Toward multi-command auditory brain computer interfacing using speech stimuli

Shuho Yoshimoto,* Yoshikazu Washizawa,* Toshihisa Tanaka,† Hiroshi Higashi,† and Jun Tamura‡
* The University of Electro-Communications, Tokyo, Japan.
E-mail: y0910134@edu.cc.uec.ac.jp, washizawa@uec.ac.jp
† Tokyo University of Agriculture and Technology, Tokyo, Japan.
E-mail: tanakat@cc.tuat.ac.jp, higashi@sip.tuat.ac.jp
‡ Nara Institute of Science and Technology, Nara, Japan.
E-mail: jun-ta@is.naist.jp

***Abstract*—Brain-computer interfaces (BCIs) based on event-related potentials (ERP) are promising tools to communicate with patients suffering from some severe disabled diseases. ERP is evoked by various stimuli such as auditory, olfactory, and visual stimuli. Some auditory based BCIs with certain synthetic tone have been proposed, however, it is still challenging to increase the number of commands in auditory-based BCIs, since it is usually difficult for users to remember and distinguish multiple tones that corresponds to commands. We propose a new auditory BCI framework using speech stimuli. It is easier for users to distinguish different speech stimuli than different simple tones. We show experimental results of four-command BCI. The proposed speech-based BCI achieved a classification accuracy of more than 70 percents.**

I. Introduction

BCI aims to operate a computer with only signals measured from the brain without any muscular movements [3]. Non-invasive measuring devices e.g., Electroencephalograph (EEG), Magnetoencephalography (MEG), and Functional Magnetic Response Imaging (fMRI) are used for the observation of brain activity [3]. Especially, EEG is widely used because of its simplicity in measurement. An EEG cap is often used for EEG measuring. The EEG has the advantage that subjects do not develop claustrophobia interfere with the concentration, unlike fMRI, subjects shall be measured in confined space. In addition, the costs of measuring and the equipments are cheaper than that of many of the other measurement methods. The EEG is available for implementing BCIs that typically use steady state visual evoked potentials (SSVEP) induced by periodic visual stimulation [4], auditory stady state response (ASSR) induced by periodic auditory stimuli [5], desynchronization of μ rhythm [3], and event related potentials (ERPs) caused by external or internal events.[3].

Some BCIs require training for users to control computers and/or devices. The goal of BCI research is to develop a user-friendly system that needs no training. We focused on the use of for the construction of that system. ERP comes out when a subject reacts to every stimulus such as auditory, olfactory, and visual stimuli. ERP based BCIs have the following advantages. First, ERP is not much influenced by feelings. ERP can be observed from the subjects who are excited a little. The mind does not need to be meditative state to measure ERP. However, a little concentration is required. Second, ERP based BCIs are versatile. It is easy to make ERP based BCIs for various applications because a subject has only to react to specific stimuli. Therefore, the BCI with ERP does not require the training and we can analyze easily. Third, ERP can be measured from most people. Furthermore, ERP is useful to verify whether such as a developmental disorder or an autism for whom difficult to communicate, can understand meaning of the stimuli or feel the stimuli or not.

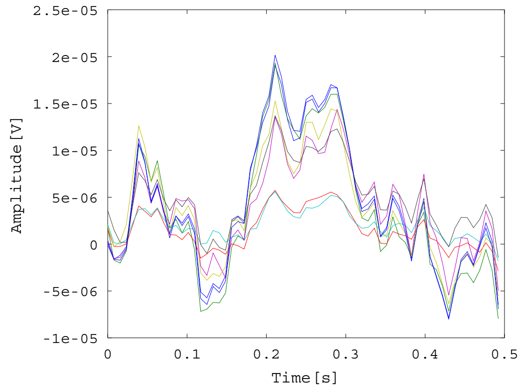

Fig. 1. The example of ERP. Averaged EEG over 50 times after stimulus has been presented. The subject pays attention to target stimuli. There are 8 channels. P300 is observed

P300 is a well-known ERP which comes out after 300 milli seconds from the presentation of a stimuli. “P” stands for a positive potential. P300 is a stable EEG, and has the evident form [2]. In addition, P300 has the feature of occurring a larger peak when the subject has been more attention [2]. The peak makes easy to classify EEG signal. Therefore, P300 is a useful component in ERP based BCIs. We show the waveform of the example of ERP in Fig. 1. Each signal is averaged over 50 times. There are eight channels. Stimuli are presented at the time of 0 seconds. We can see the wave forms in all channels that has the positive peaks around 300ms

A number of ERP-BCIs adopt visual stimuli [4]. However, it is difficult for patients of late stage of Amyotrophic Lateral

Sclerosisln to use these BCIs because their motor function such as the eye movement has been considerably reduced. Moreover, visual sense is typically decined with age because visual sense is susceptible to muscle weakness. Insead of visual stimuli, in this paper, we deal with auditory stimuli for our BCI. Auditory BCIs have following advantages. First, hearing sense is less susceptible to muscle weakness. Second, auditory stimuli can be presented by simple devices such as headphones or speakers compared with the other stimuli. Therefore, it is easy to make an experiment environment. Third, subjects do not need to switch his/her point of view or attitude. Therefore, the brain wave is not affected by muscular movements such as eye movements.

Recently several auditory BCIs with ASSR have been proposed [4][5]. These BCIs use a phenomenon induced brain waves which vibrate at a certain frequency when a subject hears the sound of a pure tone or its modulation. However, when we increase the number of commands of auditory BCIs with ASSR, these BCIs have to change the sound stimulus for each command, and it is difficult for the subject to understand these stimuli because stimuli can sound very similar to the subject. In addition, the subject has to learn which stimulus corresponds to a desired command. Further, currently, many auditory BCIs use binary class signal and contain silence or a dummy stimulus between stimuli [1][6][7]. The dummy stimulus does not give effect to the results even if subjects react it. Due to the dummy stimulus or silence, they have a limitation of information transmission rate.

We aim to increase the amount of information transmission by increasing the number of commands, presenting a stimulus continuously. In addition, we use a speech signals as auditory stimuli for easy understanding. Therefore, subjects do not have to learn the command corresponding to the operation of BCI. The subjects has only to know the meanings of the stimuli to operate the BCI.

We show our system is able to classify in high accuracy by using the Principal Components Analysis (PCA) and the Multiple linear Discriminant Analysis (MDA) even if the command is multi-class .

II. Experimental procedure

We proposed the following experiment procedure to increase the amount of information transmission, and cause a P300.

First, we measured the brain wave by using an active EEG (g.GAMMAcap[2], g.LADYbird (active), g.GAMMAbox manufactured by Guger technologies). The number of channels is 16. The electrodes are located on FCz, FC2, FC1, Cz, CP1, CP2, Pz, POz, P3, P4, TP8, TP7, C3, C4, C5 and C6, the ground is AFz, and the reference is A2 in Fig. 2. Most of channels are placed in parietal to observe ERP, and the remaining channels are placed near the parietal. In addition, we located electrodes near temporal lobe areas related to consciousness. We amplified EEG signal by the biological signal amplifier (BA 1008, Digitex) Second, we used four speech signals, “jou”, “ge”, “sa”, and “yu” as speech stimuli.

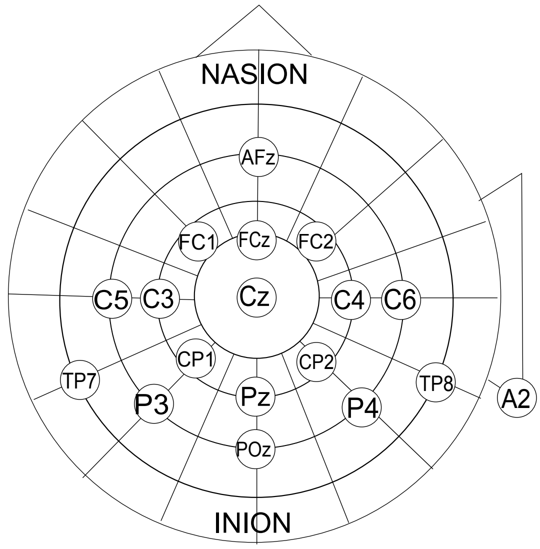

Fig. 2. Location of the electrodes. Conform to the extended 10-20 system. Ground is AFz. Reference is A2. Placed around the Cz

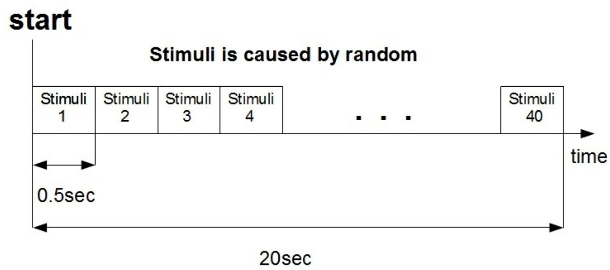

Fig. 3. Presentation scheme of the stimuli. There are four kind of stimulus. Each stimulus is presented ten times in one trial

These respectively mean “up”, “down”,“left”, and “right” in Japanese. The lengths of these stimuli are 0.5 seconds, and presented randomly 40 times in one trial (each stimulus is presented ten times in one trial), i.e., one trial is 20 seconds duration. 50 trials were recorded for each subject. We show the presentation scheme of the stimuli in Fig. 3. Volume of the stimuli is adjusted to a listener-friendly by the subject.

Third, we experimented on four subjects (four males, 19-23 years old). They were asked to close their eyes during the experiment, decide the target stimuli from among the four stimulus for each trial, pay attention to the target stimuli during trial, and count the number of the target stimuli as far as possible.

Fourth, we applied 0.5Hz analog low-cut filter and 100Hz analog high-cut filter by the amplifier. The sampling frequency is 512Hz. We used MATLAB as measuring software and AI 1664 LAX-USB (manufactured by Contec) for the A/D converter.

III. Averaging processing

We did average processing to verify whether we can classify the observed EEG under our experimental scheme or not.

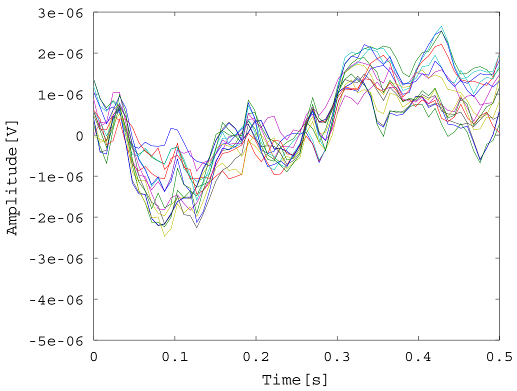

Fig. 4. Averaged waveforms of subject 3 when the subject pays attention to “jou”, and “jou” is presented. There are 16 channels. ERP (N100, P300) is observed.

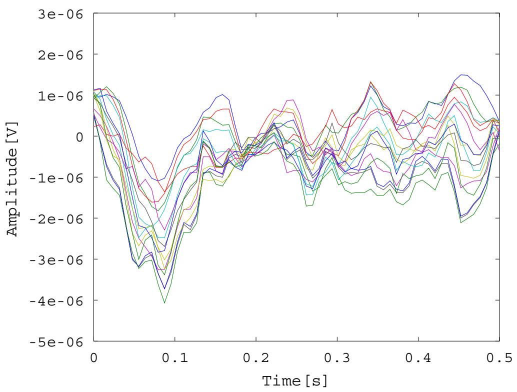

Fig. 5. Averaged waveform of subject 3 when the subject does **NOT** pay attention to “yu” and “yu” is presented. There are 16 channels. N100 is observed. P300 is not observed.

Figs. 4 and 5 show examples of observed waveform in our experiment. Fig.4 is the case that the target stimuli is presented to the subject, and Fig.5 is the case that the other stimuli is presented to the subject. N100 is observed from the both figures. N100 appears after 100 milli seconds from the presentation of the stimulus. “N” means a negative potential. N100 is not related to the conscious, and comes out when subjects receive any stimuli. On the other hand, P300 comes out only in Fig. 4.

IV. Pattern Classification

We conducted off-line analysis for the recorded data. We applied a 50Hz notch filter to remove the noise from outlet, then applied a band-pass filter whose pass-band is $[f_l \ \ f_h]$, where f_h is set to be less than 50Hz. Then the signal was down-sampled to 128Hz.

For each trial, we have 40 ERP signals. We obtained averaging signals for each class, denoted by $x_j(i,t,c)$, where $j = 1,\ldots,50$ is the index of trial, $i = 1,\ldots,16$ is the index of the electrode, $t = 0,\ldots,63$ is the time index, and $c = 1,\ldots,4$ is the class index. 4096-dimensional feature vector $\boldsymbol{x}_j$ is obtained by vectorizing $x_j(i,t,c)$. The target class of the jth trial is denoted by $y_j \in \{1,2,3,4\}$.

We performed five-fold cross validation, i.e., we randomly separated the feature vectors to five subsets, and one subset was used for the validation, and the remaining four subsets were used for the training. The learning-validation procedure was performed five times using different validation set.

Since the input dimension is very large, we performed the PCA to reduce the dimensionality. The projector of PCA is obtained from the feature vectors of all classes in the training set. The number of the reduced dimension r_{PCA} is obtained from several hyper-parameters set. The transformed feature vector is denoted by $\tilde{\boldsymbol{x}}_j$.

After the dimensional reduction, the MDA was used to extract the features. For each class, the mean vector and the variance-covariance matrix were estimated from the training feature vectors.

$$\boldsymbol{m}_c = \frac{1}{N_c} \sum_{\{j|y_j=c\}} \tilde{\boldsymbol{x}}_j \tag{1}$$

$$\boldsymbol{\Sigma}_c = \frac{1}{N_c} \sum_{\{j|y_j=c\}} (\tilde{\boldsymbol{x}}_j - \boldsymbol{m}_c)(\tilde{\boldsymbol{x}}_j - \boldsymbol{m}_c)^\top, \tag{2}$$

where N_c is the number of training vectors of the cth class. Then we obtained the within-class and between-class scatter matrices,

$$\boldsymbol{S}_w = \sum_{c=1}^{4} \boldsymbol{\Sigma}_c \tag{3}$$

$$\boldsymbol{S}_b = \sum_{c=1}^{4} \sum_{c'=1}^{c-1} (\boldsymbol{m}_c - \boldsymbol{m}_{c'})(\boldsymbol{m}_c - \boldsymbol{m}_{c'})^\top. \tag{4}$$

MDA obtains the transform $\boldsymbol{W}$ that maximizes the generalized Rayleigh quotient,

$$J(\boldsymbol{W}) = \frac{\mathrm{Trace}[\boldsymbol{W}^\top \boldsymbol{S}_b \boldsymbol{W}]}{\mathrm{Trace}[\boldsymbol{W}^\top \boldsymbol{S}_w \boldsymbol{W}]}. \tag{5}$$

The problem is reduced to the generalized eigenvalue problem,

$$\boldsymbol{S}_b \boldsymbol{w} = \lambda \boldsymbol{S}_w \boldsymbol{w}, \tag{6}$$

and the generalized eigenvectors corresponding to the largest eigenvalues are the solution of eq. (5). We used three (the number of classes minus one) eigenvectors for the transform, $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3]$. Both the training and the validation vectors were transformed into the three dimensional subspace by $\boldsymbol{W}$.

Finally, we assumed that the three dimensional transformed vectors are Gaussian distributed, and performed the Baysian classification. We obtained the values of probability density functions (PDF), $f_c(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{W}^\top \boldsymbol{m}_c, \boldsymbol{W}^\top \boldsymbol{\Sigma}_c \boldsymbol{W})$, where $\mathcal{N}$ denotes the PDF of the Gaussian distribution. The jth validation vector $\tilde{\boldsymbol{x}}_j$ is classified to the class whose logarithmic value of PDF is maximum,

$$\operatorname*{argmax}_{c=1,\ldots,4} \ \log f_c(\boldsymbol{W}^\top \tilde{\boldsymbol{x}}_j). \tag{7}$$

TABLE I
RESULTS OF OUR EXPERIMENT

Subject	Classification rate(%)	Standard deviation
1	52	24
2	74	11
3	54	15
4	60	12
Average	60	16

TABLE II
PARAMETERS THAT SHOW THE HIGHEST ACCURACIES

Subject	r	f_l [Hz]	f_h [Hz]
1	30	0.05	35
2	25	0.1	40
3	40	0.2	20
4	35	0.05	30

V. CLASSIFICATION RESULT

The classification results are shown in Table I, and parameters are shown in Table II. P300 was observed from all subjects. The results of classification are satisfying because there are no big variations. From the definition of the transfer-bit rate [8], 50 to 70 percent classification accuracy of the four command BCI is equivalent to 76 to 93 percent classification accuracy of two command BCI,

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2[(1 - P)/(N - 1)], \quad (8)$$

where B is the amount of transfer-bit, P is the probability that the desired selection, and N is the number of commands.

The relation between the rank of PCA and the classification accuracies are shown in Figure 6. When the rank of PCA is between 25-35, accuracies are the highest.

VI. CONCLUSIONS

We have proposed a new framework of auditory BCI using ERP. These classification rates are satisfactory compared with the other auditory BCI experiment performed in binary class. The classification rates of the subject 2 was the highest. The performance may be higher if the speech stimuli is slower. The relation between the margin of the stimuli and the transfer-bit rate should be investigated for future works.

This pattern classification was carried out off-line. For real applications, real time analysis is necessary. Therefore, for future works, the system should be extended to real-time classification.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 24700163 and 24360146.

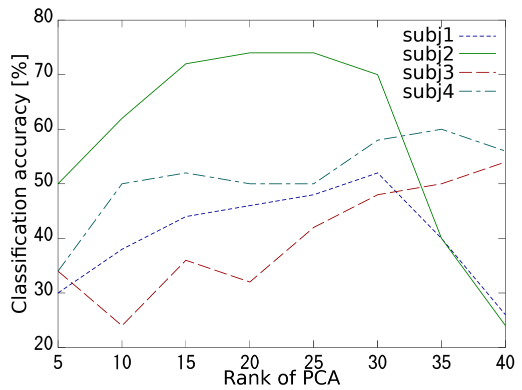

Fig. 6. Rank of PCA and classification accuracies. 50 to 70 percent classification accuracy of the four command BCI is equivalent to 76 to 93 percent classification accuracy of two command BCI.

REFERENCES

[1] S. Kanoh, K. Miyamoto and T. Yoshinobu, "A brain-computer interface (BCI) system based on auditory stream segregation," *30th Annual International IEEE Engineering in Medicine and Biology Society (EMBS) Conference*, Vancouver, British Columbia, Canada, pp. 642-645, August, 2008.

[2] W. S. Pritchard, "Psychophysiology of P300," *Psychological Bulletin*, vol. 89, no. 3, pp. 506-540, 1981.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767-791, 2002.

[4] M. Middendorf, M. McMillan, G. Calhoun, and K. S. Jones, "Brain computer interfaces based on the steady-state visual-evoked response," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 211-214, 2000.

[5] H. Higashi, T. M. Rutkowski, Y. Washizawa, A. Cichocki, and T. Tanaka, "EEG auditory steady state responses classification for the novel BCI,"*33rd Annual International Conference of the IEEE EMBS*, Boston, Massachusetts USA , pp. 4576-4579, August, 2011.

[6] D. Kim, J. Cho, H. Hwang, J. Lim, and C. Im, "A vision-free brain-computer interface (BCI) paradigm based on auditory selective attention," *33rd Annual International Conference of the IEEE EMBS*, Boston, Massachusetts USA, pp. 3684-3687, August, 2011.

[7] H. Gao, M. Ouyang, D. Zhang, B. Hong, "An auditory brain-computer interface using virtual sound field," *33rd Annual International Conference of the IEEE EMBS*, Boston, Massachusetts USA, pp. 4568-4571, August, 2011.

[8] J. R. Wolpaw, N.Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan,"Brain-computer interface technology: a review of the first international meeting," *IEEE Trance on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164-172, June, 2000.