Research report

Regional heterogeneity in the processing and the production of speech in the human planum temporale

Pascale Tremblay a,*,1, Isabelle Deschamps b,c and Vincent L. Gracco b,c,d

a Center for Mind & Brain Sciences (CIMeC), The University of Trento, Italy
b McGill University, Faculty of Medicine, School of Communication Sciences and Disorders, Montreal, Canada
c Centre for Research on Brain, Language & Music, McGill University, Canada
d Haskins Laboratories, New Haven, CT, USA

ABSTRACT

Introduction: The role of the left planum temporale (PT) in auditory language processing has been a central theme in cognitive neuroscience since the first descriptions of its leftward neuroanatomical asymmetry. While it is clear that PT contributes to auditory language processing there is still some uncertainty about its role in spoken language production.

Methods: Here we examine activation patterns of the PT for speech production, speech perception and single word reading to address potential hemispheric and regional functional specialization in the human PT. To this aim, we manually segmented the left and right PT in three non-overlapping regions (medial, lateral and caudal PT) and examined, in two complementary experiments, the contribution of exogenous and endogenous auditory input on PT activation under different speech processing and production conditions.

Results: Our results demonstrate that different speech tasks are associated with different regional functional activation patterns of the medial, lateral and caudal PT. These patterns are similar across hemispheres, suggesting bilateral processing of the auditory signal for speech at the level of PT.

Conclusions: Results of the present studies stress the importance of considering the anatomical complexity of the PT in interpreting fMRI data.

Keywords:
Posterior temporal plane
Word production
Word repetition
Whispering
Auditory feedback

1. Introduction

The planum temporale (PT) is a large sheet-like cortical area with a roughly triangular shape located on the superior temporal gyrus (STG), posterior to the primary auditory area (PAC) which is located on Heschl’s sulcus (von Economo and Horn, 1930; Pfeifer, 1936; Galaburda and Sanides, 1980). The contribution of the left PT to auditory language processing has been a central theme in cognitive neuroscience since the first descriptions of its leftward asymmetry (Geschwind and Levitsky, 1968; Galaburda et al., 1978; Steinmetz and Galaburda, 1991; for a systematic review, see Shapleske...
et al., 1999). It has been hypothesized that this asymmetry reflects a functional specialization of the left cerebral hemisphere for language (Goullven et al., 2003; Doraisaint-Pierre et al., 2006; but see also Marshall, 2000). Additional support for this hypothesis comes from lesions studies that have shown that damage in or around PT is associated with auditory language comprehension deficits (Tanaka et al., 1987; Praamstra et al., 1991; Caplan et al., 1995). Furthermore, Jacquemot et al. (2003) have shown that the left STG/lateral PT selectively responds to language-specific phonological manipulations. In other neuroimaging experiments PT activation has been observed during passive listening to speech sounds (Petersen et al., 1988; Wise et al., 1991, 2001; Zatorre et al., 1996; Mazoyer et al., 1993; Binder et al., 1996, 1997, 2000; Friederici et al., 2000; Belin et al., 2002; Callan et al., 2006; Wilson et al., 2008) with increased activation in background noise (Wang et al., 2008). However, PT also responds to non-language perceptual processing including melody (Griffiths et al., 1998; Zatorre et al., 1994), tones (Specht and Reul, 2003; Hudgahl et al., 1999; Jancke et al., 2003; Binder et al., 1996), musical instruments (Hudgahl et al., 1999) and masking noise (Wise et al., 2001). Unilateral and/or bilateral STG lesions that include PT result in impaired recognition of a wide range of non-verbal environmental sounds (including human non-verbal sounds such as crying), manmade and natural non-living sounds and animal sounds (Schnider et al., 1994). In terms of a possible functional specialization, passive and active listening results in greater PT activation bilaterally for tones compared to words, suggesting a general auditory processing function rather than language-specific processing (Binder et al., 1996, 1997). Consistent with these and other more recent findings (e.g., Warren et al., 2005a, 2005b; Kumar et al., 2007) it has been suggested that the bilateral PT acts as a computational hub whose function is to disambiguate complex sounds by isolating different properties of the acoustic objects (e.g., temporal and spectral information) and matching them to stored templates (Griffiths and Warren, 2002). Despite some findings incompatible with the computational hub hypothesis (see for example Hall and Plack, 2009), it is clear that PT is involved in the processing of sounds, including speech, a function consistent with its connectivity pattern and anatomical location.

In addition to serving as a computational hub for auditory processing, PT may be involved in sensorimotor transformation. According to this view, sensorimotor transformation is necessary to establish parity between motor and perceptual units for speech (as in the Motor Theory of Speech Perception; Liberman and Mattingly, 1985). Parity functions to link motor (articulatory) representations to their auditory consequences. Sensorimotor transformation is necessary for both speech perception (as in the Motor Theory of Speech Perception; Liberman and Mattingly, 1985) and speech production (Warren et al., 2005a, 2005b; Hickok and Poeppel, 2007). This hypothetical function is most likely accomplished through connections to and from motor regions through the arcuate fasciculus (AF) (i.e., the dorsal stream network) (Hickok et al., 2000, 2003; Buchsbaum et al., 2001; Warren et al., 2005a, 2005b; Saito et al., 2006; see also Hickok and Poeppel, 2000, 2004, 2007; Schmahmann and Pandya, 2006; Schmahmann et al., 2007; Saur et al., 2008) as well as other networks that have yet to be identified. Support for the role for PT in sensorimotor transformation comes mainly from imaging studies showing PT activation during overt speech production (Petersen et al., 1988; Wise et al., 1991; Karbe et al., 1998; Schulz et al., 2005; Saito et al., 2006; Tourville et al., 2008; Bohland and Guenther, 2006; Riecker et al., 2008; Zheng et al., 2010; Peschke et al., 2009; Dhanjal et al., 2008), but also during silent speech production or speech rehearsal, which does not involve self-generated auditory feedback (Hickok et al., 2000, 2003; Buchsbaum et al., 2001; Wise et al., 2001; Huang et al., 2001; Papathanassiou et al., 2000; Okada et al., 2003; Shergill et al., 2002; Callan et al., 2006; Pa and Hickok, 2008). Furthermore, it has been shown that an area of the posterior temporal–parietal junction, often referred to as area Spt (which includes the caudal part of PT) responds more strongly to sub-vocal rehearsal of auditory stimuli than to the perception of auditory stimuli (e.g., Buchsbaum et al., 2001; Hickok et al., 2003; Hickok et al., 2009). Consistent with these results, left PT activation co-variates with changes in the left motor cortex activation during whispered speech (Paus et al., 1996), providing indirect evidence of a connection between motor regions and PT, and support for a role for PT in sensorimotor transformation for speech.

While brain regions can support more than one function, it is possible that the apparent heterogeneity of functions supported by PT is related to the fact that the human PT, like the non-human primate PT, contains several cortical fields (e.g., von Economo and Horn, 1930; Galaburda and Sanides, 1980; Rivier and Clarke, 1997; Tardif and Clarke, 2001; Scheich et al., 1998; Sweet et al., 2005; Fullerton and Pandya, 2007). Just posterior to the transverse temporal sulcus, in a region corresponding to the rostral PT, two regions have been identified by Rivier and Clarke (1997), the lateral area (LA) and the posterior (and more medial) area (PA). Areas LA and PA are both part of the internal and external parabelt areas identified for example by Sweet et al. (2005) in humans. The caudal most region of PT, often referred to as the temporoparietal area (Tpt), and sometimes as the vertical planum temporale (vPT), is located outside of the auditory parabelt (Sweet et al., 2005). Since different cortical fields are likely to support different functions, it seems reasonable to hypothesize that PT is a functionally heterogeneous region, a claim that is supported by the literature. In sum, PT appears to be involved in auditory speech processing and in transforming auditory signals into speech motor representations. In addition, PT may also be activated as a result of auditory re-afference to monitor feedback during speech production. In order to understand the role of PT in speech processing, the contribution of each of these three sources of activation needs to be evaluated.

In the present study, we evaluated the hypothesis that the anatomically distinct regions of the left and right PT are differentially activated depending on the source and function of the auditory processing involved. We examined, in two complementary studies, the contribution of exogenous and endogenous auditory input on PT activation under different speech processing and production conditions. Our main objective was to investigate whether different regions of the human PT

\[2\) It should be noted that Tpt and vPT are not identical: the human Tpt extends onto the parietal convexity (Galaburda et al., 1978; Scopter et al., 2010), while the vPT does not.\]
contribute to the perception and production of speech in distinct ways. In the first study, we examined the processing of internally generated auditory stimulation by comparing whispered speech production with and without feedback. In the second study, we compared the processing of internally and externally generated auditory stimulation by comparing auditory word repetition (AWR) and word reading aloud. In general, PT activation appears to be functionally differentiated suggesting unique contributions to speech perception and production.

2. Experiment 1

2.1. Material and methods

2.1.1. Participants

Ten healthy adults (mean 25.1 ± 3.8 years) balanced for gender comprised the experimental group. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and were native speakers of Canadian English. All participants had normal or corrected-to-normal vision, and reported no history of speech, language or learning difficulties. Participants were screened for any contraindication to magnetic resonance imaging (MRI). Informed written consent was obtained from each participant. The study was approved by the Magnetic Resonance Research Committee (MRRC) and the Montreal Neurological Institute (MNI) Research Ethics Committee.

2.1.2. Experimental procedures

The experiment consisted of two active speech conditions that were presented in pseudo-random blocks of two or three trials. Subjects were instructed to whisper a printed word presented on the back-projected screen under either normal feedback (Whisper) or in the presence of masking noise (Whisper masked). The masking noise level was adjusted for each participant such that it was impossible for him or her to hear his or her own voice while the masking was present. The noise begun at the time participants were instructed to start whispering. The rationale for using whispered speech as a proxy of natural speech production, is that (1) in whispered speech subjects articulate in a near normal manner, and (2) the self-generated auditory feedback can be completely masked with noise, because of the absence of bone-conducted feedback. These two speaking conditions involved similar motor processes but differed in terms of the presence of auditory feedback. Importantly, participants were trained not to change their articulation in the masking condition, to avoid a Lombard effect, characterized by an increase in voice intensity in the presence of background noise (Lombard, 1911, as cited in H. Lane and B. Tranel, 1971). During the practice session, participants were asked to produce visually presented words in the presence of noise and no noise and asked to keep articulation and voice level constant. However, it should be noted that the acoustic recordings were not of high enough quality to allow for acoustical analyses, which would have demonstrated whether or not there was a Lombard effect. We are nevertheless confident that subjects maintained their voice constant between the two conditions because we did not observe differences in motor cortex activation between the tasks. Increased speaking intensity would typically translate into more effort and lead to increased motor activation – this was not observed. The lack of increased activation in the primary motor area in the Whisper Masked condition is shown in Supplementary Fig. S2.

The printed stimuli used for the speech condition were three-syllable nouns presented on a screen one at a time (see Supplemental material for the list of all stimuli). Each experimental trial began with the presentation of the visual instruction (e.g., Whisper “tomato”) presented during the volume acquisition (3.3 sec) using SuperLabPro (Cedrus Corp, CA, USA). The visual instruction remained on the screen for the duration of the volume acquisition in order to limit the need for participants to sub-vocally rehearse the words. Participants were instructed to wait until the stimulus disappeared to respond, which occurred at the end of the volume acquisition. All responses occurred during a silent interval of 6.2 sec – a delay that was chosen such that at the time of the consecutive volume acquisition, the hemodynamic response to speech production would be maximal (that is, 5–6 sec after onset of speech production).

Fifty trials of the Whisper condition, and 50 trials of the Whisper Masked conditions were acquired in total. The baseline condition, which was interleaved with the experimental conditions, was a visual fixation condition during which participants fixated on the word “rest” (30 trials). This was included for two reasons: (1) to control for the visual stimuli that occurred during the experimental conditions, (2) to allow for an examination of the effects of each tasks against a non-speech task, as opposed to looking at contrasts only. A passive listening to masking noise condition (MN) was also added to the experimental paradigm (30 trials). During this task, participants fixated on the word “rest”. The noise was presented during the delay in TR only. During the scanning session, participants also produced additional tasks that were not analyzed for the current study (tongue raising and lowering, jaw openings and closings, silent articulation).

2.1.3. Image acquisition

The data were acquired on a 1.5 T Siemens Sonata MRI scanner at the MNI (Montreal, Canada). In order to eliminate movement artifacts associated with producing speech in the scanner, a sparse image acquisition technique (Eden et al., 1999; Edmister et al., 1999; Gracco et al., 2005) was used. A silent period (6.2 sec) was interleaved between each volume acquisition. All responses occurred during the silent period. The subjects wore MR compatible headphones (Commander XG, Resonance technology, CA, USA) and their responses were recorded though a MR compatible microphone attached to the headphones and digitized directly onto a Toshiba laptop computer. Thirty-nine axial slices (whole brain coverage) oriented parallel to the AC–PC line (thickness = 4 mm, no gap, FOV = 256 × 256 mm², matrix = 64 × 64) were acquired in 3.3 sec using a multi-slice echo-planar imaging (EPI) sequence (TE = 50 msec, TR = 9.5 sec, delay in TR = 6.2 sec). The delay in TR occurred following each volume acquisition. The slices had a spatial resolution of 4 × 4 × 4 mm. Two experimental runs (13 min each) resulted in the acquisition of 266 T2*-weighted BOLD images acquired in a descending order. High-resolution T1-weighted volumes were acquired for
anatomical localization (matrix 256 × 256 mm, 176 slices, 1 × 1 × 1 mm, no gap, TE = 9.2 msec, TR = 22 msec). The participants' head was immobilized by means of a vacuum-bag filled with polystyrene balls, which was fitted around the subject's head.

2.1.4. Image analysis
Images were spatially registered, motion-corrected, despiked, mean-normalized and smoothed with a Gaussian 6-mm FWHM filter using AFNI (Cox, 1996). There were separate regressors for the experimental conditions (Whisper, Whisper masked), as well as the control condition (Listen to masking noise). Additional regressors were the mean, linear, and quadratic trend components, the six motion parameters (x, y, z and roll, pitch and yaw). A linear least squares model was used to establish a fit to each time point of the HRF for each condition. For the second level (group) analysis, subject data were transformed into stereotaxic space using the MNI305 template (Collins et al., 1994), and re-sampled to 3 × 3 × 3 mm. The group analyses were performed on the subjects’ beta values resulting from the first level analysis. We conducted a two-way mixed model ANOVA with repeated measurements on the task (Whisper, Whisper Masked). Subjects were entered in the analysis as a random factor. A cluster correction for multiple comparisons across the brain was implemented in the AFNI program 3dClusterSim. Based on 3dClusterSim results, we determined that a family-wise error (FWE) rate of $p < .05$ is achieved with a minimum cluster size of 44 contiguous voxels each significant at $p < .01$.

2.1.5. Region of interest (ROI) analysis
Three anatomically defined regions of interest were created to parcellate the left and the right PT. Anatomical definitions were determined with reference to Rivier and Clarke (1997). For each subject, PT was manually segmented into three non-overlapping regions (medial, lateral and caudal) in the following manner: PT was first split into two regions (rostral and caudal) with the transverse temporal sulcus as the rostral boundary. If two transverse temporal gyri were present, we used the gray matter caudal to the second gyri. The rostral region was then split into two halves (lateral and medial), roughly corresponding to Rivier and Clarke (1997) areas LA and PA respectively. The caudal boundary of the caudal PT was the ascending ramus of the Sylvian fissure. An example of a parcellation is shown on Fig. 1(A, B). For each subject, we calculated the mean percentage of signal change for each condition and PT region, as well as difference scores between the two whispering conditions (Whisper—Whisper Masked) and entered these values into a 2-way ANOVA with repeated measurements, with factor ROI (Medial, Lateral, Caudal PT) and Hemisphere (left, right). FDR corrected ($i = 3; q = .05$) post hoc tests (paired sample t-tests) were used to further examine the effect of ROI.

2.1.6. SNR analyses
To evaluate the sensitivity of our experimental design, following Parrish et al. (2000) and Dick et al. (2009) we conducted a series of simulations that determined, for various SNR values, the probability of identifying a signal change of magnitude of .5% and 1%. For each participant, we then computed a signal-to-noise ratio (SNR) map of PT using AFNI. SNR was defined here as the ratio of the signal’s mean to the standard deviation (SD) of the signal. The simulation was conducted using R. We derived a synthetic model of the expected activity by convolving the boxcar design with a canonical HRF (mean $\approx 1$). To this, varying amounts of smoothed noise (modeled by an autoregressive moving average) were added to obtain different SNR ratios. This process was repeated for 10,000 iterations for each noise level, where each iteration assigned a different randomly generated noise value. We then determined the probability of finding a reliable

*Fig. 1 – Example of a representative PT parcellation. A. Axial view of subject 8 brain showing the superior temporal plane. B. Zoomed view on the left superior temporal area showing the parcellation in PT. Legend: mPT = medial PT, IPT = lateral PT, cPT = caudal PT, TTG = transverse temporal gyrus, TTS = transverse temporal sulcus.*
correlation for that noise level by evaluating the “boxcar-predictor-plus-noise” data against the original boxcar predictor via regression (see Parrish et al., 2000 for details of method). This simulation method indicates the power to reliably detect a given MR percent signal change as a function of SNR. Such a simulation determines whether the design has sufficient power to detect results in cases where null results are reported.

2.2. Results

2.2.1. Behavioral data

No errors were committed in the whispering tasks.

2.2.2. Neuroimaging results

2.2.2.1. SNR ANALYSES. The mean SNR in the bilateral PT for the group was 69.8 (±9.7 SD). The power simulations determined that, given a power of .80, the minimum SNR to detect a MR signal change of .5% was 56, whereas the minimum SNR to detect a MR signal change of 1% was 27 (Supplementary Fig. S1). These results show that the actual SNR of the study was sufficient to detect signal changes in PT as low as .5%.

2.2.2.2. Voxel-based analyses. As illustrated in Fig. 2, the contrast of each whispering condition with the resting baseline revealed activation in the ventral precentral gyrus (including PMv and M1) bilaterally, left IFG pars opercularis (IFGop), left insula, transverse temporal gyrus and PT bilaterally, and cerebellum, lobules V and VI, bilaterally. In order to identify regions active for Whisper Masked above and beyond listening to masking noise alone, we also subtracted the Noise condition from Whisper Masked. This contrast revealed activation in the ventral precentral gyrus (including PMv and M1) bilaterally, left IFGop. The contrast of the masking noise condition with the resting baseline revealed strong activation in transverse temporal gyrus and PT bilaterally. A list of all areas of activation, for each condition, is provided in Table 1.

2.2.2.3. ROIs. First, we computed a difference score between the two whispering conditions (Whisper–Whisper Masked), for each ROI and each hemisphere. These scores were submitted to a 2-way ANOVA with repeated measurements, with factor ROI (lateral PT, medial PT, caudal PT) and Hemisphere (left, right). Results show a main effect of Region \( F(2, 9) = 5.91, p = .011 \), and a main effect of Hemisphere \( F(1, 9) = 19.56, p = .002 \), but no interaction \( F(2, 18) = .14, p = .88 \). FDR-corrected paired sample t-tests indicate that the medial PT was significantly different from the lateral PT \( t(9) = 3.11, p = .016 \). While the lateral PT was significantly more active for Whisper than Whisper masked, the opposite effect was found at the level of the medial PT. There was no difference between the two whisper conditions at the level of the caudal PT. These results are illustrated in Fig. 3. The laterality effect was due to an overall stronger difference between the Whisper conditions in the left PT compared to the right PT.

We also computed the difference between the Whisper Masked condition and the Noise condition (Whisper Masked–Noise listening). These difference scores were submitted to a 2-way ANOVA with repeated measurements, with factor ROI (lateral PT, medial PT, caudal PT) and Hemisphere (left, right). Results show a marginally significant laterality effect, with an overall stronger difference (Whisper Masked > Noise listening) in the left than in the right PT \( F(1, 9) = 5.03, p = .052 \).

Fig. 2 – Cluster-corrected group-level activation (FWE rate of \( p < .05 \), achieved with a minimum cluster size of 44 contiguous voxels each significant at \( p < .01 \)) from Experiment 1 for the contrast of each Whispering condition against a resting baseline. Data are overlaid on axial views of a participant’s T1-weighted MRI transformed into stereotaxic space using the MNI305 template. All coordinates are in MNI space.
Table 1 – Table of statistically significant activation at the whole-brain level, corrected for multiple comparison (FWE rate of $p < .05$, achieved with a minimum cluster size of 44 contiguous voxels each significant at $p < .01$), for Experiment 1, for each condition compared against a resting baseline: (1) Whisper, (2) Whisper Masked, (3) Listening to masking noise, and (4) for the conjunction of the two whispering condition.

<table>
<thead>
<tr>
<th>Description</th>
<th>Hemi</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Voxels</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Whisper</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus and sulcus, postcentral gyrus (including PMv and M1/S1), extending caudally into the supramarginal gyrus, and rostrally into the inferior frontal gyrus, pars opercularis and medially into the insula.</td>
<td>L</td>
<td>−50</td>
<td>−17</td>
<td>33</td>
<td>156</td>
<td>.00008</td>
</tr>
<tr>
<td>Precentral gyrus and central sulcus (M1/S1).</td>
<td>R</td>
<td>44</td>
<td>−11</td>
<td>42</td>
<td>104</td>
<td>.00017</td>
</tr>
<tr>
<td>Inferior frontal gyrus, pars opercularis and triangularis, and transverse temporal gyrus.</td>
<td>L</td>
<td>−53</td>
<td>2</td>
<td>12</td>
<td>80</td>
<td>.00256</td>
</tr>
<tr>
<td>Posterior cingulate gyrus.</td>
<td>R</td>
<td>2</td>
<td>−8</td>
<td>27</td>
<td>76</td>
<td>.00000</td>
</tr>
<tr>
<td>Anterior cingulate, medial frontal gyrus (pre-SMA).</td>
<td>L</td>
<td>−5</td>
<td>−8</td>
<td>63</td>
<td>48</td>
<td>.00005</td>
</tr>
<tr>
<td><strong>2. Whisper with masked feedback</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial occipital lobe including calcarine fissure.</td>
<td>R</td>
<td>17</td>
<td>−83</td>
<td>−1</td>
<td>492</td>
<td>.00000</td>
</tr>
<tr>
<td>Precentral gyrus and sulcus, postcentral gyrus (including PMv and M1/S1), extending rostrally into the inferior frontal gyrus, pars opercularis.</td>
<td>L</td>
<td>−50</td>
<td>−23</td>
<td>39</td>
<td>137</td>
<td>.00004</td>
</tr>
<tr>
<td>Supplementary motor area.</td>
<td>L</td>
<td>−5</td>
<td>−5</td>
<td>60</td>
<td>117</td>
<td>.00002</td>
</tr>
<tr>
<td>Precentral gyrus and sulcus, postcentral gyrus (including PMv and M1/S1), extending rostrally into the inferior frontal gyrus, pars opercularis.</td>
<td>R</td>
<td>59</td>
<td>2</td>
<td>18</td>
<td>90</td>
<td>.00002</td>
</tr>
<tr>
<td>Cerebellum, lobules V and VI.</td>
<td>R</td>
<td>17</td>
<td>−59</td>
<td>−19</td>
<td>57</td>
<td>.00001</td>
</tr>
<tr>
<td>Inferior frontal gyrus, pars opercularis.</td>
<td>L</td>
<td>−56</td>
<td>−2</td>
<td>9</td>
<td>51</td>
<td>.00016</td>
</tr>
<tr>
<td>Superior temporal sulcus and gyrus, caudal PT.</td>
<td>R</td>
<td>59</td>
<td>−41</td>
<td>12</td>
<td>46</td>
<td>.00004</td>
</tr>
<tr>
<td><strong>3. Whisper masked minus listening to masking noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus and sulcus, postcentral gyrus (including PMv and M1/S1), extending caudally into the supramarginal gyrus, rostrally into the inferior frontal gyrus, pars opercularis.</td>
<td>L</td>
<td>−53</td>
<td>−8</td>
<td>39</td>
<td>74</td>
<td>.000299577</td>
</tr>
<tr>
<td><strong>4. Listen to masking noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STG, including the planum temporale and transverse temporal gyrus.</td>
<td>R</td>
<td>50</td>
<td>−26</td>
<td>18</td>
<td>67</td>
<td>.00030</td>
</tr>
<tr>
<td>STG, including the planum temporale and transverse temporal gyrus.</td>
<td>L</td>
<td>−38</td>
<td>−32</td>
<td>12</td>
<td>53</td>
<td>.00071</td>
</tr>
<tr>
<td>Frontal operculum, anterior insula.</td>
<td>L</td>
<td>−29</td>
<td>23</td>
<td>15</td>
<td>48</td>
<td>.00003</td>
</tr>
<tr>
<td>Calcarine fissure, occipito-temporal sulcus, lingual gyrus.</td>
<td>L</td>
<td>−14</td>
<td>−65</td>
<td>−1</td>
<td>45</td>
<td>.00003</td>
</tr>
<tr>
<td>Left and right precuneus.</td>
<td>R</td>
<td>2</td>
<td>−47</td>
<td>54</td>
<td>45</td>
<td>.00018</td>
</tr>
<tr>
<td><strong>5. Whisper ∩ whisper masked</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus and sulcus, postcentral gyrus (including PMv and M1/S1), extending caudally into the supramarginal gyrus, and rostrally into the inferior frontal gyrus, pars opercularis and medially into the posterior insula.</td>
<td>L</td>
<td>−49</td>
<td>−8</td>
<td>24</td>
<td>239</td>
<td></td>
</tr>
<tr>
<td>Cerebellum, lobules V and VI.</td>
<td>R</td>
<td>11</td>
<td>−64</td>
<td>−17</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>Posterior superior temporal sulcus.</td>
<td>R</td>
<td>44</td>
<td>−38</td>
<td>11</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Caudal thalamus.</td>
<td>R</td>
<td>18</td>
<td>−33</td>
<td>6</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

2.3. Discussion

In the first experiment we examined the pattern of functional activation in three PT regions (lateral, medial and caudal) during speech production with and without masking of auditory feedback to evaluate the contribution of endogenous (self-generated) and exogenous (masking noise) auditory input to PT activation. Here we used whispered speech in order to evaluate the effects of auditory feedback, which could be effectively masked through the use of masking noise. Under normal speech conditions, the use of masking noise only reduces the air-borne feedback signal with limited effect on bone-conducted feedback. The results of this study suggest, based on magnitude of activation in PT, that the left PT might be more strongly involved in processing speech than the right PT, although the activation patterns, overall, were similar. There was, however, evidence of functional heterogeneity within hemisphere. The ROIs differed in their response to the presence of auditory feedback. For the whispered speech conditions (masked and not masked), the lateral PT was most strongly activated for auditory feedback when it was perceptible, that is when subjects were able to hear what they produced; in contrast, the medial PT exhibited a decrease in activation in the presence of perceptible feedback. Finally, the caudal PT was not modulated by the presence or absence of auditory feedback. When the masking noise was combined with the self-generated feedback there was no significant increase in activation magnitude in any PT region over and above the auditory input from an exogenous source (passive listening to masking noise). Hence, it appears that the PT, as a whole, does not respond selectively to auditory input that is of functional relevance, in this case, self-generated feedback. In sum, our results demonstrate some degree of specialization within each PT region.

Interestingly, while our results suggest regional specializations within PT, they do not support the hypothesis of a functional specialization of the left PT in the processing of speech and language. We found similar activation patterns in...
the left and right PT, with an overall greater activation magnitude in the left PT. In the following experiment we extend the investigation of potential regional and hemispheric specialization in PT during speech perception and production with a more naturalistic set of conditions comparing overt speaking (with voiced feedback) when elicited from written or auditory stimuli.

### 3. Experiment 2

#### 3.1. Materials and method

##### 3.1.1. Participants

Participants were ten healthy adults (mean 27.3 ± 3.8 years; 5 females). All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and were native speakers of Canadian English. All participants had normal or corrected-to-normal vision, and reported no history of speech, language or learning difficulties. Participants were screened for any contraindication to MRI. Informed written consent was obtained from each participant. The study was approved by the MRRC and the MNIMNI Research Ethics Committee.

##### 3.1.2. Experimental procedures

The experiment consisted of six experimental conditions that were presented in pseudo-random blocks of 12 or 13 trials. The conditions included (1) silent word reading (RS), (2) reading aloud (RA), (3) auditory word listening (AWL), (4) AWR, (5) picture viewing and (6) picture naming. Each condition was repeated 50 times each. Only the first four tasks were analyzed for the present experiment, which resulted in a $2 \times 2$ design: modality (visual, auditory) and task (passive, active).

The stimuli were 50 concrete words derived from a set of objects pictures of chosen from Snodgrass and Vanderwart (1980) (see Supplemental material for the list of all stimuli). The words were highly concrete (mean 604 ± 18), highly imageable (mean 599 ± 25) and highly familiar (mean 560 ± 37). The mean written frequency of the words was 40 ± 29 (range 10–127) (Francis and Kucera, 1982). The mean number of letter was 4.6 ± 1.2 and the mean number of syllables was 1.3 ± .5.

Each block began with the presentation of a visual instruction, which was followed by the presentation of an auditory or a visual stimulus for 700 msec. A sparse sampling image acquisition protocol was used in which, at the end of each volume acquisition, a 2.36 sec period of silence occurred during which the gradients were switched off. The stimuli were presented, and the overt responses produced, during the silent interval. The subjects wore MR compatible headphones (Commander XG, Resonance technology, CA, USA) and their verbal responses were recorded through a MR compatible microphone attached to the headphones and digitized directly onto a laptop computer. The baseline condition, which was also presented in blocks of 12 or 13 trials, was a visual fixation condition during which participants stared at the word “rest”.

All stimuli were presented through Presentation® software (Neurobehavioral Systems, Albany, CA, USA). The experiment lasted approximately 40 min.

#### 3.1.3. Image acquisition

The data were acquired on a 3 T Siemens Trio MR scanner (Siemens AG, Erlangen, Germany) at the MNI. Thirty-six axial slices (whole brain coverage) oriented parallel to the AC–PC line (thickness = 4 mm, no gap, FOV = 256 × 256 mm², matrix = 64 × 64) were acquired in 2.64 sec using a multi-slice EPI sequence (TE = 50 msec, TR = 5 sec, delay in TR = 2.36 sec). The delay in TR occurred following each volume acquisition. The slices had a spatial resolution of $4 \times 4 \times 4$ mm. Two experimental runs (14 min each) resulted in the acquisition of 382 T2*-weighted BOLD images acquired in descending order. High-resolution T1-weighted volumes were acquired for anatomical localization (matrix $256 \times 256$ mm, 160 slices, $1 \times 1 \times 1$ mm, no gap, TE = 9.2 msec, TR = 22 msec). Participants’ head was immobilized by means of a vacuum-bag filled with polystyrene balls and a forehead-restraining device (Hybex Innovations, St-Leonard, Qc, Canada).

#### 3.1.4. Image analysis

The same analysis flow was used in both studies. T-statistical images were computed for four contrasts: (1) AWL minus baseline; (2) AWR minus baseline; (3) RS minus baseline, and (4) RA minus baseline. A cluster correction for multiple comparisons across the brain was implemented in the AFNI program 3dClusterSim, which takes into account the amount of smoothing in the data, the number of voxels and their sizes. Based on 3dClusterSim results, we determined that a FWE rate of $p < .05$ is achieved with a minimum cluster size of 44 contiguous voxels each significant at $p < .01$.

#### 3.1.5. ROI analysis

The process of identifying and extracting data in PT was identical to what was done in Experiment 1 with a manual
segmentation of PT into three non-overlapping medial, lateral and caudal regions. We then extracted the mean percentage of signal change for each condition and each PT region, as well as a number of difference scores. All analyses focused on (1) the difference between the two reading conditions (Read aloud–Read silent), (2) the two auditory conditions (AWR–AWL) and (3) the two speaking conditions (AWR–Read aloud). For each analysis, we and conducted a 2-way ANOVA with repeated measurements, with factor ROI (Medial, Lateral, Caudal PT) and Hemisphere (left, right).

3.1.6. SNR analyses

The sensitivity of our experimental design was evaluated using the same procedure described in Experiment 1.

3.2. Results

3.2.1. Behavioral data

No errors were made in the speaking tasks.

3.2.2. Neuroimaging results

3.2.2.1. SNR analyses. The mean SNR for the group was 83.5 (±10.3 SD). Our simulations determined that, given a power of .80, the minimum SNR to detect a MR signal change of .5% was 59, whereas the minimum SNR to detect a MR signal change of .80, the minimum SNR to detect a MR signal change of .5% was 

3.2.2.2. Voxel-based analyses. As illustrated in Fig. 4, for RA, AWL and AWR, the whole brain analyses revealed significant activation in the transverse temporal gyrus of Heschl and in the PT, bilaterally. For the two speaking conditions (RA and AWR), additional activity was found in the rostral and caudal regions of the precentral gyrus, bilaterally, in the IFG, also bilaterally, as well as in the right cerebellum (lobules V and VI) and in the left the rostral cingulate gyrus and sulcus. Activation in the insula was found during AWL, ARW and RA, but not RS. A list of all areas of significant activation, for each of the condition compared with baseline, is provided in Table 2.

3.2.2.3. ROI analysis. In order to examine hemispheric and regional differences in PT activation patterns during speech production, we focused on three different contrasts: (1) Read Aloud compared to Read Silently (2) AWR compared to AWL. These contrasts isolate (1) the effect of overt versus covert word production, (2) the effect processing one’s own auditory feedback over and above the processing of an external auditory stimulus, and, finally, (3) the difference between auditorily and visually triggered word production.

In a first step, we computed a set of difference scores removing the effect of Read silently from the effect of Read aloud (Read aloud–Read silently). These scores were entered in a 2-way ANOVA with repeated measurements with factor ROI (lateral PT, medial PT, caudal PT) and Hemisphere (left, right). Results show a main effect of ROI [F(2,9) = 7.05, p = .005], no main effect of Hemisphere [F(1,9) = .04, p = .84], and no interaction [F(2,18) = .19, p = .83]. As illustrated in Fig. 6, two-tailed FDR-corrected paired sample t-tests indicate that the lateral PT showed a significantly greater increase in activation magnitude between AWL and AWR. As shown in Supplementary Fig. S3, inspection of the activation patterns reveals strong activation in both AWL and AWR.

Finally, in a third step, (1) we examined whether each of the ROIs was significantly active for both production tasks (Read aloud and AWR) (p ≤ .05), and (2) we computed a set of difference scores removing the effect of Read aloud from the effect of AWR. Results of the first analysis showed that the lateral and medial PT were significantly active, bilaterally, for both speaking conditions, while the bilateral caudal PT was only significantly active for AWR. For the second analysis, the difference scores were entered in a 2-way ANOVA with repeated measurements with factor ROI (lateral PT, medial PT, caudal PT) and Hemisphere (left, right). Results show a main effect of ROI [F(2,9) = 7.05, p = .005], no main effect of Hemisphere [F(1,9) = .04, p = .84], and no interaction [F(2,18) = .19, p = .83]. As illustrated in Fig. 6, two-tailed FDR-corrected paired sample t-tests indicate that the lateral PT showed a significantly greater increase sensitivity for this contrast when compared to the medial PT [t(9df) = 3.29, p = .016] and the caudal PT [t(9df) = 4.29, p = .033].

3.2.2.4. Linear trend analyses. In order to further examine the contribution of the different sectors of PT to feedback processing, we conducted a trend analysis across study 1 and 2, which was warranted by the fact that the same ROIs were used across studies, the signal normalized to percentage of change, and the SNRs were comparable across studies. The focus of this analysis was to test for a progressive increase in activation in each PT region, for speaking tasks with increasing level of auditory feedback. For this analysis, hence, we examined a linear trend from Whisper masked (articulation, masked voiceless feedback), to Whisper (articulation and voiceless feedback), to Read aloud (articulation, voiced feedback). This analysis was conducted separately for the left and right PT regions. p values were FDR corrected. As shown in Fig. 7, this analysis revealed that activity in the left and right lateral PT did not increase commensurate to an increase in the amount of auditory feedback [left: F(1,9) = 4.37, p = .068; right: F(1,9) = 1.58, p = .24], similar results were obtained for the caudal PT [left: F(1,9) = .90, p = .77; right: F(1,9) = .098, p = .76]. In contrast, in the medial PT, there was a significant linear trend on the left [F(1,9) = 6.99, p = .027] but not on the right [F(1,9) = .60, p = .45].
3.3. Discussion

The main focus of Experiment 2 was to extend and refine the results of Experiment 1, which suggested that different parts of PT participate differentially in speech production and possibly in speech perception. Overall, our results support and extend the results of Experiment 1 in showing regional differences in the processing of the auditory signal that accompany single word production in a variety of contexts.

Importantly, our results demonstrate that PT, as a whole, is activated during passive listening, RA and AWR, but not during silent reading. Activation of PT during RA but not during reading silently indicates that PT is activated by endogenous (self-generated) auditory feedback consistent with the results from Experiment 1 using whispered speech. However, when comparing the activation associated with normal (voiced) speech and whispered speech it becomes clear that when activated by voiced feedback, the lateral and
medial portions of PT are more responsive than the caudal portion. One finding of note related to word reading is the lack of PT activation (in all three regions) during silent reading (RS), a finding that cannot be attributed to a lack of sensitivity of our experimental design, as revealed by the SNR analyses, or our spatial smoothing window (Buchsbaum et al., 2005). The lack of activation in the silent reading condition contrasts with previous studies employing silent reading yielding

Table 2 – Table of statistically significant activation at the whole-brain level, corrected for multiple comparison (FWE rate of p < .05, achieved with a minimum cluster size of 44 contiguous voxels each significant at p < .01), for Experiment 2, for each condition compared against a resting baseline: (1) Read silently, (2) Read aloud, (3) AWL and (4) AWR.

<table>
<thead>
<tr>
<th>Description</th>
<th>Hemi</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Voxels</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Read silently</td>
<td>L</td>
<td>-8</td>
<td>-71</td>
<td>-2</td>
<td>44</td>
<td>.01000</td>
</tr>
<tr>
<td>B. Read aloud</td>
<td>R</td>
<td>50</td>
<td>-13</td>
<td>5</td>
<td>300</td>
<td>.00027</td>
</tr>
<tr>
<td>Transverse temporal gyrus and planum temporale, extending downward into the superior temporal sulcus, medially into the insula, and laterally and rostrally into the inferior frontal gyrus (pars opercularis and triangularis).</td>
<td>L</td>
<td>-43</td>
<td>-25</td>
<td>11</td>
<td>146</td>
<td>.00029</td>
</tr>
<tr>
<td>Caudal middle temporal gyrus.</td>
<td>L</td>
<td>-37</td>
<td>-49</td>
<td>-1</td>
<td>53</td>
<td>.00006</td>
</tr>
<tr>
<td>Para cingulate sulcus, and medial frontal gyrus (pre-SMA).</td>
<td>L</td>
<td>-8</td>
<td>14</td>
<td>39</td>
<td>49</td>
<td>.00002</td>
</tr>
<tr>
<td>C. AWL</td>
<td>R</td>
<td>53</td>
<td>-26</td>
<td>0</td>
<td>331</td>
<td>.00001</td>
</tr>
<tr>
<td>Transverse temporal gyrus and planum temporale, expanding medially into the posterior insula, and ventrally into the superior temporal sulcus and middle temporal gyrus.</td>
<td>L</td>
<td>-38</td>
<td>-26</td>
<td>-23</td>
<td>306</td>
<td>.00007</td>
</tr>
<tr>
<td>D. AWR</td>
<td>R</td>
<td>45</td>
<td>-17</td>
<td>4</td>
<td>673</td>
<td>.00000</td>
</tr>
<tr>
<td>Transverse temporal gyrus and planum temporale, superior temporal sulcus, expanding medially into the posterior insula and putamen and more dorsally and laterally on the pars opercularis of IFG, and on the precentral gyrus and central sulcus.</td>
<td>L</td>
<td>-32</td>
<td>-23</td>
<td>5</td>
<td>422</td>
<td>.00001</td>
</tr>
<tr>
<td>Transverse temporal gyrus and planum temporale, superior temporal sulcus, expanding medially into the posterior insula, the putamen, thalamus.</td>
<td>R</td>
<td>23</td>
<td>-53</td>
<td>-19</td>
<td>80</td>
<td>.00000</td>
</tr>
<tr>
<td>Cerebellum lobules V and VI.</td>
<td>R</td>
<td>-9</td>
<td>-26</td>
<td>29</td>
<td>49</td>
<td>.00002</td>
</tr>
<tr>
<td>Posterior cingulate gyrus.</td>
<td>L</td>
<td>-9</td>
<td>-26</td>
<td>29</td>
<td>49</td>
<td>.00002</td>
</tr>
</tbody>
</table>

Fig. 5 – Results of the ROI analyses for Experiment 2. The vertical axis represents the difference between activation levels for the reading conditions (Read aloud–Read silent) expressed as percentage of BOLD signal change, for each of the left (A) and right (B) PT regions (medial, lateral, and caudal). Asterisks indicate significant differences between PT regions.

Fig. 6 – Results of the ROI analyses for Experiment 2. The vertical axis represents the difference between activation levels for the speaking conditions (AWR–Read aloud) expressed as percentage of BOLD signal change, for each of the left (A) and right (B) PT regions (medial, lateral, and caudal). Asterisks indicate significant differences between PT regions.
activation in the left posterior PT area identified as area Spt (Hickok et al., 2000, 2003; Okada et al., 2003; Buchsbaum et al., 2005; Pa and Hickok, 2008). In previous studies, however, silent reading was combined with verbal working memory or articulatory rehearsal requirements. In the present study, participants were simply asked to read silently each visually presented word, and similar to previous observations of silent reading (see Fiez and Petersen, 1998; Turkeltaub et al., 2002; Mechelli et al., 2003) no activation in PT was observed. This finding more clearly highlights the processing requirements needed to activate the left posterior PT during reading and suggests that sound–symbol associations and sensorimotor transformations may not be part of the silent reading process.

Another interesting result from Experiment 2 is the reduced level of activity in all regions of PT for word repetition compared to word listening. Given that re-afference provides a substantial level of activation in all regions of PT, no difference in activation magnitude suggests a modulation of auditory input during speech production, reflecting a speech-induced suppression of auditory cortex activity (Numminen et al., 1999; Curio et al., 2000; Houde et al., 2002; Poulet and Hedwig, 2002). The role of speech-induced suppression is unclear but it is most likely related to changing the sensitivity of auditory neurons to self-generated speech (Behroozmand et al., 2010; Eliades and Wang, 2008; Poulet and Hedwig, 2003). It appears that the different regions of PT exhibit both idiosyncratic as well as general processing features such as suppression.

4. General discussion

In the recent years, the role of the left PT in speech and language has received much attention (e.g., Hickok and Poeppel, 2000, 2004, 2007; Warren et al., 2005a, 2005b). Despite the renewed interest in this region, the details of the functional and anatomical organization of the human PT remain to be uncovered. The present study is the first, to our knowledge, to examine the contribution of three anatomically distinct regions of the left and right PT to speech production using a set of four overt word production tasks with different levels of exogenous and endogenous auditory stimulation. Specifically, we examined the potential regional and inter-hemispheric specialization in PT for the production and perception of language. Our results demonstrate regional differences in the functional activation patterns in PT but similar inter-hemispheric trends. In the following paragraphs, we summarize and interpret these results.

4.1. Regional specializations in PT

Our results highlight important differences in the manner in which different regions of PT are modulated by different language tasks involving speech perception and speech production, and therefore emphasize the importance of considering the anatomical complexity in PT in interpreting functional magnetic imaging data (fMRI) data. Based on these findings, it appears that each region may be acting upon the incoming auditory signal in different ways and for different purposes. Here we suggest (1) that the lateral PT acts as a general auditory processing area, (2) that the medial PT is involved in processing speech feedback possibly as part of a feedback monitoring loop, and (3) that the caudal PT is involved in auditory–motor transformation for speech production.

4.1.1. Lateral PT: general auditory processing

The lateral PT was the only region whose activation magnitude increased for whispered auditory feedback relative to masked feedback. The apparent differential response to the presence of masking noise, whether applied during passive listening or in combination with self-generated feedback, suggests a stronger or more prominent role for this part of PT in auditory processing. In addition, the comparison of Experiments 1 and 2 reveals (1) that activation in lateral PT increased linearly as the amount of auditory stimulation increased, and (2) that it was more sensitive than all other ROIs to the contrast of AWR and RA, which isolates the effect of adding external auditory...
stimulation. Taken together, these findings suggest that the lateral PT is not specifically related to speech production or feedback control but instead in general auditory processing, perhaps acting, as suggested by Griffiths and Warren (2002), as a computational hub whose function is to disambiguate complex sounds, a process that would be achieved by isolating different properties of the acoustic objects and matching them to previously stored auditory templates.

4.1.3. Caudal PT and auditory—motor transformation
For the caudal PT, the lack of a linear increase in activity with increasing amount of auditory feedback during speech production is consistent with previous results showing caudal PT/Spt activation for tasks not involving auditory feedback such as silent speech (Hickok et al., 2000, 2003; Buchsbaum et al., 2001; Wise et al., 2001; Huang et al., 2001; Papathanassiou et al., 2000; Okada et al., 2003; Shergill et al., 2002; Callan et al., 2006; Pa and Hickok, 2008). Moreover, the caudal PT was not activated for RA but only for AWR, a task requiring the conversion of an auditory target into a motor program, as well as for AWL. This suggests that this region in not involved in processing self-generated feedback per se, but it may be involved in monitoring external auditory input in order to facilitate internally-generated motor commands. This interpretation is in line with a recent fMRI study in which participants were asked to listen to songs and speech, as well as to speak and sing spontaneously and in synchrony with external auditory stimuli. Results showed that activation in PT was found only in the synchronized production tasks (Saito et al., 2006). In keeping with these and previous results, here we suggest that the caudal PT may be involved in the process of integrating auditory and motor signals for sensorimotor transformation for speech production (Hickok and Poeppel, 2000, 2004, 2007), or possibly articulatory rehearsal, which would account for the activation in this region during AWL.

4.2. Inter-hemispheric patterns
As discussed in the Introduction, the potentially “special” role of the left PT to auditory language processing has been a central theme in cognitive neuroscience since the first discovery of a leftward asymmetry in this region in the nineteen sixties (Geschwind and Levitsky, 1968; Galaburda et al., 1978; Steinmetz and Galaburda, 1991). In the present study, we found a number of interesting regional differences in the processing of speech across the different PT regions. Interestingly, our results demonstrate similar functional activation patterns, and functional activation magnitude, in the left and right PT regions, which suggests a bilateral processing of incoming auditory signal, which does not support the hypothesis of a distinct contribution of the left PT to the processing and production of speech. Admittedly, however, it is possible that some regions of the left PT contribute more strongly than the their corresponding regions in the right PT during different tasks, for instance more demanding phonological tasks. This will have to be examined in future studies.

4.3. Conclusion
In summary, the present findings stress the importance of considering the anatomical complexity in PT in interpreting fMRI data. Indeed, the functional activation patterns of three PT regions suggest important functional difference between the different regions of PT. The PT, as a whole, appears to be part of multiple networks in which internal and external auditory input is acted upon in multiple ways and for multiple purposes depending on the task.

Acknowledgments
We would like to thank Anthony S. Dick for his help with the analysis of SNR and his comments on several versions of this manuscript, the research staff and MR technicians at the Montreal Neurological Institute (Montreal, Canada) for their help in collecting the data, and all the participants. This study was supported by research grants from the Canadian Institute of Health Research (CIHR) and Natural Science and Engineering Council of Canada to V.L. Gracco and a CIHR post-doctoral fellowship to P. Tremblay.

Appendix. Supplementary material
Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cortex.2011.09.004.

References
by speech and nonspeech sounds. Cerebral Cortex, 10(5): 512–528, 2000.


