

Published in final edited form as:

Neuroreport. 2006 September 18; 17(13): 1375–1379. doi:10.1097/01.wnr.0000233102.43526.e9.

Magnetoencephalographic evidence for a precise forward model in speech production

Theda H. Heinks-Maldonado^a, Srikantan S. Nagarajan^b, and John F. Houde^a

^aDepartment of Otolaryngology, University of California, San Francisco, California, USA

^bDepartment of Radiology, University of California, San Francisco, California, USA

Abstract

Electroencephalography and magnetoencephalography studies have shown that auditory cortical responses to self-produced speech are attenuated when compared with responses to tape-recorded speech, but that this attenuation disappears if auditory feedback is altered. These results suggest that auditory feedback during speaking is processed by comparing the feedback with its internal prediction. The present study used magnetoencephalography to investigate the precision of this matching process. Auditory responses to speech feedback were recorded under altered feedback conditions. During speech production, the M100 amplitude was maximally reduced to the participants' own unaltered voice feedback, relative to pitch-shifted and alien speech feedback. This suggests that the feedback comparison process may be very precise, allowing the auditory system to distinguish between internal and external sources of auditory information.

Keywords

altered feedback; efference copy; magnetoencephalography; speech production

Theoretical background

Data from a number of experiments suggest that the central nervous system processes sensory feedback from motor acts by comparing the feedback with a prediction of that feedback. This prediction, or corollary discharge, is the output of an internal (forward) model that maps motor commands (efference copy) to their expected sensory consequences. If feedback matches the prediction, the sensory response is suppressed.

Electroencephalography and magnetoencephalography (MEG) recordings have shown that auditory cortical responses to self-produced speech are attenuated when compared with responses to tape-recorded speech [1–5]. Similar phenomena are seen in the somatosensory system, where behavioral studies (e.g. [6]) provide evidence for a precise forward model in which sensory stimulation has to correspond accurately to the movement producing it in

© 2006 Lippincott Williams & Wilkins.

Correspondence and requests for reprints to Theda H. Heinks-Maldonado, PhD, Department of Otolaryngology, University of California, San Francisco, 513 Parnassus Avenue, HSE 800, Box 0444, San Francisco, CA 94143, USA, Tel: + 415 476 2512; theda_heinks@yahoo.de.

order to attenuate its perception, with the amount of perceptual attenuation being proportional to the accuracy of the sensory prediction.

An approach to testing the precision of a forward model in the auditory system is to manipulate the re-afferent auditory feedback that participants hear as they produce speech. Evidence exists from positron emission tomography studies [7] that during talking unaltered and altered auditory feedback (either by distortion or time delay) activates different brain regions. An event-related potential (ERP) study showed results that support evidence that sensory suppression indeed occurs and is maximal when the auditory feedback during speech production is unaltered [8]. Positron emission tomography studies are, however, not able to reveal the temporal dynamics of activity on a millisecond scale like MEG, and ERP studies lack the spatial resolution that MEG can provide.

The present MEG experiment was based on the finding of the previous ERP study [8]. In that study it was concluded that the suppression of the N100 represents auditory cortex activity being suppressed. It has, however, been proposed [9] that the N100 has several different possible areas of generation, even outside auditory cortex. Thus, the conclusion in the ERP study is partly an assumption based on results of animal and MEG studies, which actually localized the reduced activation during vocalization to that area of cortex.

Accordingly, our main goal was to investigate whether the suppression effect seen in the ERP study could actually be localized to auditory cortex. We further wanted to advance findings of a previous MEG study [4], in which it was found that M100 suppression during speaking relative to listening was abolished when participants heard white noise instead of the expected voice feedback. Although this result may show some evidence for specificity of the cortical response, white noise is far different from speech and produces widespread cortical activation and therefore does not necessarily show the precision of the sensory attenuation during speaking. Thus, with this study we intended to design a more thorough investigation of the precision of the forward model. To this end, we altered the re-afferent auditory feedback associated with self-produced speech, allowing us to examine the degree to which suppression of the auditory cortical response depends on the closeness of the match between the auditory feedback and the predicted feedback.

Materials and methods

Subjects and testing procedure

We recorded MEG from 14 right-handed men (mean age 34.7 years; range: 22–52) who were fluent English speakers and naïve to the purpose of this study. The participants were screened for past or present health problems that could bias the results. After signing informed consent, participants were run in two separate procedures: a structural magnetic resonance scan to acquire anatomical data for magnetic source localization and the MEG experiment.

The audio setup

In the booth, a directional microphone, to record speech produced by the participant, was placed at a safe distance from the scanner where it did not distort the recorded magnetic

fields. The microphone signal was fed through an amplifier to the analog input of a PC running Linux (Red Hat, Inc., Raleigh, North Carolina, USA). This PC ran a realtime signal processing program that detected the participant's voice onsets and, when so directed, played out to the analog output a recording of a pre-recorded speech sample of a male voice (alien voice) at the time of these onsets. Otherwise, this program simply passed the participant's voice to the analog output with a 5-ms delay. This analog output was fed to a digital signal processor (Ultraharmonizer 7000, Eventide, Inc., Little Ferry, New Jersey, USA) that, when so directed, would lower the pitch of the incoming speech by two semitones. When not lowering the pitch by two semitones, the Ultraharmonizer lowered the pitch of the incoming speech by 0.3 semitones, in order to make the speech fed back to the participant sound like they were hearing their own voice [10] (bone conduction causes speakers to hear their own speech with low frequencies more emphasized than what others hear [11]).

The pitch-shifted speech signal was output from the Ultraharmonizer with a 15-ms delay to a signal selector whose output fed back, via an attenuator and an amplifier, to the participant's earphones, where it was binaurally presented. The signal selector allowed the participant's earphone signal to be blocked or to come from either of three different sources: (1) his speech feedback from the Ultraharmonizer, (2) playback of his previously recorded feedback or (3) instructions from the experimenter, via a microphone and amplifier.

Experimental procedure

The MEG experiment comprised one run of the speaking and one run of the listening task, with each run consisting of 240 trials. In the speaking task, participants were told to utter a short [a:] about every 5s. The feedback participants heard over headphones was varied from trial to trial between their own unaltered voice (self-unaltered), their own voice pitch-shifted downward by two semitones (self-pitch-shifted), the alien unaltered voice (alien unaltered) or the alien voice pitch-shifted downward by two semitones (alien pitch-shifted). Thus, 60 trials were performed during each condition of the speaking and the listening tasks. Every 15 trials, the presentation of stimuli on the monitor stopped and indicated a break, which the participant was able to terminate via button press.

Each trial started with the appearance of three white dots in the center of a black screen. The dots disappeared one by one to simulate a countdown to prompt the participant to vocalize. In order to avoid an overlap of visual and auditory responses, the participants were instructed to speak after the disappearance of the visual cue (the last of the three dots) on the screen. Assuming an average vocal response time of about 200–500 ms, the auditory M100 would then be relatively free of the influence of visual magnetic response components.

In the listening task, the recorded feedback sounds from the speaking task were played back, and participants were instructed to merely listen. All other features remained the same as in the speaking task including use of the same visual prompts and earphone volume.

Data acquisition and processing

MEG data were collected from a 275-channel sensor array (VSM Medtech Ltd., Coquitlam, British Columbia, Canada) throughout the predefined trial length of 5s. During the experiment, trial type was encoded by two analog signals that were recorded with the data

set. The pitch-shift signal indicated whether the trial was pitch-shifted by two semitones or not. The source signal indicated whether the participant heard the alien voice or his own voice. During data analysis, these signal transitions were identified and used to mark the occurrence of two different event markers in the recorded MEG data. The onset of the speech sound was identified and served as the marker for the beginning of the epoch. Before averaging, data were band-pass filtered 2–20 Hz; the linear trend was removed and the third gradient of the magnetic field was calculated. Data were visually inspected and trials containing artifacts were removed from further processing and analysis.

For each condition, the root mean square (RMS) responses at each detector in the left and the right temporal areas were averaged across trials. The average was created for the timewindow reaching from 200 ms prespeech onset to 800 ms postspeech onset. From this average time course data, the magnetic response M100 was measured. In order to detect the different peak amplitudes, a script written in MATLAB (The Mathworks, Natick, Massachusetts, USA) was created, which searched for the largest RMS peak in a predefined time interval in the MEG averages. If no peak was found, the maximum voltage at which it was detected was recorded instead. The M100 was defined as the largest peak between 80 and 120 ms.

Statistics

M100 amplitudes were analyzed in a three-way repeated-measures analysis of variance (ANOVA) including factors of task (speaking, listening), condition (self unaltered, self pitch-shifted, alien unaltered, alien pitch-shifted) and laterality (left, right). The RMS average of 34 temporal sensors per hemisphere was included in the analysis. All results were Greenhouse-Geisser corrected for nonsphericity, when appropriate.

Neuromagnetic source reconstruction

The analysis toolbox NUTMEG (Neurodynamic Utility Toolbox for magnetoencephalography [12]) was used to reconstruct the spatiotemporal dynamics of neural activations, which were then overlaid on structural magnetic resonance images. The lead field matrix and signal components of the sensor covariance matrix were used to transform the sensor data into source space in which the activations at different locations in cortex were estimated using adaptive spatial filters [13].

Results were plotted as maps of cortical activation at the time of the M100. For the activation maps, coregistration of matching fiducials was used to compute an affine transform that projected the source activations onto the anatomical magnetic resonance imaging of the individual subject. To facilitate visual comparisons across conditions, all activation maps used the same thresholds and color scales for all conditions.

Results

Evoked magnetic fields

The ANOVA of the average M100 responses with the factors task, laterality and condition resulted in significant main effects for task [$F(1,13)=11.03$; $P=0.006$] and condition

[$F(3,39)=5.51$; $P=0.008$]. The significant interaction of task by condition [$F(3,39)=3.67$; $P=0.036$] was examined further in two separate follow-up ANOVAs for the speaking and the listening tasks.

The ANOVA for the listening task did not result in any significant main effects or interactions. In the analysis of the speaking task, however, the factor condition became highly significant [$F(3,39)=10.73$; $P=0.001$].

Post-hoc simple contrasts were calculated for the speaking task to compare the mean of the self-unaltered condition with the mean of each of the other three altered speech conditions. After conservative Bonferroni correction for multiple comparisons, the alien feedback conditions were found to be significantly different from the self-unaltered condition [vs. alien unaltered $F(15.22)$; $P=0.002$; vs. alien pitch-shifted $F(21.47)$; $P<0.001$]. The P -value of the comparison of self-unaltered vs. self-pitch-shifted ($P=0.017$) scarcely missed the corrected threshold of $P=0.0166$ but would reach significance with a less conservative correction.

Figure 1 illustrates these results showing the smallest RMS response during the self-unaltered condition during speaking.

An ANOVA performed on the M100 peak latencies showed no significant effects.

Neuromagnetic source reconstruction

Both tasks and all conditions were analyzed in the NUTMEG to determine the precise source of activation. NUTMEG found a focal peak of activation at approximately 100 ms, which corresponded to activation in the auditory cortex. For all speaking conditions, the anatomical magnetic resonance imaging of a representative participant with the brain areas of activation are shown in Fig. 2. In the listening task, the intensity of activation in auditory cortex was much larger and very similar for all conditions.

Discussion

We tested the hypothesis that a precise forward model operates in the auditory system during speech production causing maximal suppression of the auditory cortical response to the incoming sounds that most closely match the speech sounds predicted by the model (i.e. corollary discharge). In order to test the hypothesized selectivity of this feed-forward auditory cortical suppression mechanism, we manipulated the re-afferent auditory feedback that participants heard as they produced speech and assessed its evoked auditory cortical M100 response. During speech production, the M100 amplitude was maximally reduced to the participant's own unaltered voice feedback, relative to the pitch-shifted and alien speech feedback. The different feedback types did not lead to differences in M100 amplitude during the listening task.

Most importantly, however, we were able to localize the source of the speech-related cortical activation and suppression to auditory cortex. It was possible to visualize and verify that the extent of auditory cortex activation varies with the degree of match between the internally predicted and the actual auditory feedback.

The results of this study are consistent with our previous results [8]. A difference, however, was that in our ERP study the suppression effect was most pronounced and statistically significant in the left hemisphere while in this study, the suppression was found bilaterally. As mentioned above, the ERP component N100 consists of signals from different sources as compared with the M100, which represents only the supratemporal source. Thus, a different or an additional N100 source, registered with electroencephalography but not MEG, could have led to this difference.

The results of our study presented here extend further the findings of the MEG study by Houde *et al.* [4]: Firstly, we did not rely on the potentially problematic direct comparison between spoken and played-back speech and secondly we used altered speech instead of white noise (for a discussion of these points see also [8]). By changing the reafferent auditory feedback during speech production and keeping all other parameters of the forward model constant, we were able to show selective effects of a precise forward model within the speaking task itself. There was, however, still partial suppression for the altered feedback cases, while the Houde *et al.* study found no suppression to the altered feedback (white noise). This could be due to the extreme auditory activating characteristics of Houde *et al.*'s gated noise feedback alteration, or it could be due to a difference in types of altered feedback. In the present study, all the forms of altered feedback were still speech, while in the Houde *et al.* study, the altered feedback was nonspeech. Perhaps the feedback prediction generated by the forward model includes a categorical feature indicating that speech is expected, and a match with incoming feedback on this feature alone creates some amount of suppression of the auditory response.

As described above, our 'unaltered' feedback condition was actually a 0.3 semitone pitch-shift down of participants' feedback, which we used to better match the perception of the sound of their own voice [10]. The fact that we observed large suppression for our slight downshift of pitch suggests two possibilities: The 0.3 semitone pitch-shift may have indeed matched more closely the participants' subjective perception of their own speech and thus the prediction of the forward model. On the other hand, it could be that the forward model is not sensitive to 0.3 semitone changes in pitch and that we would have found the same results without changing the participants' pitch by 0.3 semitones.

It may also be that the suppression effect is saturated for large feedback alterations but has a more continuous nature in the range of small feedback alterations, in which the degree of suppression more closely matches the degree of feedback alteration. This has to be addressed by future studies.

Conclusion

In summary, our findings of selective suppression of auditory cortex to unaltered self-produced speech support a precise forward model mechanism. We assume this precise forward model mechanism that modulates auditory cortical responsiveness during speech is similar to the forward model systems shown to modulate somatosensory responses to self-generated motor acts [14–16]. Precise auditory suppression during speech allows the auditory system to distinguish between internal and external sources of auditory

information. Auditory information from selfproduced speech may be used for feedback control of ongoing speech (e.g. loudness or prosody of speech), whereas externally generated auditory information may be used primarily to recognize environmental events. Further studies are needed to explore the potential clinical relevance of feedback processing deficits in stuttering or diseases like Parkinson's and schizophrenia.

Acknowledgments

Sponsorship: This work was supported by a grant from the National Institute on Deafness and other Communication Disorders (R01DC006436).

References

1. Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci Lett*. 1999; 272:29–32. [PubMed: 10507535]
2. Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett*. 1999; 265:119–122. [PubMed: 10327183]
3. Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp*. 2000; 9:183–191. [PubMed: 10770228]
4. Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of auditory cortex during speech: an MEG study. *J Cogn Neurosci*. 2002; 14:1125–1138. [PubMed: 12495520]
5. Ford JM, Mathalon DH, Heinks T, Kalba S, Faustman WO, Roth WT. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *Am J Psychiatry*. 2001; 158:2069–2071. [PubMed: 11729029]
6. Blakemore SJ, Frith CD, Wolpert DM. Spatio-temporal prediction modulates the perception of self-produced stimuli. *J Cogn Neurosci*. 1999; 11:551–559. [PubMed: 10511643]
7. Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, et al. Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*. 1997; 8:2379–2382. [PubMed: 9243644]
8. Heinks-Maldonado T, Mathalon D, Gray M, Ford J. Fine-tuning of auditory cortex during speech production. *Psychophysiology*. 2005; 42:180–190. [PubMed: 15787855]
9. Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*. 1987; 24:375–425. [PubMed: 3615753]
10. Shuster LI, Durrant JD. Toward a better understanding of the perception of self-produced speech. *J Commun Disord*. 2003; 36:1–11. [PubMed: 12493635]
11. Shearer WM. Self-masking effects from live and recorded vowels. *J Aud Res*. 1978; 18:213–219. [PubMed: 755818]
12. Dalal SS, Zumer JM, Agrawal V, Hild KE, Sekihara K, Nagarajan SS. NUTMEG: a neuromagnetic source reconstruction toolbox. *Neurol Clin Neurophysiol*. 2004; 52:1–4.
13. Sekihara K, Nagarajan SS, Poeppel D, Marantz A, Miyashita Y. Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Trans Biomed Eng*. 2001; 48:760–771. [PubMed: 11442288]
14. Jeannerod, M. The neural and behavioural organisation of goal-directed movements. Oxford, UK: Oxford University Press; 1988.
15. Jeannerod M. The mechanism of self-recognition in humans. *Behav Brain Res*. 2003; 142:1–15. [PubMed: 12798261]
16. Wolpert D, Flanagan J. Motor prediction. *Curr Biol*. 2001; 11:729–732.

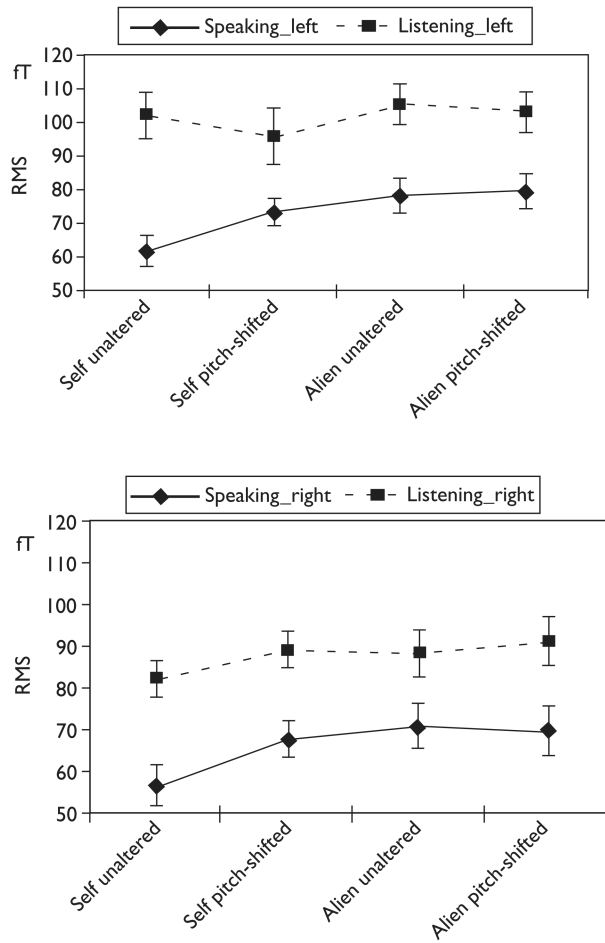


Fig. 1. Mean M100 responses to the different feedback conditions during speaking and listening. Top graph: left temporal sites; bottom graph: right temporal sites. RMS, root mean square.

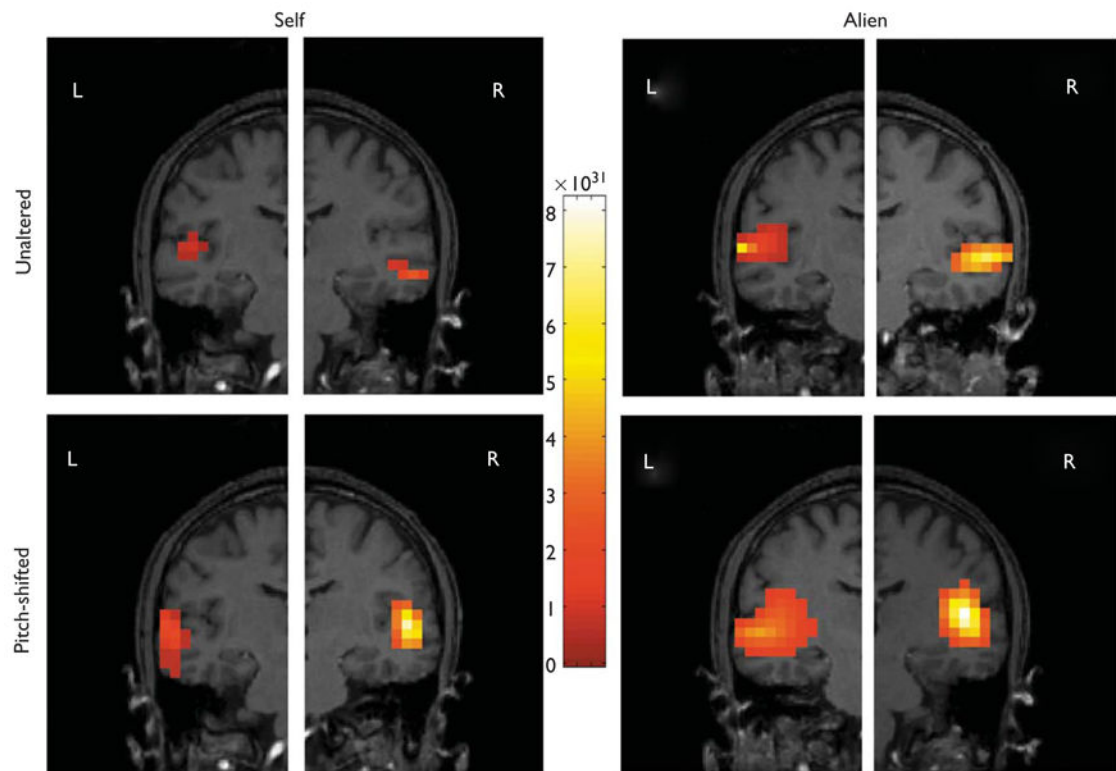


Fig. 2.

Localization of the peak magnetic response in the left and right hemispheres during speaking. Position coordinates (x, y, z) in mm: Self-unaltered left: $-45.0, -15.0, -15.0$; self-unaltered right: $55.0, -15.0, -30.0$; self-pitch-shifted left: $-55.0, -15.0, -20.0$; self-pitch-shifted right: $50.0, -10.0, -15.0$; alien unaltered left: $-60.0, -10.0, -15.0$; alien unaltered right: $55.0, -5.0, -20.0$; alien pitch-shifted left: $-45.0, -10.0, -20.0$; alien pitch-shifted right: $45.0, -15.0, -10.0$