

# Speaking Modifies Voice-Evoked Activity in the Human Auditory Cortex

Gabriel Curio,<sup>1,2\*</sup> Georg Neuloh,<sup>1,2</sup> Jussi Numminen,<sup>1</sup> Veikko Jousmäki,<sup>1</sup>  
and Riitta Hari<sup>2,3</sup>

<sup>1</sup>Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology,  
Espoo, Finland

<sup>2</sup>Neurophysics Group, Department of Neurology, Klinikum Benjamin Franklin,  
Freie Universität, Berlin, Germany

<sup>3</sup>Department of Clinical Neurosciences, Helsinki University Central Hospital, Helsinki, Finland

---

**Abstract:** The voice we most often hear is our own, and proper interaction between speaking and hearing is essential for both acquisition and performance of spoken language. Disturbed audiovocal interactions have been implicated in aphasia, stuttering, and schizophrenic voice hallucinations, but paradigms for a noninvasive assessment of auditory self-monitoring of speaking and its possible dysfunctions are rare. Using magnetoencephalography we show here that self-uttered syllables transiently activate the speaker's auditory cortex around 100 ms after voice onset. These phasic responses were delayed by 11 ms in the speech-dominant left hemisphere relative to the right, whereas during listening to a replay of the same utterances the response latencies were symmetric. Moreover, the auditory cortices did not react to rare vowel changes interspersed randomly within a series of repetitively spoken vowels, in contrast to regular change-related responses evoked 100–200 ms after replayed rare vowels. Thus, speaking primes the human auditory cortex at a millisecond time scale, dampening and delaying reactions to self-produced "expected" sounds, more prominently in the speech-dominant hemisphere. Such motor-to-sensory priming of early auditory cortex responses during voicing constitutes one element of speech self-monitoring that could be compromised in central speech disorders. *Hum. Brain Mapping* 9:183–191, 2000.

© 2000 Wiley-Liss, Inc.

**Key words:** magnetoencephalography; auditory; speech; vowel; human

---

## INTRODUCTION

Through our lifetime, we are exposed to WYHIWYS ("What You Hear Is What You Said") events. This

repeated input can imprint our auditory system so as to optimally process our own utterances. In this respect, single unit recordings in primary auditory cortex have identified specific response patterns related to replayed phonemes [Steinschneider et al., 1994; Wang et al., 1995], and, in particular, neurons in the nonprimary auditory cortex have been shown to prefer increasingly complex acoustic stimuli such as communication calls uttered by members of the social group [Rauschecker et al., 1995]. Therefore, auditory cortices hold a key position also for the self-monitoring of a speaker's own ongoing speech which is nec-

---

Contract grant sponsor: HCM Program.

\*Correspondence to: Dr. Gabriel Curio, Neurophysics Group, Department of Neurology, Klinikum Benjamin Franklin, Freie Universität, 12200 Berlin, Germany. e-mail: curio@zedat.fu-berlin.de

G. Neuloh is now at the Department of Neurosurgery, University of Bonn, 53127 Bonn, Germany

Received for publication 17 September 1999; accepted 21 December 1999

© 2000 Wiley-Liss, Inc.

essary for normal speech acquisition of infants [Deal and Haas, 1996]. Such monitoring might utilize a circuitry that matches motor speech output (“efference copy”) with its auditory sequelae (“reafference”) [Sperry, 1950; von Holst and Mittelstaedt, 1950]. Artificial interference with this matching (e.g., by delayed or deformed auditory feedback) [Chapin et al., 1981; Hirano et al., 1997; Houde and Jordan, 1998; Lee, 1950], can strongly affect speech even in normal adults. One possible mechanism for motor-to-sensory priming has been described in auditory cortices of subhuman primates where about 50% of call-responsive neurons were inhibited during phonation [Müller-Preuss and Ploog, 1981]. Some human cerebral blood flow (CBF) studies have been interpreted to indicate speech-related inhibition in human auditory cortices in parallel to these monkey data [Hirano et al., 1997; Paus et al., 1996; Wise et al., 1999] but others described undiminished activation during speech [e.g., McGuire et al., 1996].

The CBF studies evaluate task-specific regional flow modulations integrated over tens of seconds. We therefore recorded voicing-related auditory evoked fields exploiting the high time resolution of magnetoencephalography (MEG) to elucidate cortical sensorimotor interaction in humans at the millisecond time-scale, typical for ongoing speech [Hämäläinen et al., 1993; Hari, 1990]. The MEG approach allows analysis of fast intermediate steps in the cerebral processing of single phonemes. We report a transient voicing-related activation of the human auditory cortex. Specifically, we demonstrate interactions between phonation and audition to occur within the first 200 ms following a self-uttered vowel: the very act of speaking modifies responses generated in the auditory cortex in both hemispheres, but does so more strongly in the speech-dominant left hemispheres of our right-handed subjects.

## MATERIALS AND METHODS

### Subjects and task

Eight right-handed volunteers (five male, three female; 23–39 years) were studied after informed consent. They were first familiarized to a random series of frequent low (80%; 1000 Hz) and rare high (20%; 1400 Hz) squarewave sounds, presented once every 1 sec. They were then asked to utter two vowels [frequent /a/ and rare /ae/] in a comparable series but without counting. To minimize breathing artifacts, 15–25 vowels were uttered within each expiration phase, followed by a comfortable resting interval. This SPEECH

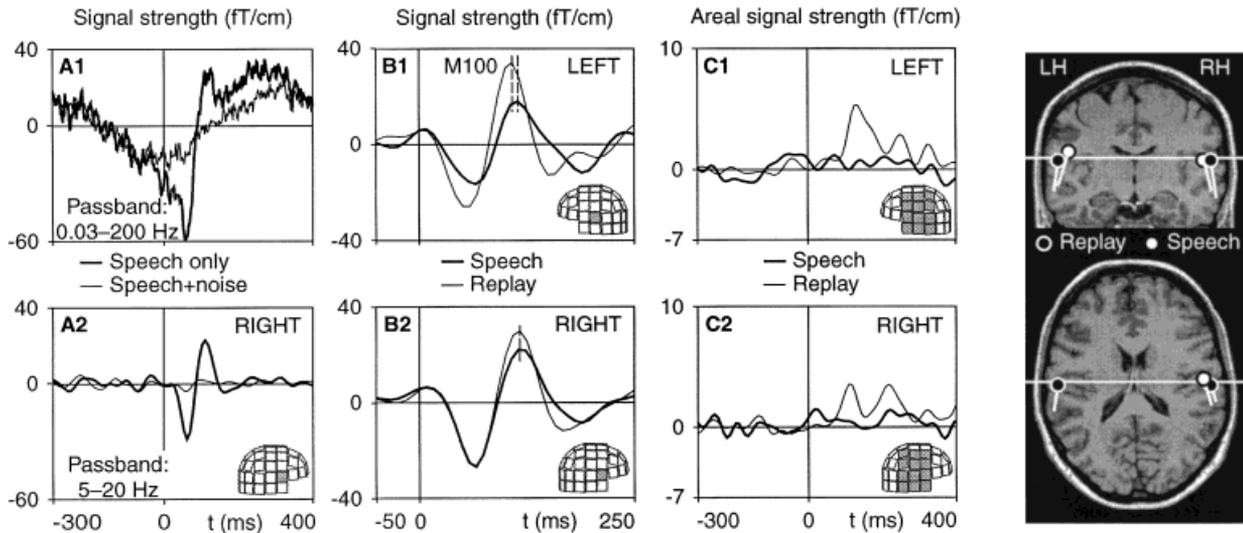
task was stopped by the experimenter after 100 rare vowels. The voice of the subject was recorded continuously on a digital audiotape (10 Hz–20 kHz; 48 kHz sampling frequency). Prior to the next condition with binaural REPLAY of these tape-recorded vowels, each subject adjusted the input volume to equalize the perceived loudness of her replayed vowels and her own voicing; the subjects optimized this comparison by speaking vowels syncopatingly between REPLAYed vowels. Then subjects were told to listen silently and with intentionally fully relaxed articulators (tongue, mandible) to the complete REPLAYed vowel series and to estimate, again without counting, whether they had performed the task as required.

### Magnetoencephalography

During both the SPEECH and REPLAY conditions, neuromagnetic fields were recorded continuously inside a magnetically shielded room with a helmet-shaped 122-channel magnetometer covering the whole scalp. The subjects were sitting upright with their heads comfortably resting against the posterior inner wall of the helmet [Ahonen et al., 1993]. The MEG recording passband was 0.03–200 Hz and the sampling rate 609 Hz. In addition to the DAT voice records used for REPLAY, the subject’s voice was recorded in an extra electric channel of the MEG acquisition system. Because this voice record was obtained simultaneously with the MEG and stored together in one data file, trigger points indicating the occurrence of vowels in relation to the ongoing MEG could be defined (see below). Because the MEG passband covered the range of fundamental voice frequencies for all speakers, it was used also for this audio record.

### Data analysis

The triggers needed for averaging the MEG responses to vowel onsets were defined as time points when the rectified audio record signal rose significantly above its noise level. These triggers were found to lag the real vowel onset—as defined by an averaged vowel signal from the audio channel—by at least 10 ms. This systematic trigger delay did not affect the analysis presented below because the trigger points were identical for SPEECH and REPLAY conditions, as they were determined from the very same audio record. Furthermore, for REPLAY, the additional delay caused by sound conduction from the mini-loudspeaker through a plastic tube to the earpieces was taken into account in the off-line analysis. Selective



**Figure 1.**

Averaged magnetic brain responses to vowels. Signals from single sensors in **(A)** and **(B)**, and spatial average from 12 sensor locations in **(C)**; see shadowed areas on the helmet surfaces. **(A1)** Wide-band auditory evoked responses of one subject to self-uttered vowels from right temporal area; thick lines illustrate responses without masking and thin lines with binaural noise masking. **(A2)** The same responses as in **(A1)** but band-pass filtered through 5–20 Hz. **(B)** Band-pass filtered responses of the same subject as in **(A)** to self-uttered (thick line; SPEECH) and replayed (thin line; REPLAY) vowels from the left **(B1)** and right

**(B2)** temporal areas. **(C)** Change-related magnetic responses evoked by the rare (20%) /ae/ vowels embedded in a series of frequent (80%) /a/ vowels during the REPLAY (thin line) and SPEECH (thick line) conditions from the left **(C1)** and right **(C2)** hemispheres. The responses are grand averages (eight subjects) of the individual spatial averages of the difference field amplitude (response to /ae/—response to /a/). Right panels: Generator sites of the M100 responses to self-uttered (white) and replayed vowels (black) displayed for one subject, superimposed on her MRI slices. Responses were modeled with current dipoles [Hari, 1990].

subaverages were then calculated for the frequent /a/ vowels and the rare /ae/ vowels. The MEG responses were band-pass filtered at 5–20 Hz to remove slow artifacts arising from head, breathing or articulator movements and from slow motor-related brain activities (Fig. 1A). The magnitudes of the local field gradients were calculated as spatial averages from 12 temporal-lobe sensor locations over the right and the left hemispheres (inset in Fig. 1C).

The first major auditory response peaking later than 80 ms following the vowel onset was labeled M100 in each of these two averaged signals. The right and left hemisphere M100 amplitudes were determined with respect to a prestimulus baseline, defined as the mean level between –300 ms and –100 ms prior to vowel onset. Given the narrow passband, M100 may include overlap from immediately preceding and following responses. M100 peak latencies and amplitudes were determined for each subject, hemisphere, and condition, and were first evaluated using a two-factor ANOVA with Condition (REPLAY vs. SPEECH) and Hemisphere (left vs. right) as factors; the final evaluation (planned comparisons) was performed using

two-tailed Wilcoxon signed-ranks tests for dependent variables.

Additional (“change-related”) responses to infrequent vowels were determined from difference curves where responses to frequent /a/ vowels had been subtracted from responses to rare /ae/ vowels; these signals were filtered at 1–20 Hz. Because of the low signal-to-noise ratio, the amplitudes for change-related responses were defined here as mean values 100–200 ms following vowel onset; this time window also covered a fair amount of possible peak latency jitter between conditions and subjects.

Because the MEG helmet system comprises planar gradiometers that detect maximal signals directly over a current dipole [Ahonen et al., 1993], M100 is supposed to be generated beneath the temporal-lobe channels showing the most prominent responses. In addition, equivalent current dipoles (ECDs) were used to model the “centers of gravity” of the distributed neuronal populations contributing to the M100 response [Hari, 1990]. Two ECDs, one for each hemisphere, were fitted independently to the field distributions in a subset of 24 channels over each temporal

lobe at the latency of the M100 peak of the mean field power in this channel subset; the time points for ECD fitting were determined separately for SPEECH and REPLAY recordings. For source identification [Hämäläinen et al., 1993], spherical head models were individually determined on the basis of magnetic resonance images (MRIs) that were available for seven of the eight subjects.

## RESULTS

### Speech performance

In the SPEECH condition, each subject was instructed to utter self-paced frequent vowels /a/ (target frequency: 80%) with randomly interspersed rare vowels /ae/ (20%) at a rate of about one vowel per second. They uttered vowels with an average time period from vowel onset to vowel onset of 700 ms, a vowel duration of about 200 ms, and an average occurrence of 18.9% deviant vowels (range 15.2–24.5%).

### MEG responses [“M100”]

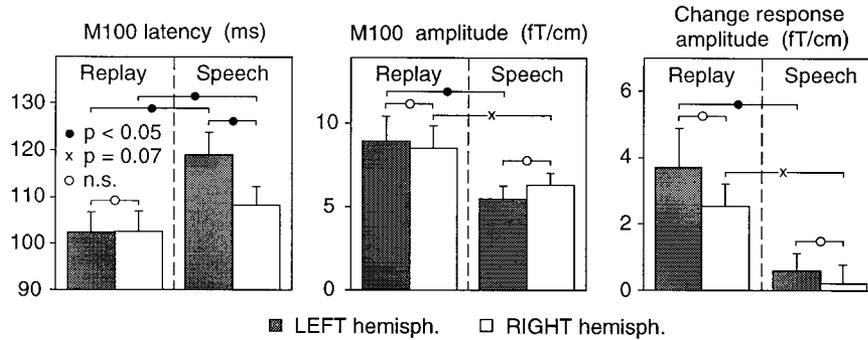
When MEG signals were averaged with respect to onsets of the REPLAYed vowels, typical M100 responses, peaking about 100 ms after the onset of frequent /a/ vowels, were observed over both temporal lobes of all eight subjects. As a novel finding, comparable MEG responses, time-locked to voice onset, were detected also during the SPEECH condition when the subjects themselves uttered the vowels; similar responses were obtained also to consonant-vowel syllables /ta/ in three subjects studied.

In control experiments, continuous binaural high-intensity white noise was used to mask the voice. The experimenters observed that all three subjects studied involuntarily increased slightly the volume of their voice during noise presentation; this consistent observation can be interpreted as the well-established “Lombard effect” [Pick et al., 1989]. Although the auditory voice feedback was therefore even louder than in the main experiment, the M100 responses disappeared completely in the one subject who perceived complete masking of his own voice (Fig. 1A) and was greatly reduced in the other two subjects who perceived only partial masking with the highest noise intensity which they still considered tolerable. Thus, the M100 responses observed during the noise-free SPEECH condition were evidently evoked via the auditory feedback pathway and not e.g., by phasic motor-related activity or by

somatosensory feedback from speech muscles; the latter two signals should not have been reduced by the masking noise.

In 11 of the 16 hemispheres of our eight subjects, it was possible to fit equivalent current dipoles to the vowel-evoked M100 field distributions during both the SPEECH and the REPLAY conditions; 13 of these 22 dipoles explained > 90% of field variance across all 122 channels, another eight dipoles explained between 85% and 90%, and one 76%. In both the SPEECH and the REPLAY condition, the neuronal generators of M100, modeled as equivalent current dipoles, were located in the auditory cortices bilaterally (Fig. 1, right). The dipole locations were compared using a repeated-measures ANOVA with Conditions (REPLAY, SPEECH, squarewaves sounds) as within-subjects factor; each coordinate (x, y, z) within each hemisphere was analysed separately. None of the main effects was statistically significant: no systematic source location difference was observed between the vowel SPEECH and REPLAY conditions, or between vowel and sound-evoked responses obtained during the task familiarisation.

A two-factor ANOVA (Conditions, Hemispheres) showed a significant main effect by Condition (REPLAY vs. SPEECH) in the analysis of both M100 amplitude [ $F(1,7) = 9.9, P < 0.02$ ] and latency [ $F(1,7) = 13.1, P < 0.009$ ], whereas the main factor Hemisphere was not statistically significant [amplitude:  $F(1,7) = 0.13, P < 0.73$ ]; latency:  $F(1,7) = 2.9, P < 0.13$ ]. There was a significant Condition  $\times$  Hemisphere interaction for peak latency [ $F(1,7) = 8.5, P < 0.03$ ] but not for peak amplitude [ $F(1,7) = 2.7, P < 0.15$ ]. Planned comparisons using two-tailed Wilcoxon signed-ranks tests for dependent variables showed that M100 peaked significantly later during the SPEECH than the REPLAY condition (Figs. 1B and 2), both in the left hemisphere (delay 16.4 ms;  $p = 0.012$ ) and the right hemisphere (5.5 ms;  $p = 0.025$ ). In addition, M100 peaked during SPEECH significantly later in the left than the right hemisphere (10.7 ms;  $p = 0.017$ ), whereas no differences were observed during REPLAY (−0.1 ms; n.s.; Fig. 2). M100 amplitudes were bilaterally smaller during SPEECH than during REPLAY (Fig. 2): in the left hemisphere, this reduction was significant (−39%;  $p < 0.025$ ) whereas in the right hemisphere, it just fell short of significance (−26%;  $p = 0.069$ ). Although the relative amplitude reduction during SPEECH thus tended to be stronger in the left than the right hemisphere, this interhemispheric difference did not reach significance ( $p = 0.21$ ).



**Figure 2.**

Latencies and amplitudes of magnetic brain responses to vowels. Mean + SEM of eight right-handed subjects. Left: M100 peak latencies of responses to self-uttered (SPEECH) and replayed

(REPLAY) vowels. The M100 peak amplitudes are presented in the middle panel, and the amplitudes of the change-related responses in the right panel.

### MEG responses related to changing vowels

During REPLAY, the brain reacted to the infrequent /ae/ vowels with an additional change-related response [Nääätänen, 1992], known to originate in the auditory cortices bilaterally [Hari, 1990]. In contrast, during the subject's own SPEECH, the infrequent /ae/ vowels among the frequent /a/ vowels did not evoke any significant change-related response (Figs. 1C and 2).

## DISCUSSION

### Modification of auditory cortex responses during speaking

The present MEG data show that (i) the subjects' own utterances activate transiently their auditory cortices, with strongest activation (M100) about 100 ms after the voice onset, (ii) this M100 response is significantly delayed in both hemispheres and dampened in the left hemisphere during SPEECH as compared with REPLAY, and (iii) speaking delayed the M100 responses of the human auditory cortex to self-generated speech sounds more prominently in the speech-dominant left hemisphere.

MEG and EEG source modelling, as well as lesion studies and intracranial recordings provide converging evidence that the neuronal generators of M100 are located in the auditory regions of the temporal lobes [Hari et al., 1980; Liegeois-Chauvel et al., 1994; Picton et al., 1999; Richer et al., 1989; Scherg and von Cramon, 1986; Scherg et al., 1989; Woods et al., 1984], and the same generation sites were confirmed here for both the SPEECH and REPLAY conditions. The specific contributions to M100 from different cytoarchitectonic

areas of the auditory cortices are still debated: source analyses of scalp potentials suggested one generator in the auditory koniokortex, and another more anteriorly [Scherg et al., 1989]. In contrast, intraoperative recordings of epileptic patients showed responses in the primary auditory cortex at Heschl's gyrus up to 50 ms whereas the 100 ms component was generated mainly at the planum temporale [Liegeois-Chauvel et al., 1994]; this finding agrees also with lesion studies reporting attenuated auditory responses at 100 ms for lesions posterior to the primary auditory cortices [Woods et al., 1987].

Given the signal-to-noise ratio of the present study, only a single equivalent current dipole was used to model the M100 generator in each hemisphere. Thereby, the "center of gravity" for M100 was found in supratemporal auditory cortices, but the possible contributions of different cortical areas could not be resolved in more detail; more widespread activations including generators within as well as outside the primary auditory cortex [Picton et al., 1999], however, might explain the fair amount of unexplained field variance.

In agreement with our finding of prominent but dampened M100 responses during SPEECH, the primary and secondary auditory cortices of awake squirrel monkeys contain a substantial number of neurons that respond to replayed vocalisations but are suppressed during phonation [Müller-Preuss and Ploog, 1981]. Speech-induced modifications of cortical excitatory postsynaptic potentials (EPSPs) are the prime candidates for explaining the observed dampening of the M100 response [Hari, 1990]. Specifically, the M100 dampening might reflect either a diminished presynaptic excitatory drive to neurons whose firing is suppressed during phonation [Müller-Preuss and Ploog,

1981], or the subsequent diminution of EPSPs in the local neuronal target population of these units. Intraoperative recordings have shown that single units in human superior temporal gyri (STG) of both hemispheres are activated differentially by a patient's own utterances and by the experimenter's tape-recorded speech [Creutzfeldt et al., 1989]. However, because the patient's own speech was not replayed, the observed differential activations could relate to different voice characteristics of the two speakers.

Thus, single-unit data from subhuman primates, and possibly also from humans, indicate motor-to-sensory priming in the form of inhibition of the auditory cortices during speaking. Such an inhibition might correlate to the speech-related modification of the M100 response. Candidate mechanisms for audiovocal interactions in primates will be discussed at cortical, subcortical, and peripheral levels.

### Cortico-cortical speech priming

In squirrel monkeys, 17% of STG neurons display convergence of acoustic and prefrontal cortical inputs; the latter either suppress the neurons' spontaneous firing or reduce their responsivity to acoustic input [Alexander et al., 1976]. Likewise, stimulation of the cingular vocalization area mainly inhibits STG neurons and suppresses their responses to vocalizations [Müller-Preuss et al., 1980]. Accordingly, the observed amplitude decrease and latency increase of M100 during SPEECH could reflect predominantly a direct cortico-cortical, motor-to-sensory inhibition representing priming of the auditory cortex by speech motor command centers. Because the cortical motor centers of speech are lateralized to the left hemisphere in right-handed humans, cortico-cortical speech priming could explain the left-hemisphere dominance of the speech-induced M100 modifications.

### Subcortical speech priming

Some marginal influences of vocalization have been described for auditory midbrain neurons (7 of 587) of subhuman primates [Kirzinger and Jürgens, 1991] as well as for human brainstem auditory evoked potentials [Papanicolaou et al., 1986]. These influences, however, do not predict the observed left-lateralized M100 delay during SPEECH. Furthermore, no differences were found in the responses of monkey collicular cells to self-produced vs. loudspeaker-transmitted calls [Müller-Preuss and Ploog, 1981].

Our data provide further evidence for mainly cortical priming during speaking: the absence of the extra

response to infrequently uttered /ae/ vowels during SPEECH is in strong contrast to the REPLAY condition during which the brain reacted to the infrequent /ae/ vowels with an additional and well-characterized change-related ["mismatch"] response, commonly interpreted to reflect automatic deviance detection against a standard template held in sensory memory [Näätänen, 1992]. This lack of change-related responses to self-uttered vowels indicates that the human auditory cortices process deviant sounds depending on the behavioral context: a change-related response was generated when vowel deviances occurred during mere listening to the REPLAY, but not when they represented an expected *sensorimotor match* between the SPEECH efference copy and its regularly upcoming auditory reafference [von Holst and Mittelstaedt, 1950]. Change-related responses (e.g., to loudness changes, are generated intracortically in monkeys) [Javitt et al., 1994]; however, in guinea pigs, responses to syllable changes can be picked up either cortically or already in nonspecific thalamic nuclei [Kraus et al., 1994a,b]. Thus, the specific contributions of thalamic vs. exclusively cortical activity to responses elicited by vowel changes remain to be clarified in subhuman primates as well as in humans.

### Audiovocal interaction in the peripheral auditory system

The sensory qualities of the vowel stimuli unavoidably were to differ between SPEECH and REPLAY conditions, mainly because of bone conduction and activation of middle ear muscles during SPEECH [Salomon and Starr, 1963]. Here, the *physical intensity* of the stimuli appears to be of particular concern because intensity influences both amplitude and latency of the auditory responses at 100 ms, with higher intensity inducing stronger and earlier responses [Elberling et al., 1981; Hegerl et al., 1994; Pantev et al., 1989; Pineda et al., 1991]. Accordingly, one could interpret the dampening and delay of M100 responses during SPEECH, compared with REPLAY, by claiming that the physical sound intensity was higher during REPLAY than during SPEECH. Therefore, one might consider determining the physical intensity of the speaker's own voice in her ear canal. This approach, however, would not take into account the additional acoustic input because of bone conduction during SPEECH or the dampening effects by concomitant activation of middle ear muscles. Thus, the net effect of SPEECH on input intensity would remain undetermined. Hence, for the present study design, we chose a "top-down" approach encompassing all those low-

level modifications and let the subjects *equalize the perceived loudness* of her REPLAYed and self-uttered vowels (see Methods).

Fortunately, the central aspects of the present results appear to be largely “immune” against eventual small differences of the effective stimulus intensity between conditions: in particular, no variations of effective stimulus intensity could explain (i) the observed interhemispheric M100 differences, or (ii) the abolition of the robust change-related response that remains invariant over even a 40-dB intensity range [Schroeger, 1994], and occurs largely independently of M100 magnitudes. Specifically, the hemispheric asymmetry in the M100 modifications can be considered safe against a suboptimal loudness matching between conditions because experimentally introduced, clearly audible variations of sound intensity (between 40 and 65 dB HL) cause no systematic interhemispheric M100 latency differences [Vasama et al., 1995].

Thus, irrespective of minor stimulus differences, the present MEG data can be interpreted to indicate that central (thalamic and/or cortico-cortical) priming of the auditory cortices during SPEECH affects both the input-related M100 responses of the auditory cortex and the later advanced processing, reflected in the lack of change-related responses.

#### Allocation of attention

Finally, one has to consider possible task-related differences in attention between the SPEECH and REPLAY conditions. A recent MEG study [Poepfel et al., 1996] has described a task-induced asymmetry for M100 responses to consonant-vowel (CV) syllables. M100 latencies and peak amplitudes were found bilaterally symmetric for passively presented stimuli; this response pattern closely resembles the findings during the REPLAY condition in the present study. When the subjects actively directed attention toward the CV stimuli, both latency and amplitude of M100 increased in the speech-dominant hemisphere, whereas the M100 amplitude decreased in the contralateral hemisphere. Three aspects of the present study argue against the possibility that the lateralized M100 modifications during SPEECH were because of a comparable effect of an increased level of attention. First, the task instructions of the present study aimed to equalize the allocation of attention during the SPEECH and REPLAY conditions. Second, the SPEECH-related left vs. right hemisphere M100 latency increase was accompanied here by an M100 amplitude decrease and, hence, cannot be directly explained by possibly higher attention during SPEECH

as compared with REPLAY. And third, although focussed attention has been found to increase responses to deviant phoneme stimuli [Aulanko et al., 1993], the change-related response was absent during SPEECH.

#### MEG and cerebral blood flow studies

The present MEG results with millisecond time resolution demonstrate for the first time an unequivocal phasic activation of human auditory cortex during SPEECH, albeit with a significant dampening and delay of these early responses when compared to a listening condition, and thereby uniquely complement recent speech-related CBF studies [Fox et al., 1996; Hirano et al., 1997; McGuire et al., 1996; Paus et al., 1996; Wise et al., 1999]. Remarkably, some of these CBF studies were interpreted to be compatible with voicing-related inhibition in human auditory cortices [e.g., Hirano et al., 1997; Paus et al., 1996; Wise et al., 1999], whereas others were understood to indicate activation [Fox et al., 1996; McGuire et al., 1996]. Basic differences in data evaluation and task design may explain these mixed results.

The CBF studies [Fox et al., 1996; Hirano et al., 1997; McGuire et al., 1996; Paus et al., 1996] used a bolus injection of a radioactive tracer and then evaluated counts summed over task periods lasting 40–120 sec. In contrast, the phasic MEG responses were averaged contingent on the occurrences of single vowels with millisecond time resolution.

CBF responses were mostly evaluated using “task minus baseline” subtractions. Therefore, brain state differences between a speech-related task (e.g., reading aloud) [Fox et al., 1996], and an “eyes-closed rest” baseline condition may include, in addition to the specific cerebral processing of voicing-related auditory input, also a general switch from a floating state of mind toward a more focussed visuo-motor-auditory network activation. Thus, CBF increases at auditory cortices (in “reading minus rest” subtraction images) could reflect a continuous attention-related activation during reading aloud where the contribution by transient specific cortical processing of auditory voice input remains undetermined; it is this latter, stimulus-driven activity and its speech-related modification that was specified in the present MEG experiments.

Finally, recent functional magnetic resonance imaging [Hickock et al., 1999] and MEG results [Numminen and Curio, 1999] provide convergent evidence for an involvement of auditory cortices specifically in left hemisphere also during subvocal (i.e., “silent,” speech). In particular, in close correspondence to the

present results, the MEG study [Numminen and Curio, 1999] found that, when a subject was articulating silently a vowel, the latency of the M100 response to a simultaneous replay of this vowel from a prerecorded tape was delayed in the left hemisphere.

### Clinical perspectives

Motor-to-sensory (phonation-to-hearing) priming might be defective, and thus worth of clinical testing, in aphasic patients known to be differentially sensitive to delayed auditory feedback [Chapin et al., 1981], in stutterers who fail to show left-lateralized auditory cortex activations during reading [Fox et al., 1996] but who improve their fluency under delayed auditory feedback [Stuart and Kalinowski, 1996], and in schizophrenics who, when experiencing voice hallucinations, may exhibit increased blood flow in Broca's area [McGuire et al., 1993] as well as modified reactivity of the auditory cortex [Tiihonen et al., 1992].

Hence, this MEG approach to analyse audiovisual interactions might provide a diagnostic tool: it is non-invasive and generally applicable, and can be repeated without any limitation in the single subject. Most importantly, MEG recordings can monitor directly neuronal correlates of cortical speech processing on the relevant millisecond time scale.

### CONCLUSIONS

The present results demonstrate interactions between audition and phonation on a millisecond time scale. The observed reactivity of the human auditory cortex to the subject's own utterances provides a basis for self-monitoring of voice intensity, pitch, and phoneme quality during speaking. Interestingly, "expected" self-produced deviances in a vowel series did not trigger change-related responses, and SPEECH-induced modifications of the 100-ms responses were more pronounced in the left auditory cortex that is specialized for phonetic aspects in syllable discriminations [Zatorre et al., 1992].

### ACKNOWLEDGMENTS

This study was supported by the EU's HCM Program through the Large-Scale Facility BIRCH at the Low Temperature Laboratory of the Helsinki University of Technology, by the Sigrid Jusélius Foundation, and by the Academy of Finland. MRIs were recorded at the Department of Radiology, Helsinki University Central Hospital. We thank A. Curio and O.V. Lounasmaa for comments on the manuscript.

### REFERENCES

- Ahonen AI, Hämäläinen MS, Kajola MJ, Knuutila JET, Laine PP, Lounasmaa OV, Parkkonen LT, Simola JT, Tesche CD. 1993. 122-channel SQUID instrument for investigating the magnetic signals from the human brain. *Phys Scripta* T49:198–205.
- Alexander GE, Newman JD, Symmes D. 1976. Convergence of prefrontal and acoustic inputs upon neurons in the superior temporal gyrus of the awake squirrel monkey. *Brain Res* 116:334–338.
- Aulanko R, Hari R, Lounasmaa OV, Näätänen R, Sams M. 1993. Phonetic invariance in the human auditory cortex. *Neuroreport* 4:1356–1358.
- Chapin C, Blumstein SE, Meissner B, Boller F. 1981. Speech production mechanisms in aphasia: a delayed auditory feedback study. *Brain Lang* 14:106–113.
- 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.
- Deal LV, Haas WH. 1996. Hearing and the development of language and speech. *Folia Phoniatr Logop* 48:111–116.
- Elberling C, Bak C, Kofoed B, Lebech J, Saermark K. 1981. Auditory magnetic fields from the human cortex. Influence of stimulus intensity. *Scand Audiol* 10:203–207.
- Fox PT, Ingham RJ, Ingham JC, Hirsch TB, Downs JH, Martin C, Jerabek P, Glass T, Lancaster JL. 1996. A PET study of the neural systems of stuttering. *Nature* 382:158–161.
- Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65:413–497.
- Hari R. 1990. The neuromagnetic method in the study of the human auditory cortex. In: Grandori F, Hoke M, Romani G, editors. *Auditory evoked magnetic fields and potentials. Advances in audiology*, Vol 6. Basel: Karger, P 222–282.
- Hari R, Aittoniemi K, Järvinen ML, Katila T, Varpula T. 1980. Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators. *Exp Brain Res* 40:237–240.
- Hegerl U, Gallinat J, Mrowinski D. 1994. Intensity dependence of auditory evoked dipole source activity. *Int J Psychophysiol* 17:1–13.
- Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, Strick PL, Ugurbil K. 1999. Auditory cortex participates in speech production. Annual Meeting, Cognitive Neuroscience Society. *J Cogn Neurosci Suppl* p 97.
- 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.
- Houde JF, Jordan MI. 1998. Sensorimotor adaptation in speech production. *Science* 279:1213–1216.
- Javitt DC, Steinschneider M, Schroeder CE, Vaughan HG Jr, Arezzo JC. 1994. Detection of stimulus deviance within primate primary auditory cortex: Intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Res* 667:192–200.
- Kirzinger A, Jürgens U. 1991. Vocalization-correlated single-unit activity in the brain stem of the squirrel monkey. *Exp Brain Res* 84:545–560.
- Kraus N, McGee T, Carrell T, King C, Littman T, Nicol T. 1994a. Discrimination of speech-like contrasts in the auditory thalamus and cortex. *J Acoust Soc Am* 96:2758–2768.

- Kraus N, McGee T, Littman T, Nicol T, King C. 1994b. Nonprimary auditory thalamic representation of acoustic change. *J Neurophysiol* 72:1270–1277.
- Lee B. 1950. Effects of delayed speech feedback. *J Acoust Soc Am* 22:824–826.
- Liegeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol* 92:204–214.
- McGuire PK, Shah GM, Murray RM. 1993. Increased blood flow in Broca's area during auditory hallucinations in schizophrenia. *Lancet* 342:703–706.
- McGuire PK, Silbersweig DA, Frith CD. 1996. Functional neuroanatomy of verbal self-monitoring. *Brain* 119:907–917.
- Müller-Preuss P, Ploog D. 1981. Inhibition of auditory cortical neurons during phonation. *Brain Res* 215:61–76.
- Müller-Preuss P, Newman JD, Jürgens U. 1980. Anatomical and physiological evidence for a relationship between the "cingular" vocalization area and the auditory cortex in the squirrel monkey. *Brain Res* 202:307–315.
- Näätänen R. 1992. Attention and brain function. Hillsdale, NJ: Erlbaum.
- Numminen J, Curio G. 1999. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Lett* 272:25–28.
- Pantev C, Hoke M, Lehnertz K, Lütkenhöner B. 1989. Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalogr Clin Neurophysiol* 72:225–231.
- Papanicolaou AC, Raz N, Loring DW, Eisenberg HM. 1986. Brain stem evoked response suppression during speech production. *Brain Lang* 27:50–55.
- Paus T, Perry D, Zatorre R, Worsley K, Evans A. 1996. Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246.
- Pick HL Jr, Siegel GM, Fox PW, Garber SR, Kearney JK. 1989. Inhibiting the Lombard effect. *J Acoust Soc Am* 85:894–900.
- Picton TW, Alain C, Woods DL, John MS, Scherg M, Valdes-Sosa P, Bosch-Bayard J, Trujillo NJ. 1999. Intracerebral sources of human auditory-evoked potentials. *Audiol Neurootol* 4:64–79.
- Pineda JA, Holmes TC, Foote SL. 1991. Intensity-amplitude relationships in monkey event-related potentials: Parallels to human augmenting-reducing responses. *Electroencephalogr Clin Neurophysiol* 78:456–465.
- Poeppel D, Yellin E, Phillips C, Roberts TP, Rowley HA, Wexler K, Marantz A. 1996. Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cogn Brain Res* 4:231–242.
- Rauschecker JP, Tian B, Hauser M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111–114.
- Richer F, Alain C, Achim A, Bouvier G, Saint-Hilaire JM. 1989. Intracerebral amplitude distributions of the auditory evoked potential. *Electroencephalogr Clin Neurophysiol* 74:202–208.
- Salomon B, Starr A. 1963. Electromyography of middle ear muscles in man during motor activities. *Acta Neurol Scand* 39:161–168.
- Scherg M, von Cramon D. 1986. Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr Clin Neurophysiol* 65:344–360.
- Scherg M, Vajsar J, Picton T. 1989. A source analysis of the late human auditory evoked potentials. *J Cogn Neurosci* 1:336–355.
- Schroeger E. 1994. Automatic detection of frequency change is invariant over a large intensity range. *Neuroreport* 5:825–828.
- Sperry W. 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol* 43:482–489.
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG Jr. 1994. Speech-evoked activity in primary auditory cortex: Effects of voice onset time. *Electroencephalogr Clin Neurophysiol* 92:30–43.
- Stuart A, Kalinowski J. 1996. Fluent speech, fast articulatory rate, and delayed auditory feedback: Creating a crisis for a scientific revolution? *Percept Motor Skills* 82:211–218.
- Tiihonen J, Hari R, Naukkarinen H, Rimon R, Jousmäki V, Kajola M. 1992. Modified activity of the human auditory cortex during auditory hallucinations. *Am J Psychiat* 149:255–257.
- Vasama JP, Mäkelä JP, Tissari SO, Hämäläinen MS. 1995. Effects of intensity variation on human auditory evoked magnetic fields. *Acta Oto-Laryngol* 115:616–621.
- von Holst E, Mittelstaedt H. 1950. Das reafferenzprinzip: Wechselwirkungen zwischen zentralnervensystem und peripherie. *Naturwissenschaften* 37:464–475.
- Wang X, Merzenich MM, Beitel R, Schreiner CE. 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. *J Neurophysiol* 74:2685–2706.
- Wise RJS, Greene J, Büchel C, Scott SK. 1999. Brain regions involved in articulation. *Lancet* 353:1057–1061.
- Woods DL, Knight RT, Neville HJ. 1984. Bitemporal lesions dissociate auditory evoked potentials and perception. *Electroencephalogr Clin Neurophysiol* 57:208–220.
- Woods DL, Clayworth CC, Knight RT, Simpson GV, Naeser MA. 1987. Generators of middle- and long-latency auditory evoked potentials: Implications from studies of patients with bitemporal lesions. *Electroencephalogr Clin Neurophysiol* 68:132–148.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A. 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.