An AdaBoost-Based Weighting Method for Localizing Human Brain Magnetic Activity

T. Takiguchi*, R. Takashima*, Y. Ariki*, T. Imada†, J.-F. L. Lin†, P. K. Kuhl†, M. Kawakatsu‡, and M. Kotani‡
* Kobe University, Japan
E-mail: takigu@kobe-u.ac.jp, takashima@me.cs.scitec.kobe-u.ac.jp, ariki@kobe-u.ac.jp
† Institute for Learning and Brain Sciences, University of Washington, USA
E-mail: imada@u.washington.edu
‡ Tokyo Denki University, Japan

***Abstract*—This paper shows that pattern classification based on machine learning is a powerful tool for analyzing human brain activity data obtained by magnetoencephalography (MEG). In our previous work, a weighting method using multiple kernel learning was proposed, but this method had a high computational cost. In this paper, we propose a novel and fast weighting method using an AdaBoost algorithm to find the sensor area contributing to the accurate discrimination of vowels. Our AdaBoost simultaneously estimates both the classification boundary and the weight to each MEG sensor, with MEG amplitude obtained from each pair of sensors being an element of the feature vector. The estimated weight indicates how the corresponding sensor is useful for classifying the MEG response patterns. Our results for vowel recognition show the large-weight MEG sensors mainly in a language area of the brain and the high classification accuracy (91.0%) in the latency range between 50 and 150 ms.**

I. Introduction

Non-invasive measurements using magnetoencephalography (MEG) have recently been used to study how stimulus features are processed in the human brain. In particular, because the neural electric activity of the brain that is associated with speech and language stimuli happens in a time frame of milliseconds, high temporal resolution of MEG is required for measuring rapid changes in brain activity during speech perception. Research carried out with MEG has reported left hemisphere dominance for processing of vowels in right-handed subjects [1], and the prominent N1m wave of the auditory-evoked field has been shown to exhibit sensitivity to a variety of acoustic attributes of the speech signal [2], as well.

Recently, the application of pattern recognition methods to neuromagnetic responses has been garnering much interest, and progress has been made through the use of machine learning, such as support vector machines (SVMs) [3][4][5]. SVMs are efficient tools for automatic recognition, but neuroscience research requires not only high-accuracy classification tools but also analysis tools that can locate both the dominant area of the brain (which shows strong activity related to speech and language), and the significant time frame (which exhibits this increased brain activity).

In our previous work [6], we presented a MEG-sensor weighting method using a multiple kernel learning (MKL) algorithm for analyzing areas of the brain that contribute to the accurate decoding of two vowels. Our subject-independent (subject-open) analysis results showed that the brain area covered by the MEG sensors with the larger weight obtained by the MKL method corresponded to the language area of the left hemisphere, and a high classification accuracy was obtained in the latency range between 100 and 200 ms. The method has a high computation cost because of the non-linear kernel process, however.

A boosting algorithm is a machine-learning-based technique for data classification in which fast and effective classifiers are produced [7]. Since 1999, several variants have been proposed, such as Real AdaBoost [8], Gentle AdaBoost [9], FloatBoost [10], and so on [11].

Boosting-based algorithms have recently been developed on a wide range of area, such as text processing [12][13], image processing [14][15], speech recognition [16][17], and so on [18]. In this paper, we present a novel and fast weighting method for the AdaBoost algorithm, where the weight is associated with each MEG sensor. In our approach, AdaBoost was applied to MEG responses or amplitudes, to localize brain areas that contribute to the accurate decoding of vowels. Sixty-one MEG amplitudes, each calculated from each of 61 pairs of MEG sensors (in total 122 MEG sensors), constituting a 61-dimension feature vector, are separately weighted. Each weight value calculated by AdaBoost indicates how useful each MEG-sensor pair is for classifying the MEG responses to vowel recognition. To identify the MEG sensors or brain areas important for vowel recognition in a subject-independent (subject-open) fashion, the weights were averaged across subjects.

II. Recording of MEG Responses to Vowels

Four right-handed volunteers (21-25 years old) were recruited as subjects after obtaining consent forms from them. All were native Japanese speakers with normal hearing.

We used two speech sounds (Japanese vowels), /a/ and /o/, to explore the subject's vowel recognition process in the brain. These 200-ms auditory stimuli were delivered to the subject's right ear through a plastic tube with a random interstimulus interval between 1,300 and 1,500 ms. The subject's task was to press a reaction key with the index finger when the subject identified the stimulus /a/ and another reaction key with the middle finger when the subject identified the stimulus /o/.

Neuromagnetic data were recorded by a 122-channel whole-scalp Neuromag MEG system in a magnetically shielded room. The MEG signal was sampled at 497 Hz for 1,200 ms including a 100-ms pre-stimulus baseline, and more than 80 epochs were averaged to increase the S/N ratio. A low-pass filter with a cutoff frequency of 40 Hz was used to calculate the feature vector. Epochs in which the magnetic signal exceeded an absolute amplitude variation of 3,000 fT/cm were discarded. Eye-movement artifacts were also automatically removed (threshold = 150 μV).

Feature extraction was applied to a 996-ms MEG signal. The mean reaction times for /a/ and /o/ were 495.1 ms (SD = 51.7) and 497.3 ms (SD = 46.8), respectively. The MEG feature vectors up to 450 ms were used to analyze the MEG response pattern to localize the brain activation during recognizing vowels.

III. Feature extraction

The signal obtained by averaging over 80 MEG epochs was converted (using a feature extraction transformation) into a representation more amenable to subject-independent recognition. As inter-subject variability in MEG signals degrades the recognition accuracy of a machine learning system, MEG magnitude was normalized by the following statistical method.

The MEG signal at time t is represented by

$$\mathbf{x}(t) = [x_1(t), \cdots, x_m(t), \cdots, x_M(t)]^T \tag{1}$$

where $x_m(t)$ denotes the observation at the m-th sensor, and the symbol M denotes the total number of MEG sensors. To avoid canceling problems due to the polarity difference between subjects, the MEG magnitude was first calculated by Eq. (2), which is a vector magnitude of paired vertical and horizontal sensors.

$$y_j(t) = \sqrt{x_i^2(t) + x_{i+1}^2(t)} \tag{2}$$

where $y_j(t)$ $(1 \leq j \leq M/2)$ is the magnitude feature.

To reduce the inter-subject variability problems in MEG magnitudes, the magnitude feature is normalized to have zero mean and unit variance.

$$\hat{y}_j(t) = \frac{(y_j(t) - \bar{y}_j)}{\sigma_j} \tag{3}$$

$$\bar{y}_j = \frac{1}{T}\sum_t y_j(t),\ \sigma_j = \sqrt{\tfrac{1}{T}\sum_t (y_j(t) - \bar{y}_j)^2} \tag{4}$$

where $\bar{y}_j$ denotes the mean magnitude feature, T denotes the total number of samples for each averaged epoch, and σ_j denotes the standard deviation. Fig. 1 shows average MEG response magnitudes from a sensor over the left language area (top figure) and the right back area (bottom figure) of a typical subject. The deflection at 100 ms is clearly strong for both stimuli /a/ and /o/, but the difference between /a/ and /o/ is also seen between 150 and 250 ms.

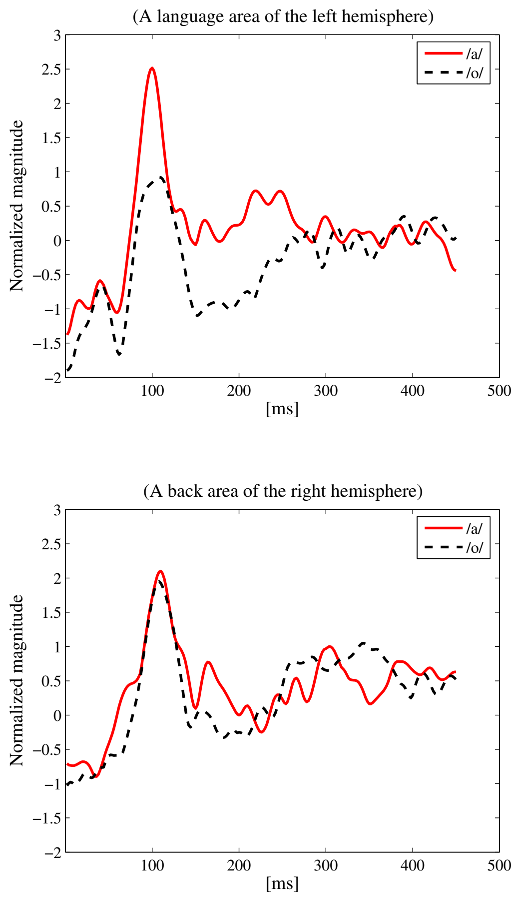

Fig. 1. Normalized MEG-magnitude features obtained at a single MEG-sensor site with a pair of MEG sensors over the left language area (top figure) and the right back area (bottom figure) of a typical subject.

The normalized MEG magnitude feature at each MEG sensor, obtained from Eq. (3), constituted a 61-dimension MEG-magnitude feature vector, as shown in Eq. (5), for further analysis or classification using an AdaBoost algorithm.

$$\hat{\mathbf{y}}(t) = [\hat{y}_1(t), \cdots, \hat{y}_{M'}(t)]^T,\ M' = M/2 \tag{5}$$

IV. MEG-sensor weighting and classification based on AdaBoost

"Boosting" is a technique in which a set of weak classifiers is combined to form one high-performance prediction rule, and AdaBoost [7] serves as an adaptive boosting algorithm in which the rule for combining the weak classifiers adapts to the problem and is able to yield extremely efficient classifiers.

In this paper, AdaBoost is developed to localize brain areas associated with the subject's task, namely the accurate decoding of vowels, by assigning independent weights to each MEG sensor, where the larger the MEG-sensor weight is, the more important the role the brain activity underneath the MEG-sensor plays is.

The AdaBoost algorithm uses a set of training data,

$$\{(\hat{\mathbf{y}}(1), c(1)), \ldots, (\hat{\mathbf{y}}(T), c(T))\} \tag{6}$$

where $\hat{\mathbf{y}}(t)$ is the t-th feature vector of the observed signal, and c is a set of possible labels. For our task, we consider just two possible labels, $c = \{-1, 1\}$, where the label, 1, means a stimulus /a/, and the label, -1, means a stimulus /o/. Next, the training data weight for the t-th training data is initialized as follows:

$$d_1(t) = \begin{cases} \frac{1}{2p}, & c(t) = 1 \\ \frac{1}{2q}, & c(t) = -1 \end{cases}$$

where p and q are the total frame number for the stimulus /a/ and for the stimulus /o/, respectively.

The weak learner on the n-th iteration generates a hypothesis h_n: $\hat{\mathbf{y}}(t) \rightarrow \{-1, 1\}$ that has a small error. In this paper, single-level decision trees (also known as decision stumps) are used as the base classifiers. To analyze areas of the brain that contributed to the accurate decoding of vowels, a decision stump is built for each dimension of the feature vector in our approach, where a dimension corresponds to an MEG sensor.

After training the weak learner of the j-th dimension on the n-th iteration, the error of $h_{n,j}$ is calculated by

$$e_{n,j} = \sum_{t:h_{n,j}(y_j(t)) \neq c(t)} d_n(t) \tag{7}$$

The decision stump which outputs the minimum error among all dimensions is defined as the n-th weak learner. Then, the dimension index which outputs the minimum error is preserved as j_n in order to calculate the weight to each MEG sensor.

$$j_n = \operatorname*{argmin}_j e_{n,j} \tag{8}$$

Next, AdaBoost sets a parameter as follows:

$$\alpha_n = \frac{1}{2} \cdot \log \left[\frac{(1 - e_n)}{e_n} \right] \tag{9}$$

Intuitively, α_n measures the importance that is assigned to h_n. Then the training data weight d_n is updated.

$$d_{n+1}(t) = \frac{d_n(t) \exp\{-\alpha_n \cdot c(t) \cdot h_n(\hat{\mathbf{y}}(t))\}}{\sum_{t=1}^{T} d_n(t) \exp\{-\alpha_n \cdot c(t) \cdot h_n(\hat{\mathbf{y}}(t))\}} \tag{10}$$

Equation (10) leads to an increase in the training data weight for the data misclassified by h_n. Therefore, the training data weight tends to concentrate on "hard" data.

In our approach, the weight for the feature (MEG-sensor weight) is calculated using

$$w_j = \textstyle\sum_n \alpha_n \delta_{j_n,j} \tag{11}$$

where $\delta_{j_n,j}$ is the Kronecker's delta, which has the value 1 if j_n is j, and 0 otherwise.

In order to classify MEG response patterns, after N-th iteration, the final hypothesis, $H(\hat{\mathbf{y}}(t))$, combines the outputs of the N weak hypotheses using a weighted majority vote.

$$H(\hat{\mathbf{y}}(t)) = sign \left\{ \textstyle\sum_{n=1}^{N} \alpha_n h_n(\hat{\mathbf{y}}(t)) \right\} \tag{12}$$

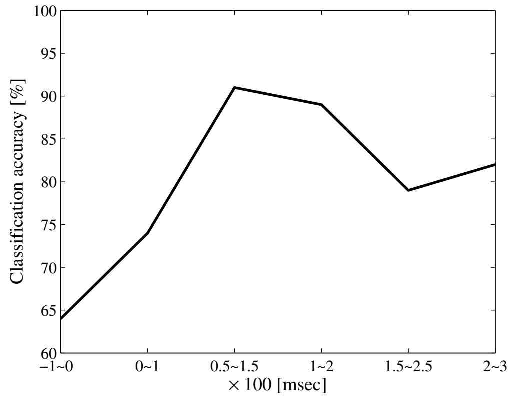

Fig. 2. Classification accuracy using AdaBoost in each 100-ms latency range.

V. Analysis of recorded MEG data

A. Analysis conditions

The AdaBoost-based analysis was evaluated on neuromagnetic responses to recognition of the vowel sounds /a/ and /o/. The total number of subjects was 4. All subjects were used for both the training and test data. The AdaBoost algorithm was independently applied to every latency range, where a single latency range contains 50 samples (about 100 ms with a sampling frequency of 497 Hz). The classification decision was made for each time instance. The frame period was set to 25 samples, meaning that the 100-ms latency range moves about every 50 ms from 0 ms until 350 ms. Since the reaction times for both speech sounds were about 500 ms, we assumed the discrimination was finished by 400 ms at latest, resulting in the final latency range from 200 ms to 300 ms for further analysis.

B. Analysis results

Fig. 2 shows the average classification accuracies. As can be seen in this figure, the classification accuracy first increased as a function of time, reached a maximum value of 91.0% in the latency range between 50 and 150 ms. The AdaBoost method gives significantly better classification performance, compared with that at the 100-ms pre-stimulus baseline, where no stimulus was presented, resulting in the resting state of the brain.

To localize the MEG sensors that are important (considered to have contributed to the processing of vowel recognition) for MEG activity pattern classification using AdaBoost, the MEG-sensor weights (w_j in Eq. (11)) obtained from the AdaBoost method are displayed on a topological plot of the scalp in Fig. 3. They show color-coded average weights for each MEG sensor in each latency range. The more important or more highly weighted MEG sensors for classifying neuromagnetic responses are shown in darker colors. The black areas indicate that this area of the brain played an important role in classification of neuromagnetic responses to vowel recognition. The larger weights in the latency range both between 50 and 150 ms and between 100 and 200 ms, where high accuracy was

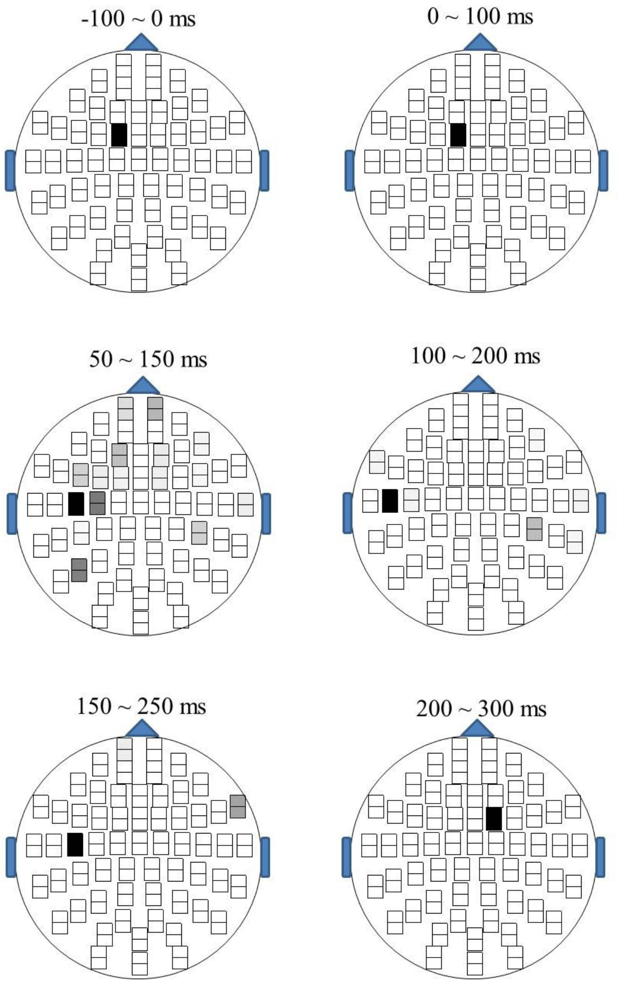

Fig. 3. MEG-sensor weighting based on AdaBoost for the subject-independent case.

achieved, are seen to be in the left language area. The weight estimated using AdaBoost seemed to be a sparse distribution, compared with that using MKL [6], but for both approaches, the larger weights in the latency range between 50 and 150 ms, between 100 and 200 ms, and between 150 and 250 ms were seen to be in the language area of the left hemisphere. Also, our AdaBoost-based weighting method could reduce the computation cost to about half that of the MKL method in this experiment.

VI. Conclusion

We presented a new MEG-sensor weighting method using an AdaBoost algorithm for analyzing areas of the brain that contributed to the accurate decoding of two vowels. Our subject-independent analysis results showed a high classification accuracy of 91.0% obtained in the latency range between 50 and 150 ms for a two-vowel recognition task. The brain area covered by the MEG sensors with the larger weight obtained by our AdaBoost method corresponded to the language area of the left hemisphere. Some differences in the brain activity area obtained from other machine learning systems will be investigated. Also, as the magnetic fields generated by brain activity are extremely weak and usually largely contaminated by external magnetic noises, we will have to develop a noise-robust feature extraction method. In addition, we will have to employ an adaptation approach to overcome the inter-subject variability, especially for the discrepancies between the subjects for training and testing.

Acknowledgments

This research was funded by the Japan Society for the Promotion of Science and a National Science Foundation Science of Learning Center grant to the University of Washington's LIFE Center (SBE-0354453 to P.K. Kuhl).

References

[1] L. Gootjes, T. Raij, R. Salmelin, and R. Hari, "Left-hemisphere dominance for processing of vowels: a whole-scalp neuromagnetic study," *NeuroReport*, vol. 10, pp. 2987-2991, 1999.

[2] I. Miettinen, H. Tiitinen, P. Alku, and P. J.C. May, "Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds," *BMC Neuroscience*, pp. 11-24, 2010.

[3] N. F. Ince, F. Goksu, G. Pellizzer, A. Tewfik, and M. Stephane, "Selection of Spectro-Temporal Patterns in Multichannel MEG with Support Vector Machines for Schizophrenia Classification," *EMBS*, pp. 3554-3557, 2008.

[4] F. Asano, M. Kimura, T. Sekiguchi, and Y. Kamitani, "Classification of Movement-Related Single-Trial MEG Data Using Adaptive Spatial Filter," *ICASSP*, pp. 357-360, 2009.

[5] J. M. Baker, A. M. Chan, K. Marinkovic, E. Halgren and S. S. Cash, "Robust Representations of Cortical Speech and Language Information," *ICASSP*, pp. 785-788, 2011.

[6] T. Takiguchi, T. Imada, R. Takashima, Y. Ariki, J.-F. L. Lin, P. K. Kuhl, M. Kawakatsu, M. Kotani, "A New Multiple-Kernel-Learning Weighting Method for Localizing Human Brain Magnetic Activity," *ICASSP*, pp. 761-764, 2012.

[7] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of computer and system sciences*, 55, pp. 119-139, 1997.

[8] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, 37(3), pp. 297-336, 1999.

[9] J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression: a statistical view of boosting," *Annals of statistics*, 38(2), pp. 337-374, 2000.

[10] S. Z. Li and Z. Q. Zhang, "FloatBoost Learning and Statistical Face Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, pp. 1112-1123, 2004.

[11] A. Grubb and D. Bagnell, "Generalized Boosting Algorithms for Convex Optimization," *ICML*, pp. 1209-1216, 2011.

[12] R. E. Schapire and Y. Singer, "BoosTexter: A Boosting-based System for Text Categorization," *Machine Learning*, 39, pp. 135-168, 2000.

[13] S. Bloehdorn and A. Hotho, "Boosting for Text Classification with Semantic Features," *WebKDD*, pp. 149-166, 2004.

[14] P. Viola and M. Jones, "Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade," *Neural Information Processing Systems*, pp. 1311-1318, 2001.

[15] A. Torralba, K. P. Murphy and W. T. Freeman, "Sharing features: efficient boosting procedures for multiclass object detection," *CVPR*, pp. 762-769, 2004.

[16] C. Dimitrakakis and S. Bengio, "Boosting HMMs with an Application to Speech Recognition," *ICASSP*, pp. 621-624, 2004.

[17] N. Miyake, T. Takiguchi, and Y. Ariki, "Noise detection and classification in speech signals with boosting," *IEEE Statistical Signal Processing Workshop*, pp. 778-782, 2007.

[18] S. W. Foo, Y. Lian, and L. Dong, "Recognition of Visual Speech Elements Using Adaptively Boosted Hidden Markov Models," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, Issue 5, pp. 693-705, 2004.