Neural mechanisms underlying auditory feedback control of speech

Jason A. Tourville[1], **Kevin J. Reilly**[1], and **Frank H. Guenther**[1,2,3,4]

Jason A. Tourville: jtour@cns.bu.edu

[1]Department of Cognitive and Neural Systems, Boston University, 677 Beacon St., Boston, MA, 02215, Telephone: (617) 353-5765, Fax Number: (617) 353-7755

[2]Division of Health Sciences and Technology, Harvard University – Massachusetts Institute of Technology

[3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital

[4]Research Laboratory of Electronics, Massachusetts Institute of Technology

Abstract

The neural substrates underlying auditory feedback control of speech were investigated using a combination of functional magnetic resonance imaging (fMRI) and computational modeling. Neural responses were measured while subjects spoke monosyllabic words under two conditions: (i) normal auditory feedback of their speech, and (ii) auditory feedback in which the first formant frequency of their speech was unexpectedly shifted in real time. Acoustic measurements showed compensation to the shift within approximately 135 ms of onset. Neuroimaging revealed increased activity in bilateral superior temporal cortex during shifted feedback, indicative of neurons coding mismatches between expected and actual auditory signals, as well as right prefrontal and Rolandic cortical activity. Structural equation modeling revealed increased influence of bilateral auditory cortical areas on right frontal areas during shifted speech, indicating that projections from auditory error cells in posterior superior temporal cortex to motor correction cells in right frontal cortex mediate auditory feedback control of speech.

Keywords

auditory feedback control; speech production; neural modeling; functional magnetic resonance imaging; structural equation modeling; effective connectivity

Introduction

While many motor acts are aimed at achieving goals in three-dimensional space (e.g., reaching, grasping, throwing, walking, and handwriting), the primary goal of speech is an acoustic signal that transmits a linguistic message via the listener's auditory system. For spatial tasks, visual feedback of task performance plays an important role in monitoring performance and improving skill level (Redding and Wallace, 2006; Huang and Shadmehr, 2007). Analogously, auditory information plays an important role in monitoring vocal output and achieving verbal fluency (Lane and Tranel, 1971; Cowie and Douglas-Cowie, 1983). Auditory feedback is crucial for on-line correction of speech production (Lane and Tranel, 1971; Xu et al., 2004; Purcell and Munhall, 2006b) and for the development and

---

Correspondence to: Jason A. Tourville, `jtour@cns.bu.edu`.

maintenance of stored motor plans (Cowie and Douglas-Cowie, 1983; Purcell and Munhall, 2006a; Villacorta, 2006).

The control of movement is often characterized as involving one or both of two broad classes of control. Under *feedback control,* task performance is monitored during execution and deviations from the desired performance are corrected according to sensory information. Under *feedforward control,* task performance is executed from previously learned commands, without reliance on incoming task-related sensory information. Speech production involves both feedforward and feedback control, and auditory feedback has been shown to impact both control processes (Houde and Jordan, 1998; Jones and Munhall, 2005; Bauer et al., 2006; Purcell and Munhall, 2006a).

Early evidence of the influence of auditory feedback on speech came from studies showing that speakers modify the intensity of their speech in noisy environments (Lombard, 1911). Artificial disruption of normal auditory feedback in the form of temporally delayed feedback induces disfluent speech (Yates, 1963; Stuart et al., 2002). Recent studies have used transient, unexpected auditory feedback perturbations to demonstrate auditory feedback control of speech. Despite being unable to anticipate the perturbation, speakers respond to pitch (Larson et al., 2000; Donath et al., 2002; Jones and Munhall, 2002; Natke et al., 2003; Xu et al., 2004) and formant shifts (Houde and Jordan, 2002; Purcell and Munhall, 2006b) by altering their vocal output in the direction opposite the shift. These compensatory responses act to steer vocal output closer to the intended auditory target.

The ease with which fluent speakers are able to coordinate the rapid movements of multiple articulators, allowing production of as many as 4–7 syllables per second (Tsao and Weismer, 1997), suggests that speech is also guided by a feedforward controller (Neilson and Neilson, 1987). Our ability to speak effectively when noise completely masks auditory feedback (Lane and Tranel, 1971; Pittman and Wiley, 2001) and the maintained intelligibility of post-lingually deafened individuals (Cowie and Douglas-Cowie, 1983; Lane and Webster, 1991) are further evidence of feedforward control mechanisms. The existence of stored feedforward motor commands that are tuned over time by auditory feedback is provided by studies of sensorimotor adaptation (Houde and Jordan, 2002; Jones and Munhall, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a). Speakers presented with auditory feedback containing a persistent shift of the formant frequencies (which constitute important cues for speech perception) of their own speech, will adapt to the perturbation by changing the formants of their speech in the direction opposite the shift. Following adaptation, utterances made immediately after removal or masking of the perturbation typically contain formants that differ from baseline formants in the direction opposite the induced perturbation (e.g., Purcell and Munhall, 2006a). These "overshoots" following adaptation indicate a reorganization of the sensory-motor neural mappings that underlie feedforward control in speech (e.g., Purcell and Munhall, 2006a) The same studies also illustrate that the feedforward speech controller continuously monitors auditory feedback and is modified when that feedback does not meet expectations.

The DIVA model of speech production (Guenther et al., 1998; Guenther et al., 2006) is a quantitatively defined neuroanatomical model that provides a parsimonious account of how auditory feedback is used for both feedback control and for tuning feedforward commands. According to the model, feedforward and feedback commands are combined in primary motor cortex to produce the overall muscle commands for the speech articulators. Both control processes are initiated by activating cells in a *speech sound map* (SSM) located in the left ventral premotor areas, including Broca's area in the opercular portion of the inferior frontal gyrus. Activation of these cells leads to the readout of excitatory feedforward commands through projections to the primary motor cortex. Additional projections from the

speech sound map to higher-order auditory cortical areas located in the posterior superior temporal gyrus and planum temporale encode auditory targets for the syllable to be spoken. The SSM-to-auditory error cell projections are hypothesized to have a net inhibitory effect on auditory cortex. The auditory targets encoded in these projections are compared to the incoming auditory signal by *auditory error cells* that respond when a mismatch is detected between the auditory target and the current auditory feedback signal. When a mismatch is detected, projections from the auditory error cells to motor cortex transform the auditory error into a corrective motor command. The model proposes that these corrective motor commands are added to the feedforward command for the speech sound so future productions of the sound will contain the corrective command. In other words, the feedforward control system becomes tuned by incorporating the commands sent by the auditory feedback control system on earlier attempts to produce the syllable.

Because the DIVA model is both quantitatively and neuroanatomically defined, the activity of model components in computer simulations of perturbed and unperturbed speech can be directly compared to task-related blood-oxygen-level-responses (BOLD) in speakers performing the same tasks. According to the model, unexpected auditory feedback should induce activation of auditory error cells in the posterior superior temporal gyrus and planum temporale (Guenther et al., 2006). Auditory error cell activation then drives a compensatory motor response marked by increased activation of ventral motor, premotor, and superior cerebellar cortex.

The current study utilizes auditory perturbation of speech, in the form of unpredictable upward and downward shifts of the first formant frequency, to identify the neural circuit underlying auditory feedback control of speech movements and to test DIVA model predictions regarding feedback control of speech. Functional magnetic resonance imaging (fMRI) was performed while subjects read aloud monosyllabic words projected orthographically onto a screen. A sparse sampling protocol permitted vocalization in the absence of scanner noise (Yang et al., 2000; Le et al., 2001; Engelien et al., 2002). An electrostatic microphone and headset provided subjects with auditory feedback of their vocalizations while in the scanner. On a subset of trials, an unpredictable real-time F1 shift was introduced to the subject's auditory feedback. Standard voxel-based analysis of neuroimaging data was supplemented with region of interest (ROI) analyses (Nieto-Castanon et al., 2003) to improve anatomical specificity and increase statistical power. Compensatory responses were also characterized behaviorally by comparing the formant frequency content of vocalizations made during perturbed and unperturbed feedback conditions. Structural equation modeling was used to assess changes in effective connectivity that accompanied increased use of auditory feedback control.

Methods

Subjects

Eleven right handed native speakers of American English (6 female, 5 male; 23–36 years of age, mean age = 28) with no history of neurological disorder participated in the study. All study procedures, including recruitment and acquisition of informed consent, were approved by the institutional review boards of Boston University and Massachusetts General Hospital. A scanner problem that resulted in the introduction of non-biological noise in acquired scans required the elimination of imaging data from one subject.

Experimental Protocol

Scanning was performed with a Siemens Trio 3T whole-body scanner equipped with a volume transmit-receive birdcage head coil (USA Instruments, Aurora, OH) at the Athinoula

A. Martinos Center for Biomedical Imaging, Charlestown, MA. An electrostatic microphone (Shure SM93) was attached to the head coil approximately 3 inches from the subject's mouth. Electrostatic headphones (Koss EXP-900) placed on the subject's head provided acoustic feedback to the subject at the beginning of each trial. Each trial began with the presentation of a speech or control stimulus projected orthographically on a screen viewable from within the scanner. Speech stimuli consisted of 8 /CɛC/ words (*beck, bet, deck, debt, peck, pep, ted, tech*) and a control stimulus (the letter string 'yyy'). Subjects were instructed to read each speech stimulus as soon as it appeared on the screen and to remain silent when the control stimulus appeared. Stimuli remained onscreen for 2 seconds. An experimental run consisted of 64 speech trials (8 presentations of each word) and 16 control trials. On a subset of speech trials, F1 of the subject's speech was altered before being fed back to the subject. Of the 8 presentations of each stimulus in an experimental run F1 was increased by 30% on 1 presentation (*shift up* condition trial), decreased by 30% on 1 presentation (*shift down* condition trial), and unaltered in the remaining 6 presentations (*no shift* condition trials). Trial order was randomly permuted within each run; presentation of the same stimulus type on more than 2 consecutive trials was prohibited as were consecutive F1 shifts in the same direction regardless of the stimulus. To allow for robust formant tracking and to encourage the use of auditory feedback mechanisms, subjects were instructed to speak each word slowly and clearly; production of each stimulus was practiced prior to scanning until the subject was able to consistently match a sample production. Still, subject mean vowel duration across all trial types ranged from 357 to 593 ms with standard deviations (SD) ranging from 44 to 176 ms). Paired *t*-tests indicated no utterance duration differences between the mean *no shift* and mean lumped *shift* responses ($df = 10$, $p = 0.79$) or between the *shift up* and *shift down* responses ($df = 10$, $p = 0.37$). Each subject performed 3 or 4 runs in a single scanning session, depending on subject fatigue and tolerance for lying motionless in the scanner. Stimulus delivery and scanner triggering were performed by *Presentation Version 0.80* (www.neurobs.com) software.

MRI Data Acquisition

A high resolution T1-weighted anatomical volume (128 slices in the sagittal plane, slice thickness = 1.33 mm, in-plane resolution = 1 mm^2, TR = 2000 ms, TE=3300 ms, flip angle = 7°, FOV = 256 mm^2) was obtained prior to functional imaging. Functional volumes consisted of 32 T2*-weighted gradient echo, echo planar images covering the whole brain in the axial plane, oriented along the bicommissural line (slice thickness = 5 mm, in-plane resolution = 3.125 mm^2, skip = 0 mm, TR = 2000 ms, TE 30 ms, flip angle = 90°, FOV = 200 mm^2).

Functional data were obtained using an event-triggered sparse sampling technique (Yang et al., 2000; Le et al., 2001; Engelien et al., 2002). The timeline for a single trial is shown in Fig. 1. Two consecutive volumes (each volume acquisition taking 2 seconds) were acquired beginning 5 seconds after trial onset. The 5 second delay period was inserted to allow collection of BOLD data at or near the peak of the hemodynamic response to speaking (estimated to occur approximately 4–7 seconds after vocalization). Auditory feedback to the subject was turned off during image acquisition to prevent transmittance of scanner noise over the headphones. The next trial started after another 3 second delay period, for a total trial length of 12 seconds and a total run length of 16 minutes. The sparse sampling design afforded several important advantages. First, it allowed subjects to speak in silence, a more natural speaking condition than speech during loud scanner noise. Second, it allowed for online digital signal processing of the speech signal to apply the perturbation, which is not possible in the presence of scanner noise. Finally, since scanning is carried out only after speech has ceased, it eliminates artifacts due to movement of the head and changing volume of the oral cavity during speech.

Acoustic Data Acquisition and Feedback Perturbation

Subject vocalizations were transmitted to a Texas Instruments DSK6713 digital signal processor (DSP). Prior to reaching the DSP board, the original signal was amplified (Behringer Eurorack UB802 mixer) and split into two channels using a MOTU 828mkII audio mixer. One channel was sent to the DSP board and the other to a laptop where it was recorded using Audacity 1.2.3 audio recording software (44,100 kHz sampling rate). Following processing, the DSP output was again split into two channels by the MOTU board, one channel was sent to the subject's headphones, the other to the recording laptop.

F1 tracking and perturbation, and signal resynthesis were carried out in the manner described by Villacorta et al.(In Press) The incoming speech signal was digitized at 8 kHz then doubled buffered; data was sampled over 16 ms blocks that were incremented every 8 ms. Each 16 ms bin was sampled at 8 kHz and then pre-emphasized to eliminate glottal roll off. To remove onset and offset transients, a hamming window was convolved with the pre-emphasized signal. An 8th order linear predictive coding (LPC) analysis was then used to identify formant frequencies. F1 was then altered according to the trial type before the signal was resynthesized and sent to the subject. A delay of 17 ms was introduced by the DSP board. Unperturbed trials were processed through the DSP in exactly the same manner as the perturbed trials except that the original F1 value was preserved, rather than shifted, during resynthesis; this was done to limit the difference in auditory feedback between perturbed and unperturbed trials to the first formant shift. The upward F1 shift had the effect of moving the vowel sound toward /æ/ (e.g., *bet→ bat*); a downward shift moved the vowel toward /ɪ/ (e.g., *bet→ bit*). When questioned following scanning, subjects reported no awareness of the feedback delay or alteration.

MRI Data Analyses

Voxel-based analysis—Voxel-based analysis was performed to assess task-related effects in a standardized coordinate frame using conventional image data analysis techniques, thereby permitting easier comparison with results from prior investigations. Image data were preprocessed using tools from the SPM2 software package provided by the Wellcome Department of Imaging Neuroscience, University College London (Friston et al., 1995b; http://www.fil.ion.ucl.ac.uk/spm/). Functional images were realigned to the mean EPI image (Friston et al., 1995a), coregistered with the T1-weighted anatomical dataset (Collignon et al., 1995), and spatially normalized into standard stereotaxic space using the EPI template provided by the Montreal Neurological Institute (MNI ICBM-152; Evans et al., 1993; Mazziotta et al., 2001). Functional images were then spatially smoothed (12 mm full-width-half-maximum Gaussian kernel) and globally scaled. Realignment parameters were included as covariates of non-interest in the study design prior to parameter estimation. Remaining global differences between the two volume acquisitions within each trial were removed during parameter estimation by a covariate that modeled these differences. The BOLD response for each event was modeled using a single-bin finite impulse response (FIR) basis function spanning the time of acquisition of the two consecutive volumes. *Shift up* and *shift down* conditions were lumped into a single shifted speech condition (hereafter referred to as *shift*) for fMRI analysis. Voxel responses were fit to a set of condition (*shift, no shift, baseline*) regressors according the general linear model.

Group statistics were assessed using fixed and mixed effects procedures. In mixed effects analysis, contrast-of-interest images were first generated for each subject by comparing the relevant condition parameter estimates on a voxel-by-voxel basis. Estimates for these analyses were obtained using a general linear model where conditions are treated as fixed effects. Group effects were then assessed by treating subjects as random effects and performing one-sample *t*-tests across the individual contrast images. The resulting group

parametric maps were thresholded by a corrected significance level to ensure a false discovery rate (FDR) < 5%. A map of normalized effect sizes for those voxels surpassing the significant *t* threshold was then created; suprathreshold voxel effects were divided by the mean significant ($p < 0.05$ uncorrected) effect of the *shift – baseline* contrast; this procedure permits assessment of relative activations. The contrast maps are shown in terms of effect size to provide a comparison of the how BOLD responses differed in the contrasted conditions.

ROI analysis—Region-of-interest (ROI) analysis was performed to test hypotheses regarding the response of specific regions to the task manipulation. According to the DIVA model, shifted feedback is expected to result in increased bilateral activation of planum temporal (PT), posterior temporal gyrus (pSTg), ventral motor and premotor cortex (vMC, vPMC), and the anterior medial cerebellum (amCB). Regional effects also served as the input for *post-hoc* tests of laterality and structural equation modeling. Regions included in these analyses were selected based on the finding of uncorrected significance in both the voxel-based and ROI-based analyses. Delineation of ROI boundaries was based on a set of *a priori* anatomical definitions and was independent of the results of the voxel-based analyses.

Cortical, subcortical, and cerebellar ROIs were created using Freesurfer (http://surfer.nmr.mgh.harvard.edu) image processing software (ROI abbreviations are provided in Table 1). Segmentation of gray and white matter structures (Fischl et al., 2002) and cortical surface reconstruction were performed on each anatomical volume (Fischl et al., 2004). Subcortical ROIs were segmented according to the Freesurfer-supplied subcortical training set. The Freesurfer cortical classifier (Fischl et al., 2002) was trained on a set of 14 manually parcellated brains. A modified version of the parcellation system defined by the Center for Morphometric Analysis (CMA) at Massachusetts General Hospital (Caviness et al., 1996) was used. This parcellation system is specifically tailored to studies of speech studies and includes both cortical and cerebellar ROIs (see Tourville, J.T. and Guenther, F.H., 2003. A cortical and cerebellar parcellation system for speech studies. Technical Report CAS/CNS-03-022, Boston University; http://speechlab.bu.edu/publications/Parcellation_TechReport.pdf). Cerebellar ROIs were parcellated by applying a cerebellar classifier to the segmented cerebellar gray and white matter in the same manner as that used for the subcortical ROIs (Fischl et al., 2002). The cerebellar classifier was based on manually parcellated cerebellums from the same 14 brains used to train the cortical classifier.

Characterization of BOLD responses within each ROI was performed according to the procedure described by Nieto-Castanon et al. (2003). Following spatial realignment, functional data was subjected to a rigid body transform and co-registered with the structural data set. The BOLD response averaged across all voxels within each ROI mask was then extracted. Regional noise temporal correlations were removed by whitening a fit of the estimated noise spectrum within each ROI. Average regional responses for each event were modeled using a single-bin FIR and fit to the same set of condition regressors (*shift, no shift, baseline*) used in the voxel based analyses.

Group effects were assessed by first computing regional contrasts for each subject. The regional contrasts were then pooled and tested for significance using one-sample *t*-tests. Regional effect sizes were normalized by the mean significant ($p < 0.05$, uncorrected) effect. In a first set of tests, the *a priori* hypotheses that responses in PT, pSTg, vMC, vPMC, and amCB are greater in the *shift* than the *no shift* conditions were tested. Probabilities were corrected to ensure FDR < 5%. Subsequent tests for significance were performed on the remaining ROIs (n = 132; only the posterior occipital cortex was excluded from ROI analysis). Regional effects associated with brain areas that were significant (uncorrected) in

the *shift – no shift* contrasts in both the ROI and voxel-based results were used as in *post-hoc* tests of laterality and structural equation modeling. Laterality effects were determined by pooling the effect for each ROI within each hemisphere across subjects ($n = 10$) and performing a paired *t*-test on the pooled data from each hemisphere.

In addition to providing average regional responses, our ROI analysis permits visualization of effects within ROIs (Nieto-Castanon et al., 2003). Briefly, voxel responses within each ROI were projected onto a spherical representation of the cortical surface. A reduced set of temporal eigenvariates was then created for each ROI by projecting the 2-D surface responses from each ROI onto a set of 15 orthogonal spatial Fourier bases and keeping only those components with low spatial frequency. The resulting set of eigenvariates for each ROI was fitted to the condition predictors and "eigenvariate contrasts" were calculated by comparing the appropriate condition effects. A spatial response profile was created for each subject by projecting the eigenvariate contrasts back onto the spherical surface using the transpose of the original set of orthogonal spatial bases. Spatial profiles were then averaged across subjects, normalized as described above, and the resulting group profile was flattened for display.

Structural equation modeling—Structural equation modeling (SEM) was performed to assess changes in effective connectivity between regions found significant in the *shift – no shift* contrast. Mean regional responses from each functional run were divided into two series, one consisting of only the 1st volume acquired for each trial, the other consisting of only the 2nd volume acquired for each trial. The two series were detrended and averaged to give a single response for each trial. Trials were then assigned to the appropriate group and concatenated within each subject. Outliers (> 3 standard deviations) were removed following mean correction, variance was unitized, and responses were concatenated across all subjects.

Covariance SEM (McIntosh and Gonzalez-Lima, 1994) was performed with AMOS 7 (http://www.spss.com/amos/index.htm) software. Path coefficients between observed variables were determined by maximum likelihood estimate. Differences in effective connectivity due to condition were assessed using the stacked model approach (Della-Maggiore et al., 2000). For a given network, the χ^2 goodness-of-fit measure was determined for a null model in which path coefficients are constrained to be equal between the two conditions and an unconstrained model in which they are allowed to vary. A comparison of the model fits, $\chi^2_{diff} = \chi^2_{null} - \chi^2_{uncon}$, was calculated using degrees of freedom equal to the difference in degrees of freedom between the two models. A significant χ^2_{diff} value was interpreted as evidence that the alternative model was a better fit than the null model and that the global effective connectivity of the network differed between the two conditions. The commonly used goodness-of-fit (GFI) and adjusted goodness-of-fit (AGFI) indices and the root mean square residual (RMR) and root mean square error of approximation (RMSEA) of differences between sampled and estimated variances and covariances were used to assess model fit (see Schumacker & Lomax (2004) and Hu and Bentler (1999) for a detailed description of these criteria). The alternative model also had to meet the AMOS 7 stability index criteria for both conditions. Connectivity between regions was constrained to meet acceptable fit and stability criteria and to produce path coefficients that were significant in at least one of the two conditions. These criteria were chosen to bias the network toward a parsimonious account of effective connectivity in the two conditions while still providing a good fit to the data. Significant connectivity was determined by converting estimated path coefficients to *z* statistics ($z =$ coefficient estimate/standard error estimate) then performing a two-tailed test that the absolute value of *z* is greater than 0 with $p > 0.05$). Comparisons of path coefficients in the two conditions were performed for the accepted

model; z statistics were calculated by dividing the difference between the estimated coefficient in each condition by the estimated standard error of the difference.

Acoustic Data Analysis

Subject responses were identified, isolated from the remainder of the acoustic recording, and resampled at 16 kHz. MATLAB software was created to identify the first two formants in the original and shifted feedback signals using LPC analysis. Analysis was performed on 20 ms samples of the speech signal, incremented every 4 ms. Vowel onset and offset, F1 and F2 contours, and signal intensity were estimated. Formant estimates for each utterance were visually inspected. If the initial formant estimation indicated a poor fit, the number of LPC coefficients was manually changed. The LPC order typically ranged between 16–18 coefficients for male speakers and 14–16 for female speakers. Onset and offset estimates were also modified as needed.

To determine whether subjects compensated for perturbations during the shift conditions, subject-specific baseline F1 traces were first created by averaging the *no shift* traces within each stimulus type. Traces were aligned to the onset of voicing. Averaging was done on a point-by-point basis and was restricted to time points that fell within the 80th percentile of all utterance lengths for a given subject to ensure a sufficient number of samples at each time point. Each shifted feedback trace was then divided by the appropriate subject- and stimulus-matched baseline *no shift* trace. The resulting F1 *shift up* and *shift down* compensation traces were then averaged across the 8 stimulus types within each subject. Characterizing the shifted feedback responses with respect to normal feedback responses was done to account for individual formant variation.

One-sample, two-tailed t-tests were performed at each time step to test for differences during the shifted conditions relative to the normal feedback condition. Compensation was detected when the null hypothesis ($H_0 = 1$) was rejected ($p < 0.05$) at a given time point and at all subsequent time points. These restrictions, which provide a conservative means for detecting compensation, necessarily overestimate the onset of compensation. Therefore, compensation response latencies were determined by fitting the mean subject compensation traces to a piecewise non-linear model of compensation (M_{Comp}). The model consisted of a constant segment (no compensation) followed by a logistic curve (compensation segment) of the form

$$M_{Comp} = \text{Max}\left[C, \frac{1 - e^{-k(t - t_0)}}{1 + e^{-k(t + t_0)}} \right],$$

where t is a vector of compensation trace time points, t_0 marks the start of the non-linear segment, i.e., the estimated onset of compensation, C is the value of the constant component, and k modulates the rate of change of the non-linear component. While C was allowed to vary during the estimation procedure, k was held constant at $k = 0.1$. C and t_0 were estimated for each subject by determining least squares fits of the model to the two F1 compensation traces, then averaging across subjects. t_0 estimates were constrained to a window between 20 and 250 ms after voicing onset; the lower limit corresponds to the earliest time point at which shifted feedback could be heard; the upper limit eliminated the influence of vowel-consonant transition effects introduced by short utterances. Confidence intervals for the t_0 fits were determined by a bootstrapping procedure. Compensation onsets were estimated from 1000 random resamples of the subject data with replacement for each condition. 95% confidence intervals were determined from the resulting t_0 distributions.

The magnitude of compensation was assessed by determining the peak response of the mean subject compensation traces in the two shift conditions. The peak response was restricted to time points following that each subject's estimated onset latency.

Results

Acoustic Responses

Subjects responded to unexpected F1-shifted auditory feedback by altering the F1 of their speech in the direction opposite the induced shift. Mean F1 traces from the *shift-up* and *shift-down* conditions expressed relative to their token-and subject-matched *no shift* responses are plotted in Fig. 2*A*. The compensation traces, averaged across subjects, demonstrate significant downward divergence in the *shift up* condition and upward divergence in the *shift down* condition compared to the *no shift* condition. One-sample *t*-tests were performed at each time point of the compensation traces to test for deviation from baseline F1 values. The tests revealed significant, sustained compensation ($df = 10$, $p < 0.05$) 176 and 172 ms after the onset of voicing in the *shift up* and *shift down* conditions, respectively.

Compensation response latencies were estimated by fitting subject compensation traces to a piece-wise non-linear model of compensation. The estimates are given with respect to the onset of perturbation. Subject latencies ranged from 87 to 235 ms (mean = 164.8 ms, SD = 43.5 ms) in the *shift up* condition and from 55 to 227 ms (mean = 107.7 ms, SD = 61.2 ms) in the *shift down* condition. Estimates of 95% confidence intervals (CI) for the mean response latencies were 145 – 186 ms and 81–139 ms in the *shift up* and *shift down* conditions, respectively. A paired *t*-test of response latencies demonstrated a significant difference between the two conditions ($df = 10$, $p = 0.01$).

Compensation magnitudes ranged from 11.1 to 59.4 Hz (mean = 30.0 Hz, SD = 14.8 Hz) in the *shift up* condition and 13.8 to 67.2 Hz (mean = 28.3 Hz, SD = 13.8 Hz) in the *shift down* condition. Expressed relative to the induced F1 perturbation, subjects compensated for 4.3% to 22.5% (mean = 13.6%, SD = 6.2%) of the upward shift and 6.3% to 25.5% (mean = 13.0%, SD = 6.7%) of the downward shift. A paired *t*-test comparing compensation magnitudes for the two shift directions indicated no difference ($df = 10$, $p = 0.75$).

The DIVA model was trained to produce the word /bɛd/ with normal feedback (see Guenther et al., 2006 for model details). Training the model consists of repeated attempts to match a dynamic auditory target formed by extracting the first 3 formants from recorded productions of the desired speech sound by an adult male speaker. Initially, the match between the model's output and the auditory target are relatively poor, resulting in corrective feedback commands. The corrective commands are used to update the model's feedforward controller (e.g., weights from a speech sound map cell to premotor cortex and auditory error cells) so that with each subsequent attempt, the model's performance is improved. As performance improves, control of the vocal tract shifts from relying primarily on the feedback control system and the model is able to reliably match the dynamic formant targets of the training example. Following successful training of word /bɛd/, the auditory perturbation task was simulated by shifting F1 feedback to the model by 30% in the upward and downward directions. The model's F1 response closely matched the experimental data (Fig 2*A*) following modification of those parameters that modulate the relative contributions of the feedforward (α_{ff}) and feedback (α_{fb}) commands to control of motor output. After training, α_{ff} and α_{fb} were 0.9 and 0.1, respectively. The F1 traces plotted in Fig. 2*B* were obtained from simulations of the upward and downward shifted F1 conditions in which $\alpha_{ff} = 0.85$ and $\alpha_{fb} = 0.15$. A static parameter governing inertial damping of the feedback command (FB_{INERT}; the relative weight of the previous feedback command value on the

current value) was also modified. The simulation results in Fig. 2C reflect a 10% increase in this damping parameter over the training value.

Neural Responses

Voxel-based analysis—BOLD responses during the normal and shifted speech conditions compared to the baseline condition are shown in Fig. 3. Voxels with peak t-statistic (peak-to-peak minimum distance = 6 mm) were assigned to anatomical regions based on cortical, subcortical, and cerebellar parcellation of the spm2 canonical brain into ROIs. For each region containing a peak response, the voxel location, t-statistic, and normalized effect of the maximum response are provided in Table 1. Talairach space (Talairach and Tournoux, 1988) coordinates reported in the table were determined using the MNI to Talairach mapping function described by Lancaster et al. (2007).

The *no shift – baseline* ($t > 3.63$, $df = 9$) contrast revealed a large network of active regions which has been previously implicated in speech production (Bohland and Guenther, 2006; Guenther et al., 2006), including bilateral peri-Sylvian auditory cortex, ventral Rolandic cortex, medial prefrontal cortex, anterior striatum, ventral thalamus, and superior cerebellum (peaking in lobule 6). Activations in the *shift – baseline* contrast ($t > 3.41$, $df = 9$) overlap those of the *no shift – baseline* contrast with two notable exceptions: extension of the superior cerebellar activation anterior-medially to include the cerebellar vermis bilaterally, and activation at the junction of the calcarine and parietal-occipital sulci.

Fig. 4A shows the results of a fixed effects analysis of the *shift – no shift* contrast ($t > 3.19$, $df = 5488$) corrected to ensure FDR < 5%. Bilateral activation of higher order auditory cortical areas in posterior superior temporal cortex, including the posterior superior temporal gyrus (pSTG) and planum temporale (PT), is consistent with the DIVA model prediction of auditory error cells in these areas (see simulation results in Fig 4B). Additional temporal lobe activity was noted in middle temporal-occipital cortex in the left hemisphere.

Responses in ventral Rolandic and lateral prefrontal cortex were noted only in the right hemisphere. In addition to ventral motor and somatosensory cortical activation, a region of activity along the ventral precentral sulcus extended to inferior frontal gyrus pars opercularis (IFo), and ventral premotor cortex (vPMC). More anteriorly, activity near the inferior frontal sulcus peaked in the inferior frontal gyrus pars triangularis (IFt). Activation in the inferior cerebellar cortex (lobule 8) was also found only in the right hemisphere.

Region of interest analysis—The *shift – no shift* results presented in Fig. 4A reflect group fixed-effects analysis; no voxels survived a mixed-effects analysis of the contrast following threshold correction (FDR < 5%). The responses shown in Fig. 4A therefore do not necessarily represent those of the general population. Failure to achieve significance at the population level can be the result of a lack of statistical power due to a limited sample size. It may also be due to anatomical variability in the subject population. To control for this possibility, we utilized a more sensitive ROI-based mixed effects analysis (Nieto-Castanon et al., 2003) on those regions to test our *a priori* hypotheses that perturbed feedback would cause increased activation in bilateral posterior auditory, ventral motor and premotor and superior cerebellar cortex.

Results from the ROI analysis of the *shift – no shift* contrast (Table 1, results in boldface type) supported our hypotheses regarding posterior auditory cortex. Significant bilateral responses ($t > 2.08$, $df = 9$, FDR < 5% for tests of 10 ROIs) were noted in pSTg and PT. However, increases in vMC, vPMC and amCB were significant only in the right hemisphere.

Tests on the remaining ROIs (n = 132) found no regions surviving a corrected significance threshold for the *shift – no shift* contrast. Several regions did survive individual (ROI-level) significance thresholds ($p < 0.05$), however, and these were consistent with the voxel-based results. Activation of a wider range of peri-Sylvian ROIs was noted bilaterally, including right inferior frontal cortex including Heschl's gyrus (Hg), planum polare (PP) and the parietal operculum (PO). Fig. 5 provides a more detailed illustration of activation within peri-Sylvian regions not directly visible in Fig. 3 and Fig. 4. The strongest responses were noted in the right hemisphere along the pSTg/PT border. An additional peak in posterior right pSTg is also visible. Peak activations in the left hemisphere were also found along the pSTg/PT border: a posterior peak in pSTg and an anterior peak in PT. Widespread activation in left PO was greater than that seen in the right hemisphere. Activation of primary auditory cortex, located in the postero-medial portion of Hg, was stronger in the right hemisphere. Other regions found active in both the voxel-based and ROI-based analyses included right inferior frontal gyrus, par opercularis and pars triangularis (iFo and iFt, respectively), right inferior posterior medial cerebellum (ipmCB), and left middle temporal-occipital cortex (MTO).

Apparent right-lateralized ventral frontal responses motivated post-hoc tests for hemispheric differences in vMC, vPMC, IFo, and IFt. Significant right lateralization in the *shift – no shift* contrast was found only in vPMC ($p = 0.03$). Significant left hemisphere lateralization of vMC ($p < 0.00$), vPMC ($p = 0.01$) and IFo ($p < 0.00$) was noted in the *no shift – baseline* contrast. vMC ($p < 0.00$) and IFo ($p = 0.01$) remained left lateralized in the *shift – baseline* contrast.

Structural equation modeling was used to compare interactions between the main regions found active in the *shift – no shift* contrast in the two speech conditions. The network shown in Fig. 6 easily met the prescribed fit and stability criteria. The unconstrained model, in which connections strengths were allowed to vary across the two speech conditions, provided a significantly better fit to the data than did the null model, in which the connections are constrained to be the same in both conditions ($\chi^2_{diff} = 19.89$, $df_{diff} = 9$, $p_{diff} = 0.02$). This indicates that the network was significantly modulated by the feedback perturbation. The unconstrained model provided an excellent fit to both the *no shift* and *shift* experimental data covariances ($\chi^2_{uncon} - 3.60$, $df_{uncon.} = 5$, $p_{uncon.} = 0.61$, $GFI_{uncon} = 1.00$, $AGFI_{uncon} = 1.00$, $RMR_{uncon} = 0.02$, $RMSEA_{uncon} = 0.00$). Pair-wise comparisons of path coefficients (Table 2) revealed that connection strengths from left pSTg to right pSTg, left pSTg to right vPMC, and from right pSTg to right IFt were significantly greater in the shift condition, indicative of greater use of these pathways when the auditory feedback control network was invoked due to the feedback shift.

Discussion

Formant shift compensation

As illustrated in Fig. 2, subjects responded to unexpected F1 shifts by altering the F1 of their speech in the direction opposite the induced shift. Computer simulations of the DIVA model verified the model's ability to account for these compensatory responses (Fig. 2*B*) following an increase in the relative contribution of auditory feedback on motor control. Adaptation to consistently applied upward or downward F1 shifts during production of /CɛC/ utterances similar to those used in the current study has been demonstrated previously (Purcell and Munhall, 2006a; Villacorta et al., In Press). In those studies, the F1 shift was presented on every trial throughout a training phase, allowing speakers to modify stored feedforward motor plans. Simulations of the DIVA model verified that the model's interactions between auditory feedback control and feedforward control could account for the adaptation results.

In the current study, the compensatory response is apparent *within an utterance* despite the unpredictable nature of the shift, and DIVA model simulations quantitatively account for this compensation.

The estimated compensation latencies (107 ms and 165 ms to the downward and upward shifts, respectively) were short enough to permit online correction within the duration of a typical /CɛC/ utterances (Hillenbrand et al., 2001; Ferguson and Kewley-Port, 2002). The estimates, particularly in the *shift down* condition, fall at the low end of ranges reported following unexpected pitch perturbation (Hain et al., 2000; Burnett and Larson, 2002; Xu et al., 2004). Faster response times may be due to differences in formant and pitch control (discussed further below) but may also reflect differences in the method used to determine response latencies. Though relatively short, the latencies are sufficiently long to allow for a cortically mediated compensatory response (see Guenther et al., 2006 for discussion) and are much longer than brainstem-mediated auditory perioral reflex responses (McClean and Sapir, 1981). A recent study used transcranial magnetic stimulation to demonstrate motor cortical involvement in phonetically specific compensation to jaw perturbation with a latency of approximately 85 ms but not during short-latency perioral reflex responses with an 18 ms latency (Ito et al., 2005).

The faster response to downward relative to upward F1 shifts during /ɛ/ production noted here has been reported previously (Purcell and Munhall, 2006b). While not conclusive, F1 may provide a more robust cue for distinguishing /ɛ/ and /ɪ/ than for /ɛ/ and /æ/ (Clopper et al., 2005), suggesting the downward F1 shift toward /ɪ/ is more likely to produce a phonemic or lexical categorical error than the upward shift toward /æ/, even if the acoustic difference is the same. The faster response to the downward shift may therefore reflect greater lexical or perceptual saliency. The impact of lexical saliency on compensatory responses is supported by a recent report describing the effects of F0 shift direction when unexpected perturbations were delivered between Mandarin bi-tonal disyllables. Shifts in the direction opposite the intended inter-syllabic tonal transition resulted in shorter latencies and larger compensations than did shifts in the same direction (Xu et al., 2004).

The auditory feedback control network

According to the DIVA model, perturbation of F1 feedback causes a mismatch between the auditory expectation for the current syllable and the auditory signal fed back to the subject (Guenther et al., 2006). This mismatch leads to activation of auditory error cells in the pSTg and PT, a prediction strongly supported by the bilateral peri-Sylvian activation noted in the *shift – no shift* contrast (Fig. 4*A*). Increased bilateral activation of posterior temporal regions during perturbation speech is consistent with previous results from studies of auditory feedback disruption, including delayed auditory feedback (Hirano et al., 1997; Hashimoto and Sakai, 2003), pitch perturbation (McGuire et al., 1996; Zarate and Zatorre, 2005; Fu et al., 2006) and noise masking (Christoffels et al., 2007)

Numerous lines of evidence support the hypothesis that the expected consequences of articulation and resulting auditory feedback are compared in posterior temporal cortex (see Guenther et al., 2006 for detailed discussion). Portions of posterior left PT and lateral pSTg bilaterally have been shown to respond during both speech perception and speech production in several studies (Hickok et al., 2003; Buchsbaum et al., 2005). Bi-directional functional connections between inferior frontal and posterior temporal cortex have been demonstrated *in vivo* using cortico-cortical evoked potentials in humans (Matsumoto et al., 2004). Attenuation of the posterior auditory cortex responses during self-produced speech has been shown in a number of studies (Paus et al., 1996; Numminen and Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Houde et al., 2002). Similar modulation of somatosensory responses to self generated movements (Blakemore et al., 1998) has been

interpreted as evidence that an efference copy from motor to sensory cortex encodes the expected sensory consequences of upcoming movements (Blakemore et al., 2000; Wolpert and Flanagan, 2001). According to these theories, the attenuation of sensory cortex by the motor plan efference copy effectively "cancels" the sensory feedback resulting from the movement. Recently, magnetoencephalography recordings demonstrated that the attenuation of auditory cortex is modulated by the how closely the auditory feedback from self-produced speech matches the expected (Heinks-Maldonado et al., 2006). Greater auditory response attenuation was noted when speakers heard normal rather than pitch-shifted auditory feedback. Work in monkeys has shown how precise the efference copy attenuation may be. Single unit recordings from monkey auditory cortex demonstrated pre-vocalization suppression of A1 and lateral belt neurons that was tightly linked to the subsequent vocal output (Eliades and Wang, 2005). Our current findings are consistent with these studies which collectively support the view that higher-level auditory cortical regions include auditory "error" cells that encode the difference between actual and expected auditory feedback during vocalization.

The DIVA model also predicts bilateral ventral precentral gyrus activation in the *shift – no shift* contrast, reflecting corrective commands sent from auditory error cells to the bilateral ventral motor cells that drive compensatory articulator movement (see Fig. 4B.; Guenther et al., 2006). The *shift – no shift* experimental results, however, revealed ventral precentral activation only in the right hemisphere in the *shift – no shift* contrast. Subsequent laterality tests on the ventral frontal responses revealed: (i) under normal auditory feedback conditions, control of the articulators is predominantly left lateralized, a finding noted previously for overt speech (Riecker et al., 2000; Sidtis et al., 2006), (ii) feedback-based articulator control relies on significantly greater involvement of right hemisphere ventral frontal regions, especially premotor and inferior frontal cortex (discussed further below), and (iii) in addition to the motor projections previously hypothesized (Guenther et al., 2006), auditory error cells appear to project to premotor and inferior prefrontal cortex in the right hemisphere. These conclusions were supported by structural equation modeling which revealed increased effective connectivity from left posterior temporal cortex to right posterior temporal and ventral premotor cortex. Reciprocal connectivity between right pSTg and vMC, though significant in both conditions, increased only modestly. Right posterior temporal cortex may exert additional influence over motor output via a connection through right IFt, which increased significantly in response to shifted feedback.

The activation of right inferior frontal gyrus, left posterior middle temporal gyrus, and right inferior cerebellum noted in the *shift – no shift* contrast were not predicted and did not reach a corrected significance threshold. However, their activation is noteworthy when considered with respect to other findings. The greater right IFt response during shifted feedback is consistent with the finding of increased BOLD response in this region when auditory feedback was delayed during speech production (Hashimoto and Sakai, 2003). The IFt increase was accompanied by increased right pSTg activation. A strikingly similar pattern of right hemisphere activation was also noted when listening to unfamiliar vs. familiar voices (Kriegstein and Giraud, 2004). Greater activation was noted near the right inferior frontal sulcus and posterior superior temporal gyrus/sulcus during the unfamiliar voice condition. A functional interaction between these regions was demonstrated only when unfamiliar voices were presented. Other studies have shown right inferior frontal activation when sensory input dictates alteration of a pre-set motor response, e.g., successful response inhibition or rapid switching of an ongoing task (see Aron et al., 2004 for review) and detection of rare ("oddball") sensory stimuli (Stevens et al., 2000). In general, right inferior frontal cortex appears to respond when sensory inputs dictate the need for increased sensorimotor processing, either due to the task definition (as in Stevens et al., 2000) or to the detection of performance errors (as in perturbed auditory feedback during speech). In the current study,

this activity may contribute to auditory feedback control by increasing the influence of sensory input on the motor output system.

Activation of right cerebellar lobule 8 has been associated with increased sequence complexity (Bohland and Guenther, 2006) and limb motor task complexity (Habas et al., 2004; Habas and Cabanis, 2006). This area has also been associated specifically with motor error correction, becoming active when unexpected execution errors were induced during a reaching task (Diedrichsen et al., 2005). This result is consistent with the present finding of increased lobule 8 activation when sensory error was introduced.

Activation in the left hemisphere posterior middle temporal cortex was found near the temporal-occipital junction in the *shift – no shift* comparison. This region has been hypothesized to serve as a lexical store (Indefrey and Levelt, 2004; Prabhakaran et al., 2006). Increased activation during the perturbation condition may therefore be the result of the F1 shift causing the speaker to hear an unanticipated word (e.g., /bɪd/ instead of /bɛd/). Further study is required to incorporate this region and the inferior cerebellum into a comprehensive model of auditory feedback control.

Implications for speech disorders

The current findings shed light on a perplexing issue in speech and language neuroscience. Functional imaging of speech production typically results in bilateral pre-frontal and sensori-motor activation (e.g. Wise et al., 1999; Bohland and Guenther, 2006; Ozdemir et al., 2006; Soros et al., 2006). These findings appear to conflict with the large body of lesion data that supports the traditional view of left hemisphere dominance of speech production (Duffy, 1995; Dronkers, 1996; Kent and Tjaden, 1997; Hillis et al., 2004). The present results reconcile these findings: while both hemispheres contribute to speech production, feedforward control is predominantly subserved by the left hemisphere whereas auditory feedback control is subserved by right hemisphere frontal regions. Inferior frontal lesions in the left hemisphere are therefore more likely to disrupt stored feedforward speech motor commands, resulting in disordered speech since the feedforward control system is more crucial for fluent speech than the auditory feedback control system (cf. Neilson and Neilson, 1987). These conclusions are supported by pitch perturbation studies which have shown greater activation in inferior frontal cortex in the Fu et al. (2006) and right posterior temporal cortex (McGuire et al., 1996). In another study of real-time pitch shifts during speech production, Toyomura et al. (2007) greater responses to the shifted feedback in inferior frontal, posterior temporal and inferior parietal cortex. The authors concluded that the feedback control of pitch is primarily influenced by the right hemisphere. The current findings suggest that auditory feedback control of speech, in general, involves a greater contribution from the right hemisphere than does feedforward control.

These results may be particularly relevant to the study and treatment of stuttering. Neuroimaging studies of speech production in persons who stutter consistently demonstrate increased right hemisphere activation in the precentral and inferior frontal gyrus regions identified in the *shift – no shift* contrast of the current study relative to normal speakers (see Brown et al., 2005 for review). It has been hypothesized that stuttering involves excessive reliance upon auditory feedback control due to poor feedforward commands (Max et al., 2004). The current findings provide support for this view: auditory feedback control during the perturbed feedback condition, clearly demonstrated by the behavioral results, was associated with increased activation of right precentral and inferior frontal cortex. According to this view, the right hemisphere inferior frontal activation is a secondary consequence of the root problem, which is aberrant performance in the feedforward system. Poor feedforward performance leads to auditory errors that in turn activate the right-lateralized auditory feedback control system in an attempt to correct for the errors This hypothesis is

consistent with the affects of fluency-inducing therapy on BOLD responses; successful treatment has been associated with a shift toward more normal, left-lateralized frontal activation (De Nil et al., 2003; Neumann et al., 2005).

Conclusions

Collectively, the behavioral and imaging results presented here provide important advancements to our understanding of the role of sensory feedback in on-line control of vocalization and the network of brain regions that support this control. Behavioral data demonstrated clear evidence of feedback-based correction of segmental vocal output. Imaging data indicated that, in the absence of feedback error, articulator control was left-lateralized in the frontal cortex. When auditory error was introduced via the F1 shift, right hemisphere frontal regions, particularly ventral precentral and inferior frontal cortex, were recruited to participate in corrective articulator movements. Increased right frontal activation is also consistently associated with the speech production of persons who stutter; thus the current findings support the theory that stuttered speech results from an over-reliance on auditory feedback-based control.

Several key aspects of the DIVA model of speech production were supported by this investigation: (i) the brain contains auditory error cells that signal differences between a speaker's auditory target and the incoming auditory signal during speech; (ii) these error cells are located in the posterior superior temporal gyrus; and (iii) unexpected perturbation of a speaker's auditory feedback results in a compensatory articulatory response within approximately 99–143 ms of the perturbation onset. The experimental results also indicate that the model should be modified to include higher-order motor corrective representations in addition to primary motor cortical cells for correcting auditory errors, and to utilize right-lateralized, rather than bilateral, motor mechanisms for auditory error correction.

Acknowledgments

References

Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex. Trends Cogn Sci. 2004; 8:170–177. [PubMed: 15050513]

Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. J Acoust Soc Am. 2006; 119:2363–2371. [PubMed: 16642849]

Blakemore SJ, Wolpert D, Frith C. Why can't you tickle yourself? Neuroreport. 2000; 11(11):R11–6. [PubMed: 10943682]

Blakemore SJ, Wolpert DM, Frith CD. Central cancellation of self-produced tickle sensation. Nat Neurosci. 1998; 1:635–40. [PubMed: 10196573]

Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. Neuroimage. 2006; 32:821–841. [PubMed: 16730195]

Brown S, Ingham RJ, Ingham JC, Laird AR, Fox PT. Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. Hum Brain Mapp. 2005; 25:105–117. [PubMed: 15846815]

Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF. Reading, hearing, and the planum temporale. Neuroimage. 2005; 24:444–454. [PubMed: 15627586]

Burnett TA, Larson CR. Early pitch-shift response is active in both steady and dynamic voice pitch control. J Acoust Soc Am. 2002; 112(3 Pt 1):1058–1063. [PubMed: 12243154]

Caviness VS, Meyer J, Makris N, Kennedy D. MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. J Cogn Neurosci. 1996; 8:566–587.

Christoffels IK, Formisano E, Schiller NO. Neural correlates of verbal feedback processing: An fMRI study employing overt speech. Hum Brain Mapp. 2007; 28:868–879. [PubMed: 17266104]

Clopper CG, Pisoni DB, de Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. J Acoust Soc Am. 2005; 118(3 Pt 1):1661–1676. [PubMed: 16240825]

Collignon, A.; Maes, F.; Delaere, D.; Vandermeulen, D.; Suetens, P.; Marchal, G. Automated multi-modality image registration based on information theory. In: Bizais, Y.; Barillot, C.; Di Paola, R., editors. Proc Information Processing in Medical Imaging. Kluwer Academic Publishers; Dordrecht, The Netherlands: 1995. p. 263-274.

Cowie, RI.; Douglas-Cowie, E. Speech production in profound post-lingual deafness. In: Lutman, ME.; Haggard, MP., editors. Hearing Science and Hearing Disorders. Academic Press; New York: 1983. p. 183-231.

Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. Hum Brain Mapp. 2000; 9:183–91. [PubMed: 10770228]

De Nil LF, Kroll RM, Lafaille SJ, Houle S. A positron emission tomography study of short- and long-term treatment effects on functional brain activation in adults who stutter. J Fluency Disord. 2003; 28:357–79. quiz 379–80. [PubMed: 14643070]

Della-Maggiore V, Sekuler AB, Grady CL, Bennett PJ, Sekuler R, McIntosh AR. Corticolimbic interactions associated with performance on a short-term memory task are modified by age. J Neurosci. 2000; 20(22):8410–8416. [PubMed: 11069948]

Diedrichsen J, Hashambhoy Y, Rane T, Shadmehr R. Neural correlates of reach errors. J Neurosci. 2005; 25(43):9919–9931. [PubMed: 16251440]

Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. J Acoust Soc Am. 2002; 111(1 Pt 1):357–366. [PubMed: 11831808]

Dronkers NF. A new brain region for coordinating speech articulation. Nature. 1996; 384(6605):159–161. [PubMed: 8906789]

Duffy, JR. Motor speech disorders: Substrates, differential diagnosis, and managment. Mosby; St. Louis, MO: 1995.

Eliades SJ, Wang X. Dynamics of auditory-vocal interaction in monkey auditory cortex. Cereb Cortex. 2005; 15(10):1510–1523. [PubMed: 15689521]

Engelien A, Yang Y, Engelien W, Zonana J, Stern E, Silbersweig DA. Physiological mapping of human auditory cortices with a silent event-related fMRI technique. Neuroimage. 2002; 16:944–953. [PubMed: 12202082]

Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305 MRI Volumes. Proceedings of the IEEE Nuclear Science Symposium on Medical Imaging. 1993; 3:1813–1817.

Ferguson SH, Kewley-Port D. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2002; 112:259–271. [PubMed: 12141351]

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33:341–355. [PubMed: 11832223]

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. Automatically parcellating the human cerebral cortex. Cereb Cortex. 2004; 14:11–22. [PubMed: 14654453]

Friston KJ, Frith CD, Frackowiak RS, Turner R. Characterizing dynamic brain responses with fMRI: a multivariate approach. Neuroimage. 1995a; 2:166–172. [PubMed: 9343599]

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R. Analysis of fMRI time-series revisited. Neuroimage. 1995b; 2:45–53. [PubMed: 9343589]

Fu CH, Vythelingum GN, Brammer MJ, Williams SC, Amaro E Jr, Andrew CM, Yaguez L, van Haren NE, Matsumoto K, McGuire PK. An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. Cereb Cortex. 2006; 16:969–977. [PubMed: 16195470]

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 2006; 96:280–301. [PubMed: 16040108]

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychol Rev. 1998; 105:611–633. [PubMed: 9830375]

Habas C, Axelrad H, Nguyen TH, Cabanis EA. Specific neocerebellar activation during out-of-phase bimanual movements. Neuroreport. 2004; 15:595–599. [PubMed: 15094459]

Habas C, Cabanis EA. Cortical areas functionally linked with the cerebellar second homunculus during out-of-phase bimanual movements. Neuroradiology. 2006; 48:273–279. [PubMed: 16465531]

Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. Exp Brain Res. 2000; 130:133–141. [PubMed: 10672466]

Hashimoto Y, Sakai KL. Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. Hum Brain Mapp. 2003; 20:22–28. [PubMed: 12953303]

Heinks-Maldonado TH, Nagarajan SS, Houde JF. Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport. 2006; 17(13):1375–1379. [PubMed: 16932142]

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci. 2003; 15:673–682. [PubMed: 12965041]

Hillenbrand JM, Clark MJ, Nearey TM. Effects of consonant environment on vowel formant patterns. J Acoust Soc Am. 2001; 109:748–763. [PubMed: 11248979]

Hillis AE, Work M, Barker PB, Jacobs MA, Breese EL, Maurer K. Re-examining the brain regions crucial for orchestrating speech articulation. Brain. 2004; 127(Pt 7):1479–87. [PubMed: 15090478]

Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H, Konishi J. Cortical processing mechanism for vocalization with auditory verbal feedback. Neuroreport. 1997; 8(9–10):2379–82. [PubMed: 9243644]

Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science. 1998; 279(5354):1213–1216. [PubMed: 9469813]

Houde JF, Jordan MI. Sensorimotor adaptation of speech I: Compensation and adaptation. J Speech Lang Hear Res. 2002; 45:295–310. [PubMed: 12003512]

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. J Cogn Neurosci. 2002; 14:1125–1138. [PubMed: 12495520]

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999; 6:1–55.

Huang VS, Shadmehr R. Evolution of motor memory during the seconds after observation of motor error. J Neurophysiol. 2007

Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. Cognition. 2004; 92(1–2):101–144. [PubMed: 15037128]

Ito T, Kimura T, Gomi H. The motor cortex is involved in reflexive compensatory adjustment of speech articulation. Neuroreport. 2005; 16(16):1791–1794. [PubMed: 16237328]

Jones JA, Munhall KG. The role of auditory feedback during phonation: studies of Mandarin tone production. J Phonetics. 2002; 30:303–320.

Jones JA, Munhall KG. Remapping auditory-motor representations in voice production. Curr Biol. 2005; 15(19):1768–1772. [PubMed: 16213825]

Kent, RD.; Tjaden, K. Brain functions underlying speech. In: Hardcastle, WJ.; Laver, J., editors. Handbook of phonetic sciences. Blackwell; Oxford: 1997. p. 220-255.

Kriegstein KV, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage. 2004; 22:948–955. [PubMed: 15193626]

Lancaster JL, Tordesillas-Gutierrez D, Martinez M, Salinas F, Evans A, Zilles K, Mazziotta JC, Fox PT. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum Brain Mapp. 2007

Lane H, Tranel B. The Lombard sign and the role of hearing in speech. J Speech Lang Hear Res. 1971; 14:677–709.

Lane H, Webster JW. Speech deterioration in postlingually deafened adults. J Acoust Soc Am. 1991; 89:859–866. [PubMed: 2016436]

Larson CR, Burnett TA, Kiran S, Hain TC. Effects of pitch-shift velocity on voice F0 responses. J Acoust Soc Am. 2000; 107:559–564. [PubMed: 10641664]

Le TH, Patel S, Roberts TP. Functional MRI of human auditory cortex using block and event-related designs. Magn Reson Med. 2001; 45:254–260. [PubMed: 11180433]

Lombard E. Le signe de l'elevation de la voix. Annales des Maladies de l'Oreille du Larynx. 1911; 37:101–119.

Matsumoto R, Nair DR, LaPresto E, Najm I, Bingaman W, Shibasaki H, Luders HO. Functional connectivity in the human language system: a cortico-cortical evoked potential study. Brain. 2004; 127(Pt 10):2316–2330. [PubMed: 15269116]

Max L, Guenther FH, Gracco VL, Ghosh SS, Wallace ME. Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. Contemp Issues Comm Sci Disord. 2004; 31:105–122.

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Feidler J, Smith K, Boomsma D, Hulshoff Pol H, Cannon T, Kawashima R, Mazoyer B. A four-dimensional probabilistic atlas of the human brain. J Am Med Inform Assoc. 2001; 8:401–30. [PubMed: 11522763]

McClean MD, Sapir S. Some effects of auditory stimulation on perioral motor unit discharge and their implications for speech production. J Acoust Soc Am. 1981; 69:1452–1457. [PubMed: 7240578]

McGuire PK, Silbersweig DA, Frith CD. Functional neuroanatomy of verbal self-monitoring. Brain. 1996; 119 (Pt 3):907–17. [PubMed: 8673501]

McIntosh AR, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional imaging. Hum Brain Mapp. 1994; 2:2–22.

Natke U, Donath TM, Kalveram KT. Control of voice fundamental frequency in speaking versus singing. J Acoust Soc Am. 2003; 113:1587–1593. [PubMed: 12656393]

Neilson M, Neilson P. Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. Speech Communication. 1987; 6:325–333.

Neumann K, Preibisch C, Euler HA, von Gudenberg AW, Lanfermann H, Gall V, Giraud AL. Cortical plasticity associated with stuttering therapy. J Fluency Disord. 2005; 30:23–39. [PubMed: 15769497]

Nieto-Castanon A, Ghosh SS, Tourville JA, Guenther FH. Region of interest based analysis of functional imaging data. Neuroimage. 2003; 19:1303–1316. [PubMed: 12948689]

Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neurosci Lett. 1999; 272:29–32. [PubMed: 10507535]

Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. Neurosci Lett. 1999; 265:119–22. [PubMed: 10327183]

Ozdemir E, Norton A, Schlaug G. Shared and distinct neural correlates of singing and speaking. Neuroimage. 2006; 33:628–35. [PubMed: 16956772]

Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. Eur J Neurosci. 1996; 8(11):2236–46. [PubMed: 8950088]

Pittman AL, Wiley TL. Recognition of speech produced in noise. J Speech Lang Hear Res. 2001; 44:487–496. [PubMed: 11407555]

Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, Britton B. An event-related fMRI investigation of phonological-lexical competition. Neuropsychologia. 2006; 44(12):2209–2221. [PubMed: 16842827]

Purcell DW, Munhall KG. Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. J Acoust Soc Am. 2006a; 120:966–977. [PubMed: 16938984]

Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am. 2006b; 119:2288–2297. [PubMed: 16642842]

Redding GM, Wallace B. Generalization of prism adaptation. J Exp Psychol Hum Percept Perform. 2006; 32:1006–1022. [PubMed: 16846294]

Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. Neuroreport. 2000; 11:1997–2000. [PubMed: 10884059]

Schumacker, R.; Lomax, R. A beginner's guide to structural equation modeling. 2. Lawrence Elrbaum; Mahwah, N.J., USA: 2004.

Sidtis JJ, Gomez C, Groshong A, Strother SC, Rottenberg DA. Mapping cerebral blood flow during speech production in hereditary ataxia. Neuroimage. 2006; 31:246–54. [PubMed: 16443374]

Soros P, Sokoloff LG, Bose A, McIntosh AR, Graham SJ, Stuss DT. Clustered functional MRI of overt speech production. Neuroimage. 2006; 32:376–87. [PubMed: 16631384]

Stevens AA, Skudlarski P, Gatenby JC, Gore JC. Event-related fMRI of auditory and visual oddball tasks. Magn Reson Imaging. 2000; 18:495–502. [PubMed: 10913710]

Stuart A, Kalinowski J, Rastatter MP, Lynch K. Effect of delayed auditory feedback on normal speakers at two speech rates. J Acoust Soc Am. 2002; 111(5 Pt 1):2237–2241. [PubMed: 12051443]

Talairach, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain. Thieme Medical Publishers; New York, NY: 1988.

Toyomura A, Koyama S, Miyamaoto T, Terao A, Omori T, Murohashi H, Kuriki S. Neural correlates of auditory feedback control in human. Neuroscience. 2007; 146:499–503. [PubMed: 17395381]

Tsao YC, Weismer G. Interspeaker variation in habitual speaking rate: evidence for a neuromuscular component. J Speech Lang Hear Res. 1997; 40:858–866. [PubMed: 9263949]

Villacorta, V. Speech and Hearing Bioscience and Technology. Massachusetts Institute of Technology; Cambridge, MA: 2006. Sensorimotor Adaptation to Perturbations of Vowel Acoustics and its Relation to Perception.

Villacorta VM, Perkell JS, Guenther FH. Sensorimotor adaptation to feedback perturbations on vowel acoustics and its relation to perception. J Acoust Soc Am. In Press.

Wise RJ, Greene J, Buchel C, Scott SK. Brain regions involved in articulation. Lancet. 1999; 353(9158):1057–61. [PubMed: 10199354]

Wolpert DM, Flanagan JR. Motor prediction. Curr Biol. 2001; 11(18):R729–32. [PubMed: 11566114]

Xu Y, Larson CR, Bauer JJ, Hain TC. Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. J Acoust Soc Am. 2004; 116:1168–1178. [PubMed: 15376682]

Yang Y, Engelien A, Engelien W, Xu S, Stern E, Silbersweig DA. A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. Magn Reson Med. 2000; 43:185–190. [PubMed: 10680681]

Yates AJ. Delayed auditory feedback. Psychol Bull. 1963; 60:213–232. [PubMed: 14002534]

Zarate JM, Zatorre RJ. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann N Y Acad Sci. 2005; 1060:404–8. [PubMed: 16597793]
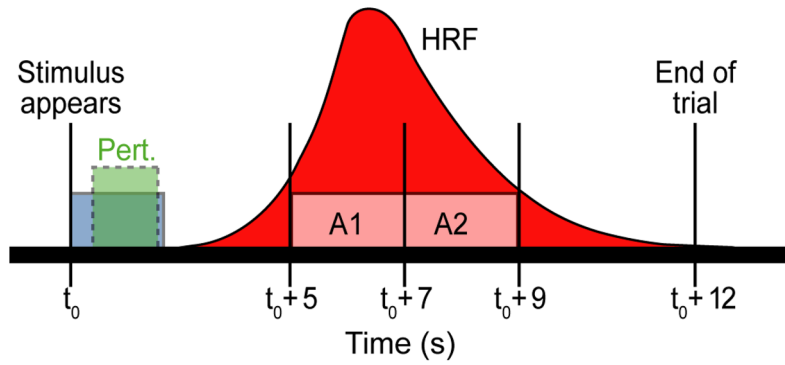
Fig. 1.
Timeline of a single trial in the event-triggered sparse sampling protocol. At the onset of each trial, the visual stimulus appeared and remained onscreen for 2 seconds (blue rectangle). On perturbed trials, auditory feedback was shifted during the subject's response (green). 3 seconds after stimulus offset, two whole-brain volumes were acquired (A1, A2). Data acquisition was timed to cover the peak of the hemodynamic response to speech; the putative hemodynamic response function (HRF), is schematized in red. The next trial started 3 seconds after data acquisition was complete, resulting in a total trial length of 12 seconds.
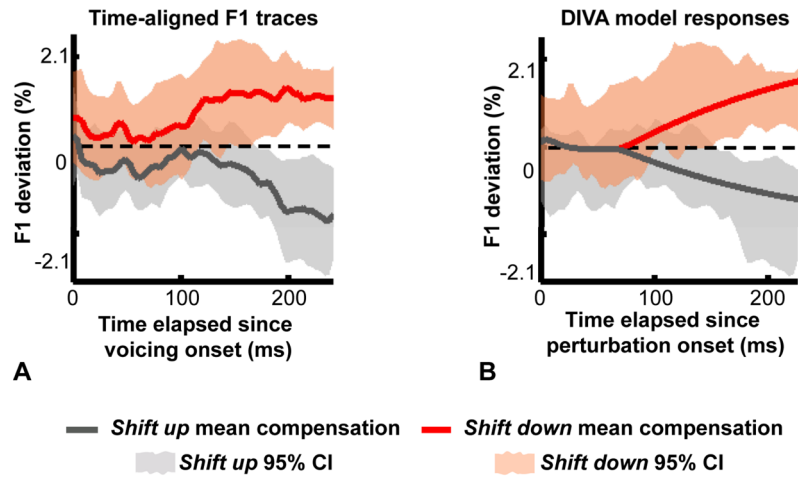
Fig. 2.

First formant response to an induced shift in F1. (A) Mean F1 compensation plotted as percent deviation from the mean F1 value in unperturbed speech. Solid lines indicate compensation in the *shift up* (gray) and *shift down* (red) conditions averaged across subjects. Dashed lines represent the matched *no shift* formant value. Shaded regions indicate 95% confidence intervals at each time point. Shifting F1 by 30% resulted in a compensatory F1 response in the direction opposite the shift *within an utterance*. (B) Comparison of experimental and simulated responses to unexpected F1 feedback perturbation during DIVA model production of the word /bɛd/. F1 traces produced by the model (lines) and 95% confidence intervals from the group experimental data (shaded regions) are aligned to the onset of perturbation.
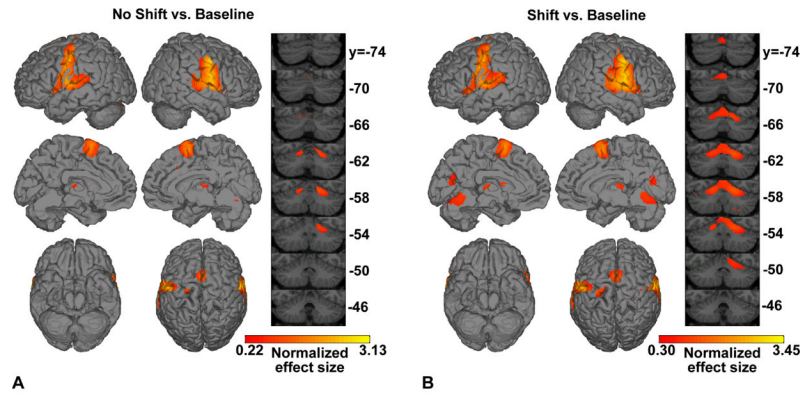
Fig. 3.
BOLD responses for unshifted speech (left) and shifted speech (right) compared to the silent baseline task. (A) Map of statistically significant normalized effect sizes ($t > 3.63$; $df = 9$; FDR < 5%) from the comparison of BOLD responses in the unshifted speech condition to the baseline condition (*No-shift – Baseline*). Coronal slices through the cerebellum are shown to the right of renderings of the lateral (top), medial (middle) and ventral (bottom) cortical surfaces of the hemispheres. Activation is found in the expected speech production network including bilateral peri-Sylvian, lateral Rolandic, and medial prefrontal cortex, superior cerebellum, ventral thalamus and anterior striatum. (B) BOLD responses during shifted speech compared to baseline (*Shift – Baseline; t > 3.41*; *df* = 9; FDR < 5%). The network of active regions during shifted feedback speech is qualitatively similar to that of normal speech, with additional activation in the superior cerebellar cortex bilaterally, in the right inferior cerebellar cortex, and in the medial parietal-occipital cortex bilaterally.
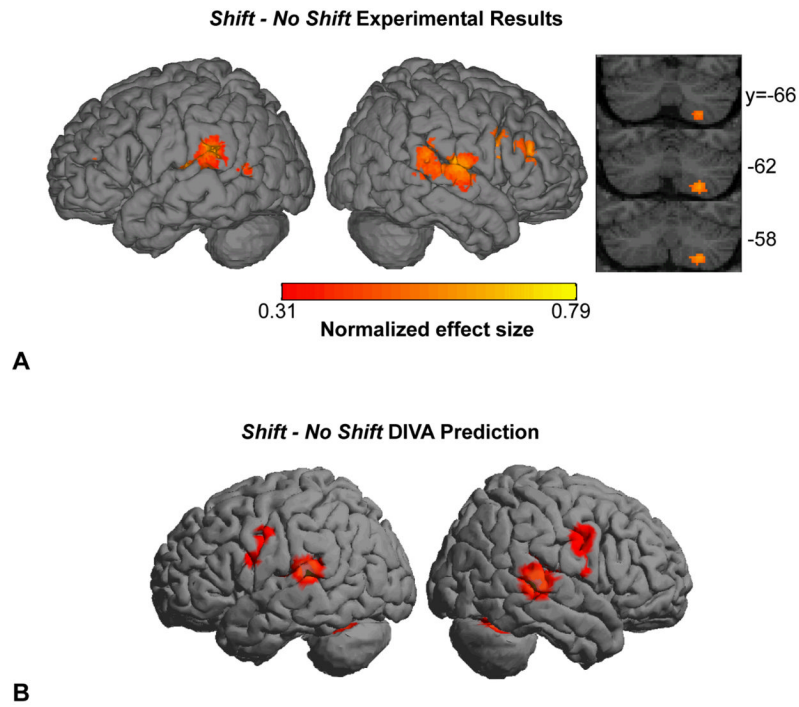
Fig. 4.
BOLD responses in the s*hift – no shift* contrast. (A) Map of statistically significant normalized effect sizes (voxel threshold: $t > 3.19$; $df = 5488$; FDR < 5%) derived from group voxel-based fixed effects analysis. Activation of the posterior peri-Sylvian region is seen bilaterally. In the right hemisphere, greater responses were found in ventral somatosensory, motor and premotor cortex, and the inferior cerebellum. (B) Simulated BOLD responses for the *shift – no shift* contrast in the DIVA model. Model cell responses from *no shift* and *shift* simulations were contrasted and normalized in the same manner applied to the experimental results and smoothed with a spherical Guassian point spread function (FWHM = 12 mm). The resulting activations were then plotted on a cortical surface rendering based on hypothesized anatomical locations (Guenther et al., 2006). The model predicts increased activation in bilateral posterior peri-Sylvian cortex (auditory error cells), ventral Rolandic cortex (motor cells and somatosensory error cells), and superior cerebellum.
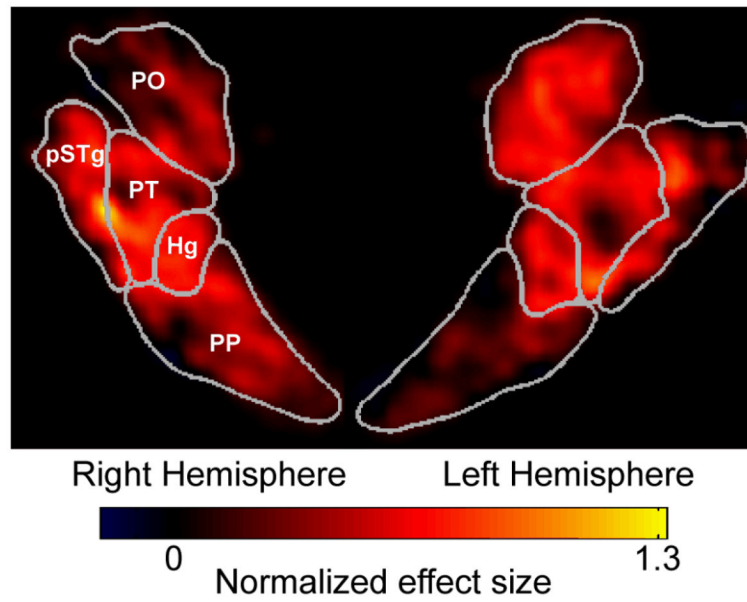
Fig. 5.
Spatial profiles of peri-Sylvian ROI responses for the *shift – no shift* contrast. Activation peaks are seen along the PT/pSTg boundary in both hemispheres. Greater response is noted in PO in the left hemisphere and Hg in the right hemisphere. Abbreviations: Hg = Heschl's gyrus; PO = parietal operculum; PP = planum polare; pSTg = posterior superior temporal gyrus; PT = planum temporale.
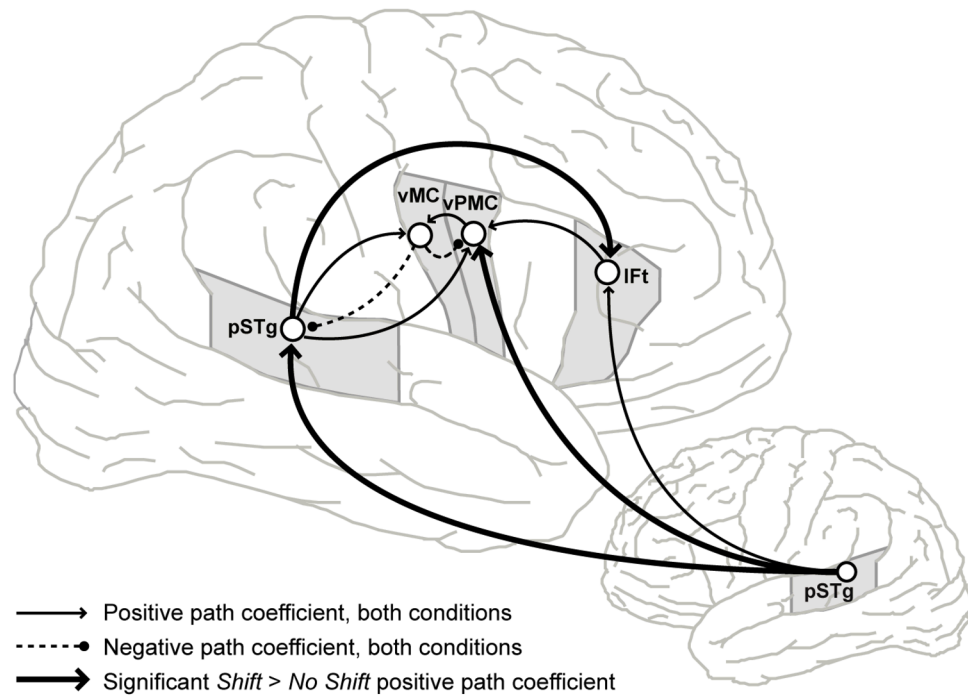
Fig. 6.
Schematic of the path diagram evaluated by structural equation modeling. Effective connectivity of the network was significantly modulated by the feedback perturbation. Path coefficients for all connections shown were significant in both conditions except Right vMC to Right vPMC (*No Shift p* = 0.07; see Table 2 for a list of all estimated path coefficients). Pairwise comparisons of path coefficients in the two conditions indicated significant increases in the positive weights from left pSTg to right vPMC and right pSTg and from right pSTg to right IFt. Abbreviations: IFt = inferior frontal gyrus, pars triangularis; pSTg = posterior superior temporal gyrus; vMC = ventral motor cortex; vPMC = ventral premotor cortex.

Table 1

Peak voxel responses for the three contrasts of interest listed by anatomical region. The location of each response in both MNI and Talairach stereotaxic space is given along with the contrast *t*-statistic and normalized effect.

Region	Region Label	No Shift - Baseline — MNI	Talairach	T	Norm. Effect	Shift – Baseline — MNI	Talairach	T	Norm. Effect	Shift - No Shift — MNI	Talairach	T	Norm. Effect	ROI T	ROI Norm. Effect
Rolandic Cortex															
Left	dMC	(−48,−8,60)	(−46,−14,56)	6.92	1.60	(−44,−8,64)	(−43,−15,60)	8.08	1.08						
	vMC													*n.s*	*n.s*
	vPMC	(−48,0,30)	(−46,−4,30)	7.06	1.05	(−60,4,8)	(−57,2,11)	8.89	1.42					*n.s*	*n.s*
	vSC	(−56,−10,42)	(−53,−14,40)	15.44	2.45	(−52,−14,38)	(−50,−18,36)	12.79	2.81					*n.s.*	*n.s.*
	dSC									(−22,−42,70)	(−22,−47,63)	3.38	0.63		
	pCO	(−62,−6,12)	(−59,−8,13)	12.09	2.42	(−60,−8,12)	(−57,−10,13)	11.23	2.69						
Right	**vMC**									(48,−10,44)	(43,−15,43)	3.25	0.54	**2.10**	**0.28**
	dMC					(44,−18,70)	(39,−25,66)	4.94	0.43						
	vPMC									(60,14,34)	(54,8,37)	3.52	0.53	**2.52**	**0.36**
	aCO	(48,8,4)	(43,5,9)	10.63	0.96										
	vSC	(54,−8,30)	(49,−12,31)	22.67	2.15	(56,−10,30)	(50,−14,31)	16.26	2.40	(70,−2,20)	(64,−6,23)	3.20	0.40	1.87	0.24
Frontal Cortex															
Left	IFo					(−60,8,2)	(−57,6,6)	8.36	1.38						
	FO	(−48,10,−2)	(−45,8,2)	5.98	1.45	(−48,10,−2)	(−45,8,2)	8.04	1.85						
	preSMA	(−2,8,62)	(−4,0,60)	10.91	1.37	(0,4,72)	(−2,−4,69)	7.61	1.70						
Right	IFo									(58,14,28)	(52,9,31)	3.64	0.62	2.13	0.32
	IFt									(56,32,24)	(51,26,29)	4.07	0.57	2.07	0.30
	FO														
	aCg	(4,20,36)	(2,14,38)	3.89	0.68	(48,12,2)	(43,9,8)	7.45	1.38						
Parietal Cortex															
Left	PO	(−44,−34,24)	(−42,−35,22)	8.85	1.01	(−40,−30,18)	(−38,−31,17)	6.98	1.99	(−54,−24,14)	(−51,−25,14)	3.98	0.67	4.47	0.63
Right	PO	(46,−26,18)	(41,−28,19)	5.42	1.26	(44,−24,18)	(39,−26,19)	5.29	1.57						
	aSMg					(72,−24,32)	(65,−27,32)	4.49	0.58						
Temporal Cortex															
Left	Hg	(−58,−10,10)	(−55,−12,11)	11.79	2.27										

| | No Shift - Baseline | | | | Shift – Baseline | | | | Shift - No Shift | | | | | |
| | Peak Voxel Location (x,y,z) | | | | Peak Voxel Stereotaxic Location | | | | Peak Voxel Stereotaxic Location | | | | | |
Region Label	MNI	Talairach	T	Norm. Effect	MNI	Talairach	T	Norm. Effect	MNI	Talairach	T	Norm. Effect	ROI T	ROI Norm. Effect
pSTg	(−64,−30,14)	(−60,−31,13)	9.70	1.31	(−62,−30,14)	(−59,−31,13)	7.93	2.00	(−66,−38,22)	(−62,−39,19)	5.25	0.59	**3.78**	**0.48**
pdSTs					(−62,−22,4)	(−58,−22,5)	4.95	1.85	(−60,−30,10)	(−57,−30,9)	3.96	0.60	2.08	0.28
PT	(−52,−34,16)	(−49,−35,15)	5.79	1.51					(−62,−24,10)	(−59,−25,10)	3.88	0.57	**3.83**	**0.65**
MTO									(−60,−62,10)	(−57,−60,7)	3.64	0.44	2.88	0.29
Right pSTg	(72,−24,6)	(65,−25,8)	7.40	1.03	(68,−16,8)	(62,−18,11)	8.66	2.31	(68,−36,18)	(62,−37,18)	4.13	0.56	**4.64**	**0.63**
pdSTs					(58,−28,6)	(53,−29,8)	5.17	1.50	(72,−40,12)	(65,−40,12)	3.94	0.44	3.15	0.31
PT	(58,−28,12)	(52,−29,13)	5.06	1.30	(64,−10,10)	(58,−12,13)	9.33	2.57	(68,−16,8)	(62,−18,11)	5.01	0.68	**4.32**	**0.49**
PP									(48,−8,−8)	(43,−9,−3)	3.59	0.63	4.07	0.45
adSTs									(56,−10,−4)	(51,−11,0)	3.31	0.55	2.18	0.39
Insular Cortex														
Left aINS	(−44,6,2)	(−42,4,6)	5.73	1.17										
aINS					(−34,−4,10)	(−33,−6,12)	4.49	0.85						
pINS					(−34,−20,8)	(−33,−21,9)	5.16	0.99						
Cerebellum														
Left **amCB, V5**													*n.s*	*n.s.*
spmCB, L6	(−16,−62,−20)	(−16,−58,−19)	5.10	0.76	(−18,−62,−20)	(−18,−58,−20)	6.72	1.08						
spmCB, V6					(−2,−78,−14)	(−3,−73,−15)	5.45	0.62						
splCB, Cr1	(−30,−86,−24)	(−29,−80,−25)	4.17	0.40										
Right **amCB, V5**													**2.25**	**0.34**
splCB, L6	(26,−58,−24)	(23,−54,−22)	6.30	0.95	(24,−56,−22)	(21,−52,−20)	9.46	1.35						
ipmCB, L8A					(28,−58,−54)	(25,−51,−49)	3.42	0.86	(26,−62,−54)	(24,−55,−49)	4.06	0.62	2.14	0.24
Subcortical Nuclei														
Left Put	(−26,−4,4)	(−25,−6,7)	7.97	0.62	(−30,−6,4)	(−29,−8,7)	4.34	0.71						
Pal	(−24,−6,−4)	(−23,−7,−0)	6.74	0.58	(−22,−2,−2)	(−21,−3,2)	5.15	0.95						
Caud					(−10,2,10)	(−10,−1,13)	3.74	0.92						
Tha, VL	(−8,−14,6)	(−9,−15,8)	4.64	0.64	(−8,−14,6)	(−9,−15,8)	3.82	0.88						
Right Put	(30,8,4)	(27,5,9)	6.13	0.35	(22,2,−2)	(19,0,3)	4.27	0.61						
Pal	(26,−2,−4)	(23,−3,1)	5.32	0.62	(14,4,10)	(12,1,14)	4.11	0.84						
Cau														

Region Label	No Shift - Baseline Peak Voxel Location (x,y,z)				Shift – Baseline Peak Voxel Stereotaxic Location				Shift - No Shift Peak Voxel Stereotaxic Location					
	MNI	Talairach	T	Norm. Effect	MNI	Talairach	T	Norm. Effect	MNI	Talairach	T	Norm. Effect	ROI T	ROI Norm. Effect
Tha, VL	(14,−12,10)	(12,−14,12)	3.98	0.62										
Tha, MD	(6,−18,8)	(4,−19,10)	4.19	0.76	(4,−22,6)	(3,−23,8)	4.50	0.98						
Occipital Cortex														
Left OC					(−2,−70,16)	(−3,−68,12)	4.94	0.81						
Right OC									(12,−84,40)	(9,−84,33)	3.39	0.57	**	**

Regional *t*-statistics and normalized effects for ROIs containing peak voxel responses are given for the *shift – no shift* contrast. Regions highlighted in boldface type were included in our initial ROI analysis of *a priori* hypotheses that each of these regions would be more active during perturbed feedback trials. The effect sizes and *t*-statistics from tests of these hypotheses are also shown in bold. When possible, a more specific label is listed in addition to the ROI label for voxels that lie within well-characterized subsets of an anatomical region (e.g., cerebellar lobule 6).

n.s. = not significant;

**
Occipital cortex was not included in the ROI analysis.

Abbreviations: amCB = anterior medial cerebellum; aCg = anterior cingulate gyrus; aMCB = anterior medial cerebellum; aMFg = anterior middle frontal gyrus; aMTg = anterior middle temporal gyrus; aSMg = anterior supramarginal gyrus; Caud = Caudate; Cr1 = cerebellar crus I; dMC = dorsal primary motor cortex; dSC = dorsal somatosensory cortex; FO = frontal operculum; Hg = Heschl's gyrus; IFo = inferior frontal gyrus, pars opercularis; IFt = inferior frontal gyrus, pars triangularis; ipmCB = inferior posterior medial cerebellum; L5 = cerebellum lobule V; L6 = cerebellum lobule VI; L8A = cerebellum lobule VIIIA; Lg = lingual gyrus; MD = mediodorsal thalamic nucleus; MTO = middle temporal occipital gyrus; OC = occipital cortex; Pal = pallidum; pCO = posterior central operculum; pdPMC = posterior dorsal premotor cortex; pdSTs = posterior dorsal superior temporal sulcus; pINS = posterior insula; PO=parietal operculum; PP = planum polare; preSMA = pre-supplementary motor area; pSMg = posterior supramarginal gyrus; pSTg = posterior superior temporal gyrus; PT = planum temporale; Put = putamen; spmCB = superior posterior medial cerebellum; splCB = superior posterior lateral cerebellum; Tha = thalamus; VL = ventrolateral thalamic nucleus; vMC = ventral primary motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex.

Table 2

Effective connectivity determined by structural equation modeling of the network shown in Figure 6 for the *no shift* and *shift* conditions. Significant increases in effective connectivity due to the F1 shift were found in connections from left pSTg to right vPMC, left pSTg to right pSTg and from right pSTg to right IFt.

Network Path	*No Shift*				*Shift*				*Shift - No Shift*	
	Path Coeff.	Standard Error	Critical Ratio	*p*	Path Coeff.	Standard Error	Critical Ratio	*p*	Critical Ratio	*p*
Left pSTg → Right vPMC	0.06	0.03	2.22	0.03	0.22	0.05	4.01	<.001	**2.64**	**0.01**
Left pSTg → Right IFt	0.25	0.02	10.61	<.001	0.26	0.04	5.84	<.001	0.12	0.90
Left pSTg → Right pSTg	0.49	0.03	15.68	<.001	0.61	0.06	10.94	<.001	**2.11**	**0.03**
Right IFt → Right vPMC	0.35	0.03	12.29	<.001	0.34	0.05	7.10	<.001	−0.03	0.98
Right pSTg → Right IFt	0.27	0.03	10.58	<.001	0.37	0.05	8.13	<.001	**2.08**	**0.04**
Right vMC → Right pSTg	−0.25	0.07	−3.53	<.001	−0.34	0.09	−3.95	<.001	−1.33	0.18
Right pSTg → Right vMC	0.36	0.06	5.96	<.001	0.40	0.07	5.33	<.001	0.58	0.56
Right vMC → Right vPMC	−0.15	0.08	−1.83	0.07	−0.26	0.09	−2.77	0.01	−1.56	0.12
Right vPMC → Right vMC	0.48	0.07	7.30	<.001	0.54	0.08	7.09	<.001	0.92	0.36
Right pSTg → Right vPMC	0.32	0.04	7.41	<.001	0.33	0.06	5.23	<.001	0.02	0.98

Abbreviations: IFt = inferior frontal gyrus, pars triangularis; pSTg = posterior superior temporal gyrus; vMC = ventral motor cortex; vPMC = ventral premotor cortex.