Behavioral/Cognitive

What Does Motor Efference Copy Represent? Evidence from Speech Production

Caroline A. Niziolek,[1] **Srikantan S. Nagarajan,**[2] **and John F. Houde**[1]

Departments of [1]Otolaryngology and [2]Radiology, University of California, San Francisco, San Francisco, California 94143

How precisely does the brain predict the sensory consequences of our actions? Efference copy is thought to reflect the predicted sensation of self-produced motor acts, such as the auditory feedback heard while speaking. Here, we use magnetoencephalographic imaging (MEG-I) in human speakers to demonstrate that efference copy prediction does not track movement variability across repetitions of the same motor task. Specifically, spoken vowels were less accurately predicted when they were less similar to a speaker's median production, even though the prediction is thought to be based on the very motor commands that generate each vowel. Auditory cortical responses to less prototypical speech productions were less suppressed, resembling responses to speech errors, and were correlated with later corrective movement, suggesting that the suppression may be functionally significant for error correction. The failure of the motor system to accurately predict less prototypical speech productions suggests that the efferent-driven suppression does not reflect a sensory prediction, but a sensory goal.

Introduction

The brain deals in predictions and is especially good at predicting the sensory consequences of well-practiced actions. Motor cortex is thought to initiate these predictions by generating an internal copy of its output, termed "efference copy" (Sperry, 1950; von Holst and Mittelstaedt, 1950), that alerts sensory cortices to upcoming feedback, changing their response properties. For example, studies of evoked potentials consistently show suppressed auditory responses to self-generated speech compared with playback of the same speech signal (Creutzfeldt et al., 1989; Curio et al., 2000; Flinker et al., 2010; Greenlee et al., 2011). This suppression is thought to reflect a partial neural cancellation of incoming sensory feedback as it is matched to the motor prediction (Bell et al., 1997; Poulet and Hedwig, 2003, 2006). Indeed, suppression is abolished when the incoming feedback has been altered to create a mismatch, as in the case of real-time auditory perturbation studies (Houde et al., 2002; Eliades and Wang, 2008; Behroozmand and Larson, 2011; Chang et al., 2013). Motor-induced suppression via an efference copy mechanism has been demonstrated across species and sensory domains (Zaretsky and Rowell, 1979; Blakemore et al., 1998; Crapse and Sommer, 2008; Cullen et al., 2011).

Given the term "copy," efference copy is widely viewed as a perfect reproduction of motor commands, and the resultant prediction (sometimes called "corollary discharge") is widely presumed to represent the precise sensory consequences of each motor act. However, not all motor acts have the consequences we intend, and the same intent can lead to variable movements. The degree to which sensory cortex "expects" this variation is still unknown. How well does efference copy, and the sensory prediction it generates, encode motor output variability? Because motor-induced suppression is thought to reflect the subtraction of predicted feedback from observed feedback, we can use it to characterize the internal sensory prediction: the greater the suppression in auditory cortex, the better the match between the two signals.

In this study, we used magnetoencephalographic imaging (MEG-I) to examine whether the accuracy of the efference copy prediction varied over repeated productions of a given vowel. Subjects produced randomized repetitions of three different vowels (speak condition); this task alternated with a nonspeaking condition in which subjects listened to playback of their utterances (listen condition). The peak auditory evoked response (M100) to subjects' own speech was subtracted from the peak response to playback of the same acoustic stimuli (listen − speak) to yield the magnitude of speaking-induced suppression (SIS). We compared the SIS of the productions nearest to each vowel's median formants with the SIS of the more outlying productions. If the efference copy prediction takes utterance-to-utterance variance into account, then SIS will not change across vowel space: every prediction will be equally accurate. However, if the prediction reflects a prototypical target at the center of the vowel distribution, then SIS will be attenuated at the vowel periphery, where the feedback least matches the prediction. By distinguishing between these alternatives, we aimed to better characterize the nature of the feedback comparison process in natural speech.

Received May 17, 2013; revised Aug. 5, 2013; accepted Sept. 1, 2013.

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

This work was supported by the National Institutes of Health Grants R01-DC010145, R21 NS 076171, and R01 DC004855, National Science Foundation Grant BCS-0926196, and National Research Service Award Grant F32-DC011249. We thank Julia Owen and Phiroz Tarapore for help with source localization analysis, Jeevit Gill for assistance in running subjects, and Susanne Honma and Danielle Mizuri for guidance in collecting the MEG data.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Caroline A. Niziolek, University of California, San Francisco, 513 Parnassus Avenue, Box 0628, San Francisco, CA 94143. E-mail: cniziolek@ohns.ucsf.edu.

DOI:10.1523/JNEUROSCI.2137-13.2013

Materials and Methods

Procedure

Fourteen subjects (eight female) with self-reported normal hearing and speech participated in the experiment. All experimental procedures were approved by the Institutional Review Board at the University of California, San Francisco. Subjects underwent MEG-I in a 275-channel CTF Omega 2000 whole-head biomagnetometer (VSM MedTech). In the MEG scanner, subjects produced 200 randomized tokens each of the words "eat," "Ed," and "add," chosen to elicit the vowels /i/, /ɛ/, and /æ/, while receiving auditory feedback of their own voices through insert earphones (speak condition). Subjects were instructed to minimize jaw movement during speech, although not at the expense of vowel quality. The speak condition was divided into two blocks, each of which was followed by a block in which the subjects heard recordings of these self-productions (listen condition). MEG traces were aligned to vowel onset in both conditions. A high-resolution T1-weighted anatomical volume was also obtained in a separate MRI session to coregister each subject's MEG source activity to a structural image of his or her own brain. Four subjects with speech-related motor artifacts who exhibited no clear M100 across all 600 speaking trials (S02, S05, S11, S14) were excluded. These subjects did exhibit normal M100 responses during passive listening trials; however, because our analysis compared the peak of the response across speaking and listening conditions, we could not assess our hypothesis in these four subjects.

Acoustic analysis

The first and second vowel formants (F1 and F2) of all recorded speech trials were tracked in mels (O'Shaughnessy, 1987), a perceptually based logarithmic frequency scale chosen because it reflects the ear's sensitivities to changes in different regions of frequency space. Formants were averaged over the first 50 ms of each production in the speak condition. This time window was chosen to ensure that only auditory information that could contribute to the M100 (i.e., that occurred before the auditory evoked response) was used to determine each trial's location in formant space. For each subject, median formants were calculated for each of the three spoken vowels, and the "center" and "peripheral" trials were, respectively, defined as the closest and farthest 100 trials with respect to their vowel's median as defined by the Euclidean distance in 2D mel frequency space (see Fig. 1*A*). This analysis subdivided the trials of each condition, yielding two speaking trial types, speak-center (speak_c) and speak-peripheral (speak_p), and two listening trial types, listen-center (listen_c) and listen-peripheral (listen_p), which consisted of the same acoustic stimuli as the two speaking trial types.

As shown in Figure 1*A*, the distributions of these two trial types were not equal: peripheral trials were spread out across frequency space, whereas center trials were more tightly clustered. To control for this difference in distributional variance (Herrmann et al., 2013), we carried out the same analysis using two different acoustic parameters, loudness and pitch, to subdivide trials into center and peripheral groups. These parameters act as controls because they were unrelated to the vowel production task. If differences in neural suppression between center and periphery were simply the result of differences in acoustic variance, these differences should also emerge in the two control conditions. For the loudness control, root mean square amplitude was tracked and averaged over the first 50 ms of each production in the speaking condition. For each subject, median amplitude was calculated for each of the three spoken vowels, and the center-loudness and peripheral-loudness trials were defined as the closest and farthest 100 trials with respect to this median. This analysis yielded four control conditions: speak-center-loudness ($\text{speak}_\text{c-dB}$), speak-peripheral-loudness ($\text{speak}_\text{p-dB}$), listen-center-loudness ($\text{listen}_\text{c-dB}$), and listen-peripheral-loudness ($\text{listen}_\text{p-dB}$). For the pitch control, fundamental frequency was tracked using an autocorrelation method and averaged over the first 50 ms of each production in the speaking condition. For each subject, median fundamental frequency was calculated for each of the three spoken vowels, and the center-pitch and peripheral-pitch trials were here defined as the closest and farthest 100 trials with respect to this median. This analysis yielded another four control conditions: speak-center-pitch ($\text{speak}_\text{c-f0}$), speak-peripheral-pitch ($\text{speak}_\text{p-f0}$), listen-center-pitch ($\text{listen}_\text{c-f0}$), and listen-peripheral-pitch ($\text{listen}_\text{p-f0}$). As pitch and loudness are normally distributed across speech productions, the acoustic distributional properties of center-loudness and center-pitch trials are much like those of formant-defined center trials, with a tighter clustering compared with trials at the periphery. Thus, the control analyses tested whether distributional properties of the sounds alone could result in differences in M100 suppression. We also tested for amplitude and pitch differences between our formant-defined center and peripheral groups using two-way ANOVAs with factors of subject and trial type (speak_c vs speak_p).

Modulatory signals such as efference copies have been theorized to enable self-monitoring for error detection. To test whether the accuracy of the efference copy prediction, as measured by SIS, has consequences for online vocal correction, we carried out an acoustic analysis to characterize formant movement during single trials. Specifically, we hypothesized that SIS might reflect a process underlying vowel "centering" (i.e., a corrective movement that causes a peripheral utterance to move closer to the center of the formant distribution). To assess centering during the course of speaking trials, we compared the centricity of formant values at the beginning to that at the middle of each trial. First, the initial distance to the median d_{init} was calculated as described above, using the average of the first 50 ms of each trial and measuring the Euclidean distance to a median derived from this time interval; that is, $d_{init} = \sqrt{(F1_{init} - median(F1_{init}))^2 + (F2_{init} - median(F2_{init}))^2}$. The mid-trial median distance d_{mid} was calculated in the same manner but used formant values averaged from the middle 50% of each trial and a median derived from this later time interval; that is, $d_{mid} = \sqrt{(F1_{mid} - median(F1_{mid}))^2 + (F2_{mid} - median(F2_{mid}))^2}$. Finally, the centering for each trial was defined as $C = d_{init} - d_{mid}$, such that positive values indicate a larger initial distance to the median and a smaller mid-trial distance to the median (i.e., movement toward the median over the course of the utterance). The centering for each trial was used as the dependent variable in a two-way ANOVA with factors of subject and trial type, a binary variable representing whether the trial was a center or peripheral trial. To ensure that differences in centering between center and peripheral trials were not merely the result of regression to the mean, we also carried out a two-way ANOVA using the absolute Euclidean distance between the starting formants and the mid-trial formants for each trial, $d_{Eucl} = \sqrt{(F1_{init} - F1_{mid})^2 + (F2_{init} - F2_{mid})^2}$, as the dependent variable; furthermore, we tested whether the average distance to the median over all trials varied from the beginning to the middle of the trial, regardless of trial type (two-way ANOVA with factors of subject and trial interval: *init* or *mid*).

MEG-I analysis

Source localization. First, the source localization algorithm Champagne (Owen et al., 2012) was used to compute source strength at each voxel in the brain. MEG sensor data were third-order gradient denoised, detrended, and filtered from 4 to 40 Hz. The 4 Hz high-pass cutoff was used to filter out low-frequency movement-related artifacts during speech, improving detection of the M100. The sensor data for all trials in the listen condition were averaged together to create a listen average. From this average, a three-component lead field was generated using the NUTMEG analysis toolbox (Dalal et al., 2004), which calculates a forward model of sensor activity given a spatially normalized MRI for each subject, coregistered using fiducial markers. The source activity was run through Champagne to compute sensor weights at each 8 mm voxel in the brain. For each subject, weights were extracted for the peak voxel in each hemisphere (Fig. 1*B*), determined by activity strength in a window around the M100 response (50–150 ms after stimulus onset). These weights, multiplied by the sensor data, gave the estimation of signal strength at the peak voxel in each hemisphere. We call these signals the "source space" data, as they represent the strength of the dipole source that generates a pattern on the sensors.

Measuring cortical suppression. In alignment with past studies (Curio et al., 2000; Ventura et al., 2009), SIS was defined as a reduction in the amplitude of the M100 response to spoken vowels compared with playback. The M100 was the largest, most robust component of the evoked response. For each hemisphere, we separately averaged the source space data for both center and peripheral trials in both speak and listen condi-

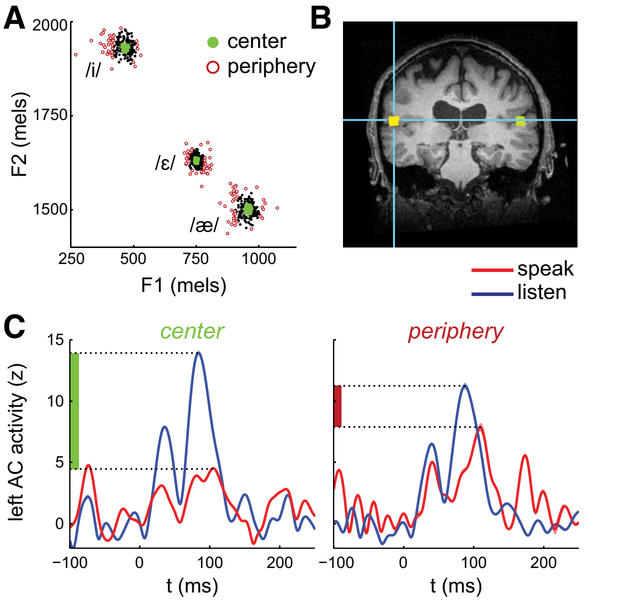

Figure 1. Center versus peripheral vowel productions in a sample subject. ***A***, Acoustic variation across repeated productions, shown in 2D formant frequency space. Green represents center productions; red represents peripheral productions; black represents remaining productions. ***B***, A source localization algorithm (Owen et al., 2012) determined the coordinates and field strength of the M100 peak (MNI −56, −24, 8 in this subject). ***C***, MEG traces aligned to vowel onset, separated into center and peripheral trials as determined by ***A***. Shaded regions represent SE ($n = 100$); vertical bars on the y-axis represent SIS magnitude.

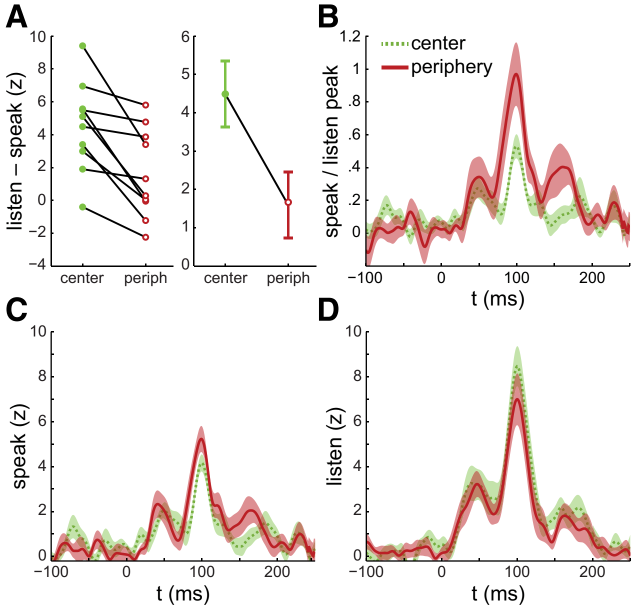

Figure 2. Modulation of SIS in the left hemisphere. ***A***, Left, SIS, defined as the difference between the M100 peaks in listen and speak conditions, separated into center and peripheral trials ($p = 0.002$). Each linked pair of points represents data from a single subject. Right, Mean SIS across all subjects for center and peripheral trials. Error bars indicate SE. ***B***, Activity in the speak condition normalized by the listen condition, averaged across subjects. ***C***, Activity during the speak condition averaged across subjects. ***D***, Activity during the listen condition averaged across subjects. All activity plots are aligned to acoustic onset ($t = 0$) and linearly scaled such that the M100 peaks across subjects are also aligned ($t = 100$ ms). Shaded regions represent SE for a random-effects analysis ($n = 10$).

tions. These average MEG traces were z-scored using the activity from a baseline of −300 to −100 ms relative to stimulus onset, calculated for each condition within each subject, to normalize baseline variability. The M100 peak in each condition (speak$_c$, speak$_p$, listen$_c$, and listen$_p$) was then defined as the peak activity between 75 and 150 ms after stimulus onset; peaks were confirmed by visual inspection. SIS in each hemisphere was calculated by taking the difference in peak amplitude between the listen and speak trials ($\text{SIS}_c = \text{speak}_c - \text{listen}_c$; $\text{SIS}_p = \text{speak}_p - \text{listen}_p$). The same analysis was carried out for the four loudness control conditions and the four pitch control conditions.

Results

The vowel production space of a sample subject (S01) is shown in Figure 1*A*. Each point represents an individual vowel production. A spread of formant values is apparent across multiple productions, with those closest to each vowel's median colored in green ("center") and those farthest colored in dark red ("periphery"). The auditory cortical responses to these two classes of speech stimuli were compared using MEG-I: the amplitude of the M100 source response at the peak voxel in each hemisphere (Fig. 1*B*) was z-scored relative to a prespeech baseline window and compared across speak and listen conditions for both center and peripheral trials.

Consistent with past studies, subjects exhibited SIS: the M100 peak was suppressed in the speak condition relative to the listen condition, especially in the left hemisphere. An example of left-hemisphere SIS in a single subject (S01) is shown in Figure 1*C*, subdivided into center (left) and peripheral (right) trials. For trials near the center of the formant distribution, the difference between z-scored speak and listen peaks is >9 SDs, or 65% of the listen peak; whereas for trials at the periphery, the difference falls off to 3 SDs, or 20% of the listen peak. These magnitudes are consistent with a mean SIS falling between 20% and 65% in past studies that averaged together the entire distribution of utterances (Curio et al., 2000; Houde et al., 2002; Ventura et al., 2009; Behroozmand and Larson, 2011) (51%, 39%, 30%, and 58%, respectively).

Every subject in the experiment showed this decrease in SIS from center to peripheral trials in the left hemisphere (Fig. 2*A*); however, this was not the case in the right hemisphere (Fig. 3*A*). A three-way ANOVA with factors of subject, hemisphere, and trial type (center vs peripheral) found a main effect of trial type ($F_{(1,9)} = 5.45, p = 0.045$) and of subject ($F_{(1,9)} = 9.09, p = 0.002$), as well as an interaction between trial type and hemisphere ($F_{(1,9)} = 14.94$, $p = 0.004$). We examined the cortical responses in each hemisphere separately to assess the differential changes in SIS. In the left hemisphere, the difference between center and periphery was robust (two-tailed paired t test, $n = 10$, $p = 0.002$). This difference remained significant when the one participant who did not show robust SIS was excluded ($p = 0.004$). The group average MEG signal from the speak activity, normalized by peak listen response, is shown in Figure 2*B*. Changes in both speak and listen conditions contributed to the change in suppression: the M100 in the speak condition was greater in peripheral trials ($p = 0.038$; Fig. 2*C*), and the M100 in the listen condition was greater in center trials ($p = 0.030$; Fig. 2*D*).

In contrast to the left hemisphere, identical analysis of activity in the right hemisphere did not yield significant changes in either SIS ($p = 0.464$; Fig. 3*A*,*B*), speak activity ($p = 0.191$; Fig. 3*C*), or listen activity ($p = 0.891$; Fig. 3*D*) between center and peripheral trials. One potential basis for a hemisphere-specific effect is the weak right hemisphere source amplitude, significantly smaller than the source amplitude in the left (two-tailed paired t test, $n = 10$, $p = 0.038$); there was also significantly smaller SIS in the right hemisphere (two-tailed paired t test, $n = 10$, $p = 0.036$), as has been reported previously (Curio et al., 2000; Houde et al., 2002).

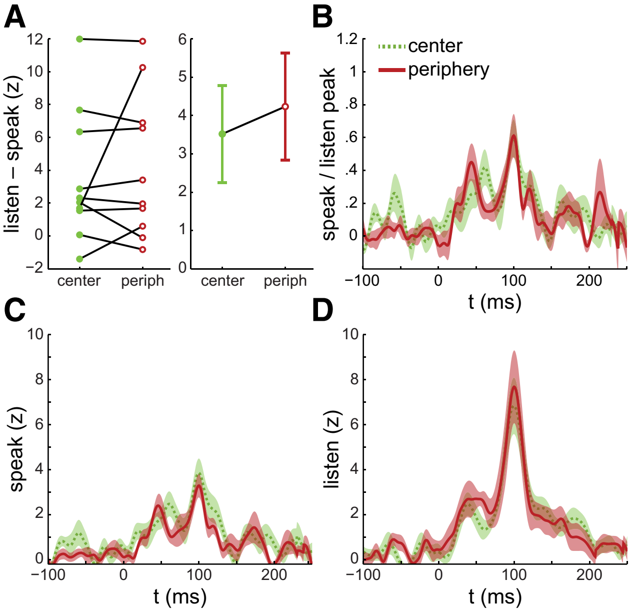

Figure 3. Modulation of SIS in the right hemisphere. ***A***, Left, SIS separated into center and peripheral trials, as in Figure 2. ***B***, Activity in the speak condition normalized by the listen condition, averaged across subjects. ***C***, Activity during the speak condition averaged across subjects. ***D***, Activity during the listen condition averaged across subjects. For all activity plots, scaling and error bars are as in Figure 2.

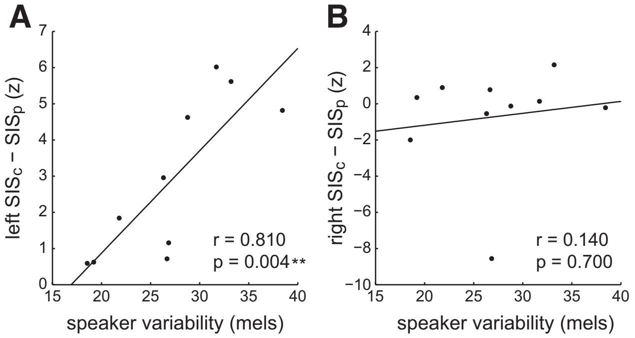

Figure 4. Speech production variability predicts modulation of SIS. Correlation of per-speaker formant variability and SIS fall-off (center − periphery) in left (***A***) and right (***B***) hemispheres.

Hemispheric differences were also found in M100 latency effects across conditions. Interestingly, only in the right hemisphere, there was a significantly earlier M100 in the center speak trials than in the peripheral speak trials (two-tailed paired *t* test, $n = 10$, $p = 0.023$; center latency mean = 99.3 ± 4.6 ms; peripheral latency mean = 112.8 ± 3.7 ms); there was no significant latency difference in the listen condition (center latency mean = 85.3 ± 3.5 ms; peripheral latency mean = 103.4 ± 6.4 ms). No differences in M100 latency were found in the left hemisphere, either between center and peripheral trials or between speak and listen conditions; all mean latencies in the left hemisphere fell between 94.0 and 98.7 ms.

Because center trials were more tightly clustered than peripheral trials (Fig. 1*A*), vowel formants in the region of the median were more frequently heard than those in any other same-sized acoustic region. To ensure that the observed differences in SIS were indeed the result of the deviation from the median formant production and not of other possible acoustic confounds, we carried out the same analysis using speech amplitude (loudness) and fundamental frequency (pitch), instead of formant frequency, to subdivide trials into center and peripheral groups. Amplitude and fundamental frequency over the first 50 ms of voicing were normally distributed (Kolmogorov–Smirnov test); across subjects, amplitude had a mean range of 2.8 dB, whereas fundamental frequency had a mean range of 136 mels. No difference in SIS was observed between trials near a subject's median amplitude (vowels produced with average loudness) and those at the edges of the amplitude distribution (the quietest and loudest vowels) (two-way ANOVA, no main effect of trial type or hemisphere; main effect of subject: $F_{(1,9)} = 11.69$; $p < 0.001$). Similarly, no difference in SIS was observed between trials near a subject's median fundamental frequency (vowels produced with average pitch) and those at the edges of the frequency distribution (the highest- and lowest-pitched vowels) (two-way ANOVA, no main effect of trial type; main effect of hemisphere; $F_{(1,9)} = 7.39$; $p = 0.024$; main effect of subject: $F_{(1,9)} = 17.93$; $p < 0.001$; interaction between subject and hemisphere: $F_{(1,9)} = 5.5$; $p = 0.009$). A two-way ANOVA found no differences in speech amplitude ($F_{(1,9)} = 1.31$, $p = 0.253$) or fundamental frequency ($F_{(1,9)} = 0.7$, $p = 0.404$) between center and peripheral trials as defined by formant frequency.

Furthermore, the spread of formant values varied across subjects; some speakers were more variable in their utterances than others. Speakers with a larger formant spread (measured as the average mel distance-to-median of all trials) showed a larger left-hemisphere decrease in SIS between center and peripheral trials (Fig. 4*A*; Pearson's correlation, $n = 10$, $r = 0.810$, $p = 0.004$; Spearman's rank correlation, $n = 10$, $r = 0.855$, $p = 0.004$), illustrating that speakers with greater variability in their speech show greater changes in their left-hemisphere auditory cortical suppression. This relationship was not significant in the right hemisphere (Fig. 4*B*; Pearson's correlation, $n = 10$, $r = -0.263$, $p = 0.464$; Spearman's rank correlation, $n = 10$, $r = 0.139$, $p = 0.707$); a two-way ANOVA with hemisphere as a categorical factor and speaker variability as a continuous factor showed no interaction between these factors in their prediction of changes in SIS ($F_{(1,16)} = 1.45$, $p = 0.246$).

Finally, because center and peripheral trials evoked different responses in auditory cortex, we examined subsequent changes in acoustic output that may have been caused by the cortical response. For each trial, we compared the formant frequencies at the beginning of the utterance (first 50 ms) with those at the middle of the utterance (middle 50%) to give a measure of how much acoustic change occurred over the course of a single vowel production. Peripheral trials underwent more acoustic change than center trials as measured by Euclidean distance in format space (two-way ANOVA with factors of subject and trial type, $F_{(1,9)} = 53.07$, $p < 0.001$), moving an average of 51 mels (92 Hz) from their starting point, compared with 42 mels (73 Hz) for center trials. We additionally examined the direction of this movement with respect to the median. A two-way ANOVA revealed a main effect of trial type: peripheral trials, but not center trials, underwent a "centering" from the beginning to the middle of the utterance; that is, the distance to the median decreased over time by an average of 18 mels, or 33 Hz ($F_{(1,9)} = 877.87$, $p < 0.001$; Fig. 5*A*). The centering behavior was not merely the result of regression to the mean, as the average distance from the median for all trials (not just peripheral trials) also decreased over time (two-way ANOVA, $F_{(1,9)} = 12.1$, $p < 0.001$); that is, peripheral centering was not canceled out by outward movements in center trials. Furthermore, centering in peripheral trials was cor-

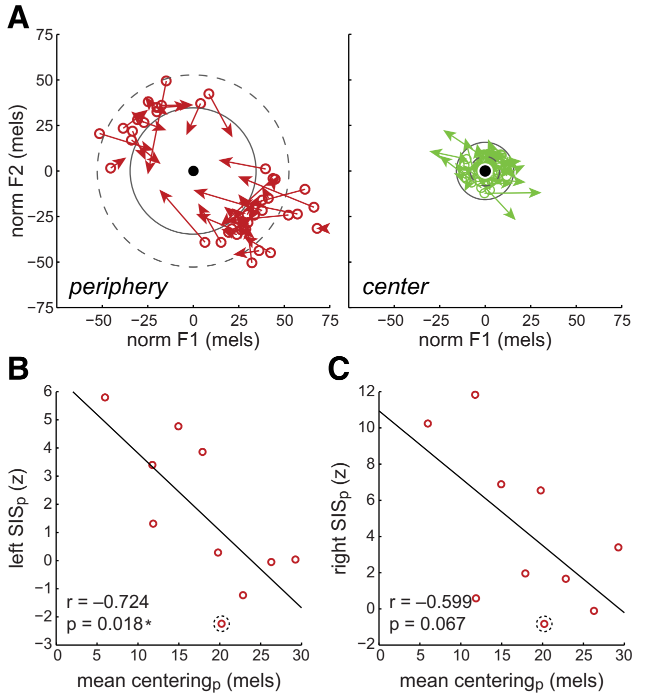

Figure 5. SIS predicts subsequent corrective behavior. ***A***, Vowel production data from a single subject demonstrating formant changes in individual trials from the first 50 ms of vocalization (open circles) to mid-utterance (arrowheads). The radii of the gray circles represent the mean distance to median (filled black circle) during the first 50 ms (dashed line) and mid-utterance (solid line), showing an average corrective movement toward the median in the peripheral trials. ***B***, ***C***, Per-subject correlations between SIS in peripheral trials and subsequent corrective changes in acoustic output in those peripheral trials for the left (***B***) and right (***C***) hemispheres. The circled data point in each plot indicates the subject shown in ***A***.

related with the amount of SIS on a subject-by-subject basis (Fig. 5*B*,*C*; Pearson's correlation in the left hemisphere: $r = -0.724$, $p = 0.018$; in the right hemisphere: $r = -0.599$, $p = 0.067$, not significant; Spearman's rank correlation in the left hemisphere: $r = -0.782$, $p = 0.012$; in the right hemisphere: $r = -0.588$, $p = 0.080$, not significant). In other words, the smaller the neural suppression in peripheral trials for a given subject, the more that subject subsequently "corrected" the acoustic output during those peripheral trials, making it less peripheral (closer to the median). In center trials where there was no corrective action, there were no significant correlations between amount of centering and SIS in the left ($r = -0.276$, $p = 0.440$) or right ($r = -0.039$, $p = 0.914$) hemispheres. We found no interaction between the correlations in the two hemispheres (two-way ANOVA with SIS as a dependent variable, hemisphere as a categorical factor, and centering as a continuous factor; main effect of centering: $F_{(1,16)} = 10.44$, $p = 0.005$; interaction of hemisphere and centering: $F_{(1,16)} = 0.24$, $p = 0.631$).

Discussion

Cortical responses to incoming sensory feedback are modulated by the motor system via an efference copy prediction of that feedback. Is this prediction precise enough to take into account the variability in sensory outcomes that is inherent to repeated motor acts? We used the SIS of auditory cortical responses to characterize the internal prediction of auditory feedback. The present results show that sensory predictions do not accurately track feedback variability across repetitions of the same motor task. Specifically, speaking generated less auditory suppression when vowels were farther from a speaker's median production. The decreased suppression in these peripheral trials suggests a decreased overlap between the prediction and the observed feedback (i.e., a bigger prediction error), even though efference copy is thought to be based on the very motor commands that generated the peripheral vowels. Furthermore, this prediction error was strongly correlated with subsequent corrective changes in speech output, providing the first evidence that SIS may have behavioral consequences for speakers.

These findings have important implications for the nature of efference copy signals and the representations they evoke in sensory cortex. First, the efference copy signals responsible for motor-induced suppression are not always well matched to the sensory consequences of movement: they do not predict all of the variability in the acoustic output. There are several explanations for the lack of precision in the prediction. It is possible that the central efference copy itself is more or less invariant for a given motor act in a fixed context, causing sensory cortex to always expect the desired or prototypical consequences of that invariant act. This is consistent with an efference copy that reflects a motor plan, not outgoing motor commands, and that is generated upstream of primary motor cortex, in supplementary motor (Haggard and Whitford, 2004) or other premotor (Voss et al., 2006) areas. If this were the case, variability in output could be introduced downstream of the efference copy, in the form of central noise in primary motor cortex and peripheral noise at the neuromuscular junction. However, evidence from single-unit recordings in rhesus monkeys suggests that at least some of the variability is generated during motor preparation, as variation in premotor as well as primary motor activity during a preparatory period predicted variations in reaching movements (Churchland et al., 2006). Alternatively, therefore, motor efference copy may faithfully encode precise motor commands, but precision may be lost in its translation to the resulting corollary discharge in sensory cortex, at which point precise details are no longer encoded. In either case, the resulting sensory prediction is naive to output variability.

Second, our findings liken peripheral speech productions to errors. Real-time vowel feedback perturbation (Purcell and Munhall, 2006; Tourville et al., 2008) offers a window into error-correction processes by causing a prediction-feedback mismatch that simulates a speech error. Such mismatches result in reduced suppression (Houde et al., 2002; Behroozmand and Larson, 2011), or even enhancement (Eliades and Wang, 2008; Behroozmand et al., 2009; Chang et al., 2013), of cortical responses to sensory feedback, as well as causing compensatory behavioral changes that partially counteract the perceived error. In the current study, vowel productions near the periphery also appeared to generate imperfect matches between feedback and prediction; furthermore, the greater the eccentricity of the peripheral vowels, the larger the mismatch (Fig. 4*A*). These outlying vowels behaved as mild or potential errors both neurally and behaviorally: suppression was reduced, as is seen during a feedback perturbation, and subsequent speech output moved closer to the median, resembling a compensatory behavioral response to a perturbation. The negative correlation between behavioral movement and neural suppression, together with the fact that the movement follows the neural activity in time, suggests that the suppression may reflect an error-detection/correction process. It has been previously suggested that this suppression could be used in self-monitoring (Eliades and Wang, 2003), but our study provides the first evidence in natural speech production that the degree of suppression conveys information that allows speakers to detect and correct deviations from an auditory target. Error-like responses to peripheral productions that were not realized as errors suggests that the efference copy prediction may reflect higher-level (e.g., phonemic) properties of those target sounds, not simply the specific motor commands used to generate them.

Our current findings are consistent with an efferent-evoked sensory prediction that represents a sensory goal. In the speaking task used here, one candidate for a higher-level sensory goal is a prototypical production at the center of a vowel's formant distribution. Using this prototype as a putative target, we grouped trials based on their proximity to the median formant values at the start of the utterance. The center of the formant distribution is not the only possible candidate—a target somatosensory configuration would also be compatible with our results—but it is a reasonable surrogate for a sensory goal, as there is strong evidence that the goals of speech movements are regions in auditory-perceptual space (Perkell et al., 1997; Guenther et al., 1998; Perkell, 2012) with the prototype at the center of these regions (Kuhl, 1991). The correlation between speech production variability and changes in suppression (Fig. 4*A*) suggests that the absolute acoustic distance from this central target region may determine the goodness of the match. For speakers with large variability, peripheral trials are acoustically farther from the median than their counterparts in more precise speakers; utterances that fall in this more distant periphery may thus be perceived as a greater mismatch, resulting in a larger decrease in suppression from the central trials. Furthermore, a recent feedback perturbation study using fMRI provides additional evidence that feedback may be matched to a prototypical or "best" vowel target region within the context of an utterance: trials that fell closer to vowel boundaries (farther from the center) had larger auditory cortical and behavioral compensatory responses to perturbation, even when the perturbation magnitude and direction were held constant (Niziolek and Guenther, 2013).

Some past research has questioned whether sensory suppression phenomena such as SIS are specific to motor-driven predictability or whether they reflect more general attentional processes (Schafer et al., 1981; Cardoso-Leite et al., 2010; Lange, 2011; Sowman et al., 2012; SanMiguel et al., 2013). For example, in the speak condition, auditory feedback onset is simultaneous with action onset, whereas in the listen condition, the temporal expectancy of the stimuli is less certain, as they are externally triggered. Thus, differences in brain activity between the two conditions could reflect a more general prediction of when the stimulus would occur. In the current study, however, the central and peripheral trials were prompted identically; thus, changes in SIS between center and peripheral trials cannot be ascribed to differences in temporal expectancy or attentional orienting. Similarly, these changes cannot be explained by the slight differences in the spoken and recorded auditory signals resulting from bone conduction, as this was constant across all trials.

In models of motor control, external sensory feedback is compared with an internal prediction (Miall and Wolpert, 1996; Golfinopoulos et al., 2010; Friston, 2011; Houde and Nagarajan, 2011). In feedback perturbation studies designed to test these models, external perturbations cause a change in feedback without modifying the commands sent by the motor system. It is therefore difficult to determine whether cortical responses to perturbation reflect a mismatch between feedback and motor commands or between feedback and sensory goals: both are out of alignment when feedback is artificially altered. We therefore undertook to probe the nature of internal predictions using natural production variation. Here, we provide evidence that these predictions are a better match for utterances that are closer to a sensory prototype. This finding is most consistent with models whose predictions reflect a desired sensory target.

References

Behroozmand R, Larson CR (2011) Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci 12:54. CrossRef Medline

Behroozmand R, Karvelis L, Liu H, Larson CR (2009) Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin Neurophysiol 120:1303–1312. CrossRef Medline

Bell C, Bodznick D, Montgomery J, Bastian J (1997) The generation and subtraction of sensory expectations within cerebellum-like structures. Brain Behav Evol 50:17–31. CrossRef Medline

Blakemore SJ, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. Nat Neurosci 1:635–640. CrossRef Medline

Cardoso-Leite P, Mamassian P, Schütz-Bosbach S, Waszak F (2010) A new look at sensory attenuation action-effect anticipation affects sensitivity, not response bias. Psychol Sci 21:1740–1745. CrossRef Medline

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci U S A 110:2653–2658. CrossRef Medline

Churchland MM, Afshar A, Shenoy KV (2006) A central source of movement variability. Neuron 52:1085–1096. CrossRef Medline

Crapse TB, Sommer MA (2008) Corollary discharge across the animal kingdom. Nat Rev Neurosci 9:587–600. CrossRef Medline

Creutzfeldt O, Ojemann G, Lettich E (1989) Neuronal activity in the human lateral temporal lobe: II. Responses to the subjects own voice. Exp Brain Res 77:476–489. CrossRef Medline

Cullen KE, Brooks JX, Jamali M, Carriot J, Massot C (2011) Internal models of self-motion: computations that suppress vestibular reafference in early vestibular processing. Exp Brain Res 210:377–388. CrossRef Medline

Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. Hum Brain Mapp 9:183–191. CrossRef Medline

Dalal SS, Zumer JM, Agrawal V, Hild KE, Sekihara K, Nagarajan SS (2004) NUTMEG: a neuromagnetic source reconstruction toolbox. Neurol Clin Neurophysiol 2004:52. Medline

Eliades SJ, Wang X (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. J Neurophysiol 89:2194–2207. CrossRef Medline

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

Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. J Neurosci 30:16643–16650. CrossRef Medline

Friston K (2011) What is optimal about motor control? Neuron 72:488–498. CrossRef Medline

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

Greenlee JD, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA 3rd (2011) Human auditory cortical activation during self-vocalization. PLoS One 6:e14744. CrossRef Medline

Guenther FH, Hampson M, Johnson D (1998) A theoretical investigation of reference frames for the planning of speech movements. Psychol Rev 105:611–633. CrossRef Medline

Haggard P, Whitford B (2004) Supplementary motor area provides an efferent signal for sensory suppression. Cogn Brain Res 19:52–58. CrossRef Medline

Herrmann B, Henry MJ, Obleser J (2013) Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. J Neurophysiol 109:2086–2096. CrossRef Medline

Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front Hum Neurosci 5:82. CrossRef Medline

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

Lange K (2011) The reduced N1 to self-generated tones: an effect of temporal predictability? Psychophysiology 48:1088–1095. CrossRef Medline

Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural Netw 9:1265–1279. CrossRef Medline

Niziolek CA, Guenther FH (2013) Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. J Neurosci 33:12090–12098. CrossRef Medline

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

Owen JP, Wipf DP, Attias HT, Sekihara K, Nagarajan SS (2012) Performance evaluation of the Champagne source reconstruction algorithm on simulated and real M/EEG data. Neuroimage 60:305–323. CrossRef Medline

Perkell JS (2012) Movement goals and feedback and feedforward control mechanisms in speech production. J Neurolinguistics 25:382–407. CrossRef Medline

Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J, Guiod P (1997) Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. Speech Comm 22:227–250. CrossRef

Poulet JF, Hedwig B (2003) A corollary discharge mechanism modulates central auditory processing in singing crickets. J Neurophysiol 89:1528–1540. CrossRef Medline

Poulet JF, Hedwig B (2006) The cellular basis of a corollary discharge. Science 311:518–522. 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

SanMiguel I, Todd J, Schröger E (2013) Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. Psychophysiology 50:334–343. CrossRef Medline

Schafer EW, Amochaev A, Russell MJ (1981) Knowledge of stimulus timing attenuates human evoked cortical potentials. Electroencephalogr Clin Neurophysiol 52:9–17. CrossRef Medline

Sowman PF, Kuusik A, Johnson BW (2012) Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. Exp Brain Res 222:149–157. CrossRef Medline

Sperry RW (1950) Neural basis of the spontaneous optokinetic response produced by visual inversion. J Comp Physiol Psychol 43:482–489. CrossRef Medline

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

Ventura MI, Nagarajan SS, Houde JF (2009) Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci 10:58. CrossRef Medline

von Holst E, Mittelstaedt H (1950) The reafference principle: interaction between the central nervous system and the periphery. Die Naturwissenschaften 37:464–476. CrossRef

Voss M, Ingram JN, Haggard P, Wolpert DM (2006) Sensorimotor attenuation by central motor command signals in the absence of movement. Nat Neurosci 9:26–27. CrossRef Medline

Zaretsky M, Rowell CH (1979) Saccadic suppression by corollary discharge in the locust. Nature 280:583–585. CrossRef Medline