Classifying P300 Responses to Vowel Stimuli for Auditory Brain-Computer Interface

Yoshihiro Matsumoto*, Shoji Makino*, Koichi Mori[†] and Tomasz M. Rutkowski*[‡]
* Multimedia Laboratory, TARA Center, University of Tsukuba, Tsukuba, Japan
E-mail: tomek@tara.tsukuba.ac.jp
[†] Research Institute of National Rehabilitation Center for Persons with Disabilities, Tokorozawa, Japan
[‡] RIKEN Brain Science Institute, Wako-shi, Japan

***Abstract*—A brain-computer interface (BCI) is a technology for operating computerized devices based on brain activity and without muscle movement. BCI technology is expected to become a communication solution for amyotrophic lateral sclerosis (ALS) patients. Recently the BCI2000 package application has been commonly used by BCI researchers. The P300 speller included in the BCI2000 is an application allowing the calculation of a classifier necessary for the user to spell letters or sentences in a BCI–speller paradigm. The BCI–speller is based on visual cues, and requires muscle activities such as eye movements, impossible to execute by patients in a totally locked-in state (TLS), which is a terminal stage of the ALS illness. The purpose of our project is to solve this problem, and we aim to develop an auditory BCI as a solution. However, contemporary auditory BCI–spellers are much weaker compared with a visual modality. Therefore there is a necessity for improvement before practical application. In this paper, we focus on an approach related to the differences in responses evoked by various acoustic BCI–speller related stimulus types. In spite of various event related potential waveform shapes, typically a classifier in the BCI speller discriminates only between targets and non–targets, and hence it ignores valuable and possibly discriminative features. Therefore, we expect that the classification accuracy could be improved by using an independent classifier for each of the stimulus cue categories. In this paper, we propose two classifier training methods. The first one uses the data of the five stimulus cues independently. The second method incorporates weighting for each stimulus cue feature in relation to all of them. The results of the experiments reported show the effectiveness of the second method for classification improvement.**

I. Introduction

A brain–computer/machine interface (BCI/BMI) is a technology that allows any computerized device to be operated without muscle activity [1]. This technology is expected to enable ALS patients to communicate independently. A state-of-the-art BCI is implemented on the basis of the classification of feature values extracted from brain activity, but their accuracy is currently insufficient for a wide range of applications. Therefore, there is a need to improve the classification accuracy if we are to realize a better quality of interface. Recent hot research projects designed to achieve this improvement focus on an auditory BCI (aBCI), which is a method of applying auditory evoked responses (AER) generated by corresponding sound stimuli [2].

There are several types of BCI, as shown in Figure 1 [3]. The aBCI is grouped under non–invasive BCI and stimulus–driven BCI. An invasive BCI requires cranial surgery to embed the electrodes on/in the surface of the cortex. It enables the BCI system to measure brain activity in a high signal to noise ratio (SNR) condition with higher infection and side effect complication risk. In contrast, the non–invasive BCI can be used by attaching electrodes to the surface of the human head, but the admissible SNR is lower than with an invasive system. Therefore, the development of accurate applications with a non–invasive BCI is more challenging. The next difficulty is related to the design of simple interface paradigms, such as an imagery paradigm based on a user's intentional brain wave modulation induced by the imagining/planning of certain actions (imagery of movement, etc.). A solution to this problem is a method that applies brain responses to artificial sensory stimuli to generate commands. This is called a stimulus–driven BCI (see Figure 1). For example, a user chooses to concentrate on one stimulus from multiple stimuli presented sequentially. This evokes a specific event related potential (ERP) pattern, called the "aha–response" or P300 (see Figure 2), since it is a positive electroencephalography (EEG) deflection around 300 ms after the onset of the stimulus. The BCI application classifies the responses and translates them into computerized commands. Many conventional studies about stimulus–driven BCI focus on a visual modality, which uses evoked responses to visual stimuli such as flashing or switching ON/OFF of a signal. However, visual BCI requires eye movement for selecting commands and is difficult for advanced ALS patients.

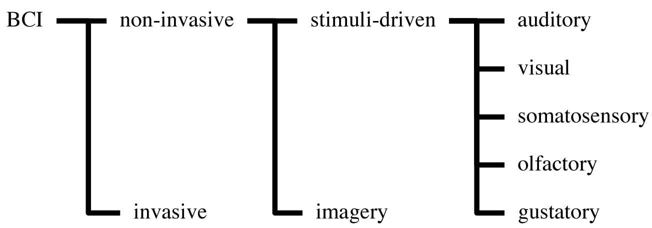

Fig. 1. A tree diagram of the BCI types. BCI's can be divided into invasive and non–invasive, based on their implementation. Furthermore, the non–invasive BCI can be divided into imagery and stimuli–driven, depending on whether the sensory stimulus is used or not. Following the general division of the senses into five categories, there are also five types of the stimuli–driven BCI: visual; auditory; somatosensory; olfactory; and gustatory.

The aBCI, which requires no muscle activity, is one of the best solutions to the problem. Even for healthy users, aBCI

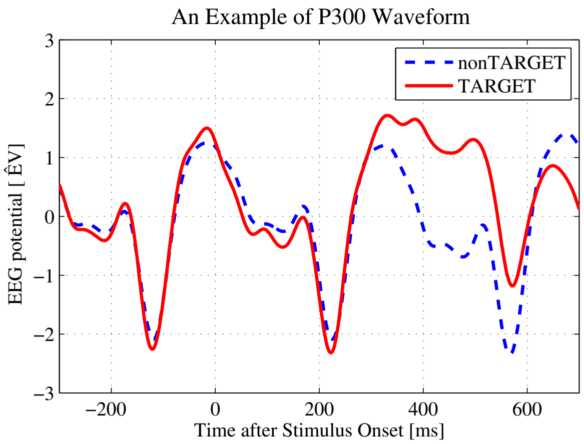

Fig. 2. An example of P300 response. The horizontal axis represents the time course after the stimulus onset, and the vertical depicts the EEG evoked potential in μV. The blue dashed line represents the ERP in response to non–target stimuli and the red solid line represents the targets. Each waveform shows the averaging of 2700 epochs. The target response has a positive ERP deflection after 300 ms compared with the non–targets.

has advantages, as follows:

- no need to fix the viewing direction;
- ease in preparing stimulation devices;
- no known cases of epilepsy evoked by auditory stimuli.

However, the performance of aBCI is still not good for practical use. Even though recent studies have reported reliable classification accuracy, there is still plenty of room for improvement.

Recently, a BCI speller, which is included in BCI2000, has been a commonly used application. A recent study by Chang et al. [4] made a comparison between auditory– and visual–BCI (vBCI) spellers. It reported that the accuracy of an aBCI speller is lower than of a vBCI speller in spite of almost the same experimental conditions. One of the reasons for this, which we now focus on, is that the waveform of AER differs depending on the auditory stimuli. For example, EEG responses differ for sound stimuli from the right or the left [6], [7], or depending whether sound images are virtual or real sound sources [5]. The state–of–the-art BCI2000 classifier is common for all stimulus types. This can limit the accuracy in training the classifier to the features of only responses to all the auditory stimuli together. We propose a method for the improvement of the auditory BCI–speller performance by solving this problem.

II. Methods

Usually the standard BCI2000 package is used, as in the following procedures:

- a subject uses a P300 based BCI–speller for typing several letters or a short text sentence in order to record training data;
- the user executes the P300 classifier in order to generate classification parameters from the training data obtained in the previous step;
- next the classifier parameters are loaded to the BCI–speller in order to run a test or an online spelling session.

The following sections explain details of the usual BCI2000 classification procedure and our two proposed methods.

A. BCI2000 EEG Recording and Processing Environment

In the preprocessing stages of BCI2000, the EEG data is filtered in order to remove direct current and power line related alternate current interference. The filtered data are next segmented into epochs referring to each stimulus onset. Each epoch has information about the stimulus training section labeled as target or non–target. The test section uses a stimulus code with labels of presented stimuli. The above process is referred to as epoch segmentation.

In the next step, feature extraction is conducted in which epoch–segmented data are filtered with a moving average filter within the BCI2000 environment [8]. After this, the data are decimated by factor f_d. The number of averaged samples of the moving average and f_d decimation are the same. When the decimation procedure is finished for all M channels, the following matrix is obtained:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,L} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M,1} & x_{M,2} & \cdots & x_{M,L} \end{bmatrix}, \tag{1}$$

where L is the number of remaining samples after decimation. A feature vector is obtained by reshaping the above matrix into $M \times L$ dimensional vector $\mathbf{x}$:

$$\mathbf{x} = [x_{1,1}, \ldots, x_{1,L}, x_{2,1}, \ldots, x_{2,L}, \ldots, x_{M,L}]. \tag{2}$$

A classification method used in BCI2000 is a stepwise linear discriminant analysis [9]. In the training session, the classifier generates a weight vector $\mathbf{w}$ as in the following equation:

$$\mathbf{w} = (\mathbf{C}^{(1)} + \mathbf{C}^{(0)})^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(0)}), \tag{3}$$

where $\bar{\mathbf{x}}^{(k)}$ is the average over the epochs in label group k; $k = 1$ stands for targets; and $k = 0$ for non–targets. $\mathbf{C}^{(k)}$ is a covariance matrix of feature vector variables in the group labeled as k.

The variables from the vector feature are next selected with a stepwise method [9], which subsequently leads to a classification function output score obtained from each epoch input. The predicted target is selected, based on averages for each coded stimulus and the maximum result is selected. Thus, the predicted score $\hat{y}$ is calculated as in the following equation using $\mathbf{w}$ and $\bar{\mathbf{x}}_s$ values, which is the averaged feature vectors obtained from responses to the stimulus coded as s:

$$\hat{y} = \arg\max_s \mathbf{w}^T \bar{\mathbf{x}}_s. \tag{4}$$

B. The Proposed Methods

In the classical BCI2000 [8] implementation, the P300 classifier is trained by the set of all feature vectors X:

$$X = X_1 \cup X_2 \cup \ldots \cup X_S, \tag{5}$$

where, X_s is a set of feature vectors obtained as a response to the stimulus coded as s. The variable S is a number of stimuli, which is equal to the number of BCI commands in the current implementation. However, the features of each stimulus code usually have differences depending on stimulus features, for example, the origin of stimulus sound source directions [6], [7].

In our hypothesis, the differences in event related potential shapes interfere with the accurate BCI–speller classification and we expect that using a separate classifier $\mathbf{w}_s$ for each stimulus code s shall improve the final results. During the training session, a classifier $\mathbf{w}_s$, which is used subsequently for testing $\bar{\mathbf{x}}_s$, is trained by a set of feature vectors $X_{\mathbf{w}_s}$. In the test session, the $\bar{\mathbf{x}}_s$ is evaluated by classification of stimuli coded as s. Finally, the code of the classifier $\mathbf{w}_s$, which results in the maximum score among all the trained classifiers indicates the target stimulus.

In this paper, we propose two methods for choosing the sets of feature vectors.

1) Proposed Method 1: The first method proposed is based on a choice of only the response related features, as in equation:

$$X_{\mathbf{w}_s} = X_s \tag{6}$$

We expect that the classifier thus trained will become optimized for a certain stimulus code s, since the features from other stimuli will be not present. However, there is a weak point in this approach related to the lower overall number of training examples available for each response code s.

2) Proposed Method 2: The second method is an attempt to cover the weak point of Method 1. We propose to involve the remaining features of other stimulus codes in the classifier training process. This method is related to the conventional method and it optimizes the classifier by the weighting of features of the related stimulus. We define the training data multi-set $X_{\mathbf{w}_s}$ as follows:

$$X_{\mathbf{w}_s} = X \cup \underbrace{X_s \cup \cdots}_{a-1}, \tag{7}$$

where a is a parameter, which explains the multiplicity number of elements included in X_s. Parameter a can be optimized according to individual variation or combination of used stimuli. When $a = 1$, the model equals the conventional method.

III. Experiment

In order to verify the validity of our hypotheses and the subsequent boost in the efficiency of the BCI–speller classification by the proposed methods, we conducted a series of EEG experiments. The following sections explain details of the experimental conditions and the offline data processing stages. The procedure of the experiments refers in part to the spatial Japanese vowel BCI-speller experiments reported in [4]. All the experiments were performed at the Life Science Center of TARA, University of Tsukuba, Japan. The online EEG BCI experiments were conducted in accordance with *The World Medical Association Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.*

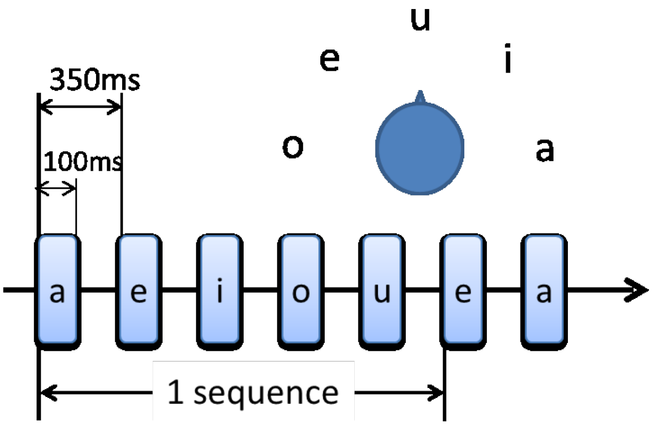

Fig. 3. An example of stimulation protocol in the case of a trial. The stimuli are localized in the upper-right figure. The axis indicates the time course and each node shows a stimulus. One sequence has five stimuli ordered randomly. The sequences come one after another and a subject attends to only one target during the trial.

A. EEG Experimental Conditions

The experiments were conducted with six healthy subjects with the BCI–speller interface included in the BCI2000 package [8]. The stimuli were five synthetic Japanese voice vowels *a, i, u, e,* and *o.* The stimuli were presented with inner ear headphones. The stimuli were virtually localized as follows: *a* from the right; *i* from the front right; *u* directly from the front, *e* from the front left; *o* from the left, respectively. The stimulus duration was set to 100 ms and the inter–stimulus–interval (ISI) was 350 ms (see Figure 3 for a diagram explaining the presentation setup). EEG was recorded with 16 channel g.ladybird active electrodes, and the signals were amplified with g.USBamp from g.tec Medical Instruments, Austria. The EEG electrodes were positioned at the following scalp locations *Cz, CPz, POz, Pz, P1, P2, C3, C4, O1, O2, T7, T8, P3, P4, F3,* and *F4* as in the *10/10 System* locations shown in Figure 4 [10]. In addition, a reference electrode was positioned on *Fpz* and the ground electrode was attached to the left earlobe of the subject. The sampling frequency was set to 512 Hz. The recorded signals were filtered with the following filters:

- notch filter order 4 and cut-off frequencies of 45 Hz and 55 Hz to eliminate first power line interference;
- high–pass filter order 2 with cut-off frequency at 1 Hz;
- low–pass filter order 6 with cut-off frequency at 20 Hz.

The responses were segmented into epochs within intervals from 0 ms to 800 ms referring to the stimulus onset. The

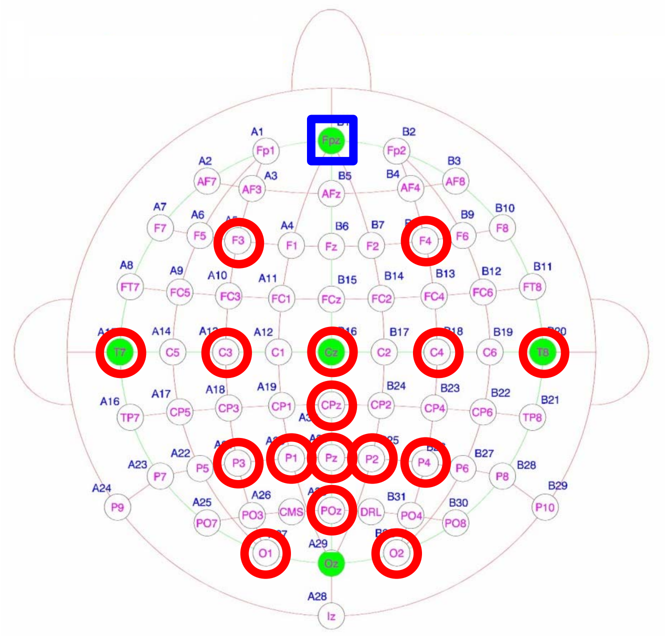

Fig. 4. The *10/10 system* locations of EEG electrodes on the scalp. The graph is based on BIOSEMI's 64-channels EEG cap layout. The red circles represent recording electrodes chosen for the experiments reported in this paper. The blue square is a ground electrode. A reference electrode was attached to the left earlobe.

decimation frequency f_d was set to 32 Hz.

B. EEG Experimental Protocol

The detailed experimental protocol applied to each subject was as follows:

step 1: The target sound stimulus was displayed to the subject;
step 2: All the sound stimuli were sequentially presented in random order (see Figure 3).
step 3: The subject paid attention only to the target stimulus, while ignoring the others.
step 4: After presenting 30 sequences, the BCI2000 application output the predicted stimulus based on a classification of all of the averages.

The above steps $1-4$ defined a single trial, and one section contained five trials. Each subject conducted three sections of the experiment in our protocol.

Based on the recorded EEG data in experiments with six subjects, we performed offline analysis of the commands classification using the conventional and proposed methods. The data were tested with a cross–validation test. The EEG datasets recorded in the first two sessions of experiments were used for classifier training and the remaining session set was evaluated as the classification test. The overall performance was evaluated as the accuracy rate defined by the number of correct outputs divided the number of all outputs.

IV. Results

The results of the classification with the methods presented above are summarized in this section. We present comparisons of results obtained with the conventional approach versus the proposed Methods 1 and 2.

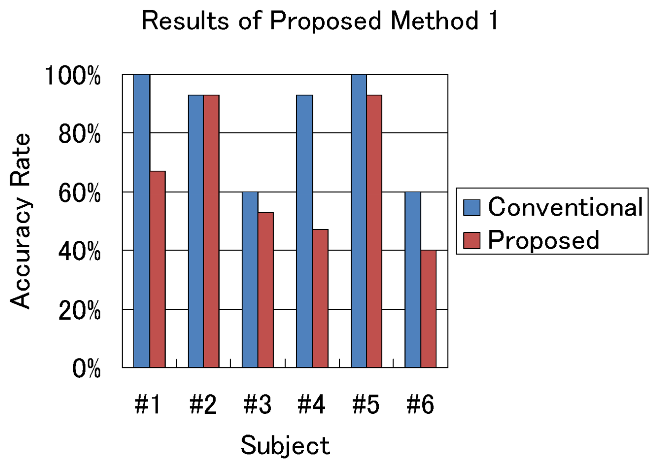

Fig. 5. The results of a comparison of the conventional method and proposed Method 1. The horizontal axis represents the subject number and the vertical axis depicts the accuracy rate. The blue bars show the results of the conventional method and the red show proposed Method 1.

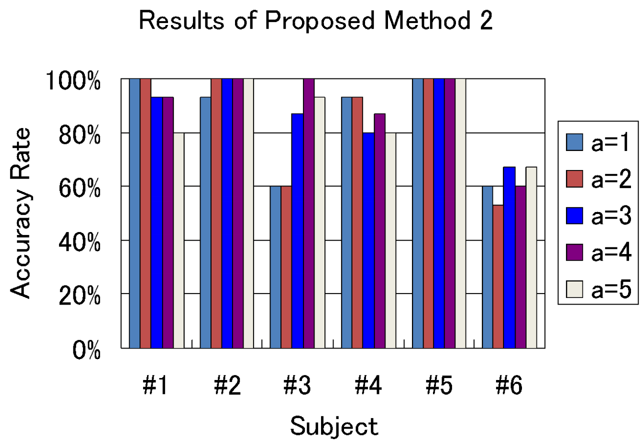

Fig. 6. The results of a comparison of the conventional method and proposed Method 2. The horizontal axis represents the subject number and vertical axis the accuracy rate. The bars show the results for each of the parameter a choices in the range of 1 to 5. The case of $a = 1$ is the same as the conventional method.

A. Results of Proposed Method 1 versus the Conventional

The results of a comparison of the conventional method versus proposed Method 1 are presented in Figure 5. The figure shows the classification accuracy rates of each of the six subjects. The blue bars in Figure 5 show results of the conventional method, while the red bars depict results of proposed Method 1. In the case of subject #2, the proposed method resulted in the same accuracy rate as for the conventional method. For the remaining subjects, the accuracy rates of the proposed method were lower than the conventional one.

TABLE I
THE RESULTS OF METHOD 2 WITH OPTIMAL a FOR EACH SUBJECT.

Subject	Conventional	Proposed 2	(Parameter a)
#1	100.0%	100.00%	($a = 1$)
#2	93.3%	100.00%	($a = 2$)
#3	60.0%	100.00%	($a = 4$)
#4	93.3%	93.33%	($a = 1$)
#5	100.0%	100.00%	($a = 1$)
#6	60.0%	66.67%	($a = 3$)
Average	85.6%	93.3%	

B. Results of Proposed Method 2 versus the Conventional

The results of a comparison between the conventional and proposed Method 2 are shown in Figure 6. This figure also reports the accuracy rates of each subject separately. The bars represent the results for the parameter a in a range $1 - 5$. Parameter setting $a = 1$ is equal to the conventional method. Subjects #1 and #4 showed the best results in case of $a = 1$, but the other subjects showed better results or attained a 100% score when an optimal parameter a was selected for each of them separately (see Table I with details).

V. CONCLUSIONS

The purpose of our project was to develop a spatial auditory BCI which advanced ALS patients could use successfully. In order to realize the purpose, we aimed to enhance the BCI–speller included in the original BCI2000 package for the spatial auditory modality. There have been conventional studies which have reported difficulty in the accuracy of the auditory ERP based classification in comparison to classic visual cases. Therefore, we proposed to improve the spatial auditory BCI by developing the novel classifier concept. This paper has reported two methods, which used an independent classifier for each vowel stimulus. We conducted experiments to verify the efficiency of such a proposal. According to the results obtained, proposed Method 1, which is the classifier training method using only related features, resulted in lower performance. This may be caused by the classifier over–fitting due to a decrease in the quantity of the available training data. However, proposed Method 2, which applied multiple parameters to train the classifier based on a variable data set, showed better results than the conventional method. In particular, subject #3 performed initially at 60% accuracy with the conventional method and later reached 100% using proposed Method 2 with the weight parameter a set to 4. This result supports the efficiency of our approach. The results of the remaining subjects also show that the final performance depended on the setting of parameter a.

However, this method still leaves room for improvement. The proposed methods use ensemble classifiers and the classification scores are unfortunately not normalized. We aim to solve this problem in our future research.

ACKNOWLEDGMENTS

This research was supported in part by the Strategic Information and Communications R&D Promotion Program (SCOPE) no. 121803027 of The Ministry of Internal Affairs and Communication in Japan, and by KAKENHI, the Japan Society for the Promotion of Science grants no. 12010738 and 24360146.

REFERENCES

[1] J. Wolpaw and E. W. Wolpaw, eds., *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, 2012.

[2] M. Schreuder, B. Blankertz, and M. Tangermann, "A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue," *PLoS ONE*, vol. 5, p. e9813, 04 2010.

[3] Y. Matsumoto, N. Nishikawa, S. Makino, T. Yamada, and T.M. Rutkowski, "Auditory steady-state response stimuli based BCI application - the optimization of the stimuli types and lengths," in *Proceedings of the Fourth APSIPA Annual Summit and Conference 2012, (Hollywood, CA, USA)*, pp. paper #285, December 3-6, 2012.

[4] M. Chang, N. Nishikawa, Z. R. Struzik, K. Mori, S. Makino, D. Mandic, and T. M. Rutkowski, "Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms," in *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*, (Asilomar Conference Center, Pacific Grove, CA USA), p. Article ID: 156, Graz University of Technology Publishing House, Austria, June 3-7, 2013.

[5] N. Nishikawa, S. Makino, and T. M. Rutkowski, "Spatial auditory BCI paradigm based on real and virtual sound images generation," in *Proceedings of the Fifth APSIPA Annual Summit and Conference (APSIPA ASC 2013)*, (Kaohsiung, Taiwan), pp. (accepted, in press), APSIPA, October 29 - November 1, 2013.

[6] M. L. Gamble and S. J. Luck, "N2ac: An ERP component associated with the focusing of attention within an auditory scene," *Psychophysiology*, vol. 48, no. 8, pp. 1057–1068, 2011.

[7] T. M. Rutkowski, A. Cichocki, and D. P. Mandic, "Spatial auditory paradigms for brain computer/machine interfacing," in *International Workshop On The Principles and Applications of Spatial Hearing 2009 (IWPASH 2009) - Proceedings of the International Workshop*, (Miyagi-Zao Royal Hotel, Sendai, Japan), p. P5, November 11-13, 2009.

[8] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw: BCI2000: A General-Purpose Brain-Computer Interface (BCI) System, IEEE Trans Biomed Eng, 51(6), June 2004.

[9] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 speller," *Journal of Neural Engineering*, vol. 3, no. 4, p. 299, 2006.

[10] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage*, vol. 34, no. 4, pp. 1600 – 1611, 2007.