# Human cortical sensorimotor network underlying feedback control of vocal pitch

Edward F. Chang[a,1,2], Caroline A. Niziolek[b,1], Robert T. Knight[a], Srikantan S. Nagarajan[c], and John F. Houde[b,2]

Departments of [a]Neurological Surgery, [b]Otolaryngology, and [c]Radiology, University of California, San Francisco, CA 94143

The control of vocalization is critically dependent on auditory feedback. Here, we determined the human peri-Sylvian speech network that mediates feedback control of pitch using direct cortical recordings. Subjects phonated while a real-time signal processor briefly perturbed their output pitch (speak condition). Subjects later heard the same recordings of their auditory feedback (listen condition). In posterior superior temporal gyrus, a proportion of sites had suppressed responses to normal feedback, whereas other spatially independent sites had enhanced responses to altered feedback. Behaviorally, speakers compensated for perturbations by changing their pitch. Single-trial analyses revealed that compensatory vocal changes were predicted by the magnitude of both auditory and subsequent ventral premotor responses to perturbations. Furthermore, sites whose responses to perturbation were enhanced in the speaking condition exhibited stronger correlations with behavior. This sensorimotor cortical network appears to underlie auditory feedback-based control of vocal pitch in humans.

A fundamental question in neuroscience is how sensory feedback is integrated into the control of complex motor actions. Auditory feedback in particular has been shown to affect the motor control of speech. For example, speakers reflexively increase their speech volume in noisy environments (1, 2). Furthermore, in experiments that manipulate individual features of audio feedback such as pitch (3–5), loudness (6, 7), formant frequencies (8, 9), and frication energy (10), speakers make very specific adjustments in vocal output to compensate for those changes. Such compensatory behavior strongly suggests the existence of feedback error-detection and -correction circuits in the speech motor control system. Indeed, past neuroimaging studies have revealed a complex brain network activated by auditory feedback manipulation (11–13), including motor, premotor, and auditory cortical areas. However, the neural mechanisms underlying vocal responses to auditory feedback remain poorly understood.

A parallel issue involves the effect of motor actions on sensory responses. Recent experimental findings have demonstrated that the act of speaking modulates cortical responses to speech (14–19). For example, speaking-induced suppression (SIS) is a specific case of motor-induced suppression (20–22) in which auditory responses to self-produced speech are suppressed (listening > speaking). Self-vocalization also can enhance auditory responses to transient perturbations in vocal pitch feedback (speaking > listening) (23), a phenomenon called "speech perturbation-response enhancement" (SPRE). However, the functional significance of auditory modulations such as SIS and SPRE, and specifically their effect on motor cortical activity and vocal behavior, remains unclear.

These phenomena have raised important questions about the role of auditory feedback during speech. Principally, what roles are played by the cortical areas found to be important for sensorimotor control of vocalization? Does modulatory activity in these regions have consequences for corrective modifications in vocal output? Given that compensatory responses to perturbations rely on auditory self-monitoring, we hypothesized that speech-driven auditory cortical modulations such as SIS and SPRE underlie the corrective vocal output.

To address these questions, we recorded directly from the peri-Sylvian speech cortices in patients undergoing electrocorticographic (ECoG) monitoring for seizure localization. These recordings offer a unique spatial scale between single units and extracranial field potentials. ECoG monitoring has the advantage of simultaneous high spatial and temporal resolution as well as the excellent signal-to-noise properties needed for single-trial analyses. During neural recording, we used a digital signal-processing device (DSP) to induce real-time pitch perturbations while subjects vocalized a prolonged vowel /ɑ/ sound (Fig. 1). The subject's microphone signal was manipulated to create 200-cent (two-semitone) upward or downward shifts in pitch (F0) and was fed back to the subject's earphones (speak condition). This pitch-shifted audio feedback was recorded and later played back to subjects (listen condition). We evaluated neural recording sites for suppression and enhancement by comparing the neural responses in the listen condition with those in the speak condition. We also correlated neural activity with the changes in vocal output elicited by the pitch perturbation.

## Results

**Acoustic Pitch Perturbations Induce Highly Variable Degrees of Vocal Compensation.** The behavioral response to a brief pitch perturbation in auditory feedback is shown in Fig. 1. In this single-trial example, the DSP perturbed the subject's vocal feedback by abruptly lowering the pitch by 200 cents, as can be seen in the narrow-band spectrogram of the acoustic recording at the earphones (Fig. 1B). In this trial, ~170 ms after the perturbation onset, the pitch of the subject's vocalization begins to deviate from baseline, and by 400 ms it has increased by ~100 cents. This increase is seen readily in the pitch track in Fig. 1B, where the red line corresponds to the pitch of the vocalization recorded at the microphone, and the blue line corresponds to the shifted pitch output heard at the earphones. As shown by the blue line, the subject acts to cancel the pitch feedback shift partially; that is, the response is compensatory.

Although, on average, all seven subjects displayed compensatory (and not following) behavior, the response to perturbations varied from trial to trial. A histogram of the compensation magnitudes across trials for a single subject (Fig. 1C) shows highly variable response magnitudes ranging from −25 to 60% compensation (coefficient of variation = 1.41), with an average compensation of 10.6%; the average compensation across subjects was 10.8%, or 21.6 cents (one-sample $t$ test, $P < 0.001$), in agreement with previous studies of similar shift magnitude (3). In some trials, no compensation or even negative compensation (i.e., following) was observed. We hypothesized that speaking-related modulation (i.e., SIS and SPRE) could explain the behavioral variability in compensation across trials.
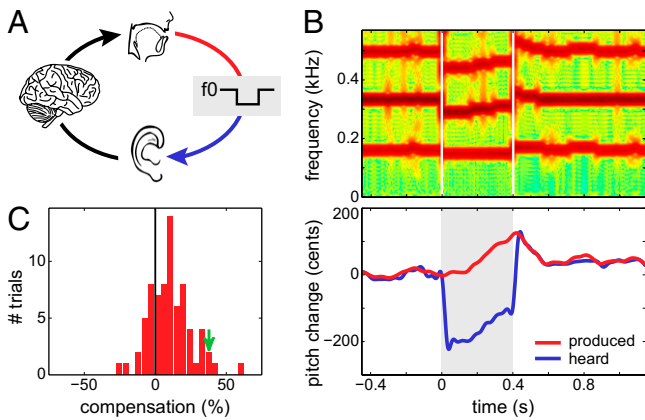
NEUROSCIENCE

**Fig. 1.** Apparatus and behavior. (*A*) Diagram of the pitch perturbation apparatus. A DSP shifted the pitch of subjects' vocalizations (red line) and delivered this auditory feedback (blue line) to subjects' earphones. (*B*) Spectrogram (*Upper*) and pitch track (*Lower*) of an example trial with pitch perturbation applied. (*C*) Histogram of compensatory responses as a percentage of pitch shift. The green arrow denotes the trial shown in *B*.

**Cortical Neurophysiology During Pitch Perturbation of Vocalization.**
We used time–frequency analyses (Hilbert transform) to extract the high-γ component of the local field potential (50–150 Hz) (19, 24, 25). This component has been found to correlate well with neuronal spiking (26, 27) and to be a reliable indicator of focal, event-related cortical activity; therefore we focused our analysis on the high-γ band. Examination of all the electrodes over the lateral hemisphere (Fig. 2*A*) revealed significant high-γ activity in the peri-Sylvian sensorimotor network for vocalization. Time–frequency spectrogram plots of the local field potential from three representative auditory posterior temporal gyrus electrodes (e21, e22, and e23) and one representative vocal motor precentral gyrus electrode (e45) from an example subject are shown in Fig. 2 to illustrate varying response types. These spectrograms, averaged across all trials, show strong evoked neural modulation in the high-γ band as well as in lower α- and β-band frequencies (e.g., e23). Furthermore, the high-γ responses, in contrast to the other bands, demonstrated a clear temporal flow of phasic activation that differed between speaking and listening. (High-γ responses for representative electrodes in all subjects are shown in Figs. S1 and S2.)

At the onset of vocalization (Fig. 2*B*), the ventral precentral gyrus electrode showed activation preceding the vocalization, consistent with anticipatory motor commands (e45; high-γ responses are at right). During the listen condition, a small activation was observed in this motor electrode following the onset, consistent with a "mirror neuron" response (24, 28). Multiple auditory electrodes showed activation increases after the onset, primarily over the posterior superior temporal gyrus (pSTG; e21 and e22) and at the temporal–parietal junction (e23) (29, 30). During the listen condition, the response magnitude was largely identical across these electrodes. In contrast, during the speak condition, we observed heterogeneous response properties, with some electrodes showing no change from the listen condition (e22) and others showing substantially suppressed activity (e21). Electrodes showing this suppressed activity were defined as "SIS" electrodes.

During the 400-ms pitch perturbation, heterogeneous response types were observed in pSTG at different latencies and amplitudes (Fig. 2*C*). e22 and e23 had low-latency, high-amplitude responses which showed significant enhancement during the speak condition (compared with the listen condition). These responses reflect an augmented sensitivity to unexpected feedback while the subject was actively vocalizing. Electrodes showing this enhanced activity were defined as SPRE electrodes. Following this auditory response, increased high-γ activity was observed in the motor electrode (e45) at ~200 ms after the perturbation onset. Additionally, a small increase in high-γ activity was observed in the listen condition, consistent with a mirror-neuron response to the speech audio.

These findings demonstrate the time course of cortical activation from the motor to auditory cortices at the onset of vocalization and vice versa during the perturbation. The auditory cortex shows bidirectional modulation of activity by speech onset (SIS) and pitch perturbation (SPRE). Importantly, auditory electrodes can show a strong SPRE effect with no SIS (as in e22), suggesting separate mechanisms for the two types of auditory modulation.

**Cortical Activity During Perturbation Predicts Compensatory Behavior.**
To probe the behavioral implications of perturbation-related neural activity, we used the trial-by-trial activity at each electrode as a predictor of compensation. Fig. 3*A* is a raster plot showing single-trial high-γ activity in the speak condition, time-aligned to peak compensation and sorted by percent compensation, for each of the electrodes shown in Fig. 2. For the correlated electrodes, the neural response is strongest at the top of the plot, where compensation is highest. Fig. 3*B* shows the behavioral compensation of each trial as a function of per-trial high-γ
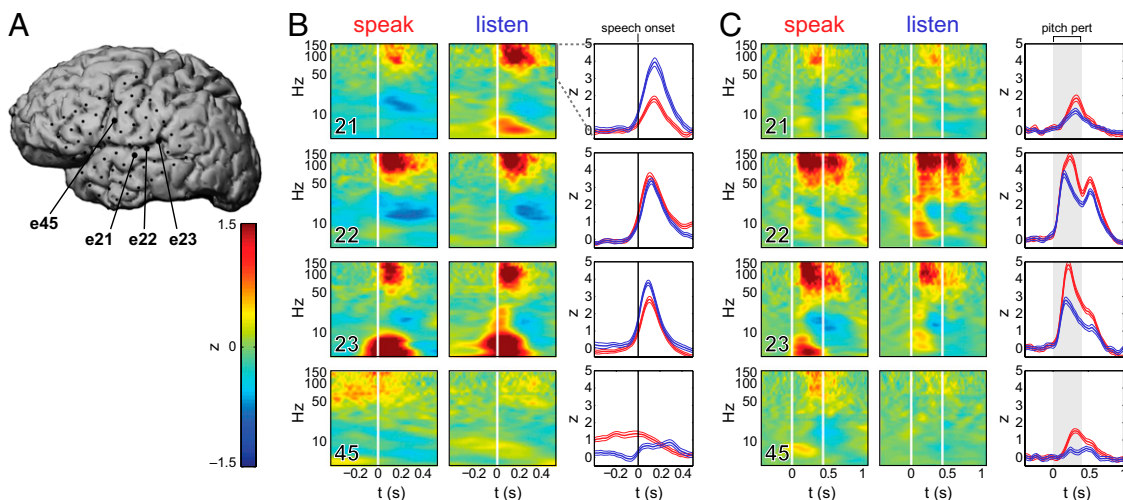


**Fig. 2.** Four ECoG channels from a single subject (GP35). (*A*) Location of the four electrodes on the cortical surface. (*B* and *C*) Spectrograms and high-γ line plots for each electrode in the speak (red) and listen (blue) conditions. Vertical lines represent speech onset in *B* and perturbation onset and offset in *C*.
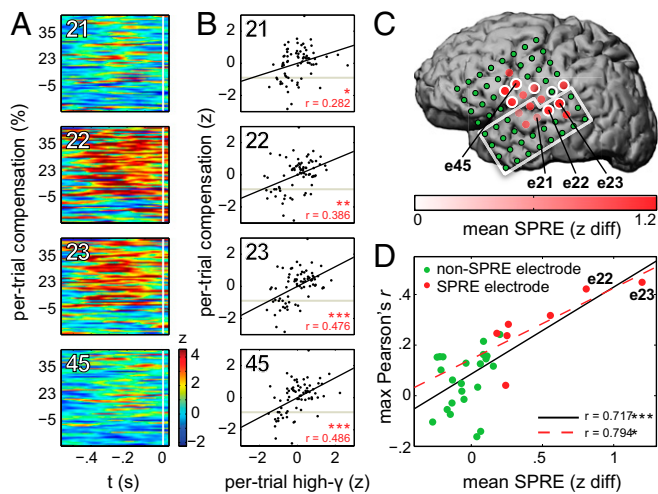
**Fig. 3.** Correlations between high-γ activity and compensation in a single subject (GP35). Asterisks denote statistical significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). (*A*) Single-trial rasters of high-γ activity, ordered by descending compensation, for the four electrodes shown in Fig. 2. The vertical white line marks the time of peak compensation. (*B*) Per-trial correlations for the same four electrodes. Gray horizontal lines indicate the zero compensation level, with compensatory responses above and following responses below the line. (*C*) Spatial distribution of significantly correlated electrodes (circled) and SPRE electrodes (red; opacity denotes degree of SPRE). The white box contains electrodes labeled "temporal" and used in the analysis in *D*. (*D*) Mean SPRE correlated with Pearson's *r* for each electrode. The solid black line is the best-fit line to all temporal electrodes ($P < 0.001$). The dashed red line is the best-fit line to SPRE electrodes alone ($P = 0.033$).

activity for the same four electrodes. (Correlations of neural activity and behavioral compensation for all other subjects are shown in Figs. S1 and S2.) Compensation was most correlated with high-γ activity for electrodes in the pSTG (e22 and e23) and ventral precentral gyrus (e45). These correlations remained significant ($P < 0.05$) even when trials with negative compensation were removed. In these electrodes, the correlation between compensation and activity was weaker in other frequency bands, including the time-locked evoked responses.

Across the entire left-hemisphere subdural grid of our example subject, correlated electrodes clustered in the ventral premotor cortex as well as in the posterior temporal-inferior parietal cortex, close to auditory sites exhibiting SIS and SPRE. Fig. 3*C* illustrates the considerable overlap between the pattern of significantly correlated electrodes (white circles) and that of the SPRE electrodes (red dots). This overlap suggested that the activity of premotor electrodes during perturbation is indicative of compensatory commands to laryngeal muscles and led us to investigate whether SPRE in auditory cortical electrodes also co-occurs with neural–behavioral correlations. In the temporal cortex (the white-outlined box in Fig. 3*C*), electrodes that exhibited SPRE showed stronger correlations between activity and compensation than those that did not (unpaired two-sample *t* test, $n = 30$, $P < 0.001$). Furthermore, the degree of enhancement (SPRE) for an electrode was predictive of the correlation between that electrode's activity and the compensatory pitch change (Fig. 3*D*). However, the same analysis using SIS as a covariate did not show any difference in correlation strength (unpaired two-sample *t* test, $n = 30$, $P = 0.49$), suggesting that SPRE, and not SIS, is a marker for influence on the corrective motor signal.

Across subjects, the same pattern holds across temporal electrodes in four left-hemisphere and three right-hemisphere grids: SPRE electrodes showed stronger behavioral correlations than non-SPRE electrodes [three-way ANOVA of Fisher *z*-transformed correlation values, $F(1,6) = 38.58$, $P < 0.001$; see Fig. 4 *A* and *D* for left- and right-hemisphere grids, respectively]. SIS did not affect correlation strength significantly [$F(1,6) = 3.26$, $P =$

0.073]. There were no significant interactions between any of the factors of SIS, SPRE, and subject.

Because SPRE is defined as a speaking-related enhancement, all SPRE electrodes have a significant response to perturbation in the speak condition. To ensure that the differences in correlation strength were not caused merely by differences in activity during speaking, we divided the non-SPRE group based on each electrode's response to perturbation. Fig. 4*B* shows the population of temporal electrodes across all left-hemisphere subjects sorted into three groups: electrodes with no response to perturbation (green), electrodes with a response to perturbation but no enhancement from speaking (blue), and electrodes with an enhanced response in the speak condition (SPRE; red). The SPRE electrodes had the highest correlations with compensation; auditory electrodes that responded to perturbation but lacked speaking-related enhancement had weaker correlations [one-way ANOVA, $F(2,153) = 20.05$, $P < 0.001$]. In other words, taking into account the difference between speak and listen conditions increases predictive power. Furthermore, as shown in Fig. 4*C*, the more an auditory electrode showed an enhanced response to perturbation during speaking, the more that electrode correlated with compensatory behavior (Pearson's correlation, $n = 154$, $r = 0.437$, $P = 0.001$). A one-way ANCOVA ensured that this result was not an effect of subject [$F(1,3) = 19.86$, $P < 0.001$; individual subject correlations shown in Fig. S3]. Results showed a similar trend in the right hemisphere (Fig. 4 *E* and *F*) but were underpowered because the grid placement limited the coverage in temporal and ventral premotor areas in these subjects. For this reason, we have focused subsequent analyses on the four subjects with left-hemisphere grids having coverage relevant to the task.

**Spatial Distribution of SIS and SPRE Across Subjects.** SPRE electrodes for all left-hemisphere subjects clustered mostly in the ventral premotor cortex and in the posterior superior temporal cortex, including the temporal–parietal junction, with additional SPRE responses found along the anterior extent of the superior temporal gyrus. SIS responses covered similar cortical territory but typically were not seen in the SPRE electrodes, suggesting separate neural populations (Fig. 5). Furthermore, in all our subjects, the degrees of
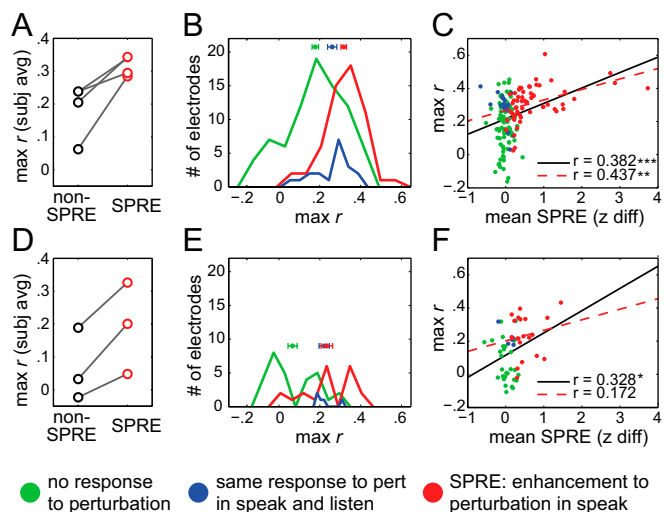
**Fig. 4.** Correlations between high-γ activity and compensation. (*A* and *D*) Per-subject correlation scores averaged across non-SPRE and SPRE temporal electrodes for the left- (*A*) and right- (*D*) hemisphere grids. Each linked pair of points represents data from a single subject. (*B* and *E*) Histogram of electrodes categorized by response properties for the left (*B*) and right (*E*) hemisphere. Error bars show SE. (*C* and *F*) Mean SPRE correlated with maximum Pearson's *r* for each electrode for the left (*C*) and right (*F*) hemisphere. Asterisks denote statistical significance as in Fig. 3, and the black and red lines are best-fit lines.
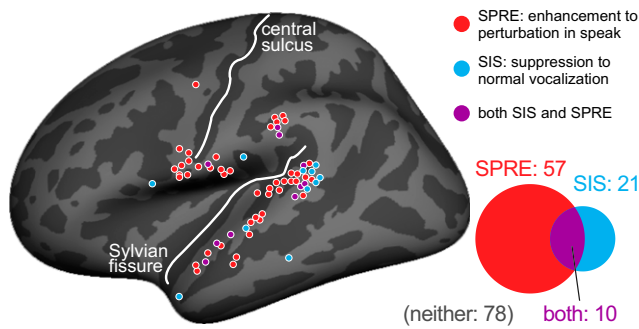
Chang et al.

**Fig. 5.** Spatial distribution of SPRE and SIS electrodes. Points were mapped from individual subject's brains to an average surface; any electrodes that appear to be positioned in the sulci are the result of surface coregistration inaccuracies. Gyri are light gray; sulci are dark gray.

SIS and SPRE in any given electrode were not significantly correlated (Pearson's correlation of SIS and SPRE in all temporal electrodes; in left-hemisphere grids: $n = 156$, $r = -0.02$, $P = 0.78$; in right-hemisphere grids: $n = 54$, $r = -0.16$, $P = 0.25$), further suggesting separable mechanisms for suppression of predicted and enhancement of unpredicted speech auditory feedback.

## Discussion

Rapid compensatory responses to auditory perturbation are evidence for an auditory–motor feedback loop for the online control of speech. We explored the cortical basis of feedback compensation by recording directly from peri-Sylvian speech cortices while applying pitch perturbations to the auditory feedback signal. We assessed the role of the modulatory effects of vocalization by comparing neural responses during speech with those evoked by listening to recordings of the same auditory stimuli. Consistent with past studies, we found that the act of speaking can induce bidirectional modulation of auditory cortex: suppression during normal vocalization, when the acoustic targets meet motor-generated expectations (15–18, 31, 32), and enhancement during vocalization with pitch-altered feedback, when they do not (23). However, with the high spatial resolution and single-trial specificity of intracranial recordings, we were able to relate the two phenomena, demonstrating that suppression is not predicted by enhancement. Moreover, here we present directly recorded electrophysiological evidence that activity from both motor and auditory cortices is correlated with subsequent behavioral motor compensation on a per-trial basis. In particular, correlations in auditory cortex were highest for sites with strong enhancement (SPRE). Although correlated activity was not limited to these enhanced sites, the greater the enhancement in a given site, the more likely was its activity to be predictive of compensatory behavior. These results support a model of human vocal motor control with a strong contributory role of auditory cortex to motor-driven compensation.

In many current models of motor control, a forward model encodes the predicted sensory consequences of motor commands via efference copy (33). In the speech domain, the motor cortex projects a neural representation of the intended speech signal to auditory and somatosensory cortices. This efference copy allows a selective SIS suppression of the neural response to the resulting feedback sensations through a comparison with, or subtraction of, the predicted feedback (34, 35). It has been theorized that such suppression affords a mechanism to distinguish between sensations that come from the speaker and those that are external. Self-generated (and therefore well-predicted) sounds give rise to suppressed responses and are thereby "tagged as self," allowing speakers to attend better to sounds from the external acoustic environment. However, the comparison between efference copy and external feedback also may play another important role: It may enable speakers to detect mismatches between intended and observed sensory outcomes.

We have provided evidence that speech-related enhancement is a hallmark of auditory influence on motor output. We suggest that this enhancement has a corrective function: It underlies the self-monitoring of one's own vocalizations for online modification and control.

A pitch perturbation alters auditory feedback so that it does not match our internal predictions. Recent models of speech motor control postulate an auditory cortical mechanism for encoding this prediction error (29, 36–39) and can be viewed as special cases of predictive coding (40, 41) in which top-down predictions enable auditory regions to compute the error, which then is passed back to higher levels to refine the predictions. The prediction error is thought to be encoded by superficial pyramidal cells (42) that tend to fire and show spike-field coherence in the γ frequency range (43). A predictive coding account is compatible with our high-γ ECoG data and is consistent with a state-feedback model for speech motor control, and these speech models predict many of the results discussed here, such as the network of cortical areas activated during auditory feedback perturbation (ventral premotor cortex, ventral primary motor cortex, and pSTG) and the temporal sequence of cortical activity. However, the existing implementations of simple predictive coding models for speech implicitly assume that the prediction error is derived only from the motor-based predictions that underlie SIS—that is, that the enhancement of unexpected input (SPRE) depends on the colocalized suppression of expected input (SIS). This assumption is supported by the data of Eliades and Wang (15), who demonstrated in the marmoset that cortical suppression during vocalization acted to increase the sensitivity of single neurons to vocal feedback, implying a shared mechanism. In contrast, we found a decoupling between suppression and enhancement, with most modulated electrode sites exhibiting SIS or SPRE independently rather than both (Fig. 5). In addition, we provide evidence that compensation is tied to enhancement but not to suppression. A single mechanism based on the comparison of predicted and observed feedback cannot account for this dissociation of the two responses.

One possible explanation for SIS in the absence of SPRE is that the perceptual attributes of auditory input are encoded in functionally segregated sites; specifically, some sites that show SIS may code for prediction error in aspects of the acoustic signal that were not perturbed, such as loudness or timbre, and thus would not show enhanced responses to a perturbation in pitch. However, current models that use the same population of cells for suppression and enhancement would not explain the large number of cortical sites in the present study that displayed SPRE but not SIS. The dissociation of these responses may suggest that the two have distinct purposes: SIS for tagging sensations as self, and SPRE for detecting vocal error, including corrective commands to motor cortex.

Activity in speech premotor cortex was found to correlate with trial-by-trial compensation (Fig. 3 *B* and *C*), whether that compensation was achieved by raising or lowing the pitch of the voice. This correlation suggests that the premotor cortical activity underlies the corrective adjustment of output pitch and confirms and elaborates functional imaging studies implicating the left premotor cortex in pitch shift responses (11, 13). Similar to the auditory SPRE electrodes, these correlated motor sites also showed a greater response during speaking than listening (Fig. 3*C*). (We do not refer to this response as true "SPRE," because motor cortex is expected to be more active during speech.) Partial correlation analysis showed that auditory and motor electrodes contribute distinct components to the correlation with behavior. We speculate that auditory SPRE activity signals the corrective response and that somatosensory state, additive noise, and cortical and subcortical activity outside the range of our electrode grids might account for the independent motor component.

The correlations found in frontal premotor and posterior temporal areas are consistent with well-studied anatomical connections between these areas, most notably the arcuate fasciculus (44). Auditory and motor cortical areas also are functionally

Chang et al.

connected, as measured in vivo (45) and noninvasively during speech production (46). A recent study exploring phase synchrony between electrode sites in left inferior frontal gyrus and left pSTG found increased prespeech synchrony in subjects who exhibited greater SIS (47) and hypothesized that this synchrony was the neural instantiation of efference copy. It is plausible that this circuit is a two-way loop, enabling both the delivery of predictions to auditory cortex and the "reply" of consequent feedback mismatches to motor cortex. A functional imaging study has found evidence for the auditory-to-motor reply in the form of increased effective connectivity between these regions during an auditory perturbation (12) (although these connections were from the left pSTG to right-hemisphere motor regions). Although we cannot prove causality from these data, the following four points are consistent with a causal relationship: (*i*) the temporal sequence of postperturbation cortical activity begins with auditory cortex, which is followed by motor cortex activation and then by behavioral compensation; (*ii*) the cortical activation is correlated with compensation on a trial-by-trial basis; (*iii*) the time of maximum correlation precedes the peak compensation response; and (*iv*) the correlation occurs only when the neural signals are aligned to the peak behavioral response (not to the feedback perturbation). Taken together, these observations support the interpretation that auditory responses to perturbation act to signal motor areas that mediate compensation. In our example subject, the increase in high-γ activity starts at the STG and is followed by a significant motor increase ∼100 ms later (Fig. 2*C*), implying that the corrective motor commands are driven by the enhanced auditory detection of feedback error. Further analysis is needed to elucidate the role of auditory–motor feedback loops in vocal behavior, although caution in analyses of causality is needed, given the transient nature of the neural responses to perturbations (48).

A distinct experimental advantage of ECoG is the ability to record from multiple sites simultaneously in real time, in contrast to the sampling limitations of single-unit recordings and the temporal constraints of fMRI. Nonetheless, ECoG in this experiment also had specific limitations. First, the extent of grid coverage in humans was guided by the clinical indications for their epilepsy localization and always was done unilaterally. In some cases, the standard grid on the right hemisphere did not cover both auditory and motor regions, because clinical language mapping is not evaluated routinely in the nondominant hemisphere. Therefore we were limited in our interpretation of responses from right-hemisphere cortical sites. Second, the electrode contacts are limited to the gyral cortical surface and therefore do not sample intrasulcal, cerebellar, and subcortical areas of potential interest effectively. Despite these limitations, we were able to use directly recorded high-γ oscillations to reveal the specific auditory and motor components of the cortical network involved in vocal feedback.

In summary, we probed the neural circuitry underlying auditory feedback control in speech, using a pitch perturbation to elicit a specific compensatory pitch change. Here we report evidence of neural correlations with trial-by-trial compensation, showing a contributory role of both motor and auditory cortices. Furthermore, we present a cross-subject view of the spatial distribution of functional modulations (SIS and SPRE) as well as evidence that they differentially predict compensatory behavior. These results are evidence for the sensorimotor control of vocalization in humans through the dynamic coordination of multiple cortical areas.

## Methods

The experimental protocol was approved by the University of California, San Francisco institutional review boards and Committees on Human Research. Subjects gave their informed consent before testing.

**Subjects.** The nine subjects in this study underwent surgical placement of intracranial subdural grid electrodes as part of their surgical workup for epilepsy localization. Table S1 shows the patient characteristics included in this study. All subjects underwent neuropsychological language testing and were found to be normal. The Boston naming test and verbal fluency test were used for preoperative language testing. The Wada test was used for language dominance assessment. None of the subjects reported any speech or hearing problems.

Of the nine subjects run in the study, data from one subject (GP18) contained excessive artifacts in the electrode recording and were excluded from analysis. Data from another subject (GP34) were excluded because of a lack of any pitch perturbation response: With no evidence for a reaction to the perturbation, we could not be sure that the subject had heard the feedback coming from the headphones. As a result, seven subjects' data were included for analysis: four with grids implanted in the left hemisphere and three with grids implanted in the right hemisphere. Right-hemisphere coverage of the ventral premotor and auditory cortex was limited (e.g., 54 electrodes in the right temporal cortex vs. 156 in the left temporal cortex).

**Apparatus.** The experimental apparatus consisted of a DSP, a laptop PC, a computer monitor, and a headphone–microphone headset. A microphone picked up the subject's speech and passed it to the DSP, which altered the pitch of the subject's speech in real time (12-ms feedback delay) and fed the altered speech back to the subject via the headphones. The pitch alteration process was based on the method of sinusoidal synthesis developed by McAulay and Quatieri (49). The laptop PC controlled the triggering of the DSP and the prompts for the subject to speak, shown on the monitor.

**Procedure.** The experiment consisted of a speaking condition and a listening condition, each lasting 74 trials (four blocks of 15 trials each and a final block of 14 trials). In the speaking condition, subjects phonated the vowel /a/ for roughly 3.5 s. At a random latency (1,325–1,800 ms) from the signal to begin vocalizing, the DSP perturbed the pitch of the auditory feedback by ± 200 cents (i.e., two semitones) for 400 ms. A single perturbation occurred in each trial, and equal numbers of positive and negative perturbations were distributed randomly across the 74 trials. The subjects were not explicitly instructed to maintain their pitch. In the subsequent listening condition, subjects passively listened to playback of the audio feedback they had heard during the speaking condition. We excluded trials in which the perturbation occurred less than 400 ms after the subject began vocalizing.

The electrocorticogram was recorded using a variety of multichannel subdural cortical electrode arrays. The position of the electrodes was determined exclusively by clinical criteria. The signal was recorded with a multichannel amplifier optically connected to a digital data acquisition system (Tucker-Davis Technologies) sampling at 3,052 Hz. Audio data also were recorded on this system in synchronization with the ECoG data.

**Data Analysis.** *Audio analysis.* To assess behavioral responses to the feedback perturbation, pitch-tracking analysis was performed on each subject's audio data. Voice onset was determined using the same threshold procedure for trials from the speak and listen conditions. Perturbation onset and offset were determined via an indicator signal that was output by the DSP and recorded on the ECoG data acquisition system. Pitch was estimated using the standard autocorrelation method (50). Trials with erroneous pitch tracks caused by excessive pitch variation were removed.

Mean percent compensation was calculated as −100*(cents peak response change/cents perturbation), with the minus sign introduced to make compensation a positive value.

*Compensation estimation.* Compensation for each trial was estimated by cross-correlation analysis. Each trial's pitch track was cross-correlated with the mean compensation response, and the latency of the peak cross-correlation was used to estimate the timing of that trial's compensation response relative to the mean response time. Compensation for the trial was estimated by comparing the magnitude of the peak cross-correlation with the magnitude of the peak of the mean response's autocorrelation: The ratio of the two magnitudes gave the fraction of the mean compensation that represented compensation on that trial.

*ECoG data analysis.* Trials in which any of the electrodes showed artifacts or excessive noise were removed. ECoG data were preprocessed and bandpass-filtered into 45 separate frequency bands, logarithmically spaced to cover the frequency range from 1–150 Hz. Each band then was Hilbert transformed to extract the time course of the amplitude envelope in that band. The spectrogram plots of Fig. 2 were created with these band envelope data. Finally, each band envelope time course was smoothed with a 100-ms boxcar kernel, then converted to *z* scores using the mean and variance of trial data in a baseline window extending from 1.5–1.0 s before voice onset. The normalized band envelopes then were analyzed using three alignments of the neural data: voice onset (SIS), perturbation onset (SPRE), and compensation peak (neural–behavioral correlation).

**SIS.** To calculate SIS, trials were time-aligned to voice onset. SIS was defined as the difference in the mean $z$-scored trial data between the two experimental conditions (listen − speak). Significance was calculated from a one-way ANOVA, using a $P$ value threshold determined to set the false-discovery rate (FDR) over all significance tests to less than 5%. In determining the overall SIS exhibited by each electrode, only the data up to 300 ms after voice onset were considered. Within this interval, the total SIS of an electrode then was calculated as the sum over time points of the SIS values. For the purposes of the classification analyses shown in Fig. 5, an electrode was classified as exhibiting SIS if there was at least one time point in the analysis interval that showed significant SIS (FDR-corrected $P < 0.05$).

**SPRE.** To calculate SPRE, trials were time-aligned to pitch perturbation onset. The $z$ scores were calculated from a baseline window extending from 0.4–0.1 s before perturbation onset. SPRE was defined as the difference in the mean $z$-scored trial data between the two experimental conditions (speak − listen) in a manner similar to that done to calculate SIS (FDR-corrected $P < 0.05$). In determining the overall SPRE exhibited by each electrode, only the data from 50–550 ms after perturbation onset were considered; this interval is when responses to the perturbation were expected, based on previous studies (23). Within this interval, the total SPRE of an electrode then was calculated as the sum over time points of the SPRE

values. For the purposes of classification analyses, an electrode was classified as exhibiting SPRE if there was at least one time point in the analysis interval that showed significant SPRE and at least one time point where the response in the speak condition was significantly different from zero.

**Correlation.** To determine the trial-by-trial correlation between grid electrode activity and compensation, electrode activity was time-aligned to the subjects' compensation responses (rather than to perturbation onset) and compared with the compensation value for each trial. To examine differences in correlation scores between different classes of electrodes and multiple subjects, the correlation for each electrode was Fisher $z$-transformed and then used as the dependent variable in a three-way ANOVA, with SIS, SPRE, and subject as the categorical independent variables. A one-way ANCOVA was applied to variables in Fig. 4B, with SPRE as a predictor variable, compensation as the dependent variable, and subject as the categorical group variable.

1. Lane H, Tranel B (1971) The Lombard sign and the role of hearing in speech. *J Speech Lang Hear Res* 14(4):677–709.
2. Lombard E (1911) Le signe de l'élévation de la voix. *Ann Maladies Oreille, Larynx, Nez. Pharynx* 37(2):101–119.
3. Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103(6):3153–3161.
4. Elman JL (1981) Effects of frequency-shifted feedback on the pitch of vocal productions. *J Acoust Soc Am* 70(1):45–50.
5. Jones JA, Munhall KG (2000) Perceptual calibration of F0 production: Evidence from feedback perturbation. *J Acoust Soc Am* 108(3 Pt 1):1246–1251.
6. 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(4):2363–2371.
7. Heinks-Maldonado TH, Houde JF (2005) Compensatory responses to brief perturbations of speech amplitude. *ARLO* 6(3):131–137.
8. Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. *Science* 279(5354):1213–1216.
9. Purcell DW, Munhall KG (2006) Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119(4):2288–2297.
10. Shiller DM, Sato M, Gracco VL, Baum SR (2009) Perceptual recalibration of speech sounds following speech motor learning. *J Acoust Soc Am* 125(2):1103–1113.
11. Toyomura A, et al. (2007) Neural correlates of auditory feedback control in human. *Neuroscience* 146(2):499–503.
12. Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39(3):1429–1443.
13. Zarate JM, Zatorre RJ (2008) Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 40(4):1871–1887.
14. Eliades SJ, Wang X (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol* 89(4):2194–2207.
15. Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453(7198):1102–1106.
16. Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: An MEG study. *J Cogn Neurosci* 14(8):1125–1138.
17. Flinker A, et al. (2010) Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci* 30(49):16643–16650.
18. Greenlee JDW, et al. (2011) Human auditory cortical activation during self-vocalization. *PLoS ONE* 6(3):e14744.
19. Crone NE, et al. (2001) Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* 57(11):2045–2053.
20. Aliu SO, Houde JF, Nagarajan SS (2009) Motor-induced suppression of the auditory cortex. *J Cogn Neurosci* 21(4):791–802.
21. Blakemore S-J, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. *Nat Neurosci* 1(7):635–640.
22. Blakemore SJ, Wolpert D, Frith C (2000) Why can't you tickle yourself? *Neuroreport* 11 (11):R11–R16.
23. Behroozmand R, Karvelis L, Liu H, Larson CR (2009) Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin Neurophysiol* 120(7):1303–1312.
24. Chang EF, et al. (2011) Cortical spatio-temporal dynamics underlying phonological target detection in humans. *J Cogn Neurosci* 23(6):1437–1446.
25. Edwards E, et al. (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* 50(1):291–301.
26. Ray S, Maunsell JHR (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9(4):e1000610.
27. Steinschneider M, Fishman YI, Arezzo JC (2008) Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb Cortex* 18(3):610–625.
28. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7(7):701–702.
29. Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69(3):407–422.
30. Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat Neurosci* 12(6):718–724.
31. Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp* 9(4):183–191.
32. Eliades SJ, Wang X (2005) Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb Cortex* 15(10):1510–1523.
33. Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. *Science* 269(5232):1880–1882.
34. Bell C, Bodznick D, Montgomery J, Bastian J (1997) The generation and subtraction of sensory expectations within cerebellum-like structures. *Brain Behav Evol* 50(Suppl 1): 17–31.
35. Von Holst E, Mittelstaedt H (1950) The reafference principle: Interaction between the central nervous system and the periphery. *Naturwissenschaften* 37:464–476.
36. 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(3):862–874.
37. Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96(3):280–301.
38. Ventura MI, Nagarajan SS, Houde JF (2009) Speech target modulates speaking induced suppression in auditory cortex. *BMC Neurosci* 10:58.
39. Houde JF, Nagarajan SS (2011) Speech production as state feedback control. *Front Hum Neurosci* 5:82.
40. Todorov E (2008) General duality between optimal control and estimation, in *47th IEEE Conference on Decision and Control* (Cancun, Mexico), pp 4286–4292.
41. Friston K (2011) What is optimal about motor control? *Neuron* 72(3):488–498.
42. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66(3):241–251.
43. Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. *Proc Natl Acad Sci USA* 108(27): 11262–11267.
44. Glasser MF, Rilling JK (2008) DTI tractography of the human brain's language pathways. *Cereb Cortex* 18(11):2471–2482.
45. Matsumoto R, et al. (2004) Functional connectivity in the human language system: A cortico-cortical evoked potential study. *Brain* 127(Pt 10):2316–2330.
46. Simonyan K, Ostuni J, Ludlow CL, Horwitz B (2009) Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *J Neurosci* 29(47):14912–14923.
47. Chen C-MA, et al. (2011) The corollary discharge in humans is related to synchronous neural oscillations. *J Cogn Neurosci* 23(10):2892–2904.
48. Wang X, Chen Y, Ding M (2008) Estimating Granger causality after stimulus onset: A cautionary note. *Neuroimage* 41(3):767–776.
49. McAulay R, Quatieri T (1986) Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans Acoust Speech Signal Process* 34:744–754.
50. Parsons TW (1987) *Voice and Speech Processing* (McGraw-Hill, Blacklick, OH).