31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Adaptive Active Auditory Brain Computer Interface

Bo Hong*, Bin Lou, Jing Guo, Shangkai Gao

Abstract— **An active paradigm was employed to produce reliable and prominent target response in an auditory brain computer interface (BCI), in which subject's voluntary recognition of the property of a target human voice enhances the discriminability between target and non-target EEG response. Furthermore, to adaptively decide the optimal number of trials being averaged for SVM classification, a statistical approach was proposed to convert each sample's margin in support vector space into probabilities of each voice choice being the target. In a testing of 8 subjects' EEG data from the active auditory BCI experiment, the proposed adaptive approach needs only about 4-6 trials to reach the equivalent accuracy of 15-trial averaging. The improved information transfer rate suggests the advantage of adaptive strategy in an active auditory BCI.**

Keywords—**brain computer interface, auditory, adaptive, support vector machine, late positive component**

I. Introduction

Because of reliable and easy detection of visual responses over the scalp, most brain computer interfaces (BCI) employ visual modality for presenting the stimulus[1][2]. But for a totally "locked-in" patient who has compromised vision or loses the control of eye movement, a BCI paradigm using auditory stimulus is more preferable, since hearing usually preserved in severely paralyzed patients, even in amyotrophic lateral sclerosis (ALS)[5]. Using auditory stimulus as a substitute for visual feedback, the subject can regulate either the amplitude of slow cortical potential [3] or sensorimotor rhythm [4] almost equally well for BCI purposes. But in these conditions, the auditory stream does not carry any information the subject wants to communicate. Alternatively, the BCI choices can be embedded in an auditory stimulus sequence. Sellers et al. tested an auditory BCI with an oddball paradigm containing four words. The subjects were required to attend to the target word and P300 was elicited [5], making it possible for target and non-target discrimination. Recently, we proposed a more active paradigm, with a subject's voluntary recognition of the property of a target human voice, which enhances the discriminability between brain responses to target and non-target voices [6]. The N2 and late positive component (LPC) in event related potentials (ERP) were used as features for identifying the target among multiple voice choices. In current study, this active auditory BCI paradigm was adopted, but with a new approach of adaptive classification.

For BCIs using event-related potentials, coherent averaging is an essential step to enhance the signal to noise ratio (SNR). More averaged trials increase the accuracy of target detection, with a cost of more time for selecting a target and thus lowering the speed of BCI communication. Thus, there is a tradeoff between accuracy and speed in deciding the number of trials being averaged. A common approach is to fix the number of trials empirically. In auditory BCI, more averaged trials are needed because of poor SNR of auditory EEG response. And usually, presenting an auditory stimulus needs more time than visual modality. For these reasons, deciding an optimal number of trials being averaged is even crucial in auditory BCI. In this paper, a statistical approach is proposed to adaptively decide the number of trials being averaged for a decision. The value of the discriminant function of each sample in support vector machine (SVM) classifier was converted into probabilities of each BCI choice being the target ($P*$). If the highest probability $P*$ among all BCI choices reached a pre-defined threshold estimated from training accuracy, the adaptive algorithm terminated the averaging and selected the target with highest probability $P*$. In an online experiment of 8 subjects using the aforementioned active auditory BCI paradigm, the proposed adaptive method needs only about 4-6 trials to reach the equivalent accuracy of 15-trial averaging, demonstrating the advantage of this adaptive approach.

II. Methods

A. Subjects and experimental setup

Eight subjects (six males and two females) with normal hearing were included in this study. They all gave informed consent prior to the experiments. All auditory stimuli were presented to subjects by using insert earphones (Etymotic ER2, Illinois, USA). The EEG was recorded using a standard EEG cap (Electro-Cap, Neuroscan, USA) with 30 surface electrodes, based on the International 10-20 system, referenced to linked-mastoids. The signals were digitized at a rate of 1000Hz, band-pass filtered at 0.05-200Hz (SynAmps2, Neuroscan, USA). All electrode impedances were kept below 5 kΩ during data recording.

This work was supported in part by the National Natural Science Foundation of China (60675029 and 30630022).

* B Hong is with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China (e-mail: hongbo@tsinghua.edu.cn).

B Lou, J Guo, X Gao and S Gao are with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China.

Eight digits spoken in Chinese were presented randomly to form a stimulus sequence (single trial). Each digit was 200ms in duration and adjusted to have the same intensity. The stimulus onset asynchrony (SOA) was randomized from 250 to 450ms. The laterality of the voice was totally random. The digit voices were presented either to the subject's right ear or to the left. One block consisted of 15 trials. Before each block, the subject was told which digit was the 'target'. Subjects' task was to identify the target digits in the sequence and discriminate their laterality (silently saying 'left' or 'right') while ignoring non-target voices. For each subject, a total of 8 blocks (120 trials) of stimulus were presented and corresponding EEG data were registered.

B. Preprocessing and training the classifier

Based on the spatial pattern of ERP components shown in Figure 1 (See Result section for details), P3 electrode over parietal cortex was chosen as the only one signal channel for the following classification. 800ms segments of EEG data after each stimulus onset were extracted and down-sampled to 20Hz, resulting in a feature vector with 17 sample points.

Support vector machine (SVM) is a powerful and robust method for pattern classification, showing superior performance the detection of event-related potentials in BCI[2,14]. Hence, SVM was adopted as the classification method for our active auditory system. Before training the classifier, all features were normalized to [-1, 1] and the class labels were set as $y_k \in \{-1,1\}$ representing the class of non-target and target respectively.

To improve the SNR of the features, we averaged feature vectors of three consecutive single trials to get one training sample. Since there are 15 trials in one block, 13 training samples were obtained by using one-step sliding window. Note that there are seven non-target voices and only one target voice, which leads to uneven training dataset sizes of these two classes. If we randomly choose the same number of non-target samples as the target samples, the variation of the classifier is very high. However, if non-target samples are much larger than target ones, the generalization ability of the classifier will suffer. Therefore, the non-target to target ratio (NTR) of training samples should be appropriately chosen. Empirically, NTR=3 is a tradeoff between classifier stability and generalization ability. To obtain an optimal performance, the parameters of SVM classifier were determined by 5-fold cross-validation.

C. Posterior probability calculation

Usually, the SVM classifier predicts only class label and returns a value of the discriminant function that represents the distance (margin) from the sample to the hyper-plane in the support vector space. However, the posterior probability of the sample is even more informative for classification. Platt et al. proposed a method to map the SVM outputs into posterior probabilities [10]. By using that probability mapping algorithm, after n stimulus trials, the probabilities $P\left(y_{k,n}=1 \mid \mathbf{x}_{k,n}\right)$ of feature vector $\mathbf{x}_{k,n}$ can be obtained. Then, the probability of voice k being the target is given by

$$P\left(k_n\right)=P\left(y_{k,n}=1 \mid \mathbf{x}_{k,n}\right)\cdot\prod_{i,i\neq k} P\left(y_{i,n}=-1 \mid \mathbf{x}_{i,n}\right) \quad (1)$$

It is very important to note that the subject focuses on only one target voice in each block, which adds a special condition to the above probability. Suppose that the target is detected after n trials (condition Φ_n), the probability of voice k being the target after n trials is the conditional probability

$$P^*\left(k_n\right)=P\left(k_n \mid \Phi_n\right)=\frac{P\left(k_n,\Phi_n\right)}{P\left(\Phi_n\right)}=\frac{P\left(k_n\right)}{P\left(\Phi_n\right)} \quad (2)$$

Note that condition Φ_n contains all possible choices of voice k, so we have

$$P\left(\Phi_n\right)=\sum_k P\left(k_n\right) \quad (3)$$

After n trials of stimuli, the target voice was chosen as r_n according to

$$r_n=\arg\max_k P^*\left(k_n\right) \quad (4)$$

D. Adaptive determination of the trial number

In the case of fixed number averaging, the EEG response to target voice could have already reached a discriminable level before all trials were presented. To speed up the decision, a reasonable criterion should be set to terminate the stimulus repeating. In an online system, the probability $P^*\left(k_n\right)$ shown in equation (2) reflects the current performance of the subject after n trials averaging and can be used as the termination criterion. Here, we considered the output target voice r_n after n trials as the final output of the block when $P^*\left(r_n\right)$ is higher than a predefined threshold. However, due to the variation of different subjects' performance, the threshold must be related to individual's performance. We used the cross-validation accuracy (denoted as Ac) from SVM training as the reference of each subject's performance, which is usually higher than that can be achieved in the online testing phase. A threshold θ was then set as a proportion of this cross-validation accuracy:

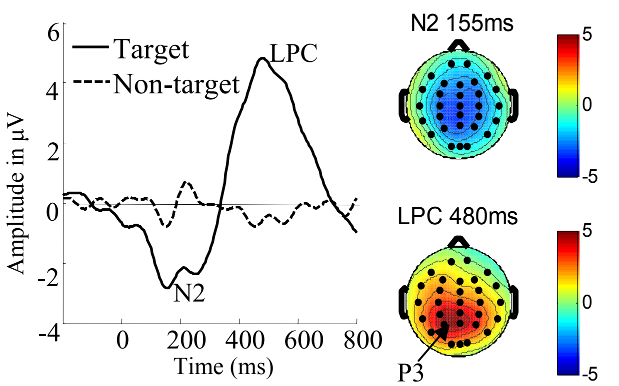
$$\theta=\lambda\cdot \mathrm{A}c \quad (5)$$

The software for the online system was developed on Visual C++ platform while the signal processing algorithm was implemented using MATLAB engine. LibSVM toolbox was employed for SVM model training [11]. Offline analysis was conducted in MATLAB with EEGLAB toolbox [7].

III. Results

A. Auditory response: ERP spatio-temporal pattern

Figure 1 shows all subjects' ground averaged event related potentials (ERP) and its amplitude topographic maps. The ERP elicited by target stimulus reveals a major negative deflection N2 in the time range of 100-300ms, with a central maximum topography, which may represent the auditory processing negativity modulated by endogenous attention [15]. A late sustained positivity across 400-700ms, could also be identified in the target ERP, which resembles the late positive component (LPC), with a parietal distribution. The LPC component reflects the subjects' voluntary response to stimulus property ('left' or 'right' laterality), which is consistent with previous findings [8]. Considering the spatial distribution of both N2 and LPC, the parietal electrode P3 was selected as the only electrode for further classification.



(a) (b)

Fig. 1. The temporo-spatial pattern of ERPs in the active auditory BCI: (a) Grand average waveform at electrode P3. (b) Average amplitude topograph at 155 ms (N2 vale) and 485 ms (LPC peak).

B. Active component: LPC for better classification

Figure 2 depicts the detection accuracies averaged across all subjects using the traditional 'area' method [2], which is calculated as a function of the number of trials averaged for each target voice. If the N2 or LPC feature was used separately, the detection accuracy was lower than in the combination case. The graph also shows that the LPC component contributes more than the N2 component. In this paradigm, the subject's voluntary response of discriminating the auditory stimulus laterality involves the mental processes of working memory and decision making, thus the LPC here is an active component. This paradigm is different from the traditional passive auditory P300, in which the subject only needs to silently count the number of target appearance.

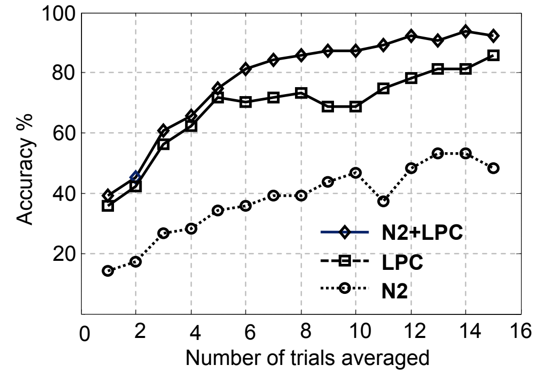


Fig.2 Average accuracy of all subjects as a function of the number of trials considered, using N2+LPC area (diamond curve), LPC area (square curve) and N2 area (circle curve), respectively.

C. Adaptive approach: track the dynamics

Figure 3 displays a typical block during which the subject's performance varied across trials. The upper plot shows the values of SVM discriminant function (DF) using fixed number of averaged trials (15 trials for this case). Evidently, the largest difference between target and non-targets does not appear at the 15th trial. This is usually caused by the fatigue and/or mental state changes of the subject during continuous auditory stimulation in a block. Our proposed adaptive approach tracks the changing state of subject using online estimation of the posterior probability $P^*(k_n)$, and outputs the result of the current block immediately after $P^*(k_n)$ is higher than the preset threshold. The black asterisks on the target curve indicate suitable number of averaging trials when $P^*(k_n)$ reached the preset threshold. In the adaptive case, 4-trial averaging achieved a better result than that of 15-trial. As shown in the upper plot, this solution can not be found by traditional method which employs the value of SVM discriminant function directly.

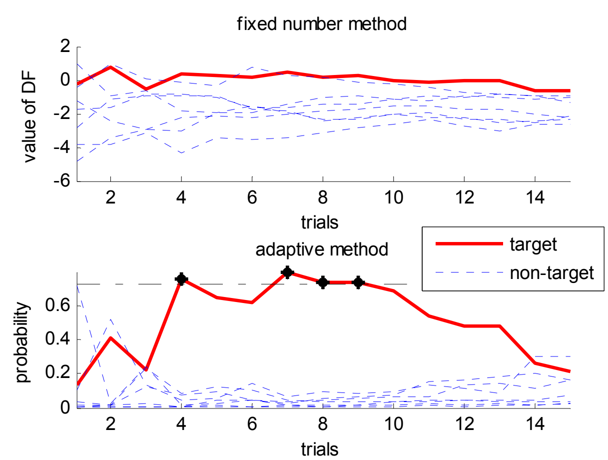


Fig.3 Dynamics of a subject's brain state in active auditory BCI shown as discriminant function value and posterior probability

D. Performance comparison

In the case of P300-based BCI, a traditional approach is to fix the number of averaging trials and combine trials by adding the discriminant function values of EEG epochs from consecutive trials. The stimulus with the highest total DF value is selected as the target. Since the possible number of trials in our case ranged from 3 to 15, we selected the fixed numbers 3, 10 and 15 for comparison. The threshold in our adaptive approach was set at 90%, 80% and 70% of the cross-validation accuracy.

The average results of eight subjects are shown in Figure 4. In one block, the mean number of averaged trials of our adaptive method was 6.0, 5.5 and 4.8 for λ=90%, 80% and 70% respectively (dark gray bars d-f in Fig4a). Even with so few trials, the online accuracy was better than that of '10 trials' and '15 trials' (light gray bars b and c in Fig4b). This leads to a significant improvement of information transfer rate (ITR) as shown in Fig4c. Although the ITR of '3 trials' (light gray bar a in Fig4c) is close to our adaptive approach, it cannot be used in a practical system because of its extremely low online accuracy.

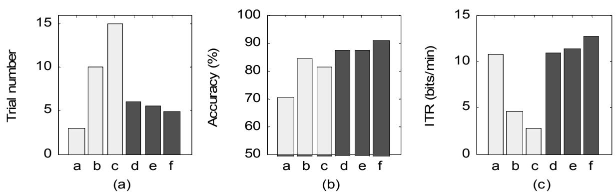


Fig. 4 Comparison of the average results of (a) trail number used per block (b) online accuracy (c) online information transfer rate. The fixed-number method is represented by the light gray bar, in which a, b, and c denotes 3, 10 and 15 trials respectively; the adaptive method is represented by the dark gray bar, in which d, e, and f denotes using 90%, 80%, and 70% of cross-validation accuracy as the threshold respectively.

IV. Discussion and Conclusion

Although our active auditory BCI paradigm shares some of the features of the classical auditory P300 paradigm, here the subject's active response to target voice elicits a prominent late positive component (LPC), which has larger amplitude and longer latency than typical P300 response [6]. Previous studies have found that the active response involving memory usually generates a LPC, which may be a subcomponent of P300 [8]. Notably, this active component is not stimulus-locked, but response-locked [9], which implies that the subject can be trained to achieve better performance in an active auditory BCI paradigm than in the classical P300 paradigm.

Recently, adaptive strategy has been adopted in some the P300-based BCIs. Lenhardt et al. estimated the score of each choice in P300 speller array being the target using training data, and then a preset empirical threshold was used to adaptively determine how many trials needed to be averaged to reach a satisfactory accuracy [12]. Zhang et al. modeled the EEG signals of three possible states (target P300, non-target P300 and non-control) by using Gaussian distribution in the margin space of support vector, and derived the likelihood of each state. The state with highest posterior probability and reached a preset threshold was selected as the most possible state. Then the target letter to be communicated was decided by traditional SVM classification [13]. Both of these two adaptive methods used an empirical threshold for adaptive determination of number of averaging trials, which has no prediction of final target detection accuracy. In our adaptive approach, the threshold was explicitly set as the probability of target detection and was estimated from training data, which is a reasonable prediction of final detection accuracy.

In conclusion, an adaptive approach of deciding the optimal number of averaging trials was showed to improve the accuracy and information transfer rate of an active auditory BCI, and its feasibility and advantage was demonstrated.

References

[1] Y. J. Wang, X. R. Gao, B. Hong, C. Jia, and S. K. Gao, "Brain Computer Interface based on visual evoked potentials," IEEE Eng Med Biol Mag, vol. 27, pp. 64-71, 2008.

[2] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, and D. J. McFarland, "A comparison of classification techniques for the P300 speller," Journal of Neuroal Engineering, vol. 3, pp. 299-305, 2006.

[3] M. Pham, T. Hinterberger, N. Neumann, and A. Kubler, "An auditory brain-computer Interface based on the self-regulation of slow cortical potentials," Neurorehabil Neural Repair, vol. 19, pp. 206-218, 2005

[4] F. Nijboer, A. Furdea, I. Gunst, and J. Mellinger, "An auditory brain-computer interface," Journal of Neuroscience Methods, vol. 167, pp. 43-50, 2008.

[5] E. W.Sellers and E. Donchin, "A P300-based brain-computer-interface: Initial tests by ALS patients," Clinical Neurophysiology, vol. 117, pp. 538-548, 2006

[6] J Guo,B Hong, F Guo,X Gao, S Gao. "An Auditory BCI Using Voluntary Mental Response," presented at the 4th International IEEE EMBS Conference on Neural Engineering, 2009.

[7] A. Delorme, and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," J. Neurosci. Methods, vol 134, pp. 9-21, 2004.

[8] J. Polich, " Updating P300:An integrative theory of P3a and P3b," Clinical Neurophysiology, vol. 118, pp. 2128-2148, 2007

[9] M. Falkenstein, J. Hohnsbein, and J. Hoormann, "Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential," Electroencephalography and Cinical Neurophysiology, vol. 92, pp. 148-160, 1994.

[10] J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," Advances in Large Margin Classifiers, Cambridge: MIT Press, 1999.

[11] C. Chang and C. Lin, LIBSVM: a Library for Support Vector Machines [Online]. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[12] A. Lenhardt, M. Kaper, and H. Ritter, "An Adaptive P300-Based Online Brain-Computer Interface," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 16, no. 2, pp. 121-130, 2008

[13] H. Zhang, C. Guan, and C. Wang, "Asynchronous P300-Based Brain--Computer Interfaces: A Computational Approach With Statistical Models," IEEE Trans. Biomed. Eng., vol. 55, no. 6, pp. 1754-1763, 2008..

[14] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI Competition 2003--Data set IIb: support vector machines for the P300 speller paradigm," IEEE Trans. Biomed. Eng., vol. 51, no. 6, pp. 1073-1076, 2004.

[15] K. Alho, Selective attention in auditory processing as reflected by event-related brain potentials. Psychophysiology, 29(3): 247-263,1992