TEMPORAL AND SPATIAL BRAIN ACTIVITIES REFLECTED IN INDEPENDENT COMPONENTS OF EVENT-RELATED MAGNETIC FIELDS DURING VOWEL PERCEPTION

Shinobu Masaki♣, Shinji Maeda♦, Kiyosi Honda♣, and Kazumi Aoyama♥
♣ATR Human Information Processing Research Laboratories, Kyoto, Japan, ♦Ecole Nationale Supérieure des Télécommunications and Centre National de la Recherche Scientifique, Paris, France, ♥Waseda University, Tokyo, Japan

ABSTRACT

Magnetoencephalographic (MEG) data of brain activities are recorded during vowel perception tasks. We discuss the location active regions estimated by an equivalent current dipole (ECD) matching, N100m latencies, and spatiotemporal decomposition using a singular value decomposition or independent component analysis (ICA) technique. Although ECDs exhibited a vowel map consistent across subjects, it didn't correspond to the phonetic pattern of vowels. So far, only the ICA seems to indicate vowel specific temporal patterns which can be interpreted by the phonological feature [front/back].

1. INTRODUCTION

The connection between perception and production in the brain constitutes as a missing link in Speech Chain. In this paper, we report our first effort toward understanding this link by means of magnetoencephalographic (MEG) observations of brain activities during vowel perception.

In this direction, we are interested in how perceived vowels are represented in the brain. For example, Ragot, et al. [9] have showed using a set of artificial vowels as stimuli that perceived pitch is manifested in the latency of N100 peak of the event-related potential. Their data show that the latency systematically gets longer as voice fundamental frequency (F0) varies from high to low. Cansino et al. [1] have estimated active regions of the cortical surface in response to pure tone stimuli of different frequencies. The active regions are represented in terms of equivalent current dipoles (ECDs). The location of ECDs is estimated by an inverse procedure from measured electoencephalographic potentials. The authors show a tonotopic organization of estimated ECDs in the cortex. In this case, it can be said that perceived pitch of a tone manifests in a specific region and that the regions of different tones are tonotopically located. For vowel perception, Diesch et al. [4] have shown that estimated ECD locations from event-related fields of different vowels can be interpreted best in terms of the phonological representation of vowels.

These studies motivated us to conduct a series of event-related MEG experiments to obtain the cortical map of vowels. As descried in the following section, we could not obtain so far a phonetically meaningful cortical map of perceived vowels. We thus ask a different question whether or not the derived event-related fields (ERFs) contain information about the identity of vowels.

2. ESTIMATED ECD LOCATIONS OF VOWELS: A CORTICAL MAP

We have conducted two different series of experiments using entirely different machine setups. We report in this section only recent one using a Neuromag 122 channels MEG machine and an associated inverse program to estimate ECD positions from the ERF data.

In this experiment, four Japanese subjects are asked to silently identify synthetic vowel tokens. The target vowels consist of [i], [e], and [ɑ]. In order to maintain subjects' attention, [o] and [u] are used as fillers. The subjects are asked to press the button [o] or [u] when they recognize these filler vowels. These vowels are synthesized from vocal-tract area functions appropriate for a male using an acoustic simulation method [5]. Their duration is fixed to 200 ms. The target vowels are repeated at least 100 times and fillers about 50 times. The stimuli are presented in random order. The inter-stimulus intervals are varied randomly between 1.1 s and 1.3 s with 0.1 s step size. The sounds are fed from the right ear through a special ear phone.

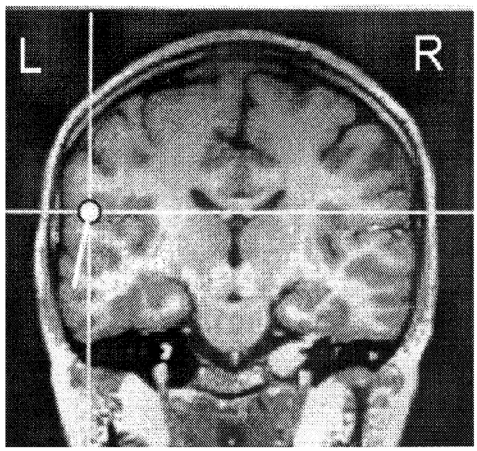

Figure 1. The deduced ECD of the vowel [ɑ] for the subject Iws is plotted on the coronal slice of the 3D MR image. This ECD, and also others shown in Figure 2, is located within the primary auditory cortex.

The ERFs of the individual subjects are obtained by aligning sets of 122 channel MEG waves at the stimulus onset and then by averaging over the repetitions of the same stimulus. The ECD locations related to the target vowels are then estimated using the inverse program from the ERF at a specified latency, i.e., the time defined by letting the stimulus onset time as the origin. The ERFs at the latency of about 100 ms where the first strong negative peak occurs, denoted as N100m, is used for the ECD estimation. In actual estimations, only ERF signals from 34 sensors positioned near the subjects' left temporal region are used. An example of the deduced ECD is depicted in Figure 1. The MRI indicates that the ECD is located at the primary auditory cortex in the left hemisphere, and that a more left of the x-coordinate means the positioning of the ECD closer to the left skull surface.

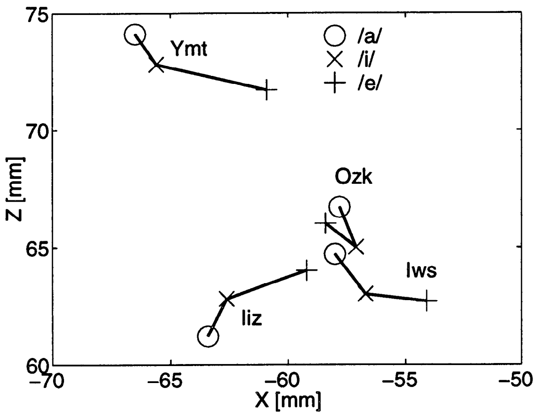

Figure 2. ECD locations of the three vowels, deduced from the ERFs at N100m, for the four subjects. The x-z plane corresponds to the coronal slice shown in Figure1.

Figure 2 illustrates the ECD locations of the three target vowels for the four listeners, estimated at the vicinity of the N100m peak. Since the ECD locations differ most importantly in the x-dimension (left-right) and relatively little in the y-dimension (anterior-posterior), they are only described in the x-z coordinates.

We are interested in, here, a topological relation of the three ECD locations for each subject. Notice in Figure 2 that the three subjects, Ymt, Iiz, and Iws, exhibit similar EDC vowel patterns: Their vowels can be ordered from the surface (left) to the inside (right) as [ɑ], [i], and [e]. It is somewhat puzzling to find out that this ordering is fairly systematic across subjects (i.e., the three out of four subjects) and that the order doesn't really correspond to that in vowel systems. If the order were to follow that of a vowel system; it should had been [ɑ], [e], and [i]. In the study of the tonotopic map of tones [1], the tones are mainly mapped also along the x-coordinate such that a low tone locates near the cortical surface and a high tone more inside. If we could assume that the vowels are mapped in the order following the second formant frequencies as an equivalence to the tone frequencies of [1], we could expect the order [ɑ], [e], [i]. In our estimated ECD vowels however, the order for [e] and [i] are reversed. Of course, there could be no real reason to expect a cortical map which topologically corresponds to the vowel system at begin with. Nevertheless, it raises an intriguing question why the three out of four subjects exhibit the same vowel order and why in that order.

It should have been interesting to investigate the vowel map in terms of the ECDs estimated at other latencies, in particular at the P200m latency. At present the ECD data are not available to us. We describe in the followings our studies motivated by somewhat more fundamental question whether or not ERF signal data contain some information specific to the individual vowels.

3. N100m LATENCY OF A TONE AND VOWELS

It is known that the latency of N100m peak varies depending on the physical characteristics of stimuli [e.g., 9, 8]. In order to assess how the latency varies depending on vowels, we prepare a set of stimuli consists of four cardinal, male and female vowels, [i], [e], [ɑ], and [u], and plus a 1 kHz tone. Vowels are synthesized specifying their vocal-tract area functions as before. The female vowels are obtained by a higher F0, a more symmetrical glottal pulses, and a shorter vocal-tract length in comparison with the male vowel counterparts. The duration is fixed to 600 ms for all the vowels, which is considerably longer than that of the vowels in the previous experiment.

The MEG data are recorded using a 37 channel BTi system. The four vowels are repeated at least 100 times for the reduction of noise by averaging. The stimuli are presented in random order with the inter-stimulus intervals which are also varied randomly between 1.8 s and 2.2 s with 0.1 s step. Two subjects participated in a series of experiments. The sounds were fed from the right ear through a piezoelectric ear phone. The 37 sensors are pointed toward the left temporal region (primary/associated auditory cortexes) of the subjects.

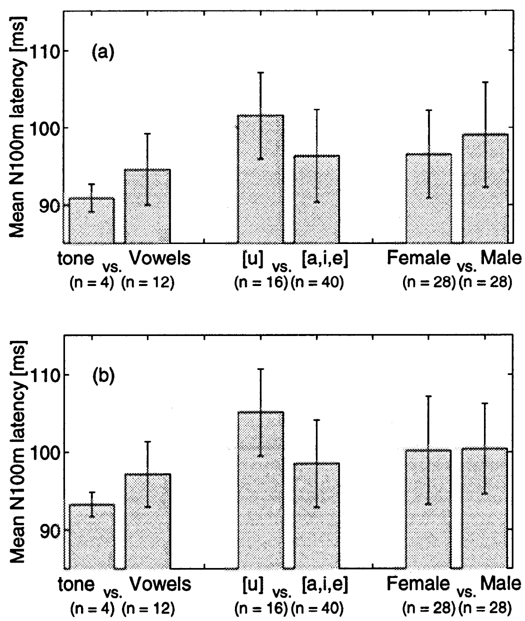

Figure 3. Mean N100m latencies compared for the three different oppositions: 1 kHz tone vs. four vowels at the left; the vowel [u] vs. the remaining three vowels at the center; female vs. male vowels at the right. For the subject Mks in (a) and Mds in (b).

The subjects are asked to silently identify the vowels. We consider four variations of the identification task:

1) vowels {[i], [e], [ɑ], [u]}; an attentive listening.

2) vowels {[i], [e], [ɑ], [u]}; a non-attentive listening.

3) vowels{[i], [ɑ], [u], and [ə]}; [ə] is used as a rare vowel (only 20 repetitions) and a subject counts the number of its occurrences. This is intended to maintain subject's attention.

4) vowels + tone {[i], [ɑ], [u], tone, and [ə]}; the same as 3).

One session consists of these four tasks with the male vowels and another session with the female vowels. Since, the two subjects participated twice in each session, in total they produced four sets of MEG data. Sessions for each subject are separated by at least one week.

The 100 repetitions of the 37 channel MEG waves for each vowel are aligned at the stimulus onset and averaged, resulting in relatively clean ERF signals. An RMS response to a vowel is then calculated by taking the ensemble average over the 37 channels of the ERF signals. The latency of the N100m peak is measured on the RMS responses. Figure 3 summarizes averaged latencies of the individual vowels and the tone gathered from various sessions for the two subjects separately.

At the left of Figures 3a and 3b, the mean latencies are obtained by opposing the 1 kHz tone against the three vowels, [i], [ɑ], and [u], (male and female) in Task 4). The latency of the tone is shorter than that of the vowels. This result corroborates a previous experiment [8], where the latency of a tone was also shorter than that of the vowel [a]. It appears that the latency becomes longer with an increase in a complexity of stimulus signal.

At the center of Figures 3a and 3b, the mean latencies are opposed by the vowel [u] vs. the remaining three vowels, [i], [e], [ɑ] in the all four tasks and male and female condition mixed. We didn't calculate the individual mean latency of the last three vowels, because there wasn't any systematic difference in the latency of these three vowels. Both subjects exhibit the longer latency of [u], but it is not clear why the latency of [u] is longer than that of the other vowels.

Finally at the right, the mean latencies are opposed by female vs. male vowels. Only the subject Msk indicates a shorter latency for the female vowel than for the male vowel. The other subject Mds doesn't show any difference. Nevertheless this result is interesting, because Ragot et al. [9] have reported that a higher F0 corresponds to a shorter latency.

It is tempting to speculate that the perceived pitch height of sounds, which is a functions of F0, spectral shapes, and intensity, is one of major factors determining the N100m latency. It is not so unreasonable to state that the pitch height of 1 kHz tone could be much higher than that of vowels; the pitch height of the vowel [u] could be lower than that of other vowels having the same F0; and obviously the perceived pitch height of female vowel should be higher than that of male vowels. If this is the case, the patterns of the paired latencies shown in Figure 3 could be qualitatively explained, at least in part, by perceived pitch height.

The latency might appear to serve for distinguishing the vowel [u] from others. Then, does a brain function exploit the latency for the distinction? We think it is hardly the case, unless the brain "knows" the stimulus onset time, which is needed to define the latency. For this reason, we seek vowel specific information in temporal and/or spatial patterns of the ERF signals, which will described in the following session.

4. SPATIOTEMPORAL DECOMPOSITION OF ERFs: SVD vs. ICA

It is neurologically accepted that the location of the dipole sources, representing active regions, remains invariant and that only their active level varies in time. This fact makes a spatiotemporal decomposition of ERFs and of event-related potentials (ERPs) attractive. A decomposition method, such as the singular value decomposition (SVD), which is identical to the principal component analysis in statistics, describes spatiotemporal data as the sum of components where each component is specified by a spatial pattern multiplied by the temporal variation of that pattern. It has been shown however that SVD cannot always recover source locations from ERFs or from ERPs, that match the actual neuronal activity regions [6, 10]. We therefore compare two different methods for analyzing the derived ERFs, the classical SVD and an independent component analysis (ICA). The ICA is constrained by higher order statistic than the SVD that is always second order by the maximum extraction of variance [e.g., 3]. We employed the so called JADE algorithm that is governed by fourth order statistics [2]. An important advantage of JADE (and generally ICA's) is the absence of the build-in orthogonality constraint present in SVD, which is only a mathematical convenience.

The temporal patterns of the first SVD component calculated from ERFs for the task 1) by subject Msk is shown in Figure 4. This first component explains about 40% of the variance for each vowel. The temporal patterns (and spatial patterns which are not shown) seem not to exhibit any vowel specific characteristic. Its absence was also true for the second component that accounts for about 20% of the variance.

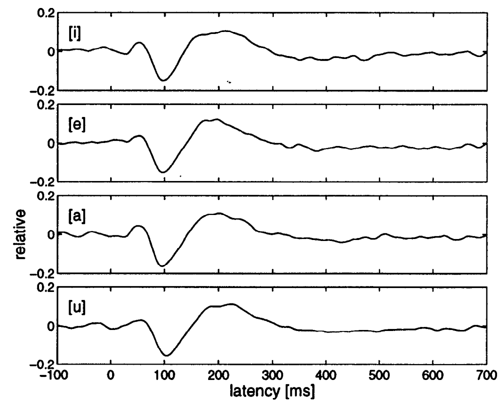

Figure 4. Temporal patterns of the first SVD component, derived for the four vowels.

Figure 6 illustrates the results of the corresponding ICA. In the analysis, the sum of the first two SVD components, explaining about 60% of the variance in total, was subjected to ICA analysis using JADE algorithm. The previous SVD procedure can be regarded as a filtering in the sense that the components higher than the second are truncated, assuming they are just noise. JADE attempts to recalculate two independent components, imposing a higher order constraint, out of the SVD filtered ERFs. The variance explained by the two independent components, therefore, is identical to that extracted by the first two SVD components, i.e., about 60%. The proportion of the variance extracted by the first and second components of ICA, of course, is different from these by SVD.

It turned out for the vowels [i] and [e] that, in contrary to the amount of the extracted variance, their second components are more meaningful to us than their first components. For example, the contour map of the spatial pattern of the first independent component for [ɑ], as shown in Figure 5a, indicates a relatively clean source (the light-gray region) and a sink (dark region) of the magnetic fields, suggesting the presence of a single dominant ECD in the cortex near the sensors. For the second independent component as shown in Figure 5b, there is only source, suggesting that this second component represents field from a distant magnetic source located far from the sensors and not within the subject's brain. For [i] and [e], the second component manifests the source-sink pair. We selected therefore their second components as the pertinent ones. As the consequence, the temporal patterns derived from ICA shown in Figure 6 correspond to the first components for [ɑ] and [u], and to the second components for [i] and [e].

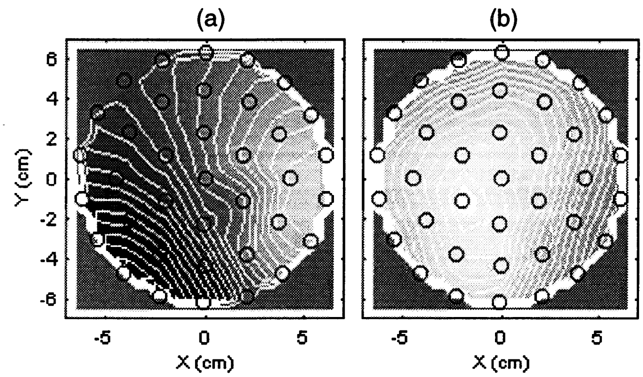

Figure 5. The spatial patterns of the first (in a) and second ICA component (in b) for the vowel [ɑ], projected on a sensor surface. The circles indicate 37 sensor positions.

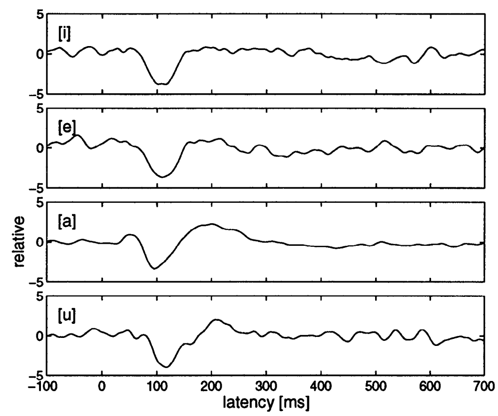

Figure 6. Temporal patterns of an ICA component, derived for the four vowels. (See text for detail.)

In contrast to the patterns derived from SVD, the waveform of the ICA derived patterns exhibits some vowel dependency. In the case of the two back vowels [ɑ] and [u], a positive peak is present at the vicinity of the latency 200 ms, presumably corresponding to p200m. Contrarily, such a peak is absent or very weak for the front vowel [i] and [e]. Also the shape of the first negative peak near the latency 100 ms, corresponding to N100m, seems to different depending on front or back feature of the vowels. The spatial patterns also exhibited, to some extent, vowel specific variations, but it was less pertinent than the temporal patterns.

5. CONCLUDING REMARKS

Admittedly, the results of the search for vowel specific patterns in the MEG records are episodic and some qualitative. We are aware of the fact that for studying a higher order brain function as the vowel identification, it could be more appropriate to use an endogenous paradigm, such as MMN [e.g., 7]. Unfortunately, MMN experiments are terribly time consuming (A session can last more than three hours). Moreover, it is not so evident to formulate a practical paradigm to study the internal organization of perceived vowels. Then, it may be worthwhile to proceed our approach to seek an effective method to extract vowel related information from observed ERFs.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Toshiaki Imada, at Communication Science Laboratories of NTT Corporation, for allowing us to use the Neuromag MEG system and for his helpful advice on the planning of experiments and data analysis. We thank also Mr. Noboru Hayashi, at High Quality Research Laboratory in Sumitomo Metal Industries Ltd., for data acquisition using the BTi MEG system. We gratefully acknowledge the continuing support of this project from Dr. Yoichi Tohkura at Communication Science Laboratories of NTT Corporation.

REFERENCES

[1] Cansino, S., Williamson, S.J. and Karron, D. 1994. Tonotopic organization of human auditory association cortex. *Brain Research*, **663**, 38-50.

[2] Cardoso, J-F. and Souloumiac, A. 1993. Blind beamforming for non Gaussian signals. *IEE-Proceedings-F*. **140** (6), 362-370.

[3] Comon, P. 1994. Independent component analysis, A new concept? *Signal Processing*. **36**, 287-314.

[4] Diesch, E., Eulitz, C., Hampson, S., and Ross, B. 1996. The neurotopography of vowels by evoked magnetic field measurements. Brain and Language, **53**, 143-168.

[5] Maeda, S. 1982. A digital simulation method of the vocal-tract system. *Speech Communication*. **1**, 199-229.

[6] Maier, J., Dagnelie, G., Spekreijse, H., and Dijk, B.W. van. 1987. Principal components analysis for source localization of VEPs in man. *Vision Research*, **27**, (2), 165-177.

[7] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J. and Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, **385**, 432-434.

[8] Obata, Y. and Masaki, S. 1995. Event related field for sinusoidal wave and the vowel [a]: Preliminary results from magnetoencephalographic experiments. *ATR Technical Report*, TR-H-173 (In Japanese).

[9] Ragot, R. and Lepaul-Ercole, R. 1996. Brain potentials as objective indexes of auditory pitch extraction from harmonics. *Neuroreport*, **7**, 4, 905-909.

[10] Rotterdam, A. van. 1970, Limitation and difficulties in signal processing by means of the principal-components analysis. *IEEE Transactions on Bio-medical Engineering*, **17**, 268-269.