33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Improved BCI Performance with Sequential Hypothesis Testing

Rong Liu, *Member, IEEE*, Geoffrey I. Newman, Sarah H. Ying, and Nitish V. Thakor, *Fellow, IEEE*

Abstract— **One of the primary challenges in noninvasive brain-computer interface (BCI) control is low information transfer rate (ITR). An approach that employs a power-based sequential hypothesis testing (SHT) technique is presented for real-time detection of motor commands. Electroencephalogram (EEG) recordings obtained during a BCI task were first analyzed with a hypothesis testing (HT) method. Using serial analysis we minimized the time to determine a cued motor imagery cursor control decision. Experimental results show that the accuracy of the SHT method was above 80% for all the subjects (n = 3). The average decision time was 3.4 s, as compared with 6.0 s for the HT method. Moreover, the proposed SHT method has three times the information transfer rate (ITR) compared with the HT method.**

I. Introduction

TO control brain-actuated devices, especially neuro-prostheses, both a stable control signal with a minimal error rate and fast decision making are critical issues [1,2]. Brain-computer interface (BCI) based on electroencephalogram (EEG) bears the advantage of low cost and wide availability, specifically, not exposing the subjects to the risk of brain surgery. However, the EEG-based methods are limited by low information transfer rate as well as susceptibility to certain artifacts such as electromyograph (EMG) [3,4].

The term "classification," used in BCI literature, generally implies decision making or response selection. EEG-based BCIs make binary or multiple decisions as they seek to recognize two or more different mental states. To facilitate decoding "motor intent", machine learning techniques are applied to find subject-specific EEG features that maximize the separation between the patterns generated by executing the mental tasks, and to train classifiers that minimize the classification error rates of these specific patterns. These methods typically reduce to training error minimization with some regularization without optimizing decision time explicitly. Examples include independent component analysis (ICA) [5], nonlinear artificial neural networks (ANNs) [6], common spatial pattern analysis (CSP) [7] and support vector machines (SVM) [8].

Hypothesis testing (HT) and sequential hypothesis testing (SHT) are two branches of statistical decision theory that provide a theoretical framework for understanding how decisions are made. HT converts a single observation into a categorical choice. SHT is a natural extension to HT that accommodates multiple pieces of evidence observed over time and minimizes the expected sample size for given error probability. Thus SHT is a valuable tool for psychophysical analysis, particularly for studying the trade-off between speed and accuracy [9]. Although SHT has been widely used to conduct binary hypothesis testing [10,11], to the best of our knowledge it has not been explored in the context of BCI control. Therefore, in this study, we investigate the effectiveness of using an SHT method for binary classification to improve the speed of BCI and evaluate the trade-off of speed and accuracy.

II. Method

A. EEG Recordings

Three subjects without BCI experience participated in this study. They gave informed consent for the study, which had been reviewed and approved by the Johns Hopkins University Institutional Review Board. EEG signals were acquired with a QuickCap 64-electrode scalp cap connected to a Neuroscan SynAmps2 amplifier channel (Neuromedical Supplies, Inc.) in the modified 10/20 International System and sampled at 250Hz. The output signals were spatially filtered using common average referencing. The weighted sum power spectrum of C3 and C4 (8-13Hz) in the mu rhythm was computed as EEG feature. The subjects performed an EEG-based cursor control experiment by analyzing the resulting EEG feature with a two-threshold HT statistical classifier.

B. Experimental Paradigm

During each training trial, a three-state (move up, move down, remain still) EEG-based BCI was used to control the vertical position of a cursor. In this program, a target was displayed either at the top (relaxation trial) or the bottom (movement trial) of the computer screen, with the imagery-controlled cursor in the middle. The subject's task was to perform mental activity of one type to drive the cursor

This work was supported in part by the U.S. National Science Foundation Cyber Enabled Discovery and Innovation (CDI) program grant to Dr. Thakor, numbers ECCS 0835632 and ECCS 0835554 and grants from the Natural Science Foundation of China (NSFC) (Grant No: 61005088), by the State Key Laboratory of Robotics and System (HIT) (Grant No: SKLRS-2010-ZD-07) and the Fundamental Research Funds for the Central Universities Grant DUT10JS03 to Dr. Liu.

Rong Liu is with the Biomedical Engineering Department, Dalian University of Technology, Dalian, Liaoning 116024 P.R.C. and the Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA (e-mail: rliu@dlut.edu.cn).

G. Newman, is with the Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA (e-mail: geoffrey@jhu.edu).

S. Ying is with the Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA (e-mail: sying@dizzy.med.jhu.edu).

N. V. Thakor is with the Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA (phone: 410-955-7093; fax: 410-955-1498; e-mail: nitish@jhu.edu).

to hit the target. A movement decision was made based on the combined power in the mu band (8-13 Hz) of the C3 and C4 electrodes with the two-threshold HT method. To determine the power, we used Burg's periodogram algorithm on window size of 1 second with 50% overlap. If the cursor touched the target (10 steps away) within 15 seconds, the trial was considered a success. The subject was then presented with a green circle for a success, or with a red circle for a failure. Complete EEG and cursor movement data were stored for later offline SHT analyses.

C. Two-threshold HT Method

We began with a labeled training set $\mathbf{S} = \{(x_1, y_1), ..., (x_L, y_L)\}$, where each training pair $(x_l, y_l) \in \chi \times \{0,1\}$. Here χ is an infinite set ($\chi = \mathbb{R}$). Each y_l can only take one of two values, represents the class label of that sample. Each training pair (x_l, y_l) is drawn independently from some unknown joint distribution P_{XY}. $p_0(x) \triangleq P_{X|Y}(x \mid y = 0)$ and $p_1(x) \triangleq P_{X|Y}(x \mid y = 1)$ denote the class conditional distribution. We take the prior probabilities for the label are uniform, i.e., $P_Y(y = 0) = P_Y(y = 1) = 1/2$.

Given S, we wish to train a model so as to classify, i.e., to assign a label of 0 or 1 to a new sample x. This sample is drawn according to the unknown distribution P_X, but its label is unavailable. If we do have access to the true conditional distributions p_0 and p_1, the decision variable f is given by

$$f = \frac{p_1}{p_0}. \tag{1}$$

A straightforward comparison with a single threshold typically results in high error rates because the two distributions overlap considerably. An alternate approach is to select two conservative thresholds η_1 and η_2. No decision is made if the sample lies in the overlapping intermediate zone. The decision rule is

$$y = \begin{cases} 1 \sim H_1, & f \geq \eta_1 \\ 0 \sim H_0, & f \leq \eta_2 \\ \textit{continue} \sim \textit{no decision} & \eta_1 < f < \eta_2 \end{cases}, \tag{2}$$

which is called two-threshold HT method.

D. SHT Method

With two thresholds, increasing η_1 or decreasing η_2 may decrease the speed or likelihood of making a decision, but increase the chance that any decision made is correct. There is an inherent trade-off between the accuracy and decision time. One powerful approach to solving this conundrum is to use SHT. A sequential hypothesis test is formulated as

$$f_n = \frac{p_1(\mathbf{X}_n \mid H_1)}{p_0(\mathbf{X}_n \mid H_0)}, \tag{3}$$

where $p_1(\mathbf{X}_n \mid H_1)$ is the likelihood function for observing the sample sequence $\mathbf{X}_n$ if the hypothesis H_1 is true, and $p_0(\mathbf{X}_n \mid H_0)$ is the likelihood function of observing $\mathbf{X}_n$ given that the hypothesis H_0 is true. To decide which hypothesis to be accepted is determined by comparing the ratio f_n with some upper and lower bound values.

Assuming the samples follow a Gaussian distribution, Eq. (3) can be rewritten as

$$\begin{aligned} 2F_n &= 2\ln\frac{p_1(\mathbf{X}_n \mid H_1)}{p_0(\mathbf{X}_n \mid H_0)} \\ &= \sum_{i=1}^{n}\left(\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right) - \frac{(x_i - \mu_1)^2}{\sigma_1^2} + \frac{(x_i - \mu_0)^2}{\sigma_0^2}\right), \end{aligned} \tag{4}$$

where μ_0 and σ_0 are the mean and standard deviation of the distribution $\hat{p}_0$ and μ_1 and σ_1 are the mean and standard deviation of the distribution $\hat{p}_1$.

Let $T_i = \ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right) - \frac{(x_i - \mu_1)^2}{\sigma_1^2} + \frac{(x_i - \mu_0)^2}{\sigma_0^2}$, we obtain the decision rule at the nth stage as

$$\hat{y} = \begin{cases} 1 \sim H_1, & \sum_{i=1}^{n} T_i \geq 2\ln(\eta_1) \\ 0 \sim H_0, & \sum_{i=1}^{n} T_i \leq 2\ln(\eta_2) \\ \textit{continue} \sim \textit{no decision} & 2\ln(\eta_1) < \sum_{i=1}^{n} T_i < 2\ln(\eta_2) \end{cases}. \tag{5}$$

The number of tests and therefore the number of data segments x_i required to terminate the test depend on the detection thresholds. For example, if the test has very conservative thresholds, a large number of observations may be required. Increasing the upper threshold may increase the number of samples required to terminate the test when H_1 is true, and decreasing the lower threshold may increase the number of samples required when H_0 is true. In this work, we choose η_1 and η_2 on the basis of the operating characteristic (OC) curves.

OC curves depict the inherent trade-off between decision time (costs) and accuracy (benefits) of a detection test. Each prediction result with one pair of thresholds represents one point in the curve. The best possible thresholds would yield a point nearest to the upper left corner or coordinate (0,1) of the space, representing 100% accuracy and no decision time.

III. Results

The following analysis illustrates the process and potential of the proposed SHT. All the subjects yield similar trends in the neurophysiological properties and feature distribution. Therefore, we show the results of a single subject for sake of clarity.

A. Neurophysiological Outcome

Fig. 1 depicts the neurophysiological properties of subject 1. It shows the EEG spectrum in 0~30Hz and the associated averaged logarithmic transform of the power in decibel topographies for each selected frequency component for each relaxation (left) versus motor imagery (right). The

10*logP scalp distributions show that there are obvious differences between the mental tasks in 8-13Hz, localized in those electrodes with higher power values in relaxation condition. This comparison reveals aspects of the EEG response common to rest and imagery, i.e. the mu rhythm decreases or desynchronizes with movement imagery.

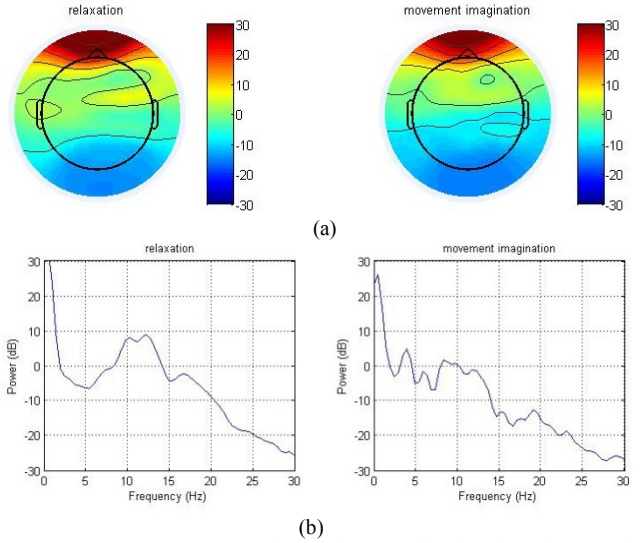

Fig. 1. (a) Average 10 Hz power during the two tasks, showing a decrease in power over motor cortex during movement imagination. (b) Average power in the classification feature, showing the decrease in power during movement imagination is specific to the mu band (8-13 Hz), and beta band (15-20Hz).

B. Normal Distribution of EEG Features

The Kolmogorov-Smirnov (KS) test results showed that the distribution of average power spectrum density (PSD) in mu-rhythm (C3 + C4) is indeed normal. We constructed the probability distributions in Fig. 2. The *x* axis in Fig. 2 displays the mean of mu rhythm power. The *y* axis displays the probability density of mu rhythm power. The average PSD of relaxation fit a Gaussian distribution with a mean, $\mu = 57.27$ dB/Hz and a standard deviation, $\sigma = 16.85$. The average PSD of motor imagery fit a Gaussian distribution with $\mu = 44.74$ dB/Hz and $\sigma = 14.98$. From this figure, we can see there is significant overlap for these two stochastic models, which results in a relatively poor discriminability for the HT method which will be shown in the next subsection.

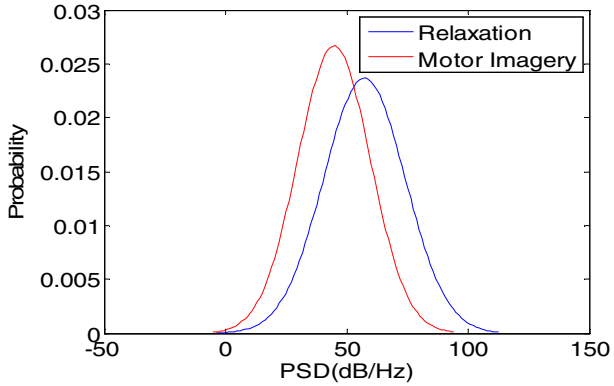

Fig. 2. Histograms show the average PSDs of mu-rhythm at Channels C3 and C4 at a fixed time window (50samples) are overlaid by two fitted normal distributions for two conditions (relaxation vs. motor imagery).

C. Slow Online Single-trial Recognition with HT

Each subject underwent different runs of 16 trials each (8 relaxation and 8 movement imagery trials), with a random order of presentation. Table I shows the performance of the online single-trial recognition with HT method, including percent of classification accuracy, mean decision time and average information transfer rate (ITR) for each run. The ITR of each trial in bits per minutes was calculated as [5]

$$ITR[bit/\min] = \left[\log_2 N + P\log_2 P + (1-P)\log_2 \frac{1-P}{N-1}\right]\frac{60}{L}$$

where N=2 is the number of classes, P (hits/number of trials) is the accuracy of classification, and L is the trial length in seconds. Each subject performed differently. Subject 1 has the highest accuracy, ITR and the shortest decision time. Although the average accuracy is more than 70%, the decision time is quite long (~ 11 s). Thus the subjects' average ITR is small. This is unacceptable for a practical BCI system.

TABLE I
PERFORMANCE (PERCENT OF CLASSIFICATION ACCURACY, MEAN TIME OF DECISION, ITR) OF HT

Subject	Accuracy	Mean Decision Time (s)	ITR(bit/min)
1	80.36%	10.63	2.86
2	54.46%	12.71	0.17
3	75.00%	10.74	2.42
Average	**69.94%**	**11.36**	**1.82**

D. Fast Recognition with SHT

The SHT procedure is illustrated in Fig. 3. The first segment of EEG is used to obtain a measurement, x_1, which in turn is used to calculate $\mathbf{T}_1$. However, it can be seen that $\mathbf{T}_1$ lies between the two decision thresholds, so no decision is made. The next segment of EEG is used to obtain a second measurement, x_2. x_1 and x_2 are used to calculate $\mathbf{T}_2$, but $\mathbf{T}_2$ also lies between the decision thresholds, so no decision is made. The next segment of EEG is used to obtain a third measurement, the same progress continues until the ninth step $x_1, x_2 \ldots x_9$, was used to calculate $\mathbf{T}_9$, which lies below the lower decision threshold. Therefore we accept H_1 and the decision is relaxation.

To assess the performance of the methods, we used the first runs as the training set and the other runs as the testing set. The results of accuracy and time for classification are summarized in Table II. Since SHT uses non-overlapping windows, we ran the HT method on non-overlapping windows of data (called HT(I)). Comparing the result of HT(I) with HT, we can see there is not only a difference between the accuracy, but the decision time has been shortened. Comparing the results for HT(I) and SHT, we can see the SHT method not only improves the classification accuracy but also yields faster decisions. Specifically, the average decision time improved to 3.40s.

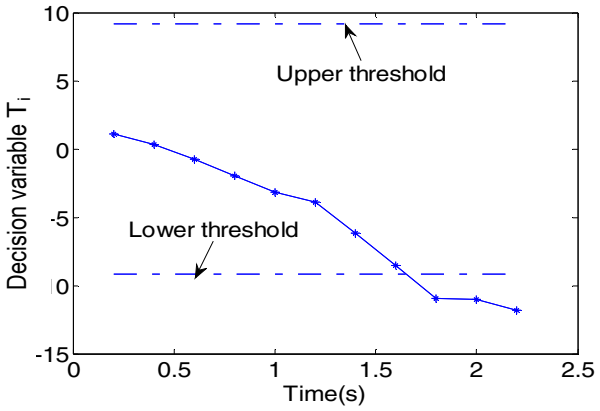

Fig. 3. Decision variable $\mathbf{T}_i$ versus time for sequential hypothesis testing.

TABLE II
ACCURACY AND TIME OF CLASSIFICATION OF HT AND SHT METHOD

Subject	HT(I)			SHT		
	Accuracy (%)	Time (s)	ITR(bit/min)	Accuracy (%)	Time (s)	ITR(bit /min)
1	78.13	5.48	2.70	83.33	2.62	8.20
2	75.00	6.23	2.17	82.29	4.43	4.52
3	68.75	6.29	1.79	81.94	3.13	7.95
Mean	**73.96**	**6.00**	**2.22**	**82.52**	**3.39**	**6.89**

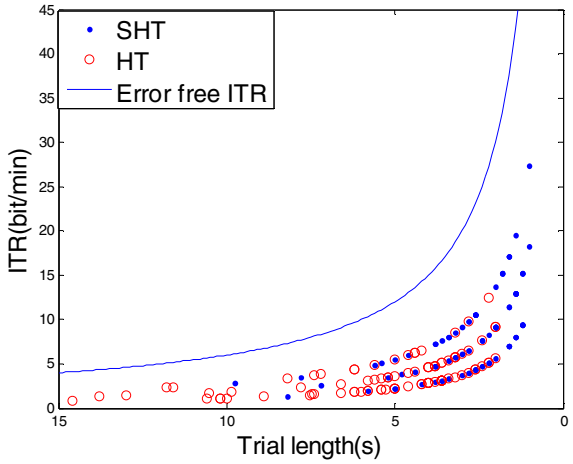

Fig. 4. Diagrams displaying the ITR in bit/min for various trial lengths for subject 1, incorporating the subject HL's success rate. The solid line represents the maximum possible ITR for an error-free classification result. The highest values were obtained for trial lengths of 1 s.

We plot the ITRs of each trial for the two methods for subject 1 (Fig. 4). The solid line in the diagrams displays the maximum ITR for an error-free classification as a function trial length. This graph represents the trade-off between having shorter trials, which would result in lower success rates, with overly long trials when have higher success rates. It can be seen that the SHT method showed improved ITR and provided the subject better control over the BCI. The results of this study show that trial lengths of 1.2 s could result in an ITR of 27.39 bit/min. As we know, there is a physiological limit when the mu rhythm is used for control. According to ref. [12], the subject requires several hundred milliseconds (roughly 500ms) to initiate motor imagery. Desynchronization and synchronization of mu oscillations need time on the order of seconds to develop because relatively large networks contribute to the generation of these rhythms. This suggests that a trial time of 1.2s may approach the limit of these physiological delays for reliable decision making.

IV. CONCLUSIONS

In this paper, a statistical algorithm for real-time detection of motor commands in a BCI system is proposed. This algorithm is based on a SHT approach in conjunction with a mu- rhythm spectral feature for motor imagery and relaxation. These features are distributed in a Gaussian shape. Post-hoc analysis showed that with the SHT method, the classification accuracy can be improved, and the decision time can be greatly decreased. The novel SHT method has advantages over conventional classification algorithms in terms of reliability, simplicity and ITR.

The algorithm is simple to implement in real time system due to its sequential nature and thus should be useful in practical BCI applications that require real-time data-acquisition, decision making within the constraints of speed and accuracy, such as for guiding wheelchair or controlling prosthetic limbs.

REFERENCES

[1] M. A. Lebedev and M. A.L. Nicolelis, "Brain-machine interfaces: past, present and future," *Trends in Neurosciences*, vol. 29, no. 9, pp. 536-546, 2006.

[2] J. R. Millan, F. Galan, D. Vanhooydonck, E. Lew, J. Philips and M. Nuttin, "Asynchronous non-invasive brain-actuated control of an intelligent wheelchair," in *proc. 31st annual international conference of the IEEE EMBS,* Minneapolis, USA, Sept. 2009, pp. 3361-3364,.

[3] P. Shenoy, M. Krauledat, B. Blankertz, R. P.N. Rao and K.-R. Muller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, pp. R13-R23, 2006.

[4] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Muller and G. Curio, "The non-invasive berlin brain-computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no.2, pp. 539–50, Mar. 2007.

[5] A. Kachenoura, L. Albera, L. Senhadji and P. Comon, "ICA: a potential tool for BCI systems," IEEE *Signal Processing Magazine*, vol. 25, no. 1, pp. 57-68, 2008.

[6] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for eeg classification," *IEEE Transactions on rehabilitation engineering*, vol. 8, pp. 457-463, 2000.

[7] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege ect., "The Berlin Brain--Computer Interface: accurate performance from first-session in BCI-naïve subjects," *IEEE Trans Biomed Eng.*, vol. 55, no. 10, pp. 2452-2462, 2008.

[8] C. J. Bell, P. Shenoy, R. Chalodhorn and R. P.N. Rao, "Control of a humanoid robot by a noninvasive brain-computer interface in humans," *J. Neural Eng.*, vol. 5, pp. 214-220, 2008.

[9] Joshua I. Gold and Michael N. Shadlen, "The neural basis of decision making," *Annu. Rev. Neurosci.* , vol. 30, pp. 535-574, 2007.

[10] N.V. Thakor, Y.-S. Zhu and K,-y. Pan, "Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm," *IEEE Trans. Biomedical Engineering*, vol.37, no. 9, pp. 837-843, 1990

[11] R. Chandramouli and N. Ranganathan," A Generalized Sequential Sign Detector for Binary Hypothesis Testing," *IEEE Signal Processing Letters*, vol. 5, no. 11, Nov. 1998.

[12] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, 1999.