Behavioral/Cognitive

Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations

Caroline A. Niziolek[1] **and Frank H. Guenther**[2,3,4,5]

[1]Department of Otolaryngology, University of California, San Francisco, San Francisco, California 94143, [2]Departments of Speech, Language, and Hearing Sciences and [3]Biomedical Engineering, Boston University, Boston, Massachusetts 02215, [4]Athinoula A. Martinos Center for Neuroimaging, Massachusetts General Hospital, Charlestown, Massachusetts 02139, and [5]Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Auditory feedback is instrumental in the online control of speech, allowing speakers to compare their self-produced speech signal with a desired auditory target and correct for errors. However, there is little account of the representation of "target" and "error": does error depend purely on acoustic distance from a target, or is error enhanced by phoneme category changes? Here, we show an effect of vowel boundaries on compensatory responses to a real-time auditory perturbation. While human subjects spoke monosyllabic words, event-triggered functional magnetic resonance imaging was used to characterize neural responses to unexpected changes in auditory feedback. Capitalizing on speakers' natural variability, we contrasted the responses to feedback perturbations applied to two classes of utterances: (1) those that fell nearer to the category boundary, for which perturbations were designed to change the phonemic identity of the heard speech; and (2) those that fell farther from the boundary, for which perturbations resulted in only sub-phonemic auditory differences. Subjects' behavioral compensation was more than three times greater when feedback shifts were applied nearer to a category boundary. Furthermore, a near-boundary shift resulted in stronger cortical responses, most notably in right posterior superior temporal gyrus, than an identical shift that occurred far from the boundary. Across participants, a correlation was found between the amount of compensation to the perturbation and the amount of activity in a network of superior temporal and inferior frontal brain regions. Together, these results demonstrate that auditory feedback control of speech is sensitive to linguistic categories learned through auditory experience.

Introduction

The speech motor system is known to rely on auditory feedback for the moment-to-moment control of speech articulators. For example, speech output is modified dynamically in response to unexpected perturbations in auditory feedback (Jones and Munhall, 2002; Bauer et al., 2006; Cai et al., 2011), with responses starting within the first few hundred milliseconds of perturbation onset. Although these responses can vary across subjects and experimental parameters (Burnett et al., 1998), the most commonly observed response is compensatory, in which the adjustments to speech output serve to counteract the imposed perturbation, raising the first formant if it has been lowered, for example (Tourville et al., 2008).

A purported goal of this online feedback system is to minimize error from an acoustic target. However, the space that defines a speech production target, and thus the magnitude of acoustic error for a given perturbation, is poorly understood. In speech models such as the Directions Into Velocities of Articulators (DIVA) model (Guenther, 1994, 1995; Guenther et al., 2006; Golfinopoulos et al., 2010), speech sounds have corresponding target regions in auditory perceptual space, and auditory feedback is used to update and refine the motor commands that guide the acoustic signal through those regions. However, evidence from the speech perception literature suggests that not all auditory space is created equal: sounds that straddle a category boundary are more discriminable than those from the same phoneme category (Liberman et al., 1957). Discrimination ability is also increased when auditory stimuli lie nearer to category boundaries, a phenomenon called the perceptual magnet effect (Kuhl, 1991; Iverson and Kuhl, 1995; Kuhl et al., 2008). These findings suggest that deviations from one's own speech production targets may also be self-perceived more clearly when they approach or cross a boundary.

The goal of the current study was to illustrate the nature of speech production target space by answering the following key question: does auditory error depend only on low-level auditory distance from an intended target, or is it enhanced by a higher-level category change? To this end, we used functional magnetic resonance imaging (fMRI) to examine the neural and behavioral responses to sudden perturbations in the auditory feedback loop as elicited by unexpected acoustic shifts. We tested the hypothesis that a feedback shift near a boundary region would evoke a greater response than a feedback shift lying safely within the accepted variability for a given speech sound. We contrasted shifts

Received Feb. 25, 2013; revised June 13, 2013; accepted June 14, 2013.

Author contributions: C.A.N. and F.H.G. designed research; C.A.N. performed research; C.A.N. analyzed data; C.A.N. and F.H.G. wrote the paper.

This research was supported in part by National Institutes of Health Grants R01 DC002852 (F.H.G.) and F32 DC011249 (C.A.N.). We thank Kevin Reilly for providing formant analysis code.

The authors declare no competing financial interests.

Correspondence should be addressed to Caroline A. Niziolek, 513 Parnassus Avenue, Box 0628, San Francisco, CA 94143. E-mail: carrien@mit.edu.

DOI:10.1523/JNEUROSCI.1008-13.2013

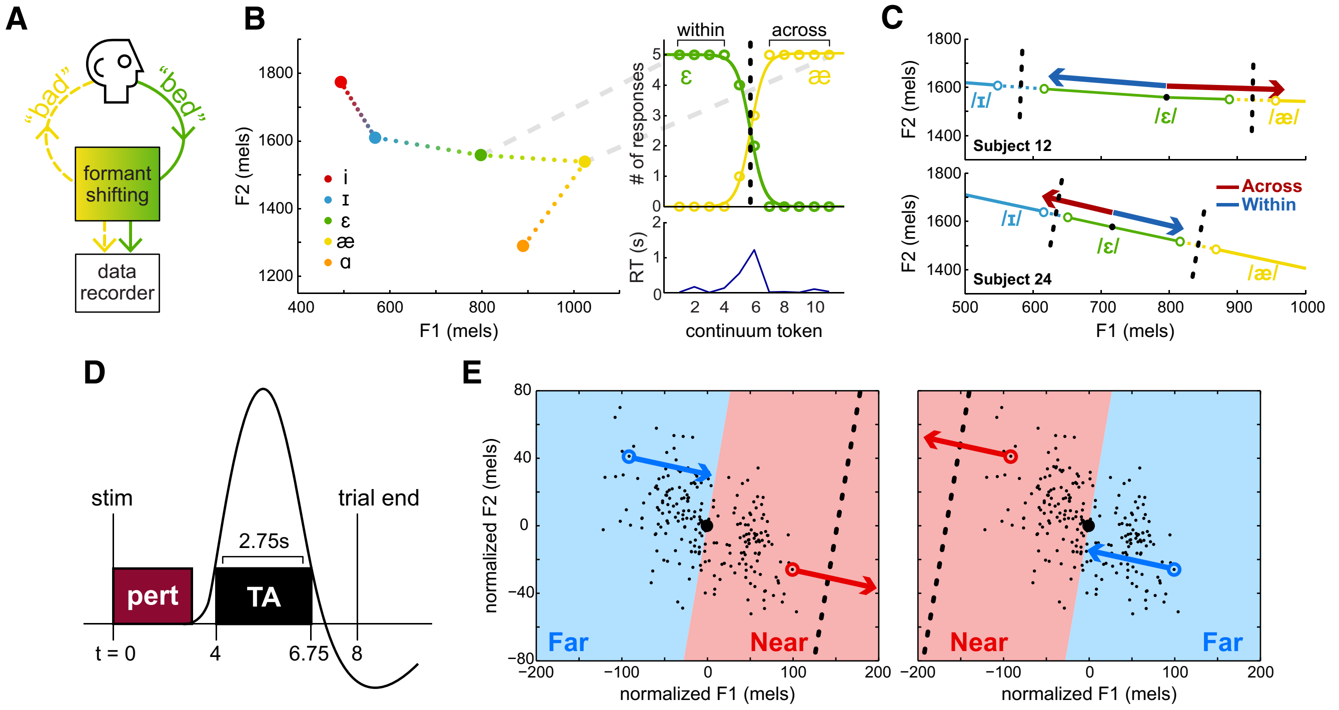

Figure 1. Experimental design. ***A***, Schematic of formant shift and recording apparatus. ***B***, Categorical perception of self-produced vowels from a sample subject. At left, the small dots between the vowels represent the formant-shifted tokens along each vowel continuum. At right, the identification responses for each vowel continuum were fit to sigmoid curves, with the estimated boundary at the intersection of the two curves (dashed line). Reaction time (RT) peaks near the category boundary. ***C***, Example of counterbalanced subjects with opposite shift patterns. The black filled circle is the median production of /ɛ/; dashed lines indicate the measured boundaries. ***D***, Trial timeline in the MRI scanner. TA, Acquisition time; pert, formant perturbation. ***E***, *Post hoc* division of trials into Near and Far conditions based on distance to the boundary (dashed line). Left and right depict the same production distribution perturbed in different directions; that is, the definition of Near and Far regions depends on the direction of formant shift.

of the same acoustic size, one that altered the phoneme of the perceived sound (e.g., /ɛ/ shifted to sound like /æ/) versus one that modified the acoustics only within category limits (e.g., /ɛ/ shifted to sound like a different version of /ɛ/). If auditory target space is modulated by category boundaries, there should be a greater response (neural and behavioral) to the cross-boundary shift; if not, the two shifts should elicit equal-magnitude responses. This perturbation paradigm offers insight into the error correction signal produced by the acoustic mismatch as well as the updated motor commands used to produce the vocal compensation.

Materials and Methods

Participants

Eighteen right-handed subjects between the ages of 19 and 33 years (mean age, 23.5 years), nine men and nine women, participated in the fMRI study. These participants were drawn from a pool of 36 subjects who completed a behavioral pretest (mean age, 23.6 years). All 18 subjects spoke American English as a first language, had no history of hearing or speech disorders, and, to be eligible for imaging, had no metal in the body. All study procedures, including recruitment and acquisition of informed consent, were approved by the institutional review boards of the Massachusetts Institute of Technology (Cambridge, MA) and Boston University (Boston, MA).

Procedure

The experiment consisted of two phases: (1) a behavioral pretest, in which subjects' production and perception spaces were assessed to set experiment parameters; and (2) an imaging session, in which subjects spoke under conditions of normal and altered auditory feedback while brain activity was measured using fMRI.

Behavioral pretest outside the scanner. The behavioral pretest determined the magnitude and direction of formant change necessary to shift each subject's perception of their own speech to a different vowel. Vowel production data were collected between the carrier consonants /b_d/, with each subject producing 10 tokens for each of the six vowels (i, ɪ, ɛ, æ, ɑ, u). The first and second formant frequencies (F1 and F2) were measured and averaged across time points. For ease of recording and subject comfort, the vowels were recorded with the subject seated at a desk, head in an upright posture. The production token closest to the two-dimensional (F1–F2) median for each vowel (the "median token") was used as input to a formant-shifting algorithm (Cai et al., 2008) that altered F1 and F2 by a constant offset while holding other acoustic properties of the sound constant (Fig. 1*A*).

For each subject, eight vowel continua were generated across the F1–F2 spectrum using the formant-shifting algorithm: i–ɪ, ɪ–i, ɪ–ɛ, ɛ–ɪ, ɛ–æ, æ–ɛ, æ–ɑ, and ɑ–æ. For vowel quality reasons, the vowel /u/ was not used. The median token from the first vowel in the pair was formant-shifted in 10 successive increments toward the other vowel in the pair (Fig. 1*B*, left). Thus, each continuum began at the median formant values of one vowel and ended at the median formant values of a neighboring vowel. The step size between each continuum token was constant on the mel scale, a perceptually based logarithmic frequency scale (O'Shaughnessy, 1987). We chose to use mels instead of hertz because the mel scale better reflects the sensitivities of the ear to changes in different regions of frequency space; that is, a fixed distance in mel space is approximately equally discriminable regardless of where along the frequency dimension it is located (Stevens et al., 1937). The tokens from all eight vowel continua were randomized and presented five times each through free-field speakers immediately after the vowel production test. Each subject heard his or her own speech and used a key press to categorize each sound as one of five possible words: bead, bid, bed, bad, or bod. Subjects' reaction time between sound presentation and key press was also measured for each token.

The categorization data were fitted to sigmoid curves to determine an approximate perceptual boundary between the vowels at the continuum endpoints (Fig. 1*B*, right), defined as the point where the two sigmoid curves crossed. Furthermore, two additional regions were defined along each continuum: (1) the "Within" region, containing all the continuum tokens that were categorized as the original vowel 100% of the time; and (2) the "Across" region, containing all the continuum tokens that were categorized as the adjacent vowel 100% of the time, that is, the perceptual judgments had consistently switched to a different vowel.

Calculating within- and cross-boundary shifts. To increase the likelihood that some formant shifts would cross a category boundary while others would not, we prespecified two types of shifts to be used in the brain imaging portion of the experiment. For each subject, a shift magnitude was chosen such that the median token shifted into the Across region if applied along one continuum, e.g., ɛ–æ (Fig. 1*C*, top, red arrow), but stayed within the Within region for another continuum, e.g., ɛ–ɪ (Fig. 1*C*, top, blue arrow). Only subjects for whom such a constant shift could be chosen went on to complete the scanning phase; these subjects made up approximately half of the total subject pool. To control for possible effects attributable to perturbation direction, 18 of these qualifying subjects were selected such that each was counterbalanced with another subject showing the opposite pattern. Thus, for every subject whose Within shift was /ɛ/–/ɪ/ (Fig. 1*C*, top), there was one whose Across shift was /ɛ/–/ɪ/ (Fig. 1*C*, bottom). Fourteen subjects produced /ɛ/ and were shifted toward /ɪ/ and /æ/, whereas four subjects produced /æ/ and were shifted toward /ɛ/ and /ɑ/. The average shift magnitude across subjects was 122.2 mels.

Brain imaging. fMRI was used to measure the blood oxygen level-dependent (BOLD) response during speech, both with and without perturbation, as well as during a non-speech baseline condition. Subjects were scanned in a 3T Siemens Tim Trio whole-body MRI machine equipped with a 32-channel volume transmit–receive birdcage head coil, located at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, Massachusetts Institute of Technology (Cambridge, MA). Images were acquired in a head-first, supine position. The subjects' speech was recorded via a custom-made MR-safe microphone, and auditory feedback was delivered via insert headphones (Stax SRS-005II Electrostatic In-The-Earspeaker). Subjects wore supra-aural ear seals surrounded by a custom-made foam helmet to insulate them from the noise of the scanner.

Functional volumes consisted of 45 T2*-weighted gradient echo, echo planar images aligned to the bicommissural line and covering the entire cortex and cerebellum in the axial plane (3 mm slice thickness; 0.3 mm gap; 2750 ms acquisition time). A high-resolution T1-weighted anatomical volume (128 slices in the sagittal plane, 1 × 1 × 1.3 mm resolution) was collected to overlay each subject's functional data on a structural image of his or her own brain.

The experiment had an event-triggered design (Fig. 1*D*), using sparse sampling and a triggering mechanism to coordinate stimulus timing with image acquisitions (Birn et al., 2004; Bohland and Guenther, 2006; Ghosh et al., 2008; Tourville et al., 2008). At the start of each trial, subjects were visually presented with a word (e.g., "bed") or a control stimulus ("***") for 2 s. Words were drawn from an eight-word list that depended on the vowel to be perturbed. These stimuli were projected in high-contrast white-on-black and displayed on a rear projection screen, visible to the subjects through a mirror mounted above the MRI head coil. Subjects were instructed to read each word aloud when it appeared on the screen, articulating their words clearly but naturally, as if speaking to someone over a noisy phone line, and to remain silent on the control trials. Immediately after each trial, a volume meter gave subjects feedback about the loudness of their speech to help keep the sound level consistent across the experiment. Each of five experimental runs consisted of 80 trials: 64 speech trials (eight presentations each of eight words) and 16 silent control trials.

Unbeknownst to the subjects, the speech trials were divided into three conditions: "NoShift" (normal speech feedback), "Within" (a shift was applied in the direction that did not cause a category change in the behavioral pretest), or "Across" (the same size shift was applied in the direction that did cause a category change in the behavioral pretest). The Within and Across trials each made up one-eighth of the experimental trials, for a total of one-fourth perturbed trials. In these randomly distributed 25% of trials, the formants were perturbed before being fed back to the subjects' headphones. The resultant perturbed trials sounded like altered pronunciations of the trial word.

After a four-second delay from the visual stimulus onset, the scanner was triggered to collect a single volume of functional data. The delay allowed volume acquisition to occur near the peak of the hemodynamic response to speech, estimated to occur ∼4–7 s after vocalization. The functional volume acquisition (acquisition time of 2.75 s) was followed by a pause of 1.25 s before the start of the next trial, for a total trial length of 8 s, to allow for the partial return of the BOLD signal to the steady state. Because the volume acquisition was timed to occur several seconds after the stimulus offset, subjects spoke in relative silence, and artifacts from tongue, jaw, or head movement were avoided. The auditory feedback to the headphones was turned off during image acquisition to prevent the transmission of scanner noise.

Auditory feedback perturbation. Subjects' recorded speech was split into two channels using a MOTU UltraLite FireWire audio interface with on-board mixer (48 kHz sampling rate). One channel of the signal was sent to the laptop to be recorded while the other was processed on the on-board sound card (Fig. 1*A*). This processed signal was resplit and sent both to the laptop and back out to the subject's headphones. Because the same procedure was used for all trials, the signal underwent the same processing delay of ∼17 ms whether or not the formants were shifted on a given trial.

Formant tracking and perturbation were performed in the manner described by Cai et al. (2008). Briefly, the speech audio signal was downsampled by a factor of four (12 kHz) and then pre-emphasized to improve formant estimation (Fant, 1960). Vowel onset and offset were detected using a root mean square (rms) threshold and rms ratio threshold. The voiced signal was then analyzed using a linear predictive coding (LPC) algorithm and the autocorrelation method to estimate the vocal tract transfer function as an all-pole model. The LPC order for each subject was determined from formant-tracking performance in the behavioral pretest (9th to 13th order). During the vowel, formants of the incoming signal were shifted by filtering the signal through a concatenation of two digital biquad infinite impulse response filters. These filters first add zeros at the detected formant frequencies to neutralize the original poles and then add new poles that are shifted in frequency by the desired amount. Finally, because the formant shift changes the gain of the spectral peaks, a gain factor was applied to the filter output before the signal was upsampled and written to the sound card output buffer. The applied formant shifts were constant in magnitude and direction over the duration of the vowel.

Acoustic analyses

The shifts for Within and Across conditions were selected so that the median pretest trial would be shifted within and across boundaries, respectively. However, there is a natural dispersion of formant values across repeated productions of a given vowel. Because of this large variance, the same shift applied to two different productions of the same vowel could result in two different category percepts. For example, consider subject 24 in Figure 1*C* (bottom), who produced the vowel /ɛ/ (as in "bed") during the imaging experiment. If a particular production of /ɛ/ happened to fall near the category boundary with /æ/ (as in "bad"), then a shift Within, toward /æ/, would push the sound over the boundary, even though this shift was intended to stay within the category. Similarly, an Across trial in which the produced vowel happened to fall nearer the /æ/ boundary would not cross the boundary with /ɪ/ (as in "bid"). Therefore, although the Across condition was meant to evoke a change in vowel category, a considerable subset of those trials unintentionally fell short of this change, and a subset of trials in the Within condition contained unintentional boundary crossings. In short, we found that vowel change was much more reliably predicted by the distance from the relevant category boundary than by the trial condition (Within vs Across).

We therefore performed a *post hoc* analysis to separate trials that likely involved a change in perceived vowel category from those that did not. Specifically, for each shift, trials were divided into two groups: the "Near" and "Far" trials were, respectively, the half of trials closest to and farthest from the category boundary that the shift pointed toward (Fig. 1*E*). We assessed distance to the boundary using the average formants during the first 20% of the acoustic signal. The goal of this analysis was to probe varying points in acoustic space for differential sensitivity to the same formant shift. For example, in Figure 1*E*, left, the production circled in blue is far from the boundary that the shift is pointing toward (dashed line at right); however, the production circled in red is very close to this boundary. We compared the compensation of trials in the blue Far region with that of trials in the red Near region. To account for any acoustic differences between the two regions, shifted productions in the Far region were compared with baseline trials that were also in the Far region, and shifted productions in the Near region were similarly compared with baseline trials in the Near region. Furthermore, these regions were counterbalanced within each subject; that is, the Near region for one shift occupies the same acoustic space as the Far region for the other shift, such that there is no acoustic or motoric difference between the two types of trials when both shifts are considered. For example, productions of the vowel /ɛ/ with a high F1 would count as Near when the shift toward /æ/ was applied (Fig. 1*E*, left) but as Far when the shift toward /ɪ/ was applied (Fig. 1*E*, right), such that both Near and Far trials cover the whole distribution. Additionally, because the Near and Far shifts are identical, they are equivalent in magnitude on any frequency scale (e.g., hertz, mels, bark, semi-tones, etc.); therefore, the choice of the mel scale does not influence the results of this analysis. Finally, this analysis does not rely on the stability of perceptual category boundaries between two vowels: the absolute location of the boundaries may vary over time, but Near and Far trials will remain Near and Far as long as the relative position of the two vowels in formant space is approximately maintained.

Acoustic data were compared across NoShift, Within, Across, Near, and Far conditions to identify the behavioral responses to shifted feedback. The first and second formants were tracked with LPC analysis and zero-phase filtered with an eight-point Hamming window. Trials containing errors in production or in formant tracking were removed. Formant values at each time point (4 ms resolution) were averaged across all NoShift trials to yield a baseline vowel trajectory in two-dimensional formant space. Averaged F1 and F2 trajectories for the shifted conditions were then compared with the baseline trajectory.

Each subject had custom-defined shift magnitudes and directions, making a simple F1 or F2 comparison across the subject population impossible. To compare behavioral responses across subjects, the difference between baseline and shifted conditions was calculated at each time point for both first and second formants. At each time point, these two formant differences were then collapsed into one dimension by computing the Euclidean distance in F1–F2 dimensions (represented in Fig. 2*A* as a green line, "deviation"). Because the magnitude of this deviation is positive whether a subject counteracts the perturbation or follows it, it is not a good measure of compensation. Therefore, the "compensation" was defined as the scalar projection of this deviation onto the inverse shift vector. In other words, the compensation is simply the component of the deviation that directly opposed the formant shift (Fig. 2*A*, blue line). A compensatory response was defined as a statistically significant deviation from the baseline trajectory in the direction opposing the formant shift ($p < 0.05$, one-tailed t test with Kolmogorov–Smirnov test for normality), using a fixed-effect analysis with each trial contributing a degree of freedom.

The mean latency of the compensatory response was defined as the first time point for which mean compensation was significant for five time points in a row. Latencies were compared across different conditions using a bootstrapping technique in which the latency was computed 1000 times using the trials from 16 randomly chosen subjects (sampled with replacement). This technique generated a latency distribution that could be compared across conditions.

Finally, the efficiency of compensation was defined as the compensation magnitude (Fig. 2*A*, blue line) as a percentage of the total deviation (Fig. 2*A*, green line). Responses that are perfectly aligned with the inverse of the shift vector (0°) have maximal efficiency (the blue and green lines are identical, so their ratio is 1:1 or 100%), whereas responses orthogonal to the shift vector (90°) have 0% efficiency (the projection of the green line on the inverse shift vector is 0). Responses in the same direction as the shift (180°) have −100% efficiency.

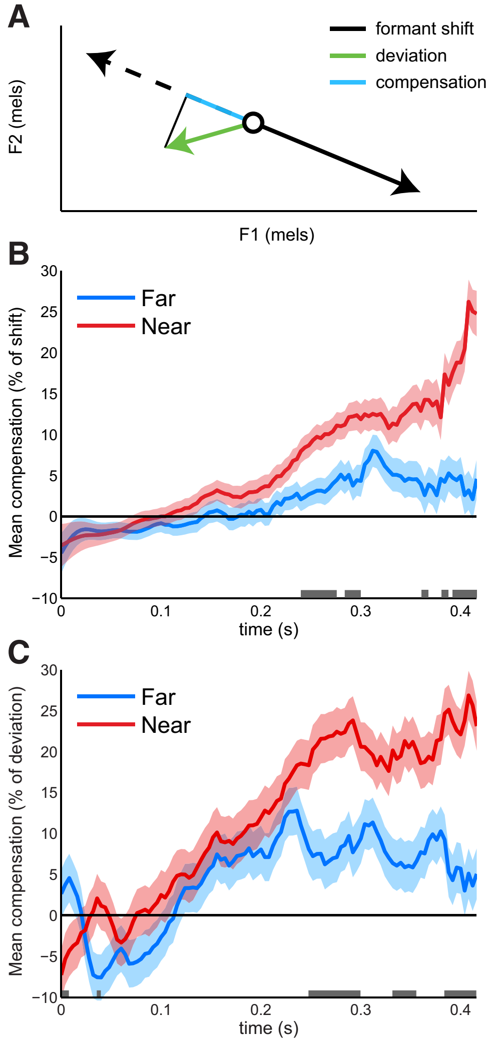

Figure 2. Behavioral responses to formant shifts. ***A***, Schematic of compensation and efficiency. The shift vector (black solid line) is reflected around the production (open circle) to give the inverse shift vector (dashed line), representing "perfect compensation." The total shift-evoked deviation (green) is projected onto the inverse shift vector to yield the compensation (blue): the component directly opposing the shift. ***B***, Compensatory changes in output evoked by formant shifts, divided into Near and Far trials. Shaded regions denote SE; gray bars denote time points at which the two conditions significantly differed from each other ($p < 0.05$). ***C***, Efficiency (ratio of compensation to total deviation) in Near and Far trials.

Functional imaging analyses

The functional imaging data were preprocessed and analyzed using Nipype (Gorgolewski et al., 2011), a pipeline platform chaining together publicly available software packages, including SPM (Friston et al., 1995), Freesurfer (Dale et al., 1999; Fischl et al., 1999), and FSL [for FMRIB (Functional MRI of the Brain) Software Library] (Smith et al., 2004; Woolrich et al., 2009) into one workflow. In the preprocessing stage, a rigid-body transformation was used to realign functional images to the mean EPI image, correcting for subject head movement. The realigned images were stripped of noncortical matter via a brain mask computed with BET (for Brain Extraction Tool) of FSL (Smith, 2002). Outliers with >2 mm of movement or with an intensity z-threshold >3 SDs from the mean were removed from the analysis using artifact detection tools (http://www.nitrc.org/projects/artifact_detect/). Images were coregistered with the T1-weighted anatomical image and spatially normalized into the Montreal Neurological Institute space (Evans et al., 1993). Freesurfer was used to segment each anatomical volume into gray and white matter structures and to perform cortical surface reconstruction. Finally, the images were smoothed on the cortical surface with a Gaussian filter (6 mm full-width at half-maximum).

For each condition of interest, a time series of finite impulses was created to represent the onsets of each event. This time series was then convolved with a canonical hemodynamic response function (HRF), generating a simulated BOLD response. The regressors for each volume were computed by sampling the height of the simulated BOLD response at the time that volume was acquired. The regressors were therefore weighted, taking into account neural responses to both the immediately preceding event and any previous events whose resulting HRFs had not entirely decayed (Ghosh et al., 2009). These regressors were used in the general linear modeling analysis.

A standard hierarchical group model approach was used to model within-subjects and between-subjects effects (Friston et al., 2005). Contrast images were generated for each subject. Conditions were treated as fixed effects. A "summary statistics" procedure was used to model the group effects, performing one-sample *t* tests across the individual contrast images. The model was applied with a *t*-value threshold of 3 and a cluster-threshold correction for multiple comparisons (75 mm^2), resulting in a false discovery rate (FDR) of <5%. A cortical labeling system tailored for studies of speech (Nieto-Castanon et al., 2003; Tourville and Guenther, 2012) was used to identify anatomical regions for active clusters in the activation maps.

Of the 18 subjects run in the study, one subject's fMRI data (subject 44) showed excessive artifacts and no significant activity in the speech − baseline contrast (a very robust contrast with strong activity expected in sensorimotor and auditory regions); therefore, this subject was excluded from additional analysis. Because the experimental design required subjects to be counterbalanced with each other, the subject paired with subject 44 (subject 46) was also excluded. As a result, 16 subjects' data were included in the final analysis.

The first-level analysis yielded *t*-contrast maps for each subject. These *t*-contrast maps were then used in a simple regression analysis with the average amount of compensation in all shifted trials as a covariate measure. The resulting *F*-contrast map shows the regions that have a statistically significant correlation with behavioral measures at the FDR-corrected $p < 0.05$ level. The FDR threshold controls the average proportion of false positives among all voxels declared active (Genovese et al., 2002).

Results

Vocal compensation to feedback shifts is greater near category boundaries

During the fMRI experiment, the auditory feedback for a randomly distributed 25% of trials was shifted in both F1 and F2 before playback. Speakers responded to the unexpected formant shifts by altering their formant trajectories away from the unshifted baseline (Fig. 2). The compensation was defined as the component of the formant deviation that is in opposition to each subject's custom shift (Fig. 2*A*). Across all subjects, the average compensation in each shifted condition was positive, although it occurred at different latencies from voicing onset for Near and Far trials (Near, 140 ms; Far, 256 ms; $p < 0.001$).

Importantly, the magnitude of compensation varied between trial types. Compensation was greater for the Near trials, which contained utterances that were more likely to be shifted into another category, than for the Far trials (Fig. 2*B*, gray bars indicate time points at which $p < 0.05$, two-tailed *t* test, FDR corrected). Notably, by 400 ms from speech onset, compensation in the Near condition reached 25% of shift magnitude, 825% larger than the concurrent compensation in the Far condition (3%) and 313% larger than the peak compensation in the Far condition (8% at 318 ms). Compensation magnitude did not significantly differ between the Within and Across conditions using the same test, although there was a trend for the compensation to be greater in Across than Within trials ($p = 0.10$); this difference attained significance under an uncorrected paired difference test (one-tailed paired *t* test across subjects, $p = 0.03$). Because the Near–Far distinction more directly accounted for category changes and because the behavioral differences were larger and more statistically robust, we focused our results on these conditions.

The efficiency was defined as the ratio of the compensation to the total deviation from baseline. Figure 2*C* shows the average efficiency across all subjects in the Near and Far conditions. The efficiency track deviates from baseline at ~150 ms after the onset of voicing (one-sample *t* test, $p < 0.05$; Near, 140 ms; Far, 152 ms). As in the compensation data, the efficiency is greater for the Near condition than for the Far condition. Furthermore, although the average efficiency over all subjects and all trials was 15%, many subjects reached near-maximal efficiency in the Near condition by vowel offset: three subjects had efficiency values >90%, with five more >80%. In other words, by the end of the utterance, these subjects altered their formants in a direction that opposed the shift almost perfectly, indicating a relatively "pure" (although incomplete) compensation. Mean vowel durations ranged from 150 to 475 ms, with a group mean ± SEM of 312 ± 26 ms.

When questioned after the scanning session, half of the subjects ($n = 8$) reported no awareness of any feedback alteration, whereas the other half ($n = 8$) did report some form of awareness of a change. Of these, many could not articulate what had changed, but others were able to specify the direction of perturbation, saying, for example, that "cab sounded like cob." Of particular note, two subjects reported that sometimes /ɛ/ sounded like /æ/, but sometimes it was like a British or Australian accent. [Anecdotally, for these two subjects, the report of vowel change (/ɛ/ to /æ/) corresponded to the Across condition, and the report of accent change corresponded to the Within condition.] An unpaired *t* test performed on the two sets of subjects, those consciously aware and those unaware of the perturbation, suggests that both groups compensated to the same degree ($p = 0.48$). Furthermore, no significant differences in brain activity were found between the two groups. However, because we assessed conscious awareness of auditory perturbations only after the imaging session was completed, not after each trial, we cannot rule out the possibility that some aspect of awareness may affect compensatory or neural responses on a trial-by-trial level. Thirteen of the 16 subjects compensated in at least one condition. No subjects had negative compensation (i.e., following) responses on average in any of the conditions.

Neural activation to feedback shifts is greater near category boundaries

Figure 3 shows the averaged BOLD activation maps for the Shift–NoShift, Near–NoShift, Far–NoShift, and Near–Far conditions using a mixed-effects analysis of surface-smoothed data. Activations for these experimental conditions are summarized in Table 1. As expected based on previous findings (Tourville et al., 2008), more cortical activation was observed for shifted conditions than for unshifted conditions (Fig. 3*A*) in bilateral posterior superior temporal gyrus (STg) and bilateral inferior frontal gyrus (IFg) pars opercularis and pars triangularis. These results are consistent with the DIVA model predictions of cells in higher-order auditory cortical areas that detect auditory error and project to premotor and inferior frontal cortices to generate corrective motor responses. Significant activity was also found in right dorsal medial prefrontal cortex, near the anterior boundary of the presupplementary motor area.

Figure 3*B*–*D* shows the Shift condition split into Near and Far trials. The Near–NoShift contrast (Fig. 3*B*) was nearly identical to the Shift–NoShift contrast, but the Far–NoShift contrast (Fig.

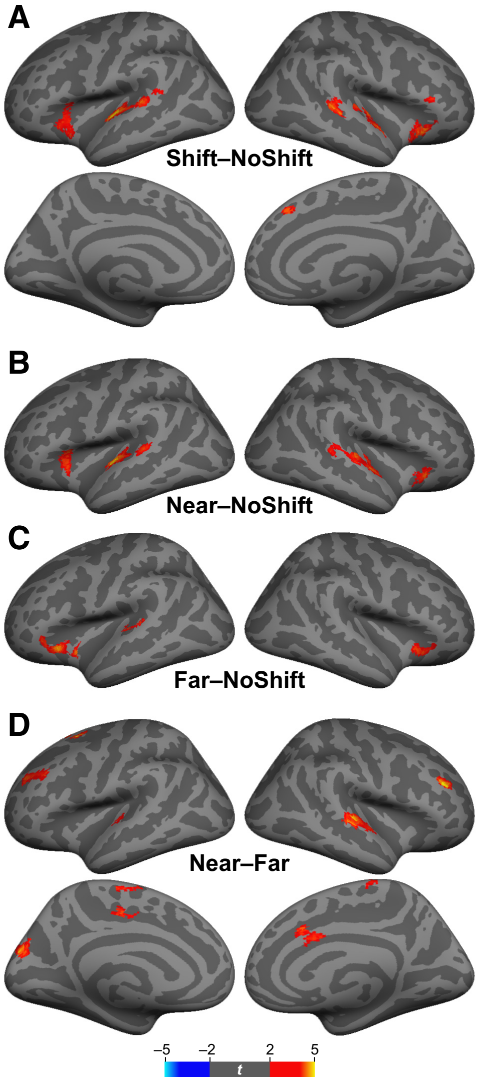

Figure 3. Brain activity evoked by formant shifts ($t > 3.08$, surface cluster threshold of 75 mm^2). ***A***, Shift–NoShift contrast. ***B***, Near–NoShift contrast. ***C***, Far–NoShift contrast. ***D***, Near–Far contrast. The medial surface had no significant clusters in ***B*** and ***C*** and is not shown for those contrasts.

Table 1. Peak voxel responses for two contrasts of interest listed by anatomical region

	Shift–NoShift		Near–Far	
Region label	Peak voxel in Talairach coordinates (x, y, z)	t	Peak voxel in Talairach coordinates (x, y, z)	t
Frontal cortex				
Left adPMC			(−15.2, 19.3, 51.7)	3.700
Left adPMC			(−19.0, 10.8, 48.3)	5.024
Left aMFg			(−24.6, 36.8, 25.8)	3.759
Left pFO	(−38.5, 15.3, 7.0)	3.543		
Left pdPMC			(−7.5, −16.0, 59.4)	3.515
Left pdPMC			(−15.2, −19.9, 41.9)	3.656
Right SFg	(8.8, 28.9, 41.2)	3.757	(12.9, 16.9, 30.8)	4.350
Right aMFg			(40.2, 36.3, 23.0)	5.859
Right dIFt	(50.2, 27.5, 8.0)	3.374		
Right FOC	(31.4, 24.7, −9.0)	4.212		
Right dMC			(3.4, −24.8, 65.1)	3.154
Temporal cortex				
Left Heschl's	(−48.1, −16.5, 0.3)	5.602	(−49.6, −19.7, 6.8)	3.182
Left STg	(−59.8, −44.1, 13.8)	3.095		
Right pdSTs	(52.8, −32.1, 4.4)	4.268		
Right PP	(52.1, −5.6, −5.5)	3.920		
Right PT	(51.8, −20.5, 3.9)	3.435	(62.5, −19.1, 5.4)	4.656
Occipital cortex				
Left OC			(−4.2, −78.6, 28.4)	4.436

Peak responses were defined as local t-statistic maxima ($t > 3.08$) in a cluster of at least 75 mm^2, determined by an FDR correction of the Shift–NoShift contrast. Each peak voxel was mapped to a cortical region (Nieto-Castanon et al., 2003; Tourville and Guenther, 2012) and is listed with the t statistic associated with that voxel. Voxel locations are provided in the Talairach stereotaxic reference frame. adPMC, Anterior dorsal premotor cortex; aMFg, anterior middle frontal gyrus; dIFt, dorsal inferior frontal gyrus, pars triangularis; dMC, dorsal motor cortex; FOC, fronto-orbital cortex; OC, occipital cortex; pdPMC, posterior dorsal premotor cortex; pdSTs, posterior dorsal superior temporal sulcus; pFO, posterior frontal operculum; PP, planum polare; PT, planum temporale; SFg, superior frontal gyrus.

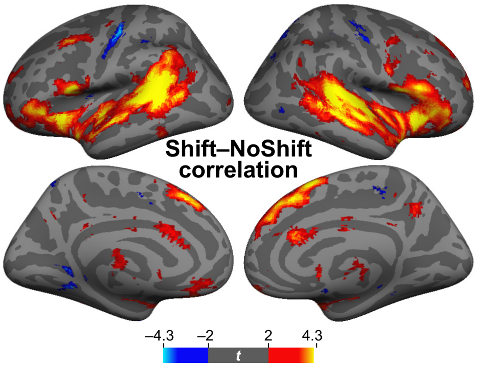

Figure 4. Brain-behavioral correlations. Correlation between behavioral compensation and activation in the Shift–NoShift contrast. $n = 16$ subjects; FDR-corrected $p < 0.05$.

3*C*) revealed significantly weaker superior temporal activations that fell below threshold in the right hemisphere. These differences survived in a direct contrast of Near–Far conditions (Fig. 3*D*), which showed increased activation of bilateral STg and supplementary motor areas (SMAs) for near-boundary trials compared with trials farther from the boundary. Thus, cortical activation was greater during trials that came closer to a category boundary, although the shifts were of identical magnitude and direction. In addition, other cortical areas that were subthreshold in the Shift–NoShift contrast showed differences in the Near and Far conditions: bilateral superior frontal gyrus (including left dorsal premotor cortex, left SMA, right dorsal motor cortex, and right cingulate), bilateral anterior middle frontal gyrus, and left cuneus.

Neural activity correlates with compensatory behavior

The regression analysis identified cortical regions whose activity correlated with a compensatory response. Because shift size and compensatory "performance" varied across subjects, the neural response to feedback perturbation may have been decreased for subjects who did not perceive the perturbations and who exhibited little compensation. We used magnitude of compensation as a regressor to help refine the cortical locations that are specifically recruited in the feedback pathway.

The t-statistic activation maps from each single-subject Shift–NoShift analysis were used in a simple regression analysis with mean compensation as a covariate measure. The F-contrast map shows the regions that have a statistically significant correlation with behavioral measures at the FDR-corrected $p < 0.05$ level (Fig. 4). This correlation analysis corroborates the results of the mean activation analysis in localizing activation primarily to IFg and STg.

Discussion

By monitoring the auditory feedback signal, the speech motor system can rapidly correct small articulatory errors in natural speech. Despite the unpredictability of the auditory perturbations in the current study, speakers altered their speech to oppose the perturbations within the first 150 ms of voicing onset (Fig. 2), as has been demonstrated previously (Xu et al., 2004; Tourville et al., 2008). However, unlike past studies that have used perturbations to low-level acoustic dimensions—for example, a decrease in pitch or an increase of the first formant—the experiments described here specifically probed the salience of shifting toward neighboring vowel targets. By capitalizing on speaker variability, we demonstrated that identical shifts can evoke systematically different response magnitudes when they occur in different parts of auditory space: specifically, shifts in the Near condition, which caused a stronger percept of vowel category change, triggered greater auditory cortical activity and greater behavioral compensation. The average compensation in the Near condition reached 25% of the shift by 400 ms, double that reported previously in studies of unexpected perturbations with shifts of this magnitude (Purcell and Munhall, 2006; Tourville et al., 2008). In contrast, the average compensation in the Far condition never exceeded 8% of the (acoustically identical) shift. This suggests that the average compensation found in past studies was driven mainly by the more salient near-boundary utterances.

These results have several important implications. First, they show that learned phoneme categories influence the perception of self-produced error, just as they influence the perception of others' speech in a passive listening setting. Second, our results are consistent with an error calculation based not just on the acoustic distance from the current planned utterance, because this distance was the same in the Near and Far conditions, but also on the distance from a prototypical or "best" vowel target region. Past studies provide evidence that discrimination ability is increased when auditory stimuli lie farther from the prototype vowel, nearer to the category boundaries (Kuhl, 1991). This effect suggests a mechanism for our current results, namely that a feedback shift applied near a boundary is more easily distinguished from the intended vowel. Our data further show that these boundary influences occur relatively rapidly, early enough in processing to allow for magnitude differences in compensatory behavior to emerge within ~200 ms of voicing onset.

In line with this idea, Mitsuya et al. (2011) showed an effect of phonological processing on speech feedback control by contrasting the responses of English and Japanese speakers to a gradual increase in F1 feedback. The English speakers, who had a neighboring vowel in the direction of the formant shift, showed a greater magnitude of adaptation than the Japanese speakers, suggesting that the vowel density of English may have resulted in greater saliency of the F1 increase. These adaptation responses may arise from somewhat different mechanisms than the rapid compensatory responses elicited by unexpected shifts in the current study (Purcell and Munhall, 2006; Katseff et al., 2012); however, this interpretation is also supported by several experiments using unexpected F0 perturbation, to both Mandarin bi-tonal disyllables (Xu et al., 2004) and prosodic contour (Chen et al., 2007). Like the cross-category formant shifts, the F0 perturbations that had linguistic relevance (opposite the intended intersyllabic tonal transition in Mandarin, or the upward intonation contour of an interrogative sentence) resulted in shorter latencies and larger compensatory responses.

In the current study, we interpret the differences in compensation between Near and Far trials as reflecting the differences in auditory salience between perturbations with and without phonetic relevance. An alternative explanation for the compensatory differences is a motoric one: for Near utterances, the direction of expected compensation (opposing the shift) is back toward the mean of the normal vowel distribution (Fig. 1*E*, compensation opposing red arrows), whereas for Far utterances, compensation would necessitate a movement even farther away from the center of the distribution (Fig. 1*E*, compensation opposing blue arrows). In the latter case, the resulting somatosensory signals would be very unlike normal productions of the vowel; thus, Near trials might be "easier" to compensate for than Far trials. Although motoric aspects may constrain compensatory responses, it remains unclear how these motoric constraints could account for the increase in auditory cortical activity seen in the Near trials compared with the Far trials (Fig. 3*D*).

Increases in neural response magnitude for near- or cross-boundary changes have also been reported in electrophysiological studies of speech perception. Auditory evoked potentials (Dorman, 1974) and mismatch negativity (MMN) responses elicited in an oddball paradigm (Sharma and Dorman, 1999; Phillips et al., 2000) were found to be larger for cross-category than within-category contrasts. Enhancement of the neural responses was also shown in the absence of a boundary crossing: in a neurophysiologic analog of the perceptual magnet effect, "good" categorizers showed poorer discrimination and weaker MMN amplitudes for contrasts involving a prototype vowel than for contrasts involving nonprototype vowels (Aaltonen et al., 1997), although all vowels tested were rated as being in the same phonetic category (/i/). In summary, these studies show the influence of category representations on the early auditory processing of speech sounds. We provide evidence here that the speech motor control system is similarly influenced by phoneme category representations in the moment-to-moment control of speech articulators.

Formant perturbation causes a mismatch between the auditory expectation for the current vowel and the perceived auditory signal. According to the DIVA model (Guenther et al., 2006), this mismatch leads to activation of auditory error cells, located bilaterally in the posterior STg. This prediction is strongly supported by the bilateral peri-Sylvian activation noted in the Shift–NoShift contrast. The activation found in these regions replicates that found by Tourville et al. (2008) in a similar fMRI study of unexpected formant perturbation and by several other studies involving pitch shifts (Zarate and Zatorre, 2005; Toyomura et al., 2007) and auditory feedback delay (Hashimoto and Sakai, 2003). Furthermore, recent work from animal model systems in vocal error processing have shown similar neural circuitry underlying auditory feedback processing (Eliades and Wang, 2008; Lei and Mooney, 2010). Together, these results are strong evidence for a neural circuit comparing speech targets and perceived auditory feedback in posterior temporal cortex. The enhancement of this auditory activity in the Near condition is consistent with a larger mismatch between perception and target, resulting in a stronger cortical error signal despite an identical acoustic discrepancy.

Furthermore, these same past studies have reported activity in posterior regions of the IFg, including the posterior portion of Broca's area, the frontal operculum, and adjacent anterior insula, which are also seen bilaterally in the current study. In the DIVA model, the left IFg is hypothesized to contain the speech sound map responsible for the generation of feedforward motor commands. No significant differences in IFg or anterior insula were observed between the Near and Far conditions, despite large differences in compensatory vocal output. However, these areas correlated strongly with compensation (Fig. 4) across subjects, suggesting that they may play a role in feedback-based corrective articulatory movements. Finally, no ventral motor cortical or cerebellar activation was found in response to perturbed conditions, in contrast with the areas of activation described by Tourville et al. (2008). However, SMAs, implicated in motor sequencing (Wildgruber et al., 1999) and articulatory planning (Indefrey and Levelt, 2004), were found to be active in the Near–Far contrast.

Half of the subjects in the current study reported some conscious awareness of the formant shift, some of whom could describe the cross-category shift as a change in vowel identity. Both those who noticed the auditory manipulation and those who did not were found to compensate to the same degree, and no differences in perturbation-evoked neural activity were found between the two groups. These findings imply that the vocal correction is automatic and that conscious awareness did not help or hinder the feedback correction response. Given that compensation is preconscious, it is also possible that the "better" compensators could counteract the perturbation before consciously hearing it.

In summary, speakers who experienced an unexpected shift of their spoken formants toward another vowel were shown to compensate whether or not the shift caused a category boundary to be crossed. However, near-boundary shifts elicited enhanced behavioral and auditory cortical responses compared with far-boundary shifts, even when the shift magnitudes and directions were acoustically identical. This enhancement is evidence that auditory feedback control of speech is influenced by learned phoneme categories. Although the compensatory responses to perturbation occur at a preconscious level, phoneme knowledge plays a role in determining the size of the corrective compensation. The categorical nature of speech sounds has been well studied in perceptual contexts; here, we provide the first demonstration of its effect on motor performance, illustrating the important role of perception in the online control of motor speech skills.

References

Aaltonen O, Eerola O, Hellström A, Uusipaikka E, Lang AH (1997) Perceptual magnet effect in the light of behavioral and psychophysiological data. J Acoust Soc Am 101:1090–1105. CrossRef Medline

Bauer JJ, Mittal J, Larson CR, Hain TC (2006) Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. J Acoust Soc Am 119:2363–2371. CrossRef Medline

Birn RM, Cox RW, Bandettini PA (2004) Experimental designs and processing strategies for fMRI studies involving overt verbal responses. Neuroimage 23:1046–1058. CrossRef Medline

Bohland JW, Guenther FH (2006) An fMRI investigation of syllable sequence production. Neuroimage 32:821–841. CrossRef Medline

Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am 103:3153–3161. CrossRef Medline

Cai S, Boucek M, Ghosh SS, Guenther FH, Perkell JS (2008) A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. Presented at the Eighth International Seminar on Speech Production, Strasbourg, France, December 8–12.

Cai S, Ghosh SS, Guenther FH, Perkell JS (2011) Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. J Neurosci 31:16483–16490. CrossRef Medline

Chen SH, Liu H, Xu Y, Larson CR (2007) Voice F[sub 0] responses to pitch-shifted voice feedback during English speech. J Acoust Soc Am 121:1157–1163. CrossRef Medline

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194. CrossRef Medline

Dorman MF (1974) Auditory evoked potential correlates of speech sound discrimination. Percept Psychophys 15:215–220. CrossRef

Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453:1102–1106. CrossRef Medline

Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM (1993) 3D statistical neuroanatomical models from 305 MRI volumes. Proc IEEE Nuclear Sci Symp Med Imaging 3:1813–1817.

Fant G (1960) Acoustic theory of speech production. The Hague, The Netherlands: de Gruyter Mouton.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207. CrossRef Medline

Friston KJ, Frith CD, Frackowiak RS, Turner R (1995) Characterizing dynamic brain responses with fMRI: a multivariate approach. Neuroimage 2:166–172. CrossRef Medline

Friston KJ, Stephan KE, Lund TE, Morcom A, Kiebel S (2005) Mixed-effects and fMRI studies. Neuroimage 24:244–252. CrossRef Medline

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870–878. CrossRef Medline

Ghosh SS, Tourville JA, Guenther FH (2008) A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. J Speech Lang Hear Res 51:1183–1202. CrossRef Medline

Ghosh S, Kovelman I, Lymberis J, Gabrieli J (2009) Incorporating hemodynamic response functions to improve analysis models for sparse-acquisition experiments. Neuroimage 47:S125. CrossRef

Golfinopoulos E, Tourville JA, Guenther FH (2010) The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage 52:862–874. CrossRef Medline

Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS (2011) Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform 5:13. CrossRef Medline

Guenther FH (1994) A neural network model of speech acquisition and motor equivalent speech production. Biol Cybern 72:43–53. CrossRef Medline

Guenther FH (1995) Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychol Rev 102: 594–621. CrossRef Medline

Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 96:280–301. CrossRef Medline

Hashimoto Y, Sakai KL (2003) Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. Hum Brain Mapp 20:22–28. CrossRef Medline

Indefrey P, Levelt WJ (2004) The spatial and temporal signatures of word production components. Cognition 92:101–144. CrossRef Medline

Iverson P, Kuhl PK (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. J Acoust Soc Am 97:553–562. CrossRef Medline

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

Katseff S, Houde J, Johnson K (2012) Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? Lang Speech 55:295–308. CrossRef Medline

Kuhl PK (1991) Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Percept Psychophys 50:93–107. CrossRef Medline

Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, Nelson T (2008) Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos Trans R Soc Lond B Biol Sci 363:979–1000. CrossRef Medline

Lei H, Mooney R (2010) Manipulation of a central auditory representation shapes learned vocal output. Neuron 65:122–134. CrossRef Medline

Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol 54:358–368. CrossRef Medline

Mitsuya T, Macdonald EN, Purcell DW, Munhall KG (2011) A cross-language study of compensation in response to real-time formant perturbation. J Acoust Soc Am 130:2978–2986. CrossRef Medline

Nieto-Castanon A, Ghosh SS, Tourville JA, Guenther FH (2003) Region of interest based analysis of functional imaging data. Neuroimage 19:1303–1316. CrossRef Medline

O'Shaughnessy D (1987) Speech communication: human and machine, pp 150. New York: Addison-Wesley.

Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T (2000) Auditory cortex accesses phonological categories: an MEG mismatch study. J Cogn Neurosci 12:1038–1055. CrossRef Medline

Purcell DW, Munhall KG (2006) Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am 119:2288–2297. CrossRef Medline

Sharma A, Dorman MF (1999) Cortical auditory evoked potential correlates of categorical perception of voice-onset time. J Acoust Soc Am 106:1078–1083. CrossRef Medline

Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155. CrossRef Medline

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23:S208–S219. CrossRef Medline

Stevens SS, Volkmann J, Newman EB (1937) A scale for the measurement of the psychological magnitude pitch. J Acoust Soc Am 8:185–190. CrossRef

Tourville JA, Guenther FH (2012) Automatic cortical labeling system for neuroimaging studies of normal and disordered speech. Soc Neurosci Abstr 38:681.06.

Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. Neuroimage 39:1429–1443. CrossRef Medline

Toyomura A, Koyama S, Miyamaoto T, Terao A, Omori T, Murohashi H, Kuriki S (2007) Neural correlates of auditory feedback control in human. Neuroscience 146:499–503. CrossRef Medline

Wildgruber D, Kischka U, Ackermann H, Klose U, Grodd W (1999) Dynamic pattern of brain activation during sequencing of word strings evaluated by fMRI. Cogn Brain Res 7:285–294. CrossRef Medline

Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM (2009) Bayesian analysis of neuroimaging data in FSL. Neuroimage 45:S173–S186. CrossRef Medline

Xu Y, Larson CR, Bauer JJ, Hain TC (2004) Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. J Acoust Soc Am 116:1168–1178. CrossRef Medline

Zarate JM, Zatorre RJ (2005) Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann N Y Acad Sci 1060:404–408. CrossRef Medline