

Evidence for Cerebellar Contributions to Adaptive Plasticity in Speech Perception

Sara Guediche^{1,2,5}, Lori L. Holt^{1,2,3}, Patryk Laurent^{1,2,6}, Sung-Joo Lim^{2,3} and Julie A. Fiez^{1,2,4}

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

Address correspondence to Sara Guediche, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA. Email: sara_guediche@brown.edu

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 neuroimaging 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 computing 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; Golfopoulos 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

(Golfinopoulos et al. 2011). Hemodynamic response changes were found in both the cerebellum and speech-related cerebral cortical areas. These findings suggest that general theories about the contributions of the cerebellum to sensorimotor adaptation can be extended to the domain of speech production.

Whereas most theories about the cerebellum emphasize its motoric role, several theoretical accounts have speculated that the cerebellum performs functions involving supervised prediction error signals for both motor and sensory tasks (Doya 2000; Ito 2008; Strick et al. 2009). For instance, Bower (1997) suggests that the role of the cerebellum is to monitor sensory information to improve the efficiency of motor, sensory, and cognitive systems. Manto et al. (2012) went so far as to describe the cerebellum as a “sensory acquisition device,” whose adaptive function in sensorimotor tasks is in “controlling sensory surfaces.” Consistent with this view, a recent neuroimaging study showed changes in cerebellar activity that reflected encoding of sensory prediction errors (Schlerf et al. 2012), and a recent neuropsychological study (Roth et al. 2013) showed that participants with cerebellar damage performed more poorly on a visual perception adaptation task when compared with matched controls. Thus, one role of the cerebellum in perception may be to contribute to supervised learning mechanisms that rely on sensory prediction error signals.

The present study examines this general question for speech perception. The ability of listeners to adapt to distorted speech signals produced by other talkers is well documented (e.g., Schwab et al. 1985; Greenspan et al. 1988; Francis et al. 2000, 2007; Fenn et al. 2003; Clarke and Garrett 2004). In general, experience with distorted speech signals improves subsequent intelligibility of the distorted speech. These improvements are bolstered by information about the correct interpretation of the distorted speech. In many studies, this information has been provided explicitly. For example, the acoustic presentation of a distorted word has been followed by the written presentation and/or clear presentation of the word target (e.g., Schwab et al. 1985; Greenspan et al. 1988; Francis et al. 2000, 2007; Fenn et al. 2003; Hervais-Adelman et al. 2008). Importantly, though, adaptive plasticity has also been observed in the absence of such external feedback (e.g., Mehler et al. 1993; Liss et al. 2002; Bradlow and Bent 2008). That is, even mere exposure to distorted speech can be sufficient to drive adaptive changes in speech perception, without any apparent external feedback.

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 containing either of 2 possible phonemes, but only one of the possible percepts is a familiar word, changes in perception favor the direction that corresponds to the word context. Studies have also shown that the degree of adaptive plasticity is related to the intelligibility of the distorted stimuli (Bradlow and Bent 2008; Guediche et al. 2009; Li and Fu 2010). Less severely distorted speech signals that are more intelligible activate lexical knowledge to a greater degree (McClelland and Elman 1986) and produce greater adaptation effects. Thus, access to lexical knowledge facilitates adaptive plasticity.

We hypothesize a role for the cerebellum in adaptive plasticity in speech perception. More specifically, we propose that the cerebellum contributes to a supervised learning

mechanism, in which discrepancies between the distorted acoustic speech input and an expected acoustic input associated with a lexical item are used to drive adaptive change. Since lexically mediated adaptive changes in perception transfer to new words (e.g., Schwab et al. 1985; Francis et al. 2000, 2007), the locus of adaptation is likely to be prelexical. Therefore, to the extent that the distorted acoustic input can at least partially activate lexical knowledge, lexical information might be used to derive an expected pattern of activation of prelexical information that can be compared with the actual pattern of activation generated from the distorted acoustic input. This would allow internally generated lexical information (derived from the distorted sensory input) to serve as a basis from which sensory prediction error signals could be computed. The resulting sensory prediction error signals could then be used to supervise and adaptively modify the mapping of the distorted acoustic signal 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.

To avoid confounds related to a hypothesized role of the cerebellum in timing processes (Ivry 1996), we examined the current hypotheses using an acoustic distortion that alters spectral properties with a minimal change to the temporal properties of the acoustic speech signal (Shannon et al. 1995; Fu and Galvin 2003; Zeng 2004). Participants experienced Pretest, Adaptation, and Posttest phases during which they attempted to recognize distorted speech with no external feedback. During the Adaptation phase, the degree of distortion was moderate, resulting in partially intelligible speech that was expected to yield more accurate internal predictions derived from lexical activation with which to drive adaptation. The effects of adaptation were tested by comparing recognition of severely distorted speech at both Pre- and Posttest. We predicted that adaptation would result in Pretest versus Posttest differences in behavior and in cerebellar activity. Furthermore, we predicted that improvements in speech perception would correspond to changes in the blood oxygen level-dependent (BOLD) signal during the Adaptation phase. In summary, we expected to find evidence for cerebellar involvement in adaptive plasticity and for a cerebro-cerebellar functional network underlying adaptive plasticity in speech perception.

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 and were paid \$60 upon study completion. After careful examination of individual results, we found compromised data (extremely low mean signal intensity) in a cerebellar region of interest in one individual. Thus, our subsequent analyses did not include the data from this participant.

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).

Signal processing to achieve the distortion was accomplished using Tiger Speech (http://www.tigerspeech.com/tst_tigerdis.html; Fu and Galvin 2003; Li and Fu 2006). Each speech token was band-pass filtered into 20 frequency bands using eighth-order Butterworth filters.

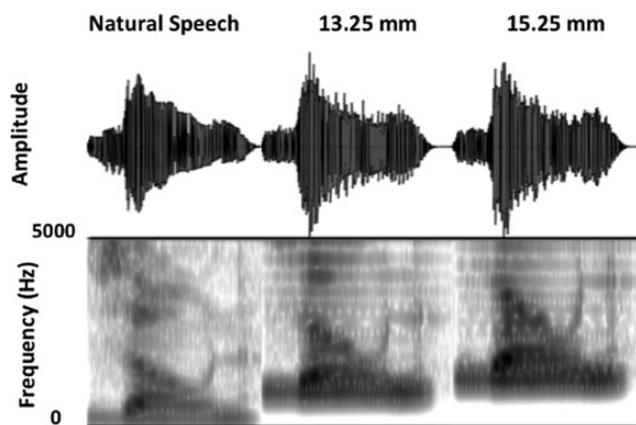


Figure 1. Waveforms (Time \times Amplitude, top) and Spectrograms (Time \times Frequency, with amplitude in gray scale, bottom) of an example word “zone.” On the left is the undistorted stimulus, in the middle the stimulus at the moderate distortion (13.25 mm), and on the right, the stimulus at the severe distortion (15.25 mm).

Following Nogaki et al. (2007), the corner frequencies of the bands were calculated using Greenwood’s (1990) formula to assure that each band was comparable in cochlear extent. Each band was half-wave rectified to extract the temporal envelope and low-pass filtered at 160 Hz. These envelopes served to modulate a sinusoidal carrier. To create the spectral shift, the carriers were frequency-shifted relative to the mean frequency of the band-pass analysis filter to either a moderate or severe degree (13.25 or 15.25 mm in terms of the Greenwood (1990) equation). These modulated carriers were summed and their overall level was adjusted to match the original speech tokens to create the compressed, spectrally shifted speech. These distortions result in poor speech intelligibility. For example, the severely distorted speech was shifted upward in frequency such that there was no spectral energy <1214 Hz. Since a great deal of information in the speech signal is carried in the lower frequencies (Fant 1949), this creates a complex mapping challenge for word recognition (see Fig. 1).

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 to examine the effects of the adaptation on Pretest versus Posttest responses to the severely distorted stimuli. Data from R6 were not included in the analyses, because the majority of the participants showed movement beyond 5 mm in any given direction during this final run.

The trial structure was identical across runs. On each 22 s trial, participants heard an acoustic stimulus through MR-compatible headphones and saw a fixation cross. This initial stimulus presentation period lasted 2 s, and was followed by an 8-s response period, and a 12-s delay (see Fig. 2). Owing to the historical focus on the cerebellum’s role in motor processes, we included 2 different response conditions to aid in differentiating cerebellar contributions to motor aspects of the task. On two-thirds of the trials (20 of 30), an on-screen cue (a question mark) prompted participants to write their response down on a note card and then turn the card, whereas on the other one-third of the trials, an on-screen cue (a “X”) indicated that they should not write a response. The duration of the response period in both cases was 8 s. The end of the response period was marked with a red fixation cross for a resting period that lasted 12 s. The nature of

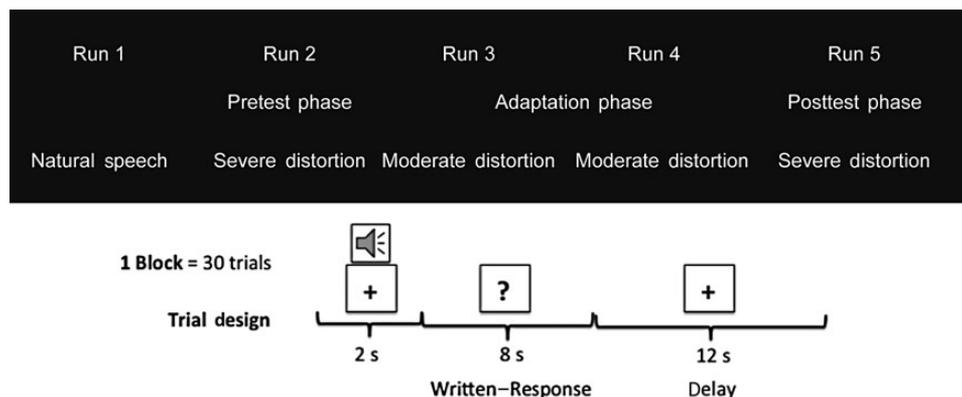


Figure 2. This figure illustrates the experimental and trial design. The main analysis compares Block 2 and Block 5. The trial design depicts a Written-Response trial. Two-thirds of the trials in each condition required a Written-Response. The other one-third of the trials did not require a Written-Response and had a fixation instead of the “question mark” after stimulus presentation.

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_2 -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 330 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 scaling the images to a global mean intensity and then smoothed using a 3-dimensional Gaussian filter (an 8-mm full-width at half maximum). The reference brain was then transformed into the Talairach space (Talairach and Tournoux 1988) using the affine transform in analysis of functional neuroimaging (AFNI).

Cerebellar regions of interest were identified by a voxel-wise repeated-measures analysis of variance (ANOVA) conducted on the Pretest (R2) and Posttest phases (R5) of the fMRI data using subjects as a random factor. The entire 22 s of each trial were used for the analysis. In a 2-way ANOVA, Phases (Pretest and Posttest) and Scan Time (each 2-s TR, 11 TRs

Table 1

Cerebral cortical regions showing the Pretest versus Posttest contrast

Region	Talairach coordinates (x, y, z)	Voxels	F-value
Left postcentral gyrus	−42, −36, 56	93	6.13
Right superior temporal gyri	44, −38, 6	38	4.92
Left superior frontal gyrus	−5, 33, 45	25	3.65
Right middle frontal gyrus	37, 22, 30	23	3.75

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

Regions showing the Pretest versus Posttest contrast that fell within the cerebellum

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

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.

per trial) were used as within-subject factors. An F-map of the ANOVA interaction effect was visualized using AFNI. Significant clusters of activation at a voxel-wise P -value of <0.001 and a contiguity threshold of 5 voxels were identified. To compute the extent threshold for significance in the language-related portions of the cerebellum, the a priori brain volume of interest, a partial cerebellum mask that encompassed regions implicated in speech and language (Stoodley and Schmahmann 2009; Keren-Happuch et al. 2012) was traced on the reference brain and used in AFNI's AlphaSim program (Ward 2000). This mask included all of the Lobules V, VI, and Crus I (1329 voxels). Using a minimum corrected cluster size for a voxel-wise P -value of <0.001 at an alpha level of 0.05, the extent threshold for significance was determined to be 5 voxels. We also computed an extent threshold for significance using the whole brain. This was determined to be 22 voxels for a minimum corrected cluster size for a voxel-wise $P < 0.001$ at an alpha level of 0.05. Clusters greater than or equal to the extent threshold are reported in Tables 1 and 2.

Next, the relationships between activity during the Adaptation phase and behavioral measures of improvement were examined for the significant cerebellar regions identified through the Pretest versus Posttest analysis. For each identified region in the cerebellum, the entire time course of signal intensity values (TRs 1–11) was extracted for each trial in the Adaptation phase (R3 and R4) for each participant. The extracted data were used to compute the average percent change in signal intensity from baseline for each participant. (The baseline consisted of averaged data acquired at the beginning [TR1] and end [TR11] of each trial.) The correlation between change in performance and % BOLD signal change during the Adaptation phase was then examined. Residual gain scores, which reduce error variance and systematic bias compared with raw gain scores (raw difference between Pretest and Posttest), were used to provide a base-free behavioral measure of change. This measure, which was calculated with a regression analysis on Posttest performance (mean partial word accuracy for Written-Response trials; 20 trials) using Pretest performance (mean partial word accuracy for Written-Response trials; 20 trials) as a predictor, is recommended for correlation analyses that use Pretest–Posttest scores and another variable (Manning and Dubois 1962; Cronbach and Furby 1970). Because the % BOLD signal change was not normally distributed, we used a rank-order transformation to conduct a nonparametric correlation analysis.

The mean percent signal change from baseline was also calculated for the Written-Response and No-Response trials, for each identified cerebellar region of interest. The mean % BOLD signal change for each participant was then used in a t -test comparison between Written-Response and No-Response trials in order to examine differences between these 2 response conditions in each cerebellar region.

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 \times 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 \times 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

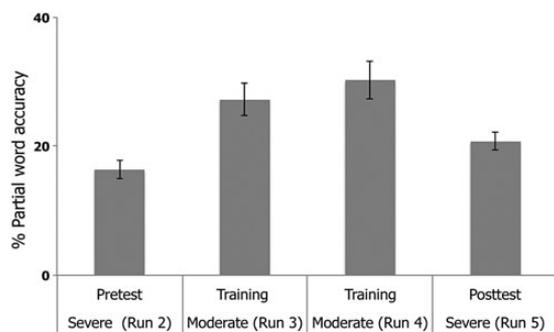


Figure 3. Partial word accuracy performance for the distorted speech conditions used in the fMRI analyses. Error bars represent standard errors of the mean. The partial word accuracy score was computed by multiplying the total number of correct (i.e., in-order matching) phonemes by the ratio of the number 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.

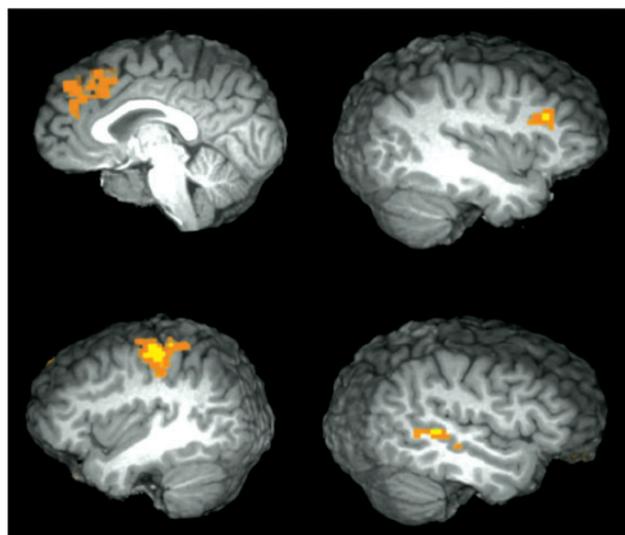


Figure 4. Significant regions in the whole-brain Pretest versus Posttest contrast at a corrected cluster size for a voxel-wise threshold $P < 0.001$, at an alpha level of 0.05 from Table 1. Significant voxel clusters were found in the left superior frontal gyrus and right middle frontal gyrus (top panel)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

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, $t_{(13)} \geq 2.5$, $P < 0.05$ (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

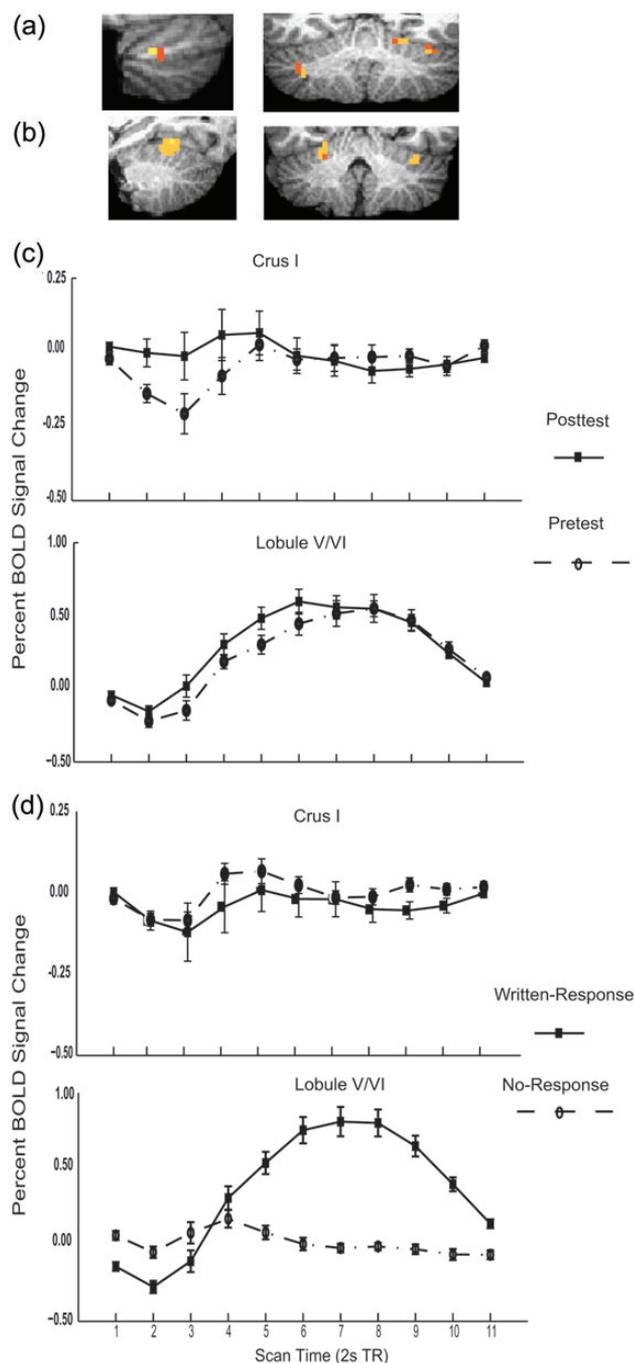


Figure 5. Sagittal and coronal slices (left is right and right is left) for significant regions in the cerebellum Pretest versus Posttest contrast at $P < 0.001$. (a) Right Crus I. On the left is a sagittal view at $x = 38$ and on the right is a coronal view at $y = -44$. (b) Regions in the left and right Lobule V/VI. On the left is a sagittal view at $x = -21$ and on the right a coronal view at $y = -36$ showing 2 regions in the left Lobule V/VI regions. (c) Time course showing a percent signal change from baseline for Pre- and Posttest in the right Crus I (top panel) and right Lobule V/VI (bottom panel). Error bars represent standard errors of the mean. (d) Time course showing a percent signal change for Written-Response compared with No-Response trials in the right Crus I (top panel) and right Lobule V/VI (bottom panel). Error bars represent standard errors of the mean.

region, 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.

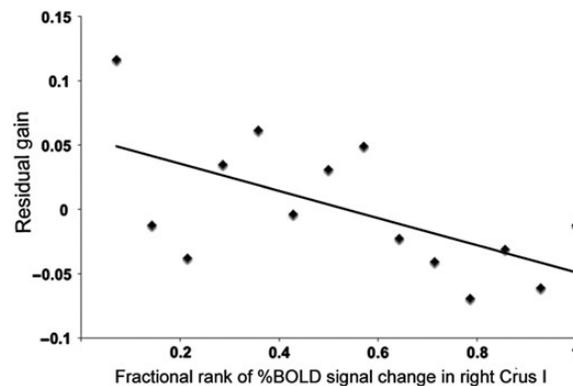


Figure 6. Scatter plot showing the relationship between behavioral adaptive plasticity measured as a residual gain of partial word accuracy scores (y-axis) and a rank of % BOLD signal from the baseline (x-axis).

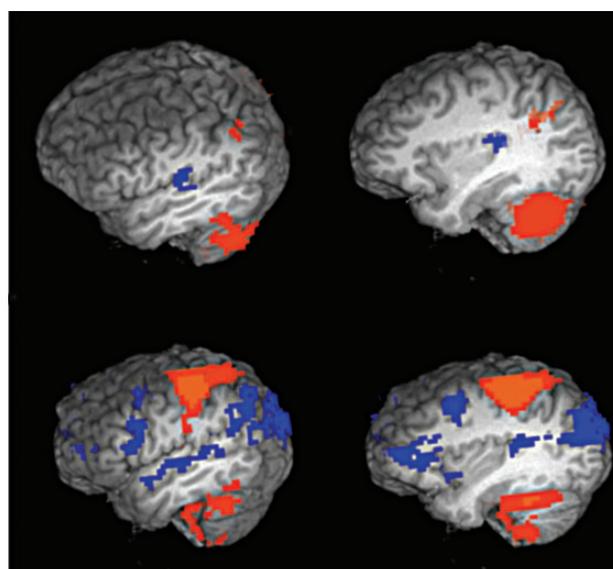


Figure 7. Functional connectivity map for the right Crus I. Sagittal slice at $x = -50$ and $x = -35$ (top panel) and right Lobule V/VI (bottom panel). Voxel-wise threshold at $P < 0.005$ (cluster threshold = 52 voxels) showing t -values from t -tests conducted on correlations of the Fisher z -transformation of the square root of R -values. Positive t -values are colored in red and negative t -values are colored in blue.

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

Table 3

Seed functional correlation for the right Crus I (35, -44, -38)

Region	Talairach coordinates (x, y, z)	# voxels	t-value
Right thalamus	22, -26, 14	173	5.2
Left angular gyrus	-37, -57, 28	83	5.8
Left insula/Heschl's gyrus/posterior Superior temporal gyrus	-38, -24, 18	76	-11.9

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

Seed functional correlation for the right Lobule V/VI (23, -35, -25)

Region	Talairach coordinates (x, y, z)	# voxels	t-value
Left inferior frontal gyrus	-49, 3, 25	1961	-19.82
Left precentral gyrus	-19, -49, 41	1865	10.18
Left postcentral gyrus	-30, -40, 56	763	13.14
Right postcentral gyrus	33, -32, 51	749	10.65
Left thalamus	-15, -21, 10	376	12.40
Right thalamus	11, -12, 0	150	12.23
Right lentiform nucleus	22, -4, 10	147	9.50
Left superior temporal gyrus/transverse temporal gyrus	-37, -34, 15	121	-6.8
Left medial frontal gyrus	-4, -15, 55	106	6.55
Right middle frontal gyrus/precentral gyrus	38, 2, 40	92	-8.17
Right inferior frontal gyrus	48, 18, 20	91	-6.47
Right fusiform gyrus	51, -11, -26	68	5.76

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.

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 aspects of speech processing. Specifically, superior temporal cortex has been associated with acoustic temporal analysis of speech signals and middle temporal cortex with lexical and semantic processing (Hickok and Poeppel 2007; Rauschecker and Scott 2009). Although both left and right temporal areas are

typically recruited during speech perception (Hickok and Poeppel 2007), in the present study, the observed changes in activity were right lateralized. Right lateralization has been attributed to a number of different factors, including processing focused on longer versus shorter timescales (Poeppel 2003), speech versus nonspeech stimuli (Molfese et al. 1975), and spectral versus temporal aspects of the acoustic signal (Zatorre and Belin 2001; Obleser et al. 2008). 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. Whereas the right Crus I

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; Eisner 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 (RO1 MH 59256, RO1 DC 004674, and NSF 1125719).

References

- Ackermann H, Graber S, Hertrich I, Daum I. 1997. Categorical speech perception in cerebellar disorders. *Brain Lang.* 60:323–331.
- Albus J. 1971. A theory of cerebellar function. *Math Biosci.* 10:25–61.
- Altmann GTM, Young D. 1993. Factors affecting adaptation to time-compressed speech. Paper Presented at the EUROSPEECH '93, Berlin, Germany.
- Baizer JS, Kralj-Hans I, Glickstein M. 1999. Cerebellar lesions and prism adaptation in macaque monkeys. *J Neurophysiol.* 81:1960–1965.
- Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex.* 19:2767–2796.
- Bower JM. 1997. Control of sensory acquisition data. *Int Rev Neurobiol.* 41:489–513.
- Bradlow A, Bent T. 2008. Perceptual adaptation to non-native speech. *Cognition.* 106:707–729.
- Buckner R, Krienen F, Castellanos A, Diaz J, Yeo B. 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol.* 106:2322–2345.
- Chang E, Niziolek C, Knight R, Nagarajan S, Houde J. 2013. Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci USA.* 110:2653–2658.
- Clarke C, Garrett M. 2004. Rapid adaptation to foreign-accented English. *J Acoust Soc Am.* 116:3647–3658.
- Clower DM, Hoffman JM, Votaw JR, Faber TL, Woods RP, Alexander GE. 1996. Role of posterior parietal cortex in the recalibration of visually guided reaching. *Nature.* 383:618–621.
- Cronbach LJ, Furby L. 1970. How we should measure “change”: or should we? *Psychol Bull.* 74:68–80.
- Cunningham HA, Welch RB. 1994. Multiple concurrent visual-motor mappings: implications for models of adaptation. *J Exp Psychol Hum.* 20:987–999.
- Davis MH, Ford MA, Kherif F, Johnsrude IS. 2011. Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *J Cog Neurosci.* 23:3914–3932.
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. 2005. Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen.* 134:222–241.
- Desmond JE, Gabrieli JD, Wagner AD, Ginier BL, Glover GH. 1997. Lobular patterns of cerebellar activation in verbal working memory and finger-tapping tasks as revealed by functional MRI. *J Neurosci.* 17:9675–9685.
- Doya K. 2000. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr Opin Neurobiol.* 10:732–739.
- Dum RP, Strick PL. 2003. An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. *J Neurophysiol.* 89:634–639.
- Egan JP. 1948. Articulation testing methods. *Laryngoscope.* 58:955–991.
- Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J Neurosci.* 30:7179–7186.
- Fant G. 1949. *Analys av de Svenska Konsonantljuden: Talets Allmänna Svängningsstruktur.* LM Ericsson, Stockholm, Sweden.
- Fenn K, Nusbaum H, Margoliash D. 2003. Consolidation during sleep of perceptual learning of spoken language. *Nature.* 425:614–616.
- Fiez JA, Petersen SE, Cheney MK, Raichle ME. 1992. Impaired non-motor learning and error detection associated with cerebellar damage. *Brain.* 115(Pt 1):155–178.
- Fiez JA, Raichle ME, Balota DA, Tallal P, Petersen SE. 1996. PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cereb Cortex.* 6:1–10.
- Francis A, Baldwin K, Nusbaum H. 2000. Effects of training on attention to acoustic cues. *Percept Psychophysiol.* 62:1668–1680.
- Francis A, Nusbaum H, Fenn K. 2007. Effects of training on the acoustic phonetic representation of synthetic speech. *J Speech Lang Hear Res.* 50:1445–1465.
- Friederici AD. 2009. Pathways to language: fiber tracts in the human brain. *Trends Cogn Sci.* 13:175–181.
- Fu Q, Galvin JJ III. 2003. The effects of short-term training for spectrally mismatched noise-band speech. *J Acoust Soc Am.* 113:1065–1072.
- Golfinopoulos E, Tourville J, Bohland J, Ghosh S, Nieto-Castanon A, Guenther F. 2011. fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage.* 55:1324–1338.
- Greenspan S, Nusbaum H, Pisoni D. 1988. Perceptual learning of synthetic speech produced by rule. *J Exp Psychol Learn.* 14:421–433.
- Greenwood DD. 1990. A cochlear frequency-position function for several species-29 years later. *J Acoust Soc Am.* 87:2592–2605.
- Guediche S, Fiez JA, Holt LL. 2009. Perceptual learning of distorted speech with and without feedback. Poster Presented at the 31st Annual Conference of the Cognitive Science Society, Amsterdam, The Netherlands.
- Guenther FH, Ghosh SS. 2003. A model of cortical and cerebellar function in speech. Proceedings of the XVth International Congress of Phonetic Sciences, Barcelona: 15th ICPhS Organizing Committee.
- Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP. 2008. Perceptual learning of noise vocoded words: effects of feedback and lexicality. *J Exp Psychol Hum Percept Perform.* 34:460–474.
- Hickok G. 2009. The functional neuroanatomy of language. *Phys Life Rev.* 6:121–143.
- Hickok G, Poeppel D. 2007. The cortical organization of speech processing. *Nat Rev Neurosci.* 8:393–402.
- Houde JF, Jordan MI. 1998. Sensorimotor adaptation in speech production. *Science.* 279:1213–1216.
- Ito M. 2008. Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci.* 9:304–313.
- Ivry RB. 1996. The representation of temporal information in perception and motor control. *Curr Opin Neurobiol.* 6:851–857.
- Kawato M. 1999. Internal models for motor control and trajectory planning. *Curr Opin Neurobiol.* 9:718–727.
- Kawato M, Wolpert D. 1998. Internal models for motor control. *Nov Found Symp.* 218:291–304.
- Kelly RM, Strick PL. 2003. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J Neurosci.* 23:8432–8444.
- Keren-Happach E, Chen SH, Desmond JE. 2012. A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Hum Brain Mapp.* doi:10.1002/hbm.22194.
- Kotz SA, Schwartze M. 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn Sci.* 14:392–399.

- Kraljic T, Samuel AG. 2006. Generalization in perceptual learning for speech. *Psychon B Rev*. 13:262–268.
- Kraljic T, Samuel A. 2011. Perceptual learning evidence for contextually-specific representations. *Cognition*. 12(3):459–465.
- Kraljic T, Samuel AG. 2005. Perceptual learning for speech: is there a return to normal? *Cogn Psychol*. 51:141–178.
- Leiner HC, Leiner AL, Dow RS. 1993. Cognitive and language functions of the human cerebellum. *Trends Neurosci*. 16:444–447.
- Li T, Fu Q. 2006. Perceptual adaptation to spectrally shifted vowels: training with nonlexical labels. *J Assoc Res Otolaryngol*. 8: 32–41.
- Li T, Fu Q. 2010. Effects of spectral shifting on speech perception in noise. *Hearing Res*. 270:81–88.
- Liss J, Spitzer S, Caviness J, Adler C. 2002. The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *J Acoust Soc Am*. 112:3022–3030.
- Manning WH, Dubois PH. 1962. Correlational methods in research on human learning. *Percept Mot Skills*. 15:187–321.
- Manto M, Bower JM, Conforto AB, Delgado-Garcia JM, Guarda SNF, Gerwig M, Habas C, Hagura N, Ivry RB, Marien P et al. 2012. Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement. *Cerebellum*. 11:457–487.
- Marr D. 1969. A theory of cerebellar cortex. *J Physiol*. 202:437–470.
- Martin TA, Keating JG, Goodkin HP, Bastian AJ, Thach WT. 1996. Throwing while looking through prisms. I. Focal olivocerebellar lesions impair adaptation. *Brain*. 119:1183–1198.
- Mathiak K, Hertrich I, Grodd W, Ackermann H. 2002. Cerebellum and speech perception: a functional magnetic resonance imaging study. *J Cogn Neurosci*. 14:902–912.
- Maye J, Aslin R, Tanenhaus M. 2008. The Weckud Wetch of the West: lexical adaptation to a novel accent. *Cognitive Sci*. 5: 543–562.
- McClelland J, Elman J. 1986. The TRACE model of speech perception. *Cogn Psychol*. 18:1–86.
- Mehler J, Sebastian N, Altmann G, Dupoux E, Christophe A, Pallier C. 1993. Understanding compressed sentences: the role of rhythm and meaning. *Ann N Y Acad Sci*. 682:272–282.
- Molfese DL, Freeman JR RB, Palermo DS. 1975. The ontogeny of brain lateralization for speech and nonspeech stimuli. *Brain Lang*. 2:356–368.
- Nogaki G, Fu QJ, Galvin JJ 3rd. 2007. Effect of training rate on recognition of spectrally shifted speech. *Ear Hear*. 28:132–140.
- Norris D, McQueen JM, Cutler A. 2003. Perceptual learning in speech. *Cogn Psychol*. 47:204–238.
- Obleser J, Eisner F, Kotz S. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci*. 28:8116–8123.
- Obleser J, Wise RJ, Alex Dresner M, Scott SK. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci*. 27:2283–2289.
- Okada K, Rong F, Venezia J, Matchin W, Hsieh I-H, Saberi K, Serences JT, Hickok G. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex*. 20:2486–2495.
- Peelle JE, Johnsrude IS, Davis MH. 2010. Hierarchical processing for speech in human auditory cortex and beyond. *Front Hum Neurosci*. 4:1–3.
- Perkell J, Lane H, Denny M, Matthies M, Tiede M, Zandipour M, Vick J, Burton E. 2007. Time course of speech changes in response to unanticipated short-term changes in hearing state. *J Acoust Soc Am*. 121:2296–2311.
- Petacchi A, Laird A, Fox P, Bower J. 2005. Cerebellum and auditory function: an ALE meta-analysis of functional neuroimaging studies. *Hum Brain Mapp*. 25:118–128.
- Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun*. 41:245–255.
- Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*. 62:816–847.
- Price CJ, Crinion JT, MacSweeney M. 2011. A generative model of speech production in Broca’s and Wernicke’s areas. *Front Psych*. 2.
- Ramrani N. 2006. The primate cortico-cerebellar system: anatomy and function. *Nat Rev Neurosci*. 7:511–522.
- Rauschecker J, Scott S. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*. 12:718–724.
- Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci*. 97:11800–11806.
- Redding GM. 2006. Generalization of prism adaptation. *J Exp Psychol Hum*. 32:1006–1022.
- Roth MJ, Synofzik M, Lindner A. 2013. The cerebellum optimizes perceptual predictions about external sensory events. *Curr Biol*. 23:930–935.
- Schlerf J, Ivry RB, Diedrichsen J. 2012. Encoding of sensory prediction errors in the human cerebellum. *J Neurosci*. 32:4913–4922.
- Schmahmann JD. 1991. Projections to the basis pontis from the superior temporal sulcus and superior temporal region in the rhesus monkey. *J Comp Neurol*. 308:224–248.
- Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M. 1999. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage*. 10:233–260.
- Schneider W, Eschman A, Zuccolotto A. 2002. E-prime user’s guide. Pittsburgh: Psychological Software Tools.
- Schultz W, Dickinson A. 2000. Neuronal coding of prediction errors. *Ann Rev Neurosci*. 23:473–500.
- Schwab E, Nusbaum H, Pisoni D. 1985. Some effects of training on the perception of synthetic speech. *Hum Factors*. 27:395–408.
- Scott S. 2012. The neurobiology of speech perception and production—Can functional imaging tell us anything we did not already know? *J Commun Disord*. 45:419–425.
- Shadmehr R, Smith M, Krakauer J. 2010. Error correction, sensory prediction, and adaptation in motor control. *Annu Rev Neurosci*. 33:89–108.
- Shannon R, Zeng F, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. *Science*. 270:303–304.
- Shiller DM, Sato M, Gracco VL, Baum SR. 2009. Perceptual recalibration of speech sounds following speech motor learning. *J Acoust Soc Am*. 125:1103–1113.
- Shum M, Shiller D, Baum S, Gracco V. 2011. Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *Eur J Neurosci*. 34:1817–1822.
- Sohoglu E, Peelle J, Carlyon R, Davis MH. 2012. Predictive top-down integration of prior knowledge during speech perception. *J Neurosci*. 32:8443–8453.
- Stoodley C, Schmahmann J. 2009. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*. 44:489–501.
- Strick P, Dum R, Fiez J. 2009. Cerebellum and nonmotor function. *Annu Rev Neurosci*. 32:413–434.
- Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain: An approach to medical cerebral imaging. Stuttgart, Germany: Thieme.
- Tian X, Poeppel D. 2012. Mental imagery of speech: linking motor and perceptual systems throughout internal simulation and estimation. *Front Hum Neurosci*. 6:314.
- Tian X, Poeppel D. 2010. Mental imagery of speech and movement implicates the dynamics of internal models. *Front Hum Neurosci*. 1:166.
- Turken AU, Dronkers NF. 2011. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Front Syst Neurosci*. 5:1–19.
- Villacorta V, Perkell J, Guenther F. 2007. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am*. 22:2306–2319.
- Vroomen J, van Linden S, deGelder B, Bertelson P. 2007. Visual recalibration and selective adaptation in auditory-visual speech perception: contrasting build-up courses. *Neuropsychologia*. 45:572–577.
- Ward D. 1998. 3dDeconvolve. Available from: <http://afni.nimh.nih.gov/afni/doc/help/3dDeconvolve.html>.

- Ward DB. 2000. Simultaneous Interference for fMRI Data. Milwaukee, WI: Biophysics Research Institute Medical College of Wisconsin.
- Wild CJ, Davis MH, Johnsrude IS. 2012. Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage*. 60:1490–1502.
- Wolpert D, Diedrichsen J, Flanagan J. 2011. Principles of sensorimotor learning. *Nat Rev Neurosci*. 12:739–751.
- Wolpert DM, Miall RC, Kawato M. 1998. Internal models in the cerebellum. *Trends Cogn Sci*. 2:338–346.
- Woods RP, Mazziotta JC, Cherry SR. 1993. MRI–PET registration with automated algorithm. *J Comput Assist Tomogr*. 17:536–546.
- Woods R, Cherry S, Mazziotta J. 1992. Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr*. 16 (4):620–633.
- Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb Cortex*. 11:945–953.
- Zeng F. 2004. Trends in cochlear implants. *Trends Amplif*. 8:1–34.