni M. 2006. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. NeuroImage 33:316-325.

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. 2001. Separate neural subsystems within 'Wernicke's area'. Brain 124(1):83-95.

Wong PCM, Nusbaum HC, Small SL. 2004. Neural bases of talker Normalization. Journal of Cognitive Neuroscience 16:1173-1184.

Zaehle T, Schmidt CF, Meyer M, Baumann S, Baltes C, Boesiger P, Jancke L. 2007. Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. NeuroImage 37:1195-1204.

Zatorre RJ, Evans AC, Meyer E. 1994. Neural mechanisms underlying melodic perception and memory for pitch. Journal of Neuroscience 14:1908-1919.

Zatorre RJ, Meyer E, Gjedde A, Evans AC. 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. Cerebral Cortex 6:21-30.

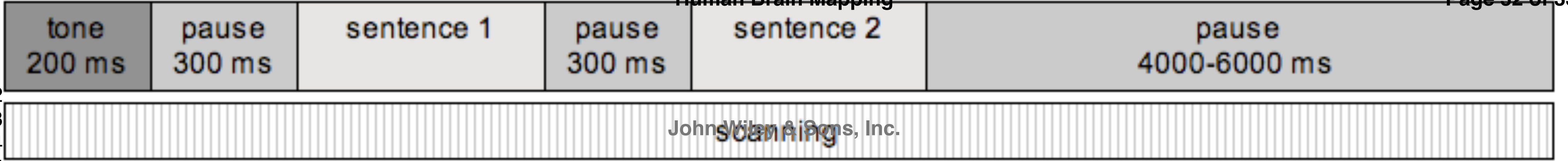
Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R. 2006. Top–down and bottom–up processes in speech comprehension. NeuroImage 32(4):1826-1836.

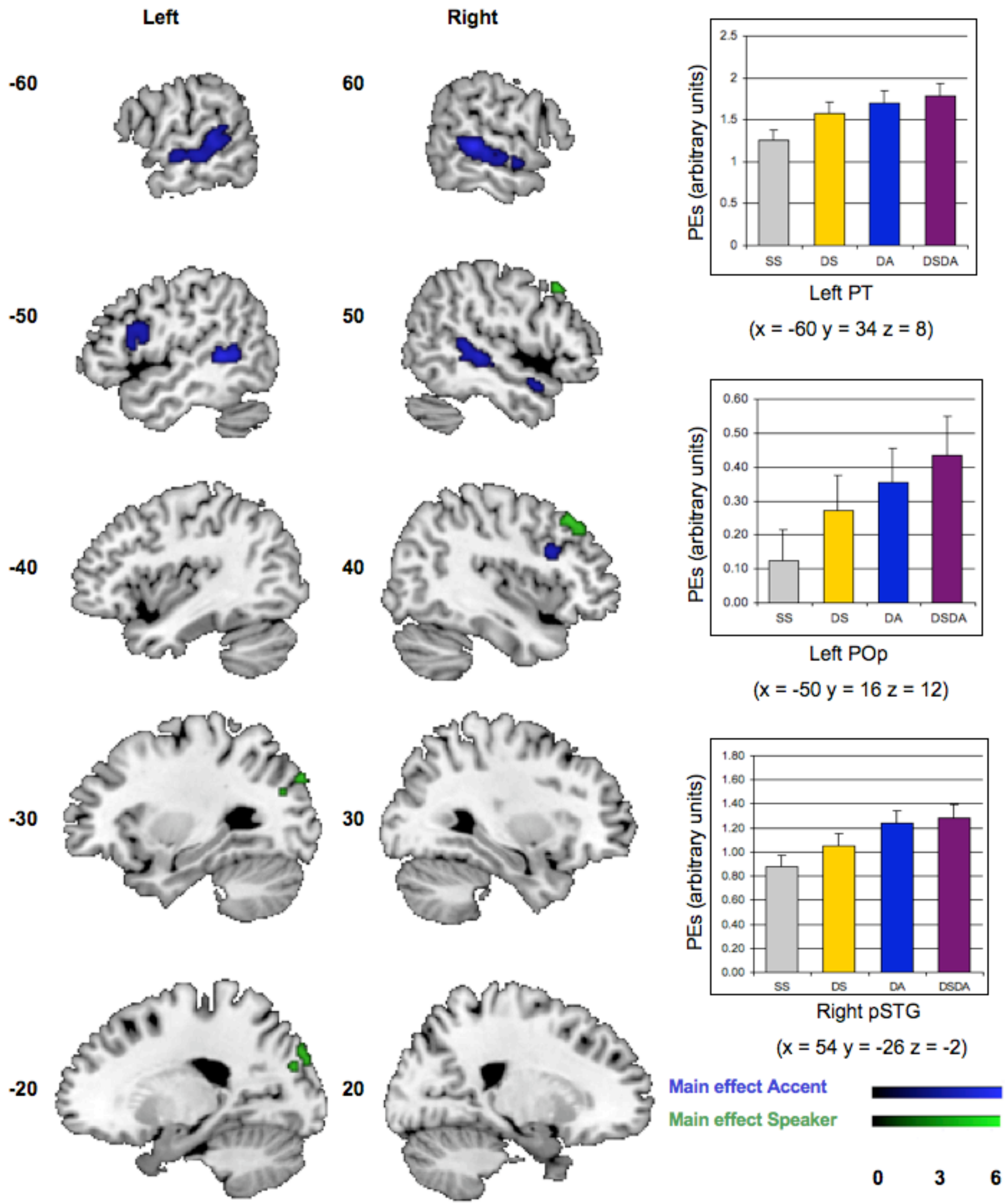
Zevin JD, McCandliss BD. 2005. Dishabituation of the BOLD response to speech sounds. Behavioral and Brain Functions 1(4).

Name	*Speaker*	*Accent*
SS	Same Speaker	Same Accent
DS	Different Speaker	Same Accent
DA	Same Speaker	Different Accent
DSDA	Different Speaker	Different Accent

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

Structure	Hemisphere	x	y	z	T-value	Z-value
		(DA + DSDA) - (SS + DS)				
posterior STG/SMG	Left	-54	40	4	5.91	5.29
Posterior STG/PT	Left	-60	-34	8	4.76	4.41
Posterior MTG	Left	-60	-26	-4	4.10	3.87
Posterior STG/SMG	Right	60	-32	2	6.10	5.43
POp/PG	Left	-50	12	24	4.04	3.81
POp/FOC	Left	-46	16	12	3.97	3.75
POp/PTr	Left	-46	16	12	3.50	3.35
Posterior STG/MTG/SMG	Right	54	-26	-2	5.07	4.65
Anterior STG/TP/MTG	Right	54	4	-16	4.87	4.49
Central Opercular Cortex	Right	38	18	26	4.36	4.08
	Right					
		(DS + DSDA) - (SS + DA)				
LOC/OP	Left	-26	-92	32	4.77	4.41
LOC	Left	-26	-76	24	4.12	3.88
LOC	Right	14	-66	66	5.47	4.96
Precuneus/LOC/SPL	Right	12	-58	60	4.32	4.05
Precuneus	Right	6	-52	50	4.06	3.83
MFG	Right	46	24	46	5.02	4.67
MFG/FP	Right	40	36	42	3.72	4.54


tone
200 ms
pause
300 ms
sentence 1
pause
300 ms
sentence 2
pause
4000-6000 ms
scanning


Left
Right
-60
60
-50
50
-40
40
-30
30
-20
20
PEs (arbitrary units)
SS
DS
DA
DSDA
Left PT
(x = -60 y = 34 z = 8)
Left POp
(x = -50 y = 16 z = 12)
Right pSTG
(x = 54 y = -26 z = -2)
Main effect Accent
Main effect Speaker
0
3
6