Spatiotemporal Properties of Magnetic Fields Induced by Auditory Speech Sound Imagery and Perception*

Shihomi Uzawa, Tetsuya Takiguchi, Yasuo Ariki and Seiji Nakagawa

Abstract— **Brain computer interface (BCI) technologies, which enable direct communication between the brain and external devices, have been developed. BCI technology can be utilized in neural prosthetics to restore impaired movement, including speech production. However, most of the BCI systems that have been developed are the "P300-speller" type, which can only detect objects that users direct his/her attention at. To develop more versatile BCI systems that can detect a user's intention or thoughts, the brain responses associated with verbal imagery need to be clarified. In this study, the brain magnetic fields associated with auditory verbal imagery and speech hearing were recorded using magnetoencephalography (MEG) carried out on 8 healthy adults. Although the magnetic fields lagged slightly and were long-lasting, significant deflections were observed even for verbal imagery, in the temporal regions, as well as for actual speech hearing. Also, sources for the deflections were localized in the association auditory cortices. Cross-correlations were calculated between envelopes of the imagined/presented speech sound and the evoked brain responses in the temporal areas. Measurable correlations were obtained for the presented speech sound; however, no significant correlations were observed for the imagined speech sound. These results indicate that auditory verbal imagery undoubtedly activates the auditory cortex, at least, and generates some observable neural responses.**

I. INTRODUCTION

Over the past decade, there has been a significant increase in the number of developments related to brain machine interfaces (BCIs). BCIs are used as tools to communicate between the brain and external devices that restore damaged human functions, like movement and speech. It is important for the successful development of BCIs to clarify the reaction patterns of the brain. For instance, it is known that the P300 response is induced when people direct their attention to a target stimulus in the oddball paradigm, in which low-probability target stimuli are mixed with high-probability standard stimuli. Most of the BCI systems being developed use this distinctive response, which reflects the process of stimulus evaluation or categorization [1]. However, with such BCI systems, users can only choose something from among what was prepared in advance. In order to develop more versatile BCI systems, it is desirable to decode brain activities associated with verbal imagery.

* Part of this research was supported by the Grants-in-Aid for Scientific Research (25282053, 26282130, 15H02771 and 26590229) from the Japan Society for the Promotion of Science.

Shihomi Uzawa is with Kobe University, Kobe, Hyogo and National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka, Japan. uzawa@me.cs.scitec.kobe-u.ac.jp

Tetsuya Takiguchi and Yasuo Ariki are with Kobe University, Kobe, Hyogo, Japan. { takigu, ariki } @kobe-u.ac.jp

Seiji Nakagawa are with National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka and Chiba University, Chiba, Japan. s-nakagawa@chiba-u.jp

Some studies on brain imaging associated with such auditory imagery has been reported in recent years. Hoshiyama et al. (2001) conducted an MEG measurement on retrieval imagery of a hammering sound. Although no obvious peaks were observed, slow deflections were elicited in the right temporal region, and sources of the responses were localized in the inferior frontal sulcus (IFS) and the insular regions [2]. Also, Jäncke and Shah observed activations of both the superior temporal gyri (STG) by imagery of syllables using functional magnetic resonance imaging (fMRI) [3]. These studies indicated that auditory imagery activated cortical regions around the auditory cortex and Broca's area. However, various areas have been reported as candidates for the "auditory imagery" cortex, and details of the spatiotemporal properties of brain activities associated with auditory imagery remain unclear.

In terms of "actual" speech hearing, a former study revealed that some correlation existed between waveforms of speech sounds and auditory evoked fields [4]. Also "internal sound" is "heard" in auditory speech sound imagery although it is not accompanied by real sounds. Therefore, inversely, it is thought that a similar type of response as "actual speech" hearing will be elicited even in the speech imagery. Some correlated brain responses with waveforms of "imagined speech" may also be observed.

In this study, to clarify the neural mechanisms of recalled speech sound processing, we attempted to observe the spatiotemporal properties of brain activities associated with speech imagery by using magnetoencephalography (MEG). Response waveform, its correlation with sounds, and source location for the speech imagery were compared to those for actual hearing.

II. MATERIALS AND METHODS

A. Stimulus

Fig. 1 shows the timing of the stimulus presentation. At the beginning of each trial, a short tone burst was presented. Single visual stimulus using Japanese syllabary characters, one of the three words such as "*amagumo*" (rain cloud), "*ibento*" (event) or "*uranai*" (fortune-telling), appears three times with intervals of 500 ms. Synchronously, auditory stimulus (4-mora speech sound corresponding to each visual word) was presented only for the first two repetitions. The speech sounds, recorded with a female voice, were taken from a commercially-available database (NTT-AT FW03) and presented diotically using earphones (ER-2, Etymotic Research, Inc., Elk Grove, IL, USA) inserted into the subjects' ear canals. After the 3rd presentation, two different intervals between trials were used depending on the method used to remove artifacts associated with blinks during the

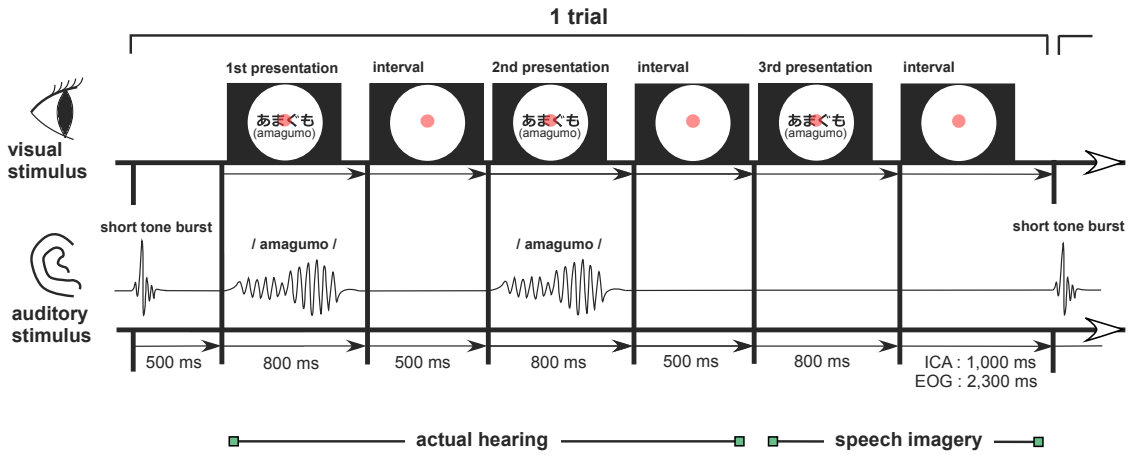

Fig. 1. *Schematic diagram of the task. 500 ms after a short tone burst, a word is shown visually (1st presentation) for about 800 ms at the center of a circle on a screen. At the same time, synchronous speech sound corresponding to the visual word was presented for about 800 ms. An interval with no visual or auditory word stimuli for 500 ms was then followed by the same word presentation (2nd presentation) and another interval. During the 3rd presentation, only the visual word appeared with no auditory stimulus. Then, the intervals between trials were set at 1,000 ms in an ICA setting and at 2,300 ms in an EOG setting, and the next trial started with a short tone burst. Each presentation in a trial consistently used a single visual word, where the 1st and 2nd presentations correspond to "actual hearing" and the 3rd presentation corresponds to "speech imagery".*

trials, namely independent component analysis (ICA) or electrooculogram (EOG).

The experimental paradigm included 2 sessions with and without auditory imagery. In the imagery session, at the same timing as the 3rd word, subjects were instructed to strongly imagine the speech sound that they heard in the 1st and 2nd presentation without moving their tongue and mouth. In the no-imagery session, subjects were instructed to just see and hear stimuli, and not to imagine the sound stimulus when seeing the 3rd word. In both sessions, subjects were also asked to push a button as quickly as possible when a sound did not correspond to the projected word. The order of the sessions was counterbalanced, and subjects took a break for about 15 minutes between sessions.

B. MEG recording

Eight healthy native Japanese speakers with normal hearing (7 males and 1 female, 20-40 years old, 6 right-handed and 2 left-handed) participated in the experiments. Necessary information regarding the experiment was given to the subjects, and informed consent was obtained prior to the experiment. The experiment was approved by the Institutional Review Board on Ergonomic Research of AIST.

MEG data were recorded using a 122-channel whole-head neuromagnetometer (Neuromag-122$^{\mathrm{TM}}$, Neuromag, Ltd., Helsinki, Finland) in a magnetically-shielded room. Visual stimuli were projected onto the center of a screen in front of the subjects using an LCD projector located outside the shielded room. Magnetic data were sampled at 400 Hz, and preprocessed by a band-pass filter between 0.03-100 Hz. Any epochs coinciding with magnetic signals exceeding 3,000 fT/cm were rejected from any further analysis.

The period between -250 and 3,000 ms, initialized by the onset of the 1st visual stimulus, was defined as the "period of actual hearing". Also, the interval between -250 and 1,600 ms, initialized by the onset of the 3rd visual stimulus, was defined as the "period of speech imagery". In both sessions, more than 80 trials were averaged at the period of actual hearing and more than 100 trials at the period of speech imagery. These responses were digitally bandpass-filtered between 0.1 and 30 Hz.

The vertical EOG of the subjects was recorded using infraorbital and supraorbital electrodes to monitor artifacts from eye blinks and eye movements in 3 of the 8 subjects. Any epochs coinciding with EOG deflections beyond 25 μ were rejected. In the other 5 subjects, the same artifacts were removed from the MEG data using ICA. The data were decomposed into 20 independent components, and some components, which were associated with eye blinks and eye movements, were removed.

III. DATA ANALYSIS

A. Latency and amplitude

Neuromag-122$^{\mathrm{TM}}$ has two pick-up coils in each position that measure two tangential derivatives, $\partial Bz/\partial x$ and $\partial Bz/\partial y$, of the field component Bz [5]. We determined:

$$B' = \sqrt{(\partial Bz/\partial x)^2 + (\partial Bz/\partial y)^2} \qquad (1)$$

as the amplitude of the response. For the 1st and 2nd word presentations with actual hearing, we employed an N1m peak latency and amplitude with the channel that showed the maximum amplitude placed over each temporal region. Since no clear response peak was observed for the 3rd presentation, the latency range in which differences of amplitude in the temporal region were estimated between the imagery and no-imagery sessions were recorded instead of peak latency.

B. Source estimation

Initially, equivalent current dipoles (ECDs) were estimated in various local regions using the Nelder-Meade simplex algorithm [6] at every 2.5 ms from the onset of the speech sound presentation to 1,600 ms. Among the calculated ECDs, those having a goodness-of-fit with 80% (1st presentation) / 70% (3rd presentation) and 3,000 mm^3 (1st presentation)

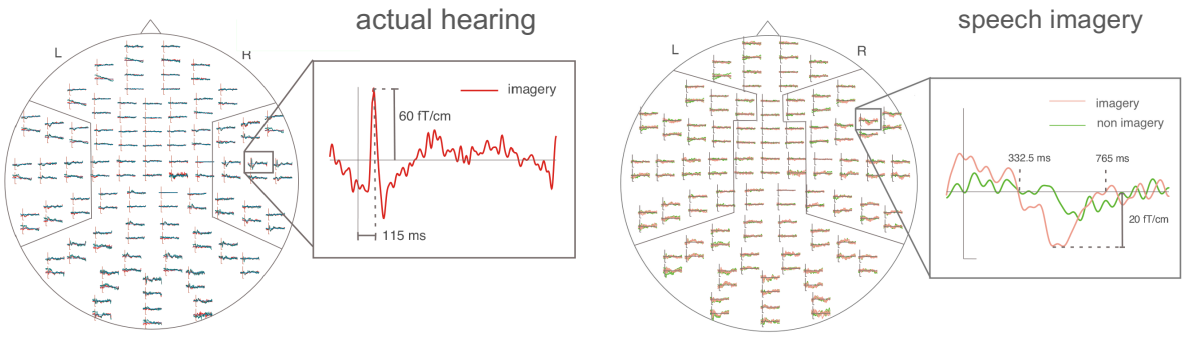

Fig. 2. *Brain magnetic field evoked by listening to stimulus (left) and imaging speech sounds (right) in subject 1.*

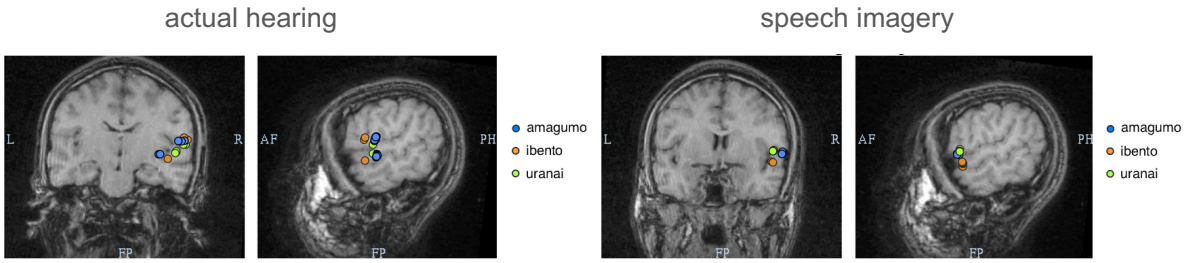

Fig. 3. *Source locations for listening to stimulus (left) and imaging speech sounds (right) in subject 1.*

/ 25,000 mm^3 (3rd presentation) confidence volume were selected. Then, continuous ECDs found in a series of more than 10 ms were defined as a source [7].

C. Estimation of cross-correlation between speech sound and brain activity

Cross-correlations between the envelope of the presented/imagined speech sound and evoked fields in the temporal region were calculated. The correlation value $r_k(\phi)$ was calculated as

$$r_k(\phi) = \frac{\sum_{i=1}^{T_s}(X_k^{i+\phi} - \overline{X}_k)(Y^i - \overline{Y})}{\sqrt{\sum_{i=1}^{T_s}(X_k^{i+\phi} - \overline{X}_k)^2}\sqrt{\sum_{i=1}^{T_s}(Y^i - \overline{Y})^2}} \quad (2)$$

The durations of the evoked fields data and speech envelopes are T_M and T_s, respectively. At the delay ϕ ($\phi = 1, 2, ..., T_M - T_s$) at the k th channel, the i th sampling (i = 1, 2, ..., N) value of the evoked fields and the envelopes is $X_k^{i+\phi}$ and Y^i, respectively. In this calculation, the extracted upper envelope of speech amplitude was down sampled from 48,000 to 400 Hz, which is a sampling frequency for magnetic data. Then, the evoked fields data were digitally band-pass filtered between 0.1 and 10 Hz in order to acquire gradual waveform.

To show differences in correlation values for combinations of envelopes and magnetic fields, the normalized correlation values between each sound envelope and the three types of waveforms were calculated. For example, normalized correlation values between a stimulus "a" and MEG waveform in presenting "a", $N_{a|MEG-a}$, were acquired as

$$N_{a|MEG-a} = \frac{r_{a|MEG-a}}{r_{a|MEG-a} + r_{a|MEG-b} + r_{a|MEG-c}} \quad (3)$$

$r_{a|MEG-b}$ is the cross-correlation value between an envelope of a stimulus "A" and magnetic evoked field induced by using a stimulus "B".

IV. RESULTS

A. MEG waveforms

Fig. 2 shows the responses evoked by what the subjects actually heard and by the speech imagery. For the actual hearing, clear deflections at a latency of about 110 ms were observed in both of the temporal regions. The RMS amplitude was normalized by the maximum value within each subject. There were significant differences among the three words for normalized RMS amplitude in the left hemisphere ($p < 0.01$) and RMS latencies in the both hemispheres ($p < 0.01$). For the speech imagery, although no steep response peak was observed, long-lasting deflections appeared over the temporal regions in 5 subjects. The latency was in a range between about 300 to 700 ms in 2 of the subjects and 400 to 950 ms in 3 of the subjects.

B. Source location

The left panel of Fig. 3 shows examples of source location for the 1st stimulus. The sources were localized in/around both auditory cortices in 7/8 subjects. The right panel shows examples of ECDs for the long-lasting responses elicited by speech imagery. In all the subjects, some ECDs were localized in the visual cortices. Additionally, 5 subjects showed some significant ECDs localized in the temporal region when they imagined the speech sound. The sources in the temporal/temporofrontal region seemed to be distributed over a relatively wide area; the STG in the right and/or left hemisphere in 6 subjects and in the inferior frontal gyrus (IFG) in the right hemisphere in 1 subject.

TABLE I

Correlation of sound envelope and MEG signal induced by hearing stimulus.

	correlation		
	amagumo	ibento	uranai
subject 1	0.595	0.514	0.369
subject 2	0.517	0.397	0.184
subject 4	0.543	0.340	0.273
subject 5	0.528	0.360	0.196
subject 6	0.434	0.289	0.246

TABLE II

Correlation of sound envelope and MEG waveform induced by speech imagery.

	correlation		
	amagumo	ibento	uranai
subject 1	0.246	0.065	0.179
subject 2	0.286	0.230	0.209
subject 5	0.139	0.096	0.162
subject 6	0.103	0.179	0.178

C. Correlation between the speech envelope and MEG waveform

Tables I and II show each subject's averaged determination coefficients of correlation between the amplitude envelopes of the presented auditory stimuli and evoked magnetic fields of actual hearing or speech imagery in the temporal regions. Fig. 4 shows the normalized correlation values for all the combinations of a presented speech envelope and the magnetic fields evoked by each speech sound. The highest values were obtained between the speech sound and corresponding magnetic field. For the 1st and 2nd stimuli with actual hearing, measurable correlations, between 0.3 and 0.6, were obtained in 5 out of 8 subjects when the envelope was delayed by 100 to 150 ms. Applying an analysis of variance to the correlation values in each subject, there were significant differences among the three words in 4 of the subjects ($p < 0.01$) and slight differences in word subject 6 ($p = 0.06$) were obtained, respectively. On the other hand, for the 3rd stimulus with speech imagery, correlations were much smaller.

V. DISCUSSION

Significant responses were observed even for the speech imagery. The response appeared in the temporal/temporofrontal channels, and their sources were localized in/around the STG and/or the IFG. These results indicate that auditory verbal imagery activates the auditory cortex, at least, and generates some observable neural responses. The activation of the IFG was observed only in the right hemisphere. Since this result corresponds to a former study [2], it appears that the left IFG plays an important role in auditory imagery.

However, the spatiotemporal properties of the evoked fields induced by the speech imagery were much different from those induced by the actual hearing. Responses for the speech imagery were smaller in amplitude, long-lasting and continued for about 400 to 500 ms, and were observed in channels that were more anterior in location. No steep peak like N1m response appeared. These differences can be explained by considering two hypotheses. First, it is thought

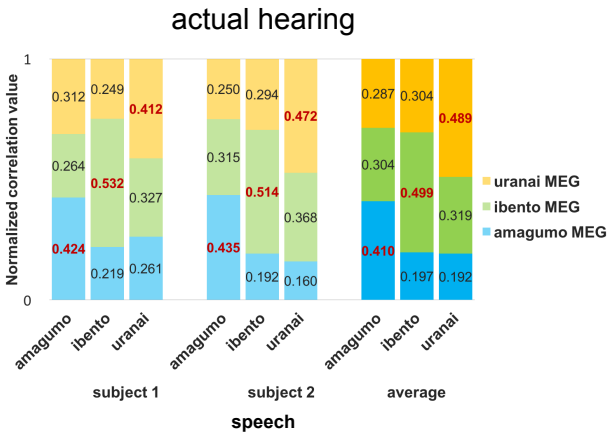

Fig. 4. *Normalized correlation values for all combinations of the presented speech envelopes and MEG waveform for each speech sound.*

that the speech imagery does not need the activation of the primary auditory cortex that processes exogenous parts of hearing. Alternatively, the activation of the association cortex seems more important for imagery. Second, fluctuations in the imagery can be a cause of the smaller amplitude and the long-lasting waveforms. For instance, it was likely that the subjects' timing of the imagery was not always the same throughout the measurement.

The correlations between the speech sound and brain activity were measurable for the presented speech sound; however, they were very small for the imagined speech sound. For the presented speech sound, higher correlations were obtained at a latency of N1m response, so it seemed that the evoked fields reflected partial physical onsets of the speech sound. This result supports the above hypothesis that speech imagery does not require the activation of the primary auditory cortex. On the other hand, the result indicated that it is difficult to estimate imagined speech sound using a simple coherence analysis. A coherence analysis using other characteristics of the speech sounds (for example time-frequency feature and f0-timecourse) and more advanced technology may be needed to clarify the relationship between auditory verbal imagery and the corresponding brain response.

REFERENCES

[1] R. Fazel-Rezai at al., "P300 brain computer interface : current challenges and emerging trends," *Frontiers in Neuroengineering*, pp. 1-15, 2012.

[2] M. Hoshiyama et al., "Hearing the sound of silence: a magnetoencephalographic study," *NeuroRep.*, Vol. 12, pp. 1097-1102, 2001.

[3] L. Jäncke and N. J. Shah, "'Hearing' syllables by 'seeing' visual stimuli," European Journal of Neuroscience, Vol. 19, pp. 2603-2608, 2004.

[4] M. Bourguignon, et al., "The pace of prosodic phrasing couples the listener's cortex to the reader's voice," *Human Brain Mapping*, Vol. 34, 314-326, 2013.

[5] A. I. Ahönen, et al., "122-channel SQUID instrument for investigating the magnetic signals from the human brain," *Physica Scripta*, Vol. 49, pp. 198-205, 1993.

[6] M. Singh, et al., "Neuromagnetic localization using magnetic resonance images," *IEEE Transactions on Medical Imaging*, Vol. 11 (1), pp. 129-134, 1992.

[7] S. Nakagawa, et al., "Spatiotemporal Source Imaging of Brain Magnetic Fields Associated with Short-term Memory by Linear and Nonlinear Optimization Methods," *IEEE Transactions on Magnetics*, Vol. 40 (2), pp. 635-638, 2004.