

# Modulation of the Auditory Cortex during Speech: An MEG Study

John F. Houde<sup>\*1</sup>, Srikantan S. Nagarajan<sup>\*1</sup>, Kensuke Sekihara<sup>2</sup>,  
and Michael M. Merzenich<sup>1</sup>

## Abstract

■ Several behavioral and brain imaging studies have demonstrated a significant interaction between speech perception and speech production. In this study, auditory cortical responses to speech were examined during self-production and feedback alteration. Magnetic field recordings were obtained from both hemispheres in subjects who spoke while hearing controlled acoustic versions of their speech feedback via earphones. These responses were compared to recordings made while subjects listened to a tape playback of their production. The amplitude of tape playback was adjusted to match the amplitude of self-produced speech. Recordings of evoked responses to both self-produced and tape-recorded

speech were obtained free of movement-related artifacts. Responses to self-produced speech were weaker than were responses to tape-recorded speech. Responses to tones were also weaker during speech production, when compared with responses to tones recorded in the presence of speech from tape playback. However, responses evoked by gated noise stimuli did not differ for recordings made during self-produced speech versus recordings made during tape-recorded speech playback. These data suggest that during speech production, the auditory cortex (1) attenuates its sensitivity and (2) modulates its activity as a function of the expected acoustic feedback. ■

## INTRODUCTION

Behavioral experiments have shown that different aspects of speech production are sensitive to auditory feedback. Adding noise to a speaker's auditory feedback results in an elevation of voice volume (Lane & Tranel, 1971). Delaying a speaker's auditory feedback by more than about 50 msec produces noticeable disruptions in speech (Yates, 1963; Lee, 1950). Shifting the spectrum of auditory feedback causes shifts in the spectrum of produced speech (Gracco, Ross, Kalinowski, & Stuart, 1994). Perturbing perceived pitch causes the subject to alter his pitch to compensate for those perturbations (Kawahara, 1993). Altering perceived formants induces compensating changes in the production of vowels (Houde & Jordan, 1998). Prolonged listening to one phoneme causes small but significant changes in the production of that phoneme (Cooper, 1979).

A number of physiological studies have shown that activity at different levels of the auditory system is modulated by the act of vocal production. Early evidence came from animal studies in bats, birds, and monkeys. In bats, responses in the lateral lemniscus of the midbrain are attenuated by about 15 dB during

vocalization (Suga & Schlegel, 1972; Suga & Shimozawa, 1974). In birds, it is now well known that vocal production is disrupted by perturbation in the auditory feedback, and that this feedback alteration affects the responses in higher-order neurons of the auditory forebrain in birds (Leonardo & Konishi, 1999; Schmidt & Konishi, 1998; McCasland & Konishi, 1981). In monkeys, activity in the auditory cortex is found to be inhibited due to either spontaneous vocalizations or to vocalizations evoked by electrical stimulation of the cingulate cortex (Muller-Preuss & Ploog, 1981). Similar, albeit less consistent results, have been reported for recordings from human temporal lobe (Creutzfeldt & Ojemann, 1989; Creutzfeldt, Ojemann, & Lettich, 1989a, 1989b). However, the exact characteristics of this attenuation have not been adequately investigated in animal studies. Furthermore, the dynamics of this attenuation is poorly understood.

Recent functional brain imaging studies in humans have investigated the activity of the auditory system during speech perception. The analysis of phonological and semantic characteristics of externally generated speech is found to engage the left inferior frontal, left temporal, and posterior parietal cortices (Demonet et al., 1992; Zatorre, Evans, Meyer, & Gjedde, 1992), while processing of prosodic qualities engages analogous regions in the right hemisphere (Zatorre et al., 1992). It has been hypothesized that similar areas also

---

<sup>1</sup>University of California, San Francisco, <sup>2</sup>Tokyo Metropolitan Institute of Technology, \*These authors contributed equally to this paper.

participate in monitoring these aspects of speech that are self-generated (Hickok & Poeppel, 2000; Levelt, 1983, 1989).

PET studies have shown that the response to primary auditory cortex (A1) to self-produced speech is minimal (Hirano et al., 1996; Hirano, Naito, et al., 1997). However, this A1 response activity increased if the subject heard an altered version of their normal speech (Hirano, Kojima, et al., 1997). In other PET studies, varying the rate of speech output while the speech-contingent auditory input was masked by constant white noise produced significant modulation of the response from the secondary auditory cortex of the left hemisphere (McGuire, Silbersweig, & Frith, 1996). Furthermore, these studies have reported that the medial prefrontal cortex and the left frontal insula/operculum were differentially activated during speech feedback alteration. However, these studies have not examined the dynamics of activity in these different auditory regions engaged during speech production.

More recently, magnetoencephalography (MEG) studies have shown that the responses from the auditory cortex to self-produced speech are attenuated in the millisecond time scale, when compared with responses from tape-recorded speech (Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Numminen & Curio, 1999; Numminen, Salmelin, & Hari, 1999). These initial MEG studies also reported a lack of a “change” response to self-produced stimuli. The study by Curio et al. (2000) examined 100 msec poststimulus (M100) response differences between speaking and tape playback conditions. The authors found significant differences in M100 amplitude across listening conditions in the left hemisphere, and significant differences in M100 latencies across listening conditions in both hemispheres. Numminen et al. have investigated this phenomena in two studies. In one study, they reported that M100 responses to 1-kHz tones were slightly delayed and significantly inhibited during overt speech as compared with silent speech (Numminen et al., 1999). In a related study, they also reported that M100 responses to short recorded vowel sounds heard while subjects also heard either self-produced or tape recorded vocalizations were delayed and dampened relative to background-free presentations (Numminen & Curio, 1999).

In this study, we sought to further examine this attenuation of responses to self-produced speech. In particular, we were interested in discerning between two possible mechanisms of the M100 response reduction. One possibility is that activity in the auditory system is generally suppressed during speech. Such “nonspecific” attenuation could be the indirect result of middle-ear muscle activity during speaking (Papanicolaou, Raz, Loring, & Eisenberg, 1986). Alternatively, the M100 response reduction during speaking could result from a comparison between actual and predicted auditory feedback—that is, an auditory version of Held’s

“reafference hypothesis” (Hein & Held, 1962). Motor system activity during speaking could generate an internal representation of the expected auditory feedback, and a match between expected and actual feedback could reduce the M100 response.

To test between the efference copy and nonspecific attenuation hypotheses, we performed a series of experiments in which MEG was used to examine M100 responses to speech in a variety of conditions. In these experiments, magnetic field recordings were obtained from the auditory cortices in both hemispheres of subjects who spoke while hearing controlled versions of their speech feedback via earphones. These responses were compared to MEG recordings made while subjects listened to a tape recording of their production.

Three experiments were conducted. First, we repeat the experiments of Numminen et al. and examined the dynamics of attenuation of responses to self-produced speech. Second, we tested the specificity of the recorded signal attenuation by examining M100 responses to tone pulses added to the speech heard by the subject. Third, we tested the “reafference hypothesis” by examining M100 responses to speech feedback altered to “mismatch” the expected speech feedback.

## RESULTS

### Experiment 1

Figure 1 shows the apparatus used in the experiments. Magnetic fields were recorded from both hemispheres in a shielded room using two 37-channel biomagnetometers. A directional microphone was placed in the chamber at a distance where it did not distort the

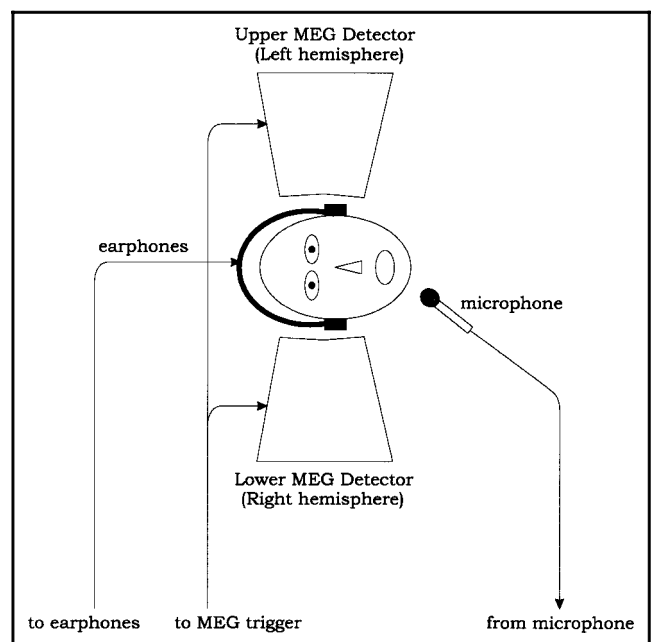


Figure 1. Apparatus used in the experiments.

recorded magnetic fields. Acoustic input was delivered to the subject via air tube-based earphones. Triggering of the data acquisition system and of stimuli delivered to the ears was controlled and differed for each of three experiments in the study.

The schematic for our first experiment is shown in Figure 2. This experiment consisted of two successive conditions: speaking and tape playback. In the speaking condition (Figure 2a), subjects were instructed to produce the short vowel sound /ə/ by phonating with their jaw and tongue relaxed and stationary. The subject's speech was picked up by a microphone and fed to a tape recorder, his earphones, and the trigger input for MEG data acquisition. A total of 100 utterances were recorded on tape. In the subsequent tape playback condition (Figure 2b), subjects were instructed to remain silent while they heard a tape-recording of their utterances from the speaking condition.

Results for one subject in the speaking condition are shown in Figure 3, which shows evoked magnetic field response recorded at each detector position, averaged over the 100 utterance trials. In the figure, the traces show average response at each detector position, aligned to the onset of vocalization (gray tick marks). For each condition (speaking, tape playback), the RMS responses at each detector were averaged together for the left and right hemispheres, as shown in Figure 4. The figure shows RMS responses averaged over detectors in the left (upper panel) and right (lower panel) hemisphere detector arrays. In each panel, the thick trace shows the average RMS response in the speaking

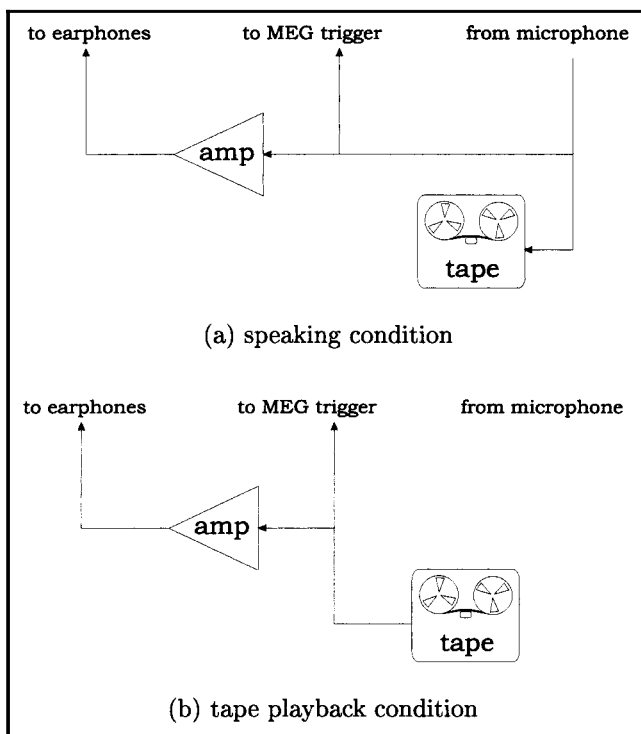


Figure 2. Setup for the two conditions of Experiment 1.

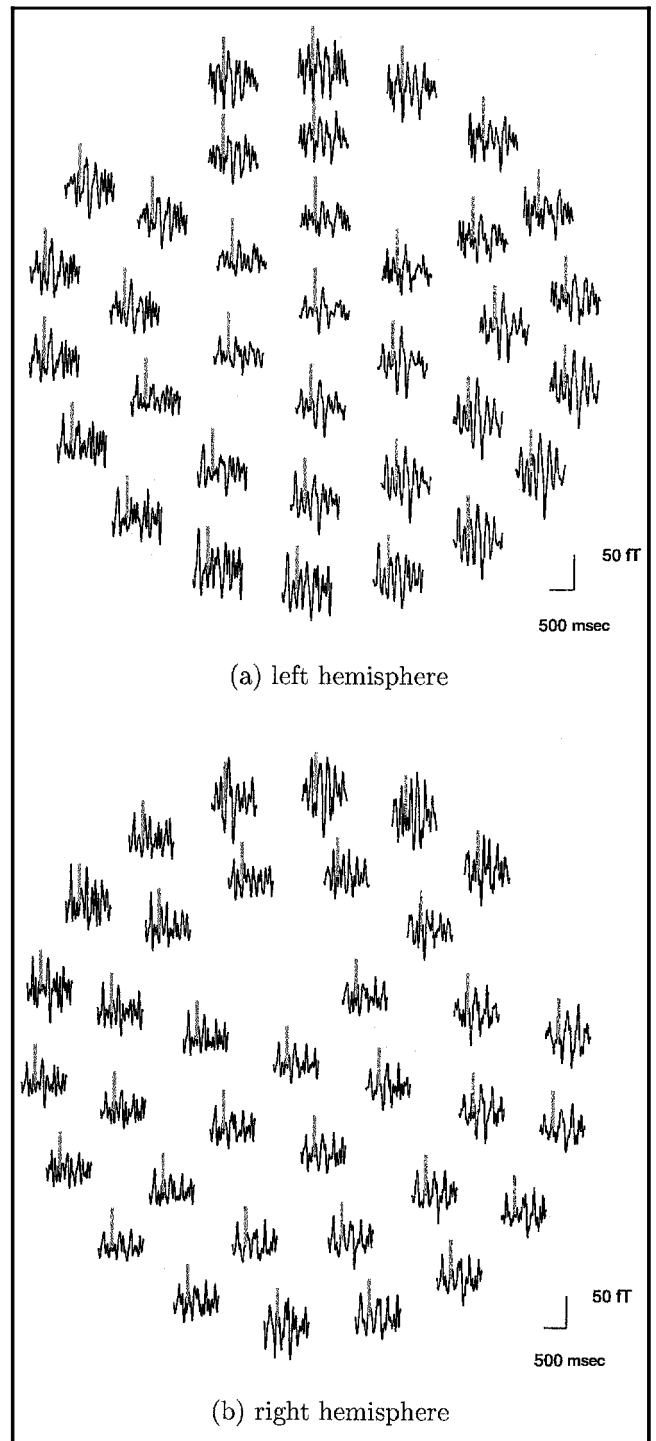
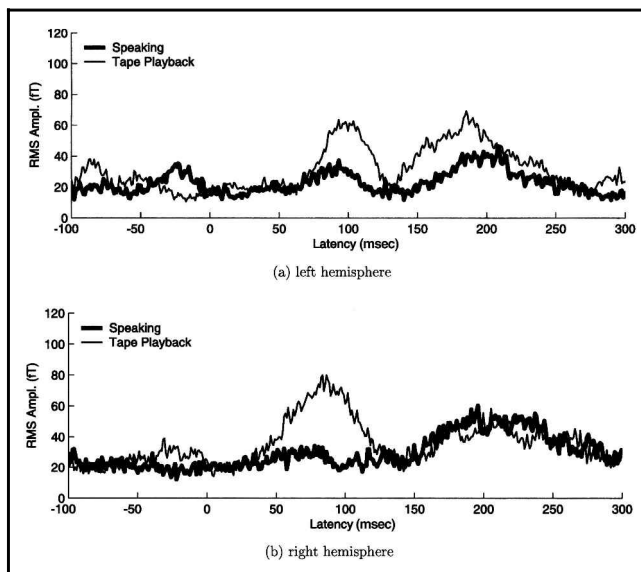


Figure 3. RMS responses across detector arrays recorded in the speaking condition of Experiment 1, from a single representative subject. The traces show average response at each detector position, aligned to the onset of vocalization (gray tick marks). (Gaps in the detector arrays are due to offline detectors.)

condition and the thin trace shows the same for the tape playback condition. The vertical axis of each panel indicates RMS field strength (fT) of the response, while the horizontal axis indicates time (msec) relative to the onset of the vocalization, which occurs at 0 msec. Many differences between responses in the speaking and



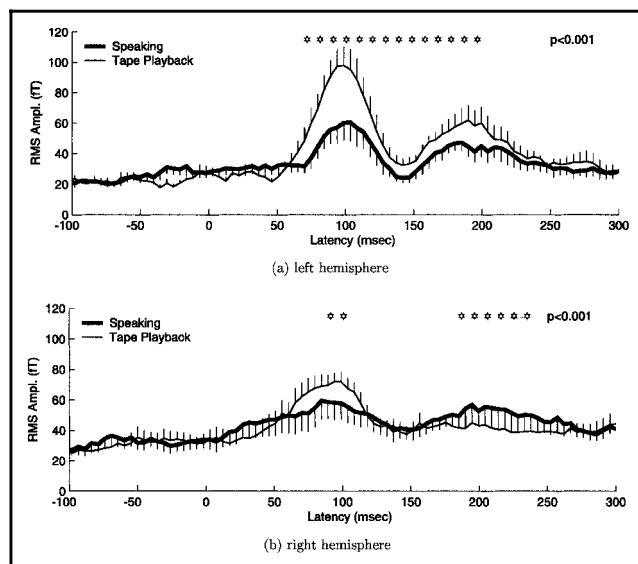
**Figure 4.** RMS responses recorded in Experiment 1, averaged over all 37 detectors in each hemisphere's detector array, from a single representative subject. In each plot, the thick traces are responses in the speaking condition, the thin traces are responses in the tape playback condition. The onset of the audio stimulus is at 0 msec.

tape playback conditions can be seen in the figure. At about 100 msec poststimulus onset, most of the traces show a peak in RMS amplitude. This peak is referred to as the M100 response, and there are large differences in its amplitude across conditions in both hemispheres for this subject. In the left hemisphere, the M100 response in the speaking condition is greatly suppressed compared to that seen in the tape playback condition. In the right hemisphere, this suppression is even greater: In the tape playback condition, there is quite a large M100 response, but in the speaking condition the M100 is essentially absent.

Figure 5 shows the same response waveforms of the previous figure, averaged across all eight subjects. In this figure, in addition to the average response waveforms, vertical bars indicating standard errors are also shown. The asterisks in each panel indicate latencies at which the speaking and tape playback responses differed significantly, relative to  $p < .001$ . These plots show that in both hemispheres, mean M100 responses were significantly smaller in the speaking condition than in the tape playback condition. It appears that left hemisphere responses are larger than right hemisphere responses in the tape playback condition, but that both hemispheres are inhibited to about the same level in the speaking condition. However, even in the speaking condition, the left hemisphere M100 response appears more focal in time than that of the right hemisphere. Less consistent are the longer latency responses. For example, the amplitude of the M200 response in the left hemisphere was significantly larger in the tape playback condition, while in the right hemisphere, it was significantly larger in the speaking condition.

The acoustic signals reaching the ears via air conduction, the so-called “side-tone” sound amplitudes, were adjusted to be the same for both the speaking and tape playback conditions; all our subjects reported that the intensity of mic and tape signals were perceptually identical. However, in the speaking condition, because the subject hears both the side tone and bone conduction of his voice, the speech heard by a subject is actually about twice the amplitude of (i.e., 3 dB louder) the speech heard in the tape playback condition (von Bekesy, 1949). Thus, a priori, we would expect this amplitude difference to be a possible confound in the experiment. The size of evoked magnetic field response in the auditory cortex and, in particular, the amplitude of the M100 response have been shown to be a monotonically increasing function of audio input level (Stufflebeam, Poeppel, Rowley, & Roberts, 1998). From this, we would expect the M100 response to be larger in the voiced condition than in the tape playback condition. However, this response difference should be small: At the audio levels used in this experiment (80 dBA at the earphones), the effect of audio amplitude on the M100 response has nearly saturated. Additionally, as we have seen, the actual M100 response recorded in the speaking condition was much less than that recorded in the tape playback condition—a result opposite of what would be predicted based on audio amplitude differences.

Source locations were performed on the spatio-temporal magnetic field responses recorded 50–200 msec following stimulus onset. Across all our stimulus conditions (mic vs. tape vs. tone response), we found no statistically significant difference ( $p > .5$ ) in the location of an equivalent dipole in each hemisphere that accounts



**Figure 5.** RMS response waveforms in Experiment 1, averaged across all subjects. The vertical bars in each trace show standard errors, while the asterisks show intervals where the response waveforms from the microphone condition (thick traces) and tape condition (thin traces) differ significantly, relative to  $p < .001$  for both hemispheres.

for these magnetic field responses. These localization data suggest that the observed response suppression/attenuation is arising from a restricted cortical region, presumably A1 and its immediate environs.

Overall, all subjects' M100 responses were localized to the temporal cortex, in the area of the auditory cortex. We infer this because lesion studies, intracranial recordings, and source modeling studies provide converging evidence that the neuronal generators of the M100 response were located in the auditory regions of the temporal lobes and include A1 and its immediate environs (Ahissar et al., 2001; Picton et al., 1999; Liegeois-Chauvel, Musolino, Badier, Marquis, & Chauvel, 1994; Reite et al., 1994; Richer, Alain, Achim, Bouvier, & Saint-Hilaire, 1989; Scherg & Von Cramon, 1985, 1986; Woods, Knight, & Neville, 1984; Hari, Aittoniemi, Jarvinen, Katila, & Varpula, 1980).

## Experiment 2

The results of Experiment 1 clearly confirmed that the M100 response is suppressed in the speaking condition. One possible mechanism that could cause this suppression is nonspecific attenuation. That is, neural signals arising as a result of activity in the motor cortex might directly and broadly inhibit auditory cortex activity.

If auditory cortex activity is nonspecifically suppressed during speaking, then it should be less responsive to all auditory signals. In Experiment 2, we tested this possibility by examining the M100 response to tones heard during speech that was either self-produced or played back from tape.

In this experiment, we measured the responses evoked by tones pips under three successive conditions, as shown in Figure 6. In the first condition (tones alone, Figure 6a), subjects remained silent and heard 1.0-kHz tone pips. In the second condition (tones and speaking, Figure 6b), subjects again heard these tones pips, but this time, the tones were presented while subjects produced long utterances of the vowel /ə/ (again, with no movement of their jaw or tongue). In the third condition (tones and tape playback, Figure 6c), subjects again remained silent and heard both tone pips and tape playback of their speech produced in the previous condition. For the last two conditions in which the subject heard tone pips and speech, the speech was attenuated by 20 dB to reduce the masking of the tone pips.

The results of Experiment 2 are summarized in Figure 7. The figure shows a plot of the mean RMS response, across detectors and subjects, for each condition and in each hemisphere. Dashed traces show subjects' responses to hearing the tones in silence (tones alone condition). Thick traces show responses to hearing the tones while subjects produced /ə/ (tones and speaking condition). Thin traces show responses to hearing the tones along with the taped speech playback (tones and tape playback condition). As in Figure 5, the vertical bars

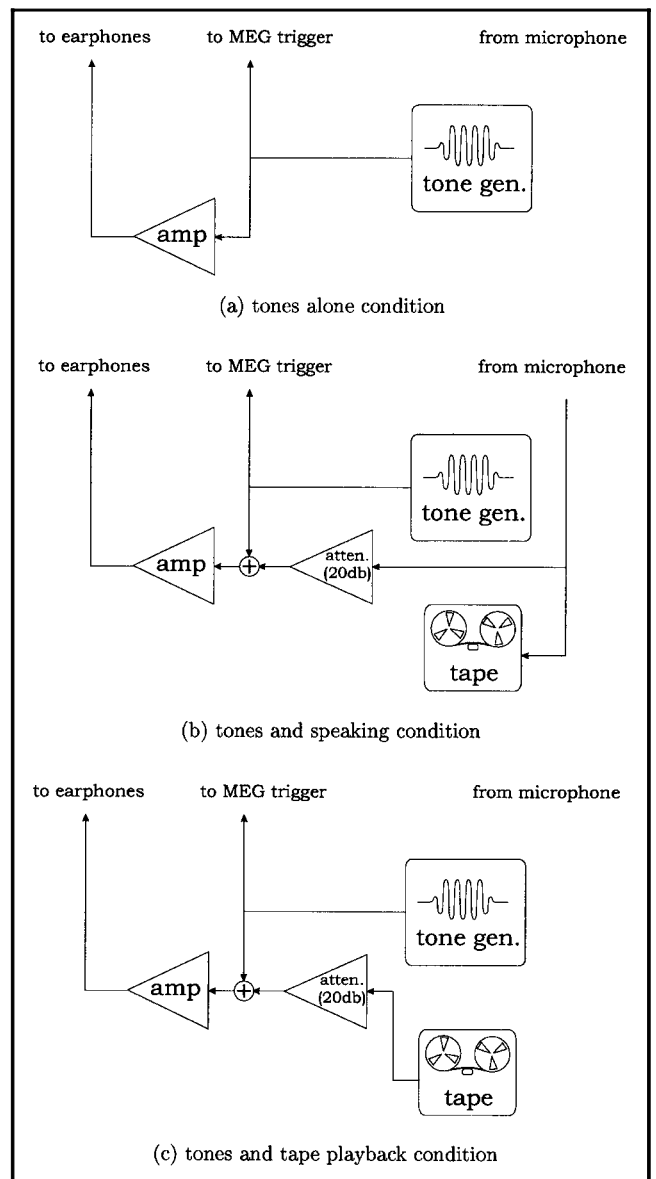
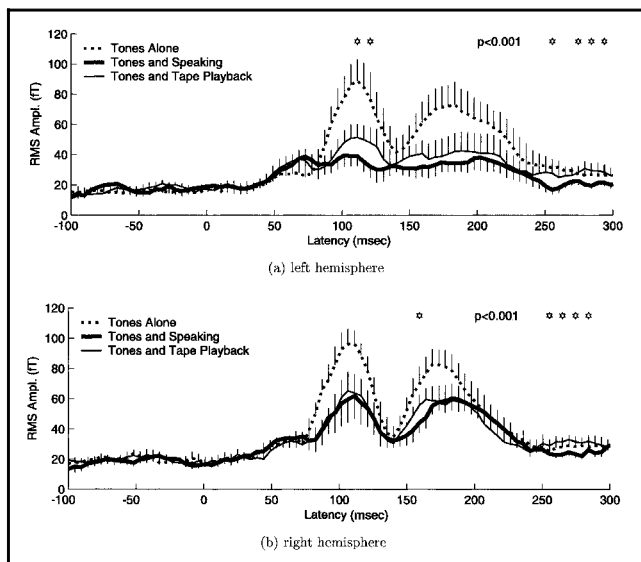


Figure 6. Setup for the three conditions of Experiment 2.

are standard errors. Asterisks indicate time intervals during which responses to tones heard while producing speech (tones and speaking condition) differed significantly ( $p < .001$ ) from the responses to tones heard with taped speech playback (tones and tape playback condition).

For most of the traces in Figure 7, the M100 and M200 responses to the tones across conditions stand out clearly. For both hemispheres, these responses are very prominent for tones heard in silence (dashed traces). In the left hemisphere, there was a reduction in M100 and M200 response to the tones in the speaking (thick traces) and tape playback (thin traces) conditions. Interestingly, the response reduction in the speaking condition appeared to be greater than the response reductions in the tape playback condition. Indeed, the statistical analysis confirmed that the M100 response in



**Figure 7.** RMS waveforms for responses to 1.0-kHz tone pips, averaged across all subjects for the three conditions of Experiment 2: the tones alone condition (dashed traces), the tones and speaking condition (thick traces), and the tones and tape playback condition (thin traces). The vertical bars in each trace show standard errors, while the asterisks show intervals where the response waveforms from the tones and speaking condition (red) and the tones and tape playback condition (blue) differed significantly, relative to  $p < .001$  for both hemispheres.

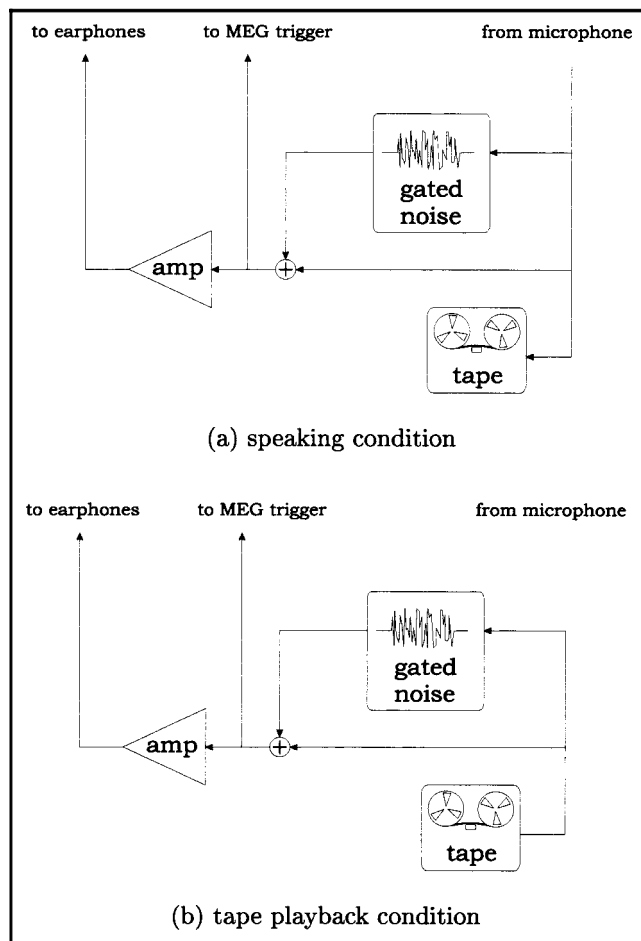
the tones and speaking condition was significantly less than the M100 response in the tones and tape playback condition. Nevertheless, the difference was small. In the right hemisphere, there was an approximately equal reduction in M100 and M200 responses to the tones in the speaking and tape playback conditions. Within the M100–M200 region, responses in these two conditions were only significantly different at one latency, as shown by the asterisk.

It appears that, at the most, the act of speaking creates only a limited degree of nonspecific attenuation to the tones in the left hemisphere that is not seen in the tape playback condition. To quantify how much nonspecific attenuation contributes to the suppression of the response to self-produced speech, we can estimate the signal reductions needed to produce the response reductions seen in these experiments. From the results of Experiment 1, we have seen that, compared to speech from the tape playback, self-production of speech created a 30% reduction in M100 amplitude in the left hemisphere and a 15% reduction in the right hemisphere, equivalent to 13 and 7 dB decreases in effective input signal, respectively.<sup>1</sup> In Experiment 2, in the left hemisphere, there was possibly an extra 7% reduction in M100 response to tones while subjects spoke as compared to the M100 response to tones during tape playback, which is equivalent to only a 3-dB reduction in effective input signal. Thus, it appears that nonspecific attenuation cannot account for all the response suppression recorded for self-produced speech.

It has also been argued that any attenuation in the responses during vocalization could occur due to bone conduction and activation of middle-ear muscles or due to inhibition at the level of the brain stem (Papanicolaou et al., 1986). Thus, to further test the hypothesis of nonspecific attenuation, we measured brainstem evoked responses to click trains in five subjects under three conditions identical to those in Experiment 2—silence, produced speech, and tape-recorded speech. In the latter two conditions, we attenuated the speech background by 20 dB to obtain a clear evoked response to clicks. For each of these conditions, we recorded the brainstem evoked response and measured the latency and amplitude of Waves I–V. The results from these experiments indicated that there were no statistical differences in either the amplitude or the latencies for all waves between the three background conditions ( $p > .2$ ). These results provide further support to the hypothesis that our observed effects were cortical in origin.

### Experiment 3

Another possible cause of the response reduction to self-produced speech is that activity in the auditory



**Figure 8.** Setup for the two conditions of Experiment 3.

cortex is the result of a comparison between the incoming signal and an internally generated prediction of what that signal would be. If this were to be the case, then altering the incoming signal should create a mismatch with that expectation that would reduce or abolish the response suppression.

In Experiment 3, we tested this possibility by altering the auditory feedback to the subject during self-production of speech. The setup for Experiment 3 is shown in Figure 8. Like Experiment 1, this experiment consisted of two conditions: speaking and tape playback. However, instead of hearing their speech (self-produced or from tape playback), subjects heard a sum of their speech plus the output of a white noise generator that produced noise bursts gated to the duration of the speech utterance. Subjects reported that they could not hear their own speech and that they only heard noise gated to their vowel production.

The results of Experiment 3 are shown in Figure 9. For both hemispheres, the thick traces show the average evoked response of subjects hearing gated noise in the speaking condition, while the thin traces show their responses to gated noise in the tape playback condition. Asterisks mark poststimulus latencies where the evoked responses in the two conditions differ significantly, relative to  $p < .001$  for both hemispheres. The figure shows subjects' evoked responses in the two conditions are significantly different at a number of latencies: In the left hemisphere, the responses in the speaking and tape playback conditions differ significantly around the M100 responses and at a latency of 150 msec. In the right hemisphere, the speaking and tape playback responses

differ significantly at the 50- and 200-msec latencies. Since we are limiting our analysis of these results to considering only M100 responses, the only response differences of interest are the significant differences seen in the M100 region of the left hemisphere responses. It is clear from the figure, however, that these response differences arise principally from a latency difference between M100 responses; the M100 response to gated noise in the tape playback condition is about 20 msec earlier than the M100 response to gated noise in the speaking condition. The amplitudes of the two M100 responses appear nearly identical.

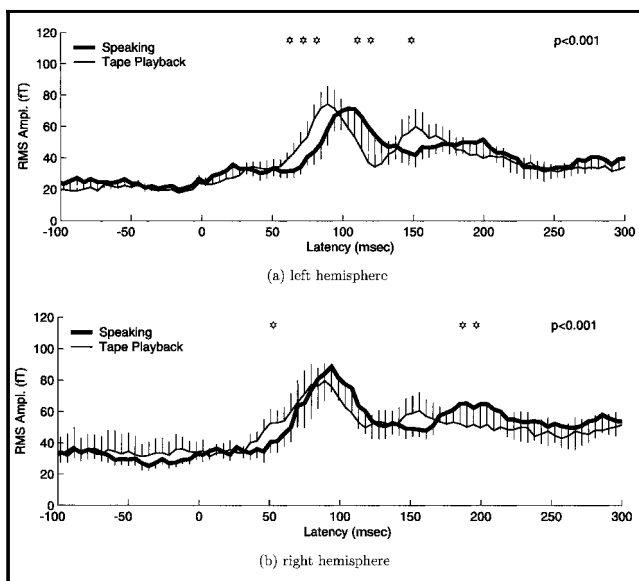
In sum, Figure 9 shows that the M100 amplitude suppression seen in Experiment 1 is abolished; in both hemispheres, the amplitude of M100 response to noise generated from subjects' speech via the microphone is as large as the M100 response to noise generated from the tape playback of their speech. Thus, when we altered subjects' speech feedback in Experiment 3, we found that the response suppression to self-produced speech disappeared, which is consistent with the suppression resulting from auditory input matching an internally expected input.

## DISCUSSION

The present study demonstrates that the human auditory cortex responds differently to self-produced speech than to externally produced speech. In Experiment 1, significant suppression in the amplitude of M100 responses was recorded in the speaking condition in both hemispheres, confirming that self-production of speech suppresses the response in the auditory cortex to that speech. In Experiment 2, M100 responses to tone pips were seen to be modestly more strongly suppressed by self-produced than by externally produced speech. However, this relatively weak extra suppression was not sufficient to explain the suppression seen in Experiment 1. On the other hand, in Experiment 3, we were able to abolish the suppression of M100 responses to self-produced speech by altering the feedback heard by subjects. This last result is consistent with the hypothesis that the suppression of response in the auditory cortex to self-produced speech results from a match with expected auditory feedback.

### Comparison with Other Studies

Experiment 1 is quite similar to the study by Curio et al. (2000), which also examined M100 response differences between speaking and tape playback conditions. The authors found M100 amplitude differences across listening conditions in both hemispheres, but, unlike the present study, these differences were not significant in the right hemisphere. Further, in contradiction to our studies, Curio et al. described significant differences in M100 latencies across listening conditions in both



**Figure 9.** RMS response waveforms in Experiment 3, averaged across all subjects. The vertical bars in each trace show standard errors, while the asterisks show intervals where the response waveforms from the speaking condition (thick traces) and tape playback condition (thin traces) differ significantly, relative to  $p < .001$  for both hemispheres.

hemispheres. Different results in M100 latencies could be accounted for by the fact that we matched the acoustic waveform presented to each subject, while Curio et al. behaviorally matched the perceptual intensity between the listening conditions. Because subjects could set the tape playback volume to be different from their speech feedback volume, differing audio levels in the two experiment conditions could account for some of the M100 latency differences that they report. With MEG data acquisition triggered by speech signal onset, triggering occurs when the audio signal exceeds some threshold voltage. As a result, triggering time is sensitive to the amplification of the signal: The more a signal is amplified, the faster it will reach the trigger threshold. Thus, if tape playback volume levels were less than speech feedback volume levels, triggering from tape playback could be delayed compared to the speaking condition. If so, then for the same true M100 latency, M100 peaks would appear to have shorter latencies in the tape playback condition.

In Experiment 3, the amplitude difference in M100 responses between the speaking and tape playback conditions was abolished by using gated noise to distort the feedback heard by subjects. Interestingly, however, this feedback manipulation created a latency difference between the two conditions: In the left hemisphere, the M100 response to gated noise triggered by tape playback was 15 msec earlier than that for gated noise triggered by the subject's own speech, while in the right hemisphere, such a latency difference is not seen. Curio et al. (2000) also found that the reduced latency of M100 responses to tape playback versus self-produced speech was significantly more pronounced in the left hemisphere. Like Curio et al., we attribute this left hemisphere latency difference to two factors. First, this latency difference may represent differential processing of externally produced versus self-produced auditory feedback. Although the gated noise altered the spectral characteristics of their speech feedback, such that it was likely to mismatch any internally generated expectations of how their feedback would sound, the gated noise did not alter the temporal characteristics of their speech feedback: Subjects still heard auditory feedback begin at the same time as they commenced phonation. This temporal predictability could potentially have allowed subjects to process differently gated noise arising from their own speech, whose onset they could predict, from gated noise arising from tape playback, whose onset they could not predict. Second, the fact that this latency difference was seen primarily in the left hemisphere is consistent with the left hemisphere's dominance in the production of speech in right-handed speakers. Other aspects of our results may also reflect this left hemisphere dominance: In Experiment 1, the M100 response to tape playback speech was larger in the left hemisphere than in the right, and the M100 response to self-

produced speech was temporally sharper in the left hemisphere than in the right. Additionally, in Experiment 2, in the left hemisphere, the M100 response to tones was significantly more suppressed by self-produced speech than by tape playback speech—a difference not seen in the right hemisphere.

Results from Experiment 2 are mostly consistent with the recent experiments by Numminen and Curio (1999). They reported that neuromagnetic responses to short recorded vowel sounds were delayed and dampened relative to background-free presentations. They did not report how much they externally attenuated the produced speech signal in order to obtain a clear response to the probe-vowel sound. In our experiments, we used probe tones instead of vowel sounds, and we attenuated the produced speech signal to the ear by 20 dB. Surprisingly, we did not observe any latency differences in the M100 responses for tones between the three conditions. We believe that the lack of a latency effect could be accounted for by the fact that we used tones instead of vowel sounds (Diesch & Luce, 1997; Poeppel et al., 1996; Kuriki & Murase, 1989). Furthermore, by an argument analogous to that discussed above for the Curio et al. (2000) latency results, it is also reasonable that any amplitude mismatches in the masking sounds could account for discrepancies in the latency effect (Hari & Makela, 1988). However, while Numminen and Curio identified auditory interference as the main cause of these modifications, their results also provide evidence against the nonspecific attenuation hypothesis, although they do not specifically discuss their results in this form. For instance, they found an additional 6–9% decrease in amplitude in response to probe-vowel sounds during produced speech background when compared with tape-recorded speech background. These amplitude differences are consistent with our Experiment 2 results.

Gunji et al. have found that vocalization-related cortical magnetic fields activated six sources that were found to temporally overlap in the period 0–100 msec after vocalization onset (Gunji, Hoshiyama, & Kakigi, 2000; Gunji, Kakigi, & Hoshiyama, 2000). Sources 1 and 2 were activated approximately 150 msec before the vocalization onset and were located in laryngeal motor areas of the left and right hemispheres respectively. Kuriki, Mori, and Hirata (1999) have also found a motor planning center for speech articulation that is activated 120–320 msec before the onset of vocalization and located in a region around the superior end of the left insula. Gunji et al. Sources 5 and 6 were located in the truncal motor area in each hemisphere and were similar to Sources 1 and 2. In contrast, their Sources 3 and 4 were located in the auditory cortices of the left and right hemisphere, respectively, and were activated after vocalization onset. However, these experiments did not explore either the modulation of the responses in the auditory or motor cortex due to vocalization or feedback alteration. In our



experiments, because we were unable to simultaneously measure activity from motor and supplementary motor areas using a 37-channel sensor array over each hemisphere, we focus here only on the activity from auditory cortices of both hemispheres.

More relevant to our results are the studies of Hirano et al. (Hirano et al., 1996; Hirano, Kojima, et al., 1997; Hirano, Naito, et al., 1997), who used PET to look at cortical activation during speaking. In their first study, Hirano et al. found a lack of activation in the auditory cortex while subjects vocalized. In their second study, Hirano et al. altered how subjects heard their own vocalizations, and this time recorded significant activation of the auditory cortex. These results are consistent with our findings in Experiments 1 and 3 of our study.

### **Why is the Response to Self-Produced Speech Suppressed?**

The present study, along with recent MEG studies by Curio et al. and other brain imaging studies that have used PET or fMRI demonstrate that during speech production, the auditory cortex suppresses its response to expected acoustic signals (Hirano et al., 1996; Hirano, Kojima, et al., 1997; Hirano, Naito, et al., 1997). Does that suppression serve any function? We speculate about two possibilities.

#### *Auditory Perception*

One possibility is that the suppression results from the process of distinguishing self-produced from externally produced sources. It would be useful for sensory systems to make this distinction since sensory information from these two types of sources is commonly used for different purposes. Sensory information from self-produced motor actions could potentially be used for feedback control. On the other hand, sensory information from external sources is primarily used for recognition.

But how could the sensory systems distinguish self-produced from externally produced sensory information? It has been proposed by previous researchers that this is done by having motor actions produce expectations of their sensory consequences (Jeannerod, 1988). These sensory outcome predictions are then compared with the actual sensory input; whatever matches the outcome predictions is inhibited. In that way, the sensory consequences of self-produced actions could be filtered from incoming sensory data, thus highlighting any unpredicted and therefore external (or novel) stimulus that should potentially be attended to.

For a number of reasons, the results of our experiments are consistent with the above account. First, although we found significant suppression of the responses of the auditory cortex to self-produced speech, we found little evidence that the act of producing speech suppresses responses to other sounds (in our

case, the 1-kHz tone). This finding is consistent with the hypothesis that only responses to self-produced sounds are being filtered from the auditory cortex. In addition, we found that altering the auditory feedback with gated noise abolished the suppression in the auditory cortex induced by speaking, consistent with the hypothesis that the filtering out of auditory information arising from self-produced speech is done by comparison with predicted auditory feedback.

Recently, Blakemore et al. have recorded the same types of response suppression phenomena in the somatosensory cortex, suggesting that the need to distinguish self-produced from externally generated sensory input may be a general property of sensory systems. In their series of experiments, Blakemore et al. used both behavioral and physiological procedures to look at responses to both self-produced and externally produced somatosensory stimulation. In one experiment they examined the well-known phenomenon that people cannot tickle themselves (Weiskrantz, Elliott, & Darlington, 1971). This phenomenon could be explained by supposing that somatosensory responses are inhibited when the incoming sensory (touch) information matches a prediction of that input generated by the person's own motor system. To test this hypothesis, these investigators had subjects tickle their palm via an apparatus that could introduce a time delay between their movement and the touch stimulation of their palm (Blakemore, Frith, & Wolpert, 1999). When this was done, subjects reported that with increasing delay, the stimulation of their palm was increasingly stimulating (ticklish). That the perceived stimulation increased with delay could be explained by supposing that the delay made the incoming sensory information mismatch, in time, the predicted sensory input. This experiment was followed up in an fMRI imaging study looking at activation of the somatosensory cortex and cerebellum to both self-produced and externally generated palm stimulation (Blakemore, Wolpert, & Frith, 1999). It was found that activation of the somatosensory cortex to self-produced palm stimulation was suppressed, compared to externally produced palm stimulation—a direct analog of the results recorded in the auditory cortex in our experiments.

#### *Speech Motor Control*

The suppression could also serve to regulate how much auditory feedback is used in the ongoing control of speech.

A problem with using auditory feedback to control speech is the delay incurred by the processing of the auditory input. To control speech, auditory feedback must convey information about the positions of the vocal tract articulators. From speech acoustics, we know that formant frequencies and pitch are the acoustic parameters that convey this information (Stevens, 1999). Extracting these parameters from the type of

signal provided by the cochlea requires significant signal processing, which, in turn, suggests the possibility of significant delay if a succession of auditory brain areas are needed for this processing (Jürgens, 2002; Burnett, Freedland, Larson, & Hain, 1998; Perkell, 1997). In addition, there will be delays associated with the generation of a response in the motor cortex. There will also be intrinsic delays in transmitting the neural signals to the appropriate muscles, as well as delays in muscle responses to these signals (Jürgens, 2002). Together, these factors could contribute to a significant feedback loop delay between auditory input and motor output.

Unfortunately, we know from control theory that, in general, using feedback to control a system with a large feedback loop delay will not work because the system becomes unstable (Franklin, Powell, & Emami-Naeini, 1991). Intuitively, this instability arises because the delay makes the feedback uncorrelated with the state of the system; sensory feedback from some time in the past does not entirely reflect the current state of the system. Yet, auditory feedback does appear to be involved in the online control of some aspects of speech production. As noted in the Introduction, if the pitch of a speaker's feedback is perturbed, they will typically compensate within 100–200 msec (Burnett et al., 1998; Kawahara, 1993; Elman, 1981). Thus, in spite of a large (100–200 msec) feedback loop delay, auditory feedback does stably regulate a speaker's pitch. How is this possible? Again, from control theory, we know that when feedback is delayed, corrupted by noise, or otherwise made uncorrelated with current motor output, a successful approach to using feedback control is to attenuate the sensory feedback to the degree that it is uncorrelated. Specifically, in feedback control based on Kalman filtering, the Kalman gain on sensory feedback is set, in part, to be inversely proportional to the degree to which sensory feedback is correlated with the current system state (Jacobs, 1993). In an analogous way, it may be that the suppression of the auditory cortex during speaking functions to attenuate the auditory feedback to a level commensurate with its delay, such that it can be used properly in the control of speech.

Finally, it is possible that both of the above accounts are true. It may be that the suppression of self-produced speech acts as a filter for incoming sensory data, allowing attention to be paid to externally produced sensations, while at the same time properly attenuating the auditory feedback for use in the online control of speech. Future experiments must be designed to distinguish between the above hypothesized possibilities.

## **METHODS**

### **Subjects**

Eight male volunteers (ages 25–40) participated in Experiments 1 and 3 of this study. Of these volun-

teers, six also participated in Experiment 2. All subjects gave their informed consent. All studies were performed with the approval of the UCSF Committee for Human Research.

### **Experimental Design**

Magnetic fields were recorded from both hemispheres in a shielded room using two 37-channel biomagnetometers with SQUID-based first-order gradiometer sensors (Magnes II, 4-D Neuroimaging, San Diego, CA). A schematic of the experimental setup is shown in Figure 1. A directional microphone was placed in the chamber at a distance where it did not distort the recorded magnetic fields. Triggering of the data acquisition system and of stimuli delivered to the ears was controlled, and it differed for each of three experiments in the study.

Fiduciary points were marked on the skin for later coregistration with structural magnetic resonance images and the head shape was digitized to constrain subsequent source modeling. The sensor was initially positioned over the estimated location of the auditory cortices in both hemispheres, such that a dipolar M100 response was evoked by single 400-msec duration tone pips (1 kHz, 5 msec rise/fall ramps, 90 dB SPL). Data acquisition epochs were 600 msec in total duration with a 100-msec prestimulus period referenced to the onset of the first speech stimulus. Data were acquired at a sampling rate of 1041 Hz. The position of the sensor was then refined so that a single dipole localization model of the averaged evoked magnetic field responses to 100 tones resulted in a correlation and goodness-of-fit greater than .95.

With satisfactory sensor positioning over the auditory cortices in both hemispheres, but before starting the experiments, subjects practiced producing the neutral vowel sound /ə/ by phonating while relaxing their upper vocal tract, generating no movement of their jaw or tongue. Preliminary experiments revealed that this form of speech production produced no movement artifacts in these magnetic field recordings.

During the speaking condition of Experiments 1 and 3, the audio onset of each utterance picked up the microphone triggered the MEG data acquisition system, which recorded 1 sec of data (300 msec pretrigger, 700 msec posttrigger) at a sampling rate of 1041 Hz. Subjects were instructed to produce short vowel sounds /ə/ by phonating their vocal chords without any movement of their jaw or tongue. Responses to 100 such utterances were recorded, and these magnetic field recordings were then averaged to obtain the evoked magnetic field responses to self-produced speech for each MEG detector channel. In the subsequent tape playback condition, the audio signal from the tape recorder was fed to the subjects earphones and the trigger input for MEG data acquisition. Again, magnetic field recordings were averaged to obtain the evoked magnetic field responses.

During the tape playback condition of Experiments 1 and 3, the audio level of the signal fed to the earphones and trigger was adjusted to be the same as that of the speaking condition. Because of this, the trigger of MEG data acquisition (which was based on an audio voltage threshold) occurred at the same time relative to utterance onset in both conditions. The equal audio levels ensured that triggering latency differences could not account for any differences between the evoked responses to self-produced and tape-recorded speech. In Experiment 1, the subject heard these equal audio levels in the earphones, which meant that his side tone amplitude was the same in both the speaking and tape playback conditions. In Experiment 3, the audio signal for each utterance was used to trigger both the MEG data acquisition system and a white noise generator (20 Hz–20 kHz bandwidth, 85 dB SPL). In this case, a sum of their audio signal and noise was presented to the subject.

In Experiment 2, onset of the tone pips triggered MEG data acquisition in each of the three conditions (tones alone, tones and speaking, and tones and tape playback). These 1-kHz tones each had a 400-msec duration, 5-msec onset and offset ramps, and a 0.8–1.2-sec ISI between tones. The subject heard the tones at 90 dB SPL. As with Experiment 1, data acquisition epochs were 1.0 sec in total duration with a 300-msec prestimulus period. Data were acquired at a sampling rate of 1041 Hz. In each condition, the evoked magnetic field responses to the tone pips were obtained by averaging 100 epochs.

Epochs 1500 msec long with a 500-msec prestimulus period were collected at a sampling rate of 1041 Hz. The raw epoch magnetic field data from these 100 utterances were then averaged to obtain the evoked magnetic field responses to gated noise during self-production.

The voltage level of the audio signal fed to earphone and trigger signal in this condition was adjusted to be identical to the voltage levels recorded during the speaking condition, ensuring that triggering differences could not account for any differences between the evoked responses to noise in the self-produced and tape playback blocks.

## Data Analysis

Data from the two sensor arrays positioned over each hemisphere were separately analyzed. From the evoked magnetic field response for each condition in each channel, the RMS value of evoked magnetic field strength averaged across channels was computed. The RMS method of averaging (square root of the mean of squared channel field strengths) was used to avoid the cancellation that occurs when the magnetic fields from detectors on opposite sides of the current dipole (whose fields are roughly equal in magnitude but opposite in sign) are averaged together. This process resulted in one RMS time waveform for each hemisphere, for each condition, for each subject. For statistical comparisons across con-

ditions, a one-way analysis of variance was performed with condition as factor and time as a repeated measure.

The RMS across channels is a convenient measure of activity that is independent of the position of individual sensors. Following optimal placement of the sensor for a significant M100 dipolar response, RMS has been shown to correlate well with the “equivalent current-dipole strength” (also called the “*Q* value”) of a dipole positioned in and around the auditory cortex. The RMS measure is also correlated with the data obtained from (1) the channel with the biggest signal-to-noise ratio, and (2) the first two principal components of the array response (Ahissar et al., 2001; Mahncke, 1998).

In sensor arrays comprising of planar gradiometers (like the Neuromag 122 or 306), the channel with largest signal-to-noise ratio is often closest to the source (as dictated by the physics). However, in the case of sensor arrays comprising of axial gradiometers (as in our Magnes II system), the channel with the largest signal-to-noise ratio reflects the peak of the magnetic field which is not the closest channel to a source. Moreover, the “best” channel could reflect either a peak “inward” or “outward” field, requiring signed inversion while averaging such data across subjects. Finally, since the sensor array position can vary considerably across subjects, it is difficult to use procedures typical for EEG and whole-head studies wherein single channel data is often picked for analysis across subjects. These considerations have led us to use RMS instead of picking specific channels for our analysis.

## Single Dipole Model Fit

A single equivalent current-dipole model was calculated separately for each hemisphere, for each poststimulus time point, using standard 4-D Neuroimaging software operating on data that had been filtered between 1 and 20 Hz. The localization algorithm used an iterative least-squares minimization to compute the strength and location of a single dipole in a spherical volume of uniform conductivity that can account for the sensor data. Dipole fits were accepted based on a local correlation maximum criteria of .95 and goodness-of-fit values greater than .95.

The MEG anatomical reference frame was established using a digital device called the “sensor position indicator.” A set of three table-mounted and one array-mounted receivers triangulates the signal from a stylus transmitter positioned at fiducial reference points on the subject’s head surface (typically nasion, left, right preauricular points, Cz, and inion). The same stylus transmitter arrangement served to define the curvature of the head by tracing the surface from which a “local sphere” model of the head was generated. To coregister the dipole locations on the MRI image, the fiducial points serve as the basis for a common coordinate system. By superimposing these fiducial landmarks on MR images

of the subject, it was possible to define the position of the computed point sources with an accuracy of 5 mm.

### Multiple Dipole Localization Analysis

Multiple dipole localization analyses of spatio-temporal evoked magnetic fields was also performed using Multiple Signal Classification (MUSIC) (Sekihara, Poeppel, Marantz, Koizumi, & Miyashita, 1997; Moshier, Lewis, & Leahy, 1992) and beamformer algorithms (Moshier et al., 1992; Sekihara, Nagarajan, Poeppel, Marantz, & Miyashita, 2001; Robinson & Vrba, 1999). MUSIC methods are based on estimation of a signal “subspace” from entire spatio-temporal MEG data using singular-value decomposition (SVD). A version of the MUSIC algorithm, referred to as the “conventional” MUSIC algorithm, was implemented in MATLAB under the assumption that the sources contributing to the MEG data arose from multiple dipoles (<37 in number) which were stationary and were located within a spherical volume of uniform conductivity (Sekihara et al., 1997). The locations of dipoles are typically determined by conducting a search over a three-dimensional grid of interest within the head. Given the sensor positions and the coordinates of the origin of a “local sphere” approximation of the head shape for each subject, a lead-field matrix was computed for each point in this 3-D grid. At each point in this grid, from the lead-field matrix and the covariance matrix of the spatio-temporal MEG data, the value of a MUSIC localizer function can then be computed (Sekihara et al., 1997, Equation (4)). Maxima of this localizer function then corresponded to the location of dipolar sources. We used the MUSIC algorithm to compute the localizer functions over a three-dimensional grid in the hemisphere contralateral to the stimulus delivery. The localizer was computed using the evoked magnetic field responses over a time window of 1000 msec following the onset of the sentences. For each subject, at each point in a 3-D grid ( $-4 < x < 6$ ,  $0 < y < 8$ ,  $3 < z < 11$ ) in the left hemisphere, the localizer function was computed using the averaged evoked auditory magnetic field responses.

### Acknowledgments

This work was funded in part by Scientific Learning Co. and a grant from the Whitaker Foundation.

Reprint requests should be sent to John F. Houde, W. M. Keck Center for Integrative Neuroscience, University of California, San Francisco, 513 Parnassus Avenue, S-762, San Francisco, CA 94143, USA, or via e-mail: houde@phy.ucsf.edu.

### Note

1. In a recent study of M100 responses to tones, it was found that, near detection thresholds, a 15-dB change in SPL caused a 34% change in M100 RMS amplitude (Stufflebeam et al., 1998).

### REFERENCES

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 13367–13372.
- Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, *11*, 551–559.
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1999). The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *Neuroimage*, *10*, 448–459.
- Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice f0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, *103*, 3153–3161.
- Cooper, W. E. (1979). *Speech perception and production: Studies in selective adaptation*. Norwood, NJ: Ablex.
- Creutzfeldt, O., & Ojemann, G. (1989). Neuronal activity in the human lateral temporal lobe: III. Activity changes during music. *Experimental Brain Research*, *77*, 490–498.
- Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989a). Neuronal activity in the human lateral temporal lobe: I. Responses to speech. *Experimental Brain Research*, *77*, 451–475.
- Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989b). Neuronal activity in the human lateral temporal lobe: II. Responses to the subjects own voice. *Experimental Brain Research*, *77*, 476–489.
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*, 183–191.
- Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, *115*, 1753–1768.
- Diesch, E., & Luce, T. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology*, *34*, 501–510.
- Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America*, *70*, 45–50.
- Franklin, G. F., Powell, J. D., & Emami-Naeini, A. (1991). *Feedback control of dynamic systems* (2nd ed.). Reading, MA: Addison-Wesley.
- Gracco, V. L., Ross, D., Kalinowski, J., & Stuart, A. (1994). Articulatory changes following spectral and temporal modifications in auditory feedback. *Journal of the Acoustical Society of America*, *95*, 2821.
- Gunji, A., Hoshiyama, M., & Kakigi, R. (2000). Identification of auditory evoked potentials of one’s own voice. *Clinical Neurophysiology*, *111*, 214–219.
- Gunji, A., Kakigi, R., & Hoshiyama, M. (2000). Spatiotemporal source analysis of vocalization-associated magnetic fields. *Cognitive Brain Research*, *9*, 157–163.
- Hari, R., Aittoniemi, K., Jarvinen, M. L., Katila, T., & Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators. *Experimental Brain Research*, *40*, 237–240.
- Hari, R., & Makela, J. P. (1988). Modification of neuromagnetic responses of the human auditory cortex by masking sounds. *Experimental Brain Research*, *71*, 87–92.
- Hein, A. V., & Held, R. (1962). A neural model for labile

- sensorimotor coordination. In E. Bernard & M. Hare (Eds.), *Biological prototypes and synthetic systems* (vol. 1, pp. 71–74). New York: Plenum.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1996). Cortical speech processing mechanisms while vocalizing visually presented languages. *NeuroReport*, 8, 363–367.
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, 8, 2379–2382.
- Hirano, S., Naito, Y., Okazawa, H., Kojima, H., Honjo, I., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1997). Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Experimental Brain Research*, 113, 75–80.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.
- Jacobs, O. L. R. (1993). *Introduction to control theory* (2nd ed.). Oxford, UK: Oxford University Press.
- Jeannerod, M. (1988). *The neural and behavioural organisation of goal-directed movements*. Oxford, UK: Oxford University Press.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, 26, 235–258.
- Kawahara, H. (1993). Transformed auditory feedback: Effects of fundamental frequency perturbation. *Journal of the Acoustical Society of America*, 94, 1883.
- Kuriki, S., Mori, T., & Hirata, Y. (1999). Motor planning center for speech articulation in the normal human brain. *NeuroReport*, 10, 765–769.
- Kuriki, S., & Murase, M. (1989). Neuromagnetic study of the auditory responses in right and left hemispheres of the human brain evoked by pure tones and speech sounds. *Experimental Brain Research*, 77, 127–134.
- Lane, H., & Tranel, B. (1971). The lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14, 677–709.
- Lee, B. S. (1950). Some effects of side-tone delay. *Journal of the Acoustical Society of America*, 22, 639–640.
- Leonardo, A., & Konishi, M. (1999). Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, 399, 466–470.
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, 92, 204–214.
- Mahncke, H. (1998). *Oscillations and temporal processing in sensory cortex*. Unpublished doctoral dissertation, University of California San Francisco.
- McCasland, J. S., & Konishi, M. (1981). Interaction between auditory and motor activities in an avian song control nucleus. *Proceedings of the National Academy of Sciences U.S.A.*, 78, 7815–7819.
- McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, 119, 907–917.
- Mosher, J. C., Lewis, P. S., & Leahy, R. M. (1992). Multiple dipole modeling and localization from spatio-temporal meg data. *IEEE Transactions on Biomedical Engineering*, 39, 541–557.
- Muller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research*, 215, 61–76.
- Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, 272, 29–32.
- Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, 265, 119–122.
- Papanicolaou, A. C., Raz, N., Loring, D. W., & Eisenberg, H. M. (1986). Brain stem evoked response suppression during speech production. *Brain and Language*, 27, 50–55.
- Perkell, J. (1997). Articulatory processes. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 333–370). Oxford, UK: Blackwell Publishers.
- Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., & Trujillo, N. J. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neuro-otology*, 4, 64–79.
- Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K., & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked m100 neuromagnetic field elicited by speech sounds. *Brain Research. Cognitive Brain Research*, 4, 231–242.
- Reite, M., Adams, M., Simon, J., Teale, P., Sheeder, J., Richardson, D., & Grabbe, R. (1994). Auditory m100 component: 1. Relationship to heschl's gyri. *Brain Research. Cognitive Brain Research*, 2, 13–20.
- Richer, F., Alain, C., Achim, A., Bouvier, G., & Saint-Hilaire, J. M. (1989). Intracerebral amplitude distributions of the auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, 74, 202–208.
- Robinson, S., & Vrba, J. (1999). Functional neuroimaging by synthetic aperture magnetometry (SAM). In Yoshimoto et al. (Eds.), *Recent advances in biomagnetism* (pp. 302–305). Sendai, Japan: Tokyo University Press.
- Scherg, M., & Von Cramon, D. (1985). Two bilateral sources of the late aep as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology*, 62, 32–44.
- Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, 65, 344–360.
- Schmidt, M. F., & Konishi, M. (1998). Gating of auditory responses in the vocal control system of awake songbirds. *Nature Neuroscience*, 1, 513–518.
- Sekihara, K., Nagarajan, S. S., Poeppel, D., Marantz, A., & Miyashita, Y. (2001). Reconstructing spatio-temporal activities of neural sources using an meg vector beamformer technique. *IEEE Transactions on Biomedical Engineering*, 48, 760–771.
- Sekihara, K., Poeppel, D., Marantz, A., Koizumi, H., & Miyashita, Y. (1997). Noise covariance incorporated meg-music algorithm: A method for multiple-dipole estimation tolerant of the influence of background brain activity. *IEEE Transactions on Biomedical Engineering*, 44, 839–847.
- Stevens, K. N. (1999). *Acoustic phonetics*. Cambridge: MIT Press.
- Stufflebeam, S. M., Poeppel, D., Rowley, H. A., & Roberts, T. P. (1998). Peri-threshold encoding of stimulus

- frequency and intensity in the m100 latency. *NeuroReport*, *9*, 91–94.
- Suga, N., & Schlegel, P. (1972). Neural attenuation of responses to emitted sounds in echolocating rats. *Science*, *177*, 82–84.
- Suga, N., & Shimozawa, T. (1974). Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science*, *183*, 1211–1213.
- von Bekesy, G. (1949). The structure of the middle ear and the hearing of one's own voice by bone conduction. *Journal of the Acoustical Society of America*, *21*, 217–232.
- Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary observations on tickling oneself. *Nature*, *230*, 598–599.
- Woods, D. L., Knight, R. T., & Neville, H. J. (1984). Bitemporal lesions dissociate auditory evoked potentials and perception. *Electroencephalography and Clinical Neurophysiology*, *57*, 208–220.
- Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60*, 213–232.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*, 846–849.