Evidence for Cerebellar Contributions to Adaptive Plasticity in Speech Perception

Sara Guediche1,2,5, Lori L. Holt1,2,5, Patryk Laurent1,2,6, Sung-Joo Lim2,3 and Julie A. Fiez1,2,4

1Center for Neuroscience at the University of Pittsburgh, Pittsburgh, PA, USA, 2Center for the Neural Basis of Cognition, Pittsburgh, PA, USA, 3Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA, 4Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA, 5Current address: Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA and 6Current address: Brain Corporation, San Diego, CA, USA

Human speech perception rapidly adapts to maintain comprehension under adverse listening conditions. For example, with exposure listeners can adapt to heavily accented speech produced by a non-native speaker. Outside the domain of speech perception, adaptive changes in sensory and motor processing have been attributed to cerebellar functions. The present functional magnetic resonance imaging study investigates whether adaptation in speech perception also involves the cerebellum. Acoustic stimuli were distorted using a vocoding plus spectral-shift manipulation and presented in a word recognition task. Regions in the cerebellum that showed differences before versus after adaptation were identified, and the relationship between activity during adaptation and subsequent behavioral improvements was examined. These analyses implicated the right Crus I region of the cerebellum in adaptive changes in speech perception. A functional correlation analysis with the right Crus I as a seed region probed for cerebral cortical regions with covarying hemodynamic responses during the adaptation period. The results provided evidence of a functional network between the cerebellum and language-related regions in the temporal and parietal lobes of the cerebral cortex. Consistent with known cerebellar contributions to sensorimotor adaptation, cerebro-cerebellar interactions may support supervised learning mechanisms that rely on sensory prediction error signals in speech perception.

Keywords: adaptation, cerebellum, fMRI, perceptual learning, supervised learning

Introduction

There is a rich literature that describes the neocortical regions involved in speech perception and production (e.g., Rauschecker and Tian 2000; Hickok and Poeppel 2007; Rauschecker and Scott 2009; Price 2012; Scott 2012). The role of subcortical regions in speech perception and production has received less attention. However, there are theoretical and empirical reasons to believe that subcortical regions, such as the cerebellum, play an important role (e.g., Fiez et al. 1992; Ackermann et al. 1997; Mathiak et al. 2002). For instance, in a meta-analysis of neuromaging studies, Stoodley and Schmahmann (2009) showed that language-related tasks engage subregions in the posterior lobe of the cerebellum, particularly Lobules VI and Crus I. Motor and sensorimotor tasks, on the other hand, engage subregions in the anterior lobe of the cerebellum, including Lobule V and adjacent portions of Lobule VI (Stoodley and Schmahmann 2009; Keren-Happuch et al. 2012). The present work investigates whether regions previously established as speech and language-related areas of the cerebellum contribute to adaptive changes in speech perception.

Historically, the cerebellum has been considered a “learning machine” that contributes to adaptive changes in behavior through supervised learning mechanisms (Marr 1969; Albus 1971). The role of the cerebellum in adaptive plasticity has extensively been studied using sensorimotor tasks, such as visually guided reaching with concurrent visual or somatosensory perturbation (Clower et al. 1996; Wolpert et al. 1998, 2011; Baizer et al. 1999; Ramnani 2006; Redding 2006; Shadmehr et al. 2010). According to supervised learning models of sensorimotor adaptation, an expected sensory consequence is derived from an intentionally planned motor action (Wolpert et al. 1998, 2011; Ramnani 2006). By making the discrepancy between the expected and actual sensory outcomes, a sensory prediction error signal can be generated. This sensory prediction error signal can guide adaptive adjustments to sensorimotor relationships and reduce the magnitude of subsequent error signals (Kawato and Wolpert 1998; Wolpert et al. 1998, 2011; Kawato 1999). Multiple lines of evidence indicate that the cerebellum participates in this supervised learning process. For instance, functional imaging studies have linked adaptive changes in sensorimotor performance with changes in cerebellar activity, and lesion studies have shown that damage to the cerebellum impairs sensorimotor adaptation (e.g., Martin et al. 1996; Baizer et al. 1999).

Adaptive plasticity in speech production has also been examined with somatosensory perturbations that affect speech movements (e.g., externally generated jaw displacements) or sensory perturbations that distort the spoken output (e.g., distortions in the timing or acoustic spectra of the produced speech (Houde and Jordan 1998; Perkell et al. 2007; Villacorta et al. 2007; Shiller et al. 2009; Gollfinopoulos et al. 2011). Models of speech production have used supervised learning mechanisms to account for adaptive plasticity (Guenther and Ghosh 2003; Kotz and Schwartz 2010; Price et al. 2011; Tian and Poeppel 2012). Details vary across models. However, a common feature is that information about the planned movement is used to generate a predicted sensory outcome. Discrepancies between the predicted and actual sensory outcomes from the speaker’s own speech are used to create sensory prediction errors that “supervise” an adaptive change in speech production. In a neurocomputational model of speech production adaptation, Guenther and Ghosh (2003) associated somatosensory and auditory sensory prediction errors with corresponding sensory cortical areas. In this model, the cerebellum interacts with the cerebral cortex in feedback and feedforward control systems, which compute predicted sensory consequences of speech production and use sensory prediction error signals to guide adaptive motor change. The model was tested in a functional magnetic resonance imaging (fMRI) study that examined compensatory speech movements in response to somatosensory perturbations of the jaw.
Evidence for Cerebellar Involvement in Adaptive Plasticity

Golfinopoulos et al. (2011) have shown that the cerebellum is involved in adaptive plasticity in speech perception, without any apparent external feedback. That is, even mere exposure to distorted speech is sufficient to drive adaptive changes in speech perception. In cases where external feedback is unavailable, listeners' word knowledge appears to play an important role in mediating adaptive plasticity (Norris et al., 2003; Kraljic and Samuel, 2005; Maye et al., 2008). For instance, when listeners are presented with an ambiguous stimulus that can be perceived as multiple meanings, the resulting sensory prediction error signals could then be used to supervise and adaptively modify the mapping of the distorted speech onto prelexical representations. This would occur through processing loops that involve cerebral cortical regions associated with speech and language, and interconnected regions in the cerebellum. The speech perception literature has alluded to the possibility that lexically mediated perceptual adaptation likely relies on a supervisory learning mechanism to produce adaptive changes in perception (Norris et al., 2003; Davis et al., 2005; Vroomen et al., 2007). However, the biological processes that generate these learning signals remain unknown.

Methods

Participants
Twenty-three healthy volunteers, all right-handed, participated in this study. Five participants were not included in the analysis due to excessive head motion, 1 was eliminated due to equipment malfunction, and 2 were removed due to incidental neurological findings in the cerebellum. Note that the high exclusion rate was not unexpected because the participants provided written responses, which contributed to movement during the scanning session. The remaining participants were used in the group analysis (6 women and 9 men; mean age 23.3 ± 0.8). The subjects provided informed consent prior to participation according to a protocol approved by the University of Pittsburgh Institutional Review Board. The remaining participants were used in the group analysis (6 women and 9 men; mean age 23.3 ± 0.8). The subjects provided informed consent prior to participation according to a protocol approved by the University of Pittsburgh Institutional Review Board.
Stimuli
A female monolingual English speaker (L.L.H.) uttered a set of phonetically balanced English monosyllabic words (Egan 1948 lists 2–8, with pronouns and plurals excluded, 293 words) into an Electrovoice RE 20 microphone connected to a digital Marantz PMC670 recorder (16-bit resolution, 22 050 Hz). These natural utterances were equated in root mean square amplitude and submitted to a filtering process to create 2 versions of the word, one with severe acoustic distortion and another with moderate distortion. The distortion involved separating the speech spectrum into a set of 20 channels, compressing all spectral detail within a channel, and shifting the channels in the frequency domain either a moderate amount, or a great deal. Without the frequency-domain shift, a 20-channel speech compression of this sort is quite intelligible (Shannon et al. 1995). The moderate spectral shift produces a moderate decline in speech intelligibility, whereas the larger shift produces a more severe decline (see Guediche et al. 2009).

Experimental Procedure
In a slow event-related design, participants completed six 11-min runs (R1–R6) consisting of 30 trials each. These runs were defined by the nature of the speech stimuli presented. Each word was randomly selected from the larger set and presented only once. Natural undistorted spoken words defined the first run (R1), to examine the response to normal, intelligible speech. In R2, spoken words with the most severe distortion were presented in a Pretest phase. Words processed with a moderate distortion were presented in R3 and R4 in an Adaptation phase. Since these less severely distorted signals were moderately intelligible, they should at least partially activate lexical knowledge and provide a source of information to compute sensory prediction error signals to drive adaptation. In R5 and R6, the words processed with the most severe distortion were presented in a Posttest phase. Words processed with a moderate distortion were presented in R3 and R4 in an Adaptation phase. Since these less severely distorted signals were moderately intelligible, they should at least partially activate lexical knowledge and provide a source of information to compute sensory prediction error signals to drive adaptation. In R5 and R6, the words processed with the most severe distortion were presented in a Posttest phase. Words processed with a moderate distortion were presented in R3 and R4 in an Adaptation phase. Since these less severely distorted signals were moderately intelligible, they should at least partially activate lexical knowledge and provide a source of information to compute sensory prediction error signals to drive adaptation.
response was pseudorandomly selected across trials with the constraint that half of each response type occurred during the first half of the run (10 Written-Response and 5 No-Response) and the other half during the second half of the run. Stimulus presentation was controlled using an E-prime Software and randomized (without replacement) (Schneider et al. 2002).

Data Acquisition

Subjects were scanned using a 3.0-T Siemens Allegra Scanner. Structural images were collected using a T_{1}-weighted pulse sequence in 38 contiguous oblique slices (3.125 mm × 3.125 mm × 3.2 mm) parallel to the anterior commissure–posterior commissure (AC–PC) line. The AC–PC slice was selected for each individual to allow for maximum coverage of the cerebellum while ensuring coverage of the temporal and parietal cortex. Thirty-eight functional slices were collected in the same location as the structural slices using a one-shot echo-planar imaging pulse sequence [epmax64] (time repetition [TR] = 2 s, time echo = 25, Field of view = 200 mm, flip angle = 70°) yielding a total of 320 volumes were acquired for each run. Sagittal high-resolution, T_{1}-weighted MP images (1 mm × 1 mm × 1 mm) were also collected at the beginning of each scan session.

Behavioral Data Analysis

Each response was phonetically coded by a trained linguist using the International Phonetic Alphabet (IPA). The coded responses (Written-Response trials) were entered into a custom-designed program that computed the phoneme accuracy of each response to derive partial word accuracy measures instead of simply scoring responses based on whole word accuracy. In this algorithm, the first phoneme in the response was labeled correct if it was also found in the first position of the target lexical response. If the first response phoneme was a correct match, the second phoneme in the response was compared with the target phoneme in that position. If the first phoneme was incorrect, it was compared with the second position and so on until a match was found. If no match was found for the first phoneme, that phoneme was labeled incorrect, and the same procedure was applied to the subsequent phonemes. From these calculations, a partial word accuracy score was computed by multiplying the total number of correct (i.e., in-order matching) phonemes by the ratio of the number of phonemes in the target stimulus to that in the elicited response, or vice versa. The numerator was always the shorter, and the denominator the longer of the two, such that partial word accuracy scores penalized both extraneous and missed phonemes (The aim of this measure was to capture interactions between the serial order of phonemes and accuracy. Example stimuli and their scores are provided in Supplementary Materials).

Imaging Analysis

The imaging data were analyzed using the Neuroimaging Software Package (NIS 3.6) developed at the University of Pittsburgh and Princeton University. Automated Image Registration (AIR 3.08) was used to reconstruct and correct for subject motion (Woods et al. 1992). Participants with movement beyond 4 mm or 4° in any direction were excluded from the analysis. The images were then detrended to adjust for scanner signal drift within runs. The skulls were removed from each structural image, and the remaining brain tissue was coregistered to a common reference brain that was chosen from among the subjects in the dataset (Woods et al. 1993). The first trial of each run was removed from the analysis to avoid contamination from the MR frequency pulse. Functional images were normalized to common intensity values by the ratio of the number of phonemes in the target stimulus to that in the elicited response, or vice versa. The numerator was always the shorter, and the denominator the longer of the two, such that partial word accuracy scores penalized both extraneous and missed phonemes (The aim of this measure was to capture interactions between the serial order of phonemes and accuracy. Example stimuli and their scores are provided in Supplementary Materials).

The imaging data were analyzed using the Neuroimaging Software Package (NIS 3.6) developed at the University of Pittsburgh and Princeton University. Automated Image Registration (AIR 3.08) was used to reconstruct and correct for subject motion (Woods et al. 1992). Participants with movement beyond 4 mm or 4° in any direction were excluded from the analysis. The images were then detrended to adjust for scanner signal drift within runs. The skulls were removed from each structural image, and the remaining brain tissue was coregistered to a common reference brain that was chosen from among the subjects in the dataset (Woods et al. 1993). The first trial of each run was removed from the analysis to avoid contamination from the MR frequency pulse. Functional images were normalized to common intensity values by the ratio of the number of phonemes in the target stimulus to that in the elicited response, or vice versa. The numerator was always the shorter, and the denominator the longer of the two, such that partial word accuracy scores penalized both extraneous and missed phonemes (The aim of this measure was to capture interactions between the serial order of phonemes and accuracy. Example stimuli and their scores are provided in Supplementary Materials).

The imaging data were analyzed using the Neuroimaging Software Package (NIS 3.6) developed at the University of Pittsburgh and Princeton University. Automated Image Registration (AIR 3.08) was used to reconstruct and correct for subject motion (Woods et al. 1992). Participants with movement beyond 4 mm or 4° in any direction were excluded from the analysis. The images were then detrended to adjust for scanner signal drift within runs. The skulls were removed from each structural image, and the remaining brain tissue was coregistered to a common reference brain that was chosen from among the subjects in the dataset (Woods et al. 1993). The first trial of each run was removed from the analysis to avoid contamination from the MR frequency pulse. Functional images were normalized to common intensity values by the ratio of the number of phonemes in the target stimulus to that in the elicited response, or vice versa. The numerator was always the shorter, and the denominator the longer of the two, such that partial word accuracy scores penalized both extraneous and missed phonemes (The aim of this measure was to capture interactions between the serial order of phonemes and accuracy. Example stimuli and their scores are provided in Supplementary Materials).

Table 1

<table>
<thead>
<tr>
<th>Region</th>
<th>Talairach coordinates (x, y, z)</th>
<th>Voxels</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left postcentral gyrus</td>
<td>−42, −36, 56</td>
<td>93</td>
<td>6.13</td>
</tr>
<tr>
<td>Right superior temporal gyr</td>
<td>44, −38, 6</td>
<td>38</td>
<td>4.92</td>
</tr>
<tr>
<td>Left superior frontal gyr</td>
<td>−5, 33, 45</td>
<td>25</td>
<td>3.65</td>
</tr>
<tr>
<td>Right middle frontal gyr</td>
<td>37, 22, 30</td>
<td>23</td>
<td>3.75</td>
</tr>
</tbody>
</table>

Note: Regions listed showed a significant Phase (Pretest and Posttest) by Scan Time (11 time points per trial) interaction at a corrected cluster size for a voxel-wise threshold P < 0.001, at an alpha level of 0.05. Peak Talairach coordinates are reported at the maximum F-value.

Table 2

<table>
<thead>
<tr>
<th>Region</th>
<th>Talairach coordinates (x, y, z)</th>
<th>Voxels</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Lobule VI</td>
<td>−21, −54, −15</td>
<td>22</td>
<td>5.0</td>
</tr>
<tr>
<td>Right Lobule VI</td>
<td>23, −35, −25</td>
<td>13</td>
<td>4.7</td>
</tr>
<tr>
<td>Left Lobule VI/Crus I</td>
<td>−40, −42, −28</td>
<td>12</td>
<td>4.4</td>
</tr>
<tr>
<td>Right Crus I</td>
<td>34, −46, −38</td>
<td>5</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Note: Regions listed showed a significant Phase (Pretest and Posttest) by Scan Time (11 time points per trial) interaction at a corrected cluster size for a voxel-wise threshold P < 0.001, at an alpha level of 0.05. Peak Talairach coordinates are reported at the maximum F-value.

Evidence for Cerebellar Involvement in Adaptive Plasticity

Guediche et al.
Functional Correlations

The functionally defined regions in the cerebellum were used as seed regions for a further analysis, in which the simple correlations between each seed region and each voxel in the brain were computed using the 3dDeconvolve AFNI command (Ward 1998). Individual general linear models for each participant were generated to obtain $R^2$ values. The square root of the $R^2$ value was then transformed using a Fisher $z$-transformation. A $t$-test was then performed on the $z$-scored correlation coefficient values using each participant to generate a group $t$-map that was visualized using AFNI at a voxel-wise $P$-value of <0.005 with a corrected cluster size that was determined to be 51 voxels using a minimum corrected cluster size at an alpha level of 0.05 with the AFNI AlphaSim program (Ward 2000).

Results

An analysis of the partial word accuracy score data in the Pretest (R2, M = 15, SEM = 1) versus Posttest (R5, M = 21, SEM = 1) phases showed a significant improvement [$t_{(13)} = 3.52, P = 0.004$]. This suggests that exposure to the moderately distorted stimuli during the intervening Adaptation phase drove an adaptive change in speech perception, even without explicit feedback. Figure 3 graphically shows mean partial word accuracy performance for the distorted speech conditions used in the fMRI analyses. The mean partial word accuracy was lower for the moderately distorted stimuli (R3, R4, M = 28, SEM = 2) than natural, undistorted speech (R1, M = 90, SEM = 1), $t_{(13)} = -24.78, P < 0.001$.

The imaging data were analyzed using voxel-wise ANOVAs, with subject as a random factor, and Phases (Pretest and Posttest) and Scan Time (TRs 1–11) as within-subject factors. Of greatest interest were the clusters that exhibited a Phase × Scan Time interaction, since this interaction pattern reflects a Pretest versus Posttest change in the hemodynamic response to the distorted stimuli. The results from the whole-brain voxel-wise ANOVAs showed a Phase × Scan Time interaction in frontal, temporal, and motor areas of the cerebral cortex (Fig. 4). The loci of peak activations for these clusters are listed in Table 1. The significant activation clusters identified in the cerebellum for this interaction are listed in Table 2, and shown in Figure 5. Four clusters were identified in the cerebellum. The peaks of these clusters fell within the following subregions as estimated by visual inspection of high-resolution magnetization-prepared 180 degrees radio-frequency pulses and rapid gradient echo images for landmarks defined by the Schmahmann cerebellar atlas (Schmahmann et al. 1999): Left Lobule V/VI, right Lobule V/VI, left Lobule VI/Crus I, and right Crus I (see Fig. 5). No additional cerebellar regions emerged at the lower voxel cluster size threshold, confirming the a priori expectation that subregions of the cerebellum previously implicated in speech and language would show changes associated with speech perception adaptation.

The right Crus I region showed a greater hemodynamic response (change in activity from baseline) in the Pretest compared with the Posttest (see Fig. 5). The experimental design included a response manipulation: for two-thirds of the trials, participants provided a Written-Response, whereas for the other one-third of trials, No-Response was required. A comparison between the mean percent signal change in the Written-Response compared with the No-Response trials revealed that the hemodynamic response in the right Crus I region was insensitive to the response demands, $t_{(13)} = -1.0, P = 0.31$ (see Fig. 5). For the other cerebellar regions, however, larger responses were observed for trials with a Written-Response, when compared with trials with No-Response, $t_{(13)} ≥ 2.5, P < 0.05$ (see Fig. 5). Thus, changes in these 3 Lobule V/VI regions may be driven in part by motoric task demands. Consequently, the extent to which these regions contribute to adaptive plasticity is unclear.

To further investigate whether any of the 4 cerebellar regions identified from the Pretest versus Posttest contrast contributed to the adaptive changes in speech perception, the relationship between the mean percent signal change during the Adaptation phase (R3 and R4) and behavioral measures of adaptation for each of these 4 regions was examined. For the right Crus I region, a significant relationship was found between the residual gain and the % BOLD signal change during the adaptation phase, $r_{(14)} = -0.60, P = 0.02$ (Fig. 6), whereas the correlation for the other 3 cerebellar regions did not reach significance ($P > 0.05$).

In a third analysis, each of the 4 cerebellar regions was used as a seed in a voxel-wise correlation analysis. For each seed

![Figure 4](http://cercor.oxfordjournals.org/)

**Figure 4.** Significant regions in the whole-brain Pretest versus Posttest contrast at a corrected cluster size for a voxel-wise threshold $P < 0.001$, at an alpha level of 0.05 from Table 1. Significant voxel clusters were found in the left superior frontal gyrus and right middle frontal gyrus (top panel) and the left postcentral gyrus and right superior temporal gyrus (bottom panel).

![Figure 3](http://cercor.oxfordjournals.org/)

**Figure 3.** Partial word accuracy performance for the distorted speech conditions used in the fMRI analyses. Error bars represent standard errors of the mean. The partial word accuracy score was computed by multiplying the total number of correct (i.e., in-order matching) phonemes by the ratio of the number of phonemes in the target stimulus to the number in the elicited response, or vice versa. For example, the score for a stimulus “Yeast” where the Response is “Least” is 0.75. The partial word accuracy scores used also penalize extraneous and missing phonemes. More examples of accuracy scores are provided in Supplementary Materials.
Each participant’s time course from the entire adaptation run was extracted and compared with the participant’s time courses for each voxel in the rest of the image volume. The computed voxel-wise correlations were used to establish potential cerebro-cerebellar connection pathways. For the right Crus I region, this analysis revealed significantly positively correlated voxel clusters in the left angular gyrus and significantly negatively correlated clusters in a left temporal area that included part of the left insula and extended into Heschl’s gyrus and the posterior temporal gyrus (see Fig. 7 and Table 3). For the remaining 3 regions, the analysis revealed a more widespread set of significantly correlated voxel clusters, including clusters within the left and right hemispheric motor and somatosensory areas (see Fig. 7 and Table 4).

Discussion
The known role of the cerebellum in sensorimotor adaptation tasks and its involvement in speech tasks motivated our hypothesis that the cerebellum contributes to adaptive...
plasticity in speech perception. Specifically, we proposed that discrepancies between the actual distorted acoustic speech input and the expected acoustic input associated with a lexical item engage cerebellar-dependent supervised learning mechanisms. Therefore, we predicted that improvements in the perception of severely distorted speech before (Pretest) compared with after adaptation (Posttest) would modulate activity in the cerebellum. A significant Pretest versus Posttest difference emerged in 4 distinct cerebellar regions. One region in the right Crus I of the cerebellum also showed a significant relationship between activity during adaptation and behavioral measures of adaptation, which provides additional evidence for cerebellar contributions to adaptive plasticity in speech perception. This region was functionally correlated with cerebral regions that encompassed portions of the left angular and left temporal gyri. The findings suggest that cerebro-cerebellar cortical interactions involving regions within the left temporal and parietal cortex, and regions within the right Crus I (and potentially Lobules V/VI), provide a functional network for achieving adaptive plasticity in speech perception.

In addition to identifying regions of significant change within the cerebellum, the Pretest versus Posttest contrast revealed differences within the cerebral cortex that localized to regions in frontal, temporal, and motor areas (see Fig. 4). These results are in line with current accounts of speech perception, which implicate frontal and temporal areas in different mechanisms. Therefore, in this study, the differences observed in the right hemisphere may reflect less reliance on processing certain aspects of the acoustic signal as the stimuli became more intelligible.

Differences in the cerebellar regions identified through the Pretest versus Posttest contrast were observed. A lateral region in the right cerebellar hemisphere, encompassing a portion of Crus I, exhibited a hemodynamic response that was insensitive to differences in motor output. Three other cerebellar regions also showed a Pretest versus Posttest difference in activity, which suggests that they may also play a role in adaptive speech perception. These regions encompassed portions of Lobule V/VI of the cerebellum, in both the right and left hemispheres. Unlike the right Crus I region, these 3 regions exhibited hemodynamic responses that were sensitive to the Written versus No-Response manipulation. This suggests that the hemodynamic changes in these regions may simply reflect changes in the motor components of the task that occurred as a consequence of adaptation (e.g., improved writing ability in the scanner with practice). A significant relationship between the activity of right Crus I during adaptation and a behavioral measure of adaptation was found. Taken together, the Pretest versus Posttest difference, the lack of sensitivity to motor task demands, and the relationship between activity during adaptation and behavioral measures of adaptation provide compelling evidence that implicates the right Crus I region in adaptive speech perception. Thus, we conclude that right Crus I plays an important role in adaptive speech perception, possibly in conjunction with other regions located in Lobules V/VI. Whereas the evidence supporting the involvement of right Crus I is straightforward, the involvement of the other regions in the adapting perception is less clear.

Findings from prior imaging studies are in accord with the general results from this study. Cerebellar activation has been reported in a number of auditory perception, speech perception, and language tasks (Fiez et al. 1992; Petacchi et al. 2005; Stoodley and Schmahmann 2009). Across neuroimaging and patient studies, the cerebellar areas recruited by perceptual and linguistic functions of speech tend to fall in either Lobule VI or Crus I, and they are distinct from Lobule V/VI regions that have been implicated in motor and sensorimotor aspects of speech (Stoodley and Schmahmann 2009; Keren-Happuch et al. 2012). The right Crus I region found in this study falls within the cerebellar territory associated with language function in 2 meta-analyses of prior neuroimaging studies, while the Lobule V/VI regions fall within the territory associated with motor function (Stoodley and Schmahmann 2009; Keren-Happuch et al. 2012). The distinctions between the Crus I and Lobule V/VI regions are also consistent with claims that the more evolutionarily recent portions of the cerebellum, which include Crus I, are involved in language and cognitive functions (Leiner et al. 1993). Our seed functional correlation analyses provided further evidence for functional distinctions between the right Crus I region and the 3 Lobule V/VI regions. Therefore, the right Crus I

### Table 3

<table>
<thead>
<tr>
<th>Region</th>
<th>Talairach coordinates (x, y, z)</th>
<th># voxels</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right thalamus</td>
<td>22, −26, 14</td>
<td>173</td>
<td>5.2</td>
</tr>
<tr>
<td>Left angular gyrus</td>
<td>−37, −57, 28</td>
<td>83</td>
<td>5.8</td>
</tr>
<tr>
<td>Left insula/Heschl’s gyrus/posterior</td>
<td>−38, −24, 18</td>
<td>76</td>
<td>−11.9</td>
</tr>
</tbody>
</table>

Note: Talairach coordinates at maximum t-value are listed for the regions that showed a significant correlation with a right Crus I seed region at a corrected cluster size for a voxel-wise threshold P < 0.005, at an alpha level of 0.05.

### Table 4

<table>
<thead>
<tr>
<th>Region</th>
<th>Talairach coordinates (x, y, z)</th>
<th># voxels</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left inferior frontal gyrus</td>
<td>−49, 3, 25</td>
<td>1961</td>
<td>−19.82</td>
</tr>
<tr>
<td>Left precentral gyrus</td>
<td>−19, −49, 41</td>
<td>1885</td>
<td>10.18</td>
</tr>
<tr>
<td>Left postcentral gyrus</td>
<td>−30, −40, 56</td>
<td>763</td>
<td>13.14</td>
</tr>
<tr>
<td>Right postcentral gyrus</td>
<td>33, −32, 51</td>
<td>749</td>
<td>10.65</td>
</tr>
<tr>
<td>Left thalamus</td>
<td>−15, −21, 10</td>
<td>376</td>
<td>12.40</td>
</tr>
<tr>
<td>Right thalamus</td>
<td>11, −12, 0</td>
<td>150</td>
<td>12.23</td>
</tr>
<tr>
<td>Right lentiform nucleus</td>
<td>22, −4, 10</td>
<td>147</td>
<td>9.50</td>
</tr>
<tr>
<td>Left superior temporal gyrus/</td>
<td>−37, −34, 15</td>
<td>121</td>
<td>−6.8</td>
</tr>
<tr>
<td>temporal gyrus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left medial frontal gyrus</td>
<td>−4, −15, 55</td>
<td>108</td>
<td>6.55</td>
</tr>
<tr>
<td>Right middle frontal gyrus/postcentral gyrus</td>
<td>38, 2, 40</td>
<td>92</td>
<td>−8.17</td>
</tr>
<tr>
<td>Right inferior frontal gyrus</td>
<td>48, 18, 20</td>
<td>91</td>
<td>−6.47</td>
</tr>
<tr>
<td>Right fusiform gyrus</td>
<td>51, −11, −26</td>
<td>68</td>
<td>5.76</td>
</tr>
</tbody>
</table>

Note: Talairach coordinates at maximum t-value are listed for the regions that showed a significant correlation with a right Lobule V/VI seed region at a corrected cluster size for a voxel-wise threshold P < 0.005, at an alpha level of 0.05.
region was significantly correlated with a left superior temporal voxel cluster (negative correlation) and with a left parietal voxel cluster (positive correlation), the 3 Lobule V/VI regions were most significantly correlated with voxel clusters in somatosensory and motor regions of the cerebral cortex. These results are consistent with neuroanatomical evidence from nonhuman primates (Dum and Strick 2003; Kelly and Strick 2003). More specifically, this literature indicates that the cerebellum receives input from the superior temporal plane and sparse input from the superior temporal sulcus (Schmahmann 1991). Connections between Lobules IV–VI and motor areas and Crus I/II and parietal cortex have been identified through neuroanatomical multisynaptic viral tracing methods (Kelly and Strick 2003). In humans, resting-state coherence measures and task-related functional connectivity measures have also revealed functional connections between Crus I and parietal cortex (Buckner et al. 2011). We conclude that the identified superior temporal and inferior parietal regions could plausibly participate in cerebro-cerebellar processing loops that support adaptive plasticity in speech perception.

Since adaptive changes in speech perception generalize to new items, it is thought that the locus of adaptive change must be relatively early within the speech processing pathway. This informs our interpretation of the cerebro-cerebellar processing loops that may drive adaptive changes in perception. The temporal area that emerged in our functional correlation analysis may be a target area that represents sensory prediction error signals. This interpretation is based on neurobiological models of speech perception, which typically posit engagement of primary auditory cortex and a belt of surrounding auditory association areas located along the superior temporal gyrus in prelexical speech processing (Rauschecker and Tian 2000; Rauschecker and Scott 2009; Okada et al. 2010; Peelle et al. 2010). Consistent with this interpretation, recent studies have shown modulation of activity in the superior temporal cortex as a function of predictive contexts (e.g., Davis 2011; Sohoglu et al. 2012; Wild et al. 2012) as well as the predictability of a sensory consequence associated with motor planning during speech production (Chang et al. 2013).

The inferior parietal area that emerged from our functional correlation analysis may be involved in guiding the supervised learning. For instance, cerebellar interactions with the angular gyrus may provide the information that is needed to compute the predicted sensory input: that is, the capacity to use the lexical-level representation of the distorted speech input to form predictions about the prelexical representation of the speech input. If the role of parietal cortex is to guide supervised learning and that of the temporal cortex is to represent the sensory prediction error signal, this could explain the opposite patterns of functional correlation with the right Crus I region. However, given the complexity of the response pattern in the cerebellar regions, any conclusions about directional differences in correlations are speculative.

The proposed lexical-mediation role for the inferior parietal region is consistent with prior findings. For instance, activity in the angular gyrus is related to improved perception of a speech distortion (e.g., responses to degraded sentences, Obleser et al. 2007; Eissner et al. 2010), and there is evidence that the angular gyrus is interconnected with areas associated with speech perception (e.g., Wernicke’s area, Friederici 2009) and lexico-semantic processing (e.g., the middle temporal gyrus, Fiez et al. 1996; Binder et al. 2009).

Prior findings also suggest that sensorimotor mechanisms could contribute to the adaptive functions of the angular gyrus. Studies of adaptive changes in speech production have provided evidence that the inferior parietal cortex provides an interface between motor and sensory signals that can be used to compute prediction error signals (Schultz and Dickinson 2000; Guenther and Ghosh 2003; Ito 2008; Shadmehr et al. 2010; Shum et al. 2011). However, the portion of the inferior parietal cortex most strongly implicated in speech production adaptation and speech monitoring is the left supramarginal gyrus (Desmond et al. 1997; Hickok 2009; Rauschecker and Scott 2009; Shum et al. 2011), whereas the findings in this study centered on the left angular gyrus. Potentially, the angular gyrus may be engaged when sensory predictions are based on internally generated lexical predictions. Binder et al. (2009) suggest that overlap between a semantic processing network and the “default network,” which includes parts of the angular gyrus, may support “processes that operate on ‘internal’ sources of information” (p. 2782). Since the angular gyrus has direct and indirect connections to areas associated with speech production (e.g., Broca’s area) (Friederici 2009; Turken and Dronkers 2011), one possibility is that internal motoric representations of the perceived lexical items could engage internal speech production mechanisms that generate predictions about the acoustic input associated with the lexical item. Although this possibility is tentative, there is evidence for auditory prediction derived from internally simulated speech (Tian and Poeppel 2010).

An attractive feature of a cerebellar-based account of adaptive plasticity is that it might allow multiple internal input-output mappings to be learned temporarily and be represented at the same time (Cunningham and Welch 1994; Kawato and Wolpert 1998; Wolpert et al. 1998, 2011; Ito 2008). Maintaining multiple mappings may be advantageous in speech perception since adaptation to some learned acoustic features generalizes, whereas adaptation to other features are specific across phonetic categories, speakers, or languages (e.g., Altmann and Young 1993; Kraljic and Samuel 2006, 2011; Bradlow and Bent 2008). However, the most important feature of a cerebellar-based account is that it provides a neurally plausible account of adaptive plasticity in speech perception.

To summarize, prior work on the role of the cerebellum in adaptive speech transformations has considered only the context of speech production. In this prior work, expected sensory outcomes are predicted based on the expected outcome of a planned movement, and used to derive the supervised prediction error signals that mediate adaptation (Wolpert et al. 1998; Doya 2000; Schultz and Dickinson 2000; Ito 2008). Our findings suggest that the cerebellum contributes to adaptive plasticity in speech perception through similar mechanisms. Adaptation-related changes in activity were found in the right Crus I region (and other regions in Lobules V/VI) of the cerebellum in response to distorted acoustic speech input that was not self-produced, and the magnitude of the Crus I response during an adaptation phase corresponded to behavioral measures of adaptive plasticity in speech perception. This perspective on speech perception adaptation forms a bridge between motor and nonmotor contributions of the cerebellum and extends understanding of speech processing network to subcortical structures. Furthermore, it offers a biologically plausible learning mechanism that could produce rapid adaptive changes in human speech perception.
Supplementary Material
Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Notes
The authors thank Fu and colleagues for providing Tigris, a program used to apply different levels of distortion to our sound files. The authors also acknowledge the contributions made by Christi Gomez for stimulus preparation and Jenna El-Wagaa and Natasha Bullock-Rest for data coding; Corrine Durisko and Kate Fissell for assistance with imaging data analysis; Andreea Bostan and Richard Dum for their expertise on cerebellar anatomy; and Peter Strick, Marc Sommer, Mark Wheeler, and Steve Small for their helpful discussions. Conflict of Interest: None declared.

Funding
This work was funded by the University of Pittsburgh, Kenneth P. Dietrich School of Arts and Sciences, and Andrew Mellon Predoctoral Fellowship (ROI MH 59256, RO1 DC 004674, and NSF 1125719).

References


