

Real-time Prediction of Fast and Slow Delivery of Mental Commands in a Motor Imagery BCI: An Entropy-based Approach

Sareh Saeedi, Ricardo Chavarriaga, José del R. Millán
Chair on Non-Invasive Brain-Machine Interface (CNBI),
Center for Neuroprosthetics, School of Engineering, EPFL
Lausanne, Switzerland
sareh.saeedi@epfl.ch

Michael C. Gastpar
Laboratory for Information in Networked Systems,
School of Computer and Communication Sciences, EPFL
Lausanne, Switzerland

Abstract—Providing adaptive shared control for Brain-Computer Interfaces (BCIs) can result in better performance while reducing the user’s mental workload. In this respect, online estimation of accuracy and speed of command delivery are important factors. This study aims at real-time differentiation between fast and slow trials in a motor imagery BCI. In our experiments, we refer to trials shorter than the median of trial lengths as “fast” trials and to those longer than the median as “slow” trials. We propose a classifier for real-time distinction between fast and slow trials based on estimates of the entropy rates for the first 2-3 s of the electroencephalogram (EEG). Results suggest that it can be predicted whether a trial is slow or fast well before a cutoff time. This is important for adaptive shared control especially because 55% to 75% of trials (for the five subjects in this study) are longer than that cutoff time.

Index Terms—Brain-Computer Interface (BCI), Shared Control, EEG, Entropy

I. INTRODUCTION

A Brain-Computer Interface (BCI) monitors the user’s brain activity and translates his/her intentions into commands to an external device, such as a wheelchair or a prosthetic device [1], [2]. In this framework, one of the main challenges is to determine the subject’s intention using an uncertain channel, such as BCI. In this respect, the use of shared control techniques results in better performance [3], [4]. In shared control systems, control commands to the device are obtained through the combination of the user’s decoded intentions with information from the device sensors [5], [6]. Some shared control BCI systems rely on discrete adaptation of the BCI commands, i.e. the command is delivered as soon as the subject completes a specific task. Examples of such systems are driving a telepresence robot or a wheelchair using brain signals in an asynchronous BCI [7], [3].

Shared control techniques usually have predefined settings based on the task and the environment in which the task is performed. In addition, the level of assistance they provide is usually constant over time. However, in order to have an effective control, this level should be adaptive so as to complement the user’s capabilities which change over time. In this way, the user always remains in control of the brain actuated device [8].

One of the main challenges of using EEG-based BCIs is the non-stationary nature of the signals [9]. During online operations, this issue can lead to changes in the accuracy and the speed in delivery of mental commands. Therefore, in order to provide adaptive shared control for such systems, online estimation of the accuracy in delivering a command as well as the time it takes are important factors.

The main goal of this work is to study the possibility of detecting slow trials (trials for which more time is needed to deliver a command) well before reaching a cutoff time. This will allow us to define the level of assistance provided for the user accordingly. For example, when the user is not able to deliver a command quickly while driving a mobile robot, the robot can slow down so that it gives more time for the user’s intention to be executed. To tackle this issue, we have used an information theoretical approach to characterize the signal. We will begin with describing our BCI system. Then, we will continue by methods used for classification of fast and slow trials. This will be followed by the results and discussion.

II. METHODS

A. BCI System

In our BCI system, the users voluntarily modulate EEG oscillatory rhythms by executing two motor imagery tasks (i.e. imagination of movements, such as right hand vs. left hand or feet vs. one of the hands). EEG was recorded with a portable 16-channel g.tec system over the sensorimotor cortex, at 512 Hz and band-pass filtered between 0.1 Hz and 100 Hz. Each channel was then spatially filtered with a Laplacian derivation before estimating its power spectral density (PSD) in the band 4-48 Hz with 2 Hz resolution over the last 1 second. The PSD was computed every 62.5 ms (i.e., 16 times per second) using the Welch method with 5 overlapped (25%) Hanning windows of 500 ms.

From the computed features (16 channels x 23 frequencies), a subset was selected based on the algorithm described in [3] to estimate the relevance of the features for discriminating the mental commands. In this algorithm, Canonical Variate Analysis (CVA) was used to extract the projection of PSD

samples onto the canonical space. The features were then ranked based on their contribution on the new space. This algorithm was run on EEG data recorded during three sessions separately and the features with high discriminant values were selected.

Online real-time classification of the selected feature vector is achieved using a Gaussian Mixture Model (GMM) framework to estimate the conditional probability distribution over the mental commands given an EEG sample [7]. In this framework, each of the two classes is represented by four Gaussian units. Then, the classifier output undergoes a rejection mechanism, where only the probability distributions of the winning class which exceed a rejection threshold (equal to 0.6) are fed to output integration level. The BCI then integrates the output of the classifier over time until it accumulated enough evidence about the user’s mental intent:

$$p(y_t) = \alpha \times p(y_t | x_t) + (1 - \alpha) \times p(y_{t-1}), \quad (1)$$

where $p(y_t | x_t)$ is the probability distribution, $p(y_{t-1})$ the previous distribution, and α the integration parameter, which was set to 0.96 in our experiments. The evidence accumulation tackles the uncertainty of single-sample classifier inference. It provides a final inference that can drastically increase the robustness of the BCI system. Moreover, it allows for providing a smooth, fast, and informative feedback to the user. In addition, it makes it possible to reconfigure the parameters according to the user’s performance.

The probabilities are integrated based on equation (1) until a class reaches a certainty threshold, which is equal to 0.7 in our experiments. The BCI command is delivered at this moment. Therefore, the time required to deliver a command changes across trials.

B. Experiment, Signal Processing and Classification

First, the five subjects underwent a training phase, in which they were asked to imagine the movement of their right hand, left hand, and feet following the relevant cue in trials with the same length. The training phase was done in a session comprising three runs. The runs consisted of 15 trials of each mental task which were randomly organized. The recorded EEG signal was assessed based on the feature selection and classification methods discussed in the previous section. Then, the two most separable mental tasks were chosen based on the classification results.

In the following sessions, the subjects were recorded in a two-class motor imagery task (e.g. hand/feet), in which they were asked to do the relevant mental task following a cue on the screen while receiving a visual feedback from the classifier outputs. In this way, they could better modulate their brain activity patterns through the feedback on their performance. The experiment was done in different sessions, comprising three runs. The runs consisted of 15 randomly organized trials of each mental task. As mentioned in the previous section, the command delivery time is not the same for all trials. In our experiments, we refer to trials shorter than the median of trial lengths as “fast” trials and to those longer than the

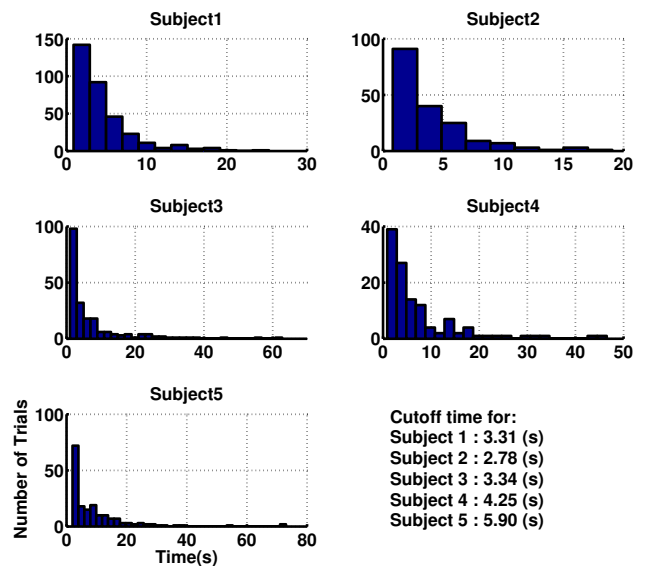


Fig. 1. Distribution of the command delivery time for each subject. Cutoff time is the median of command delivery time, based on which the trials are divided into fast and slow ones.

median as “slow” trials. The command delivery time does not seem to be correlated with the accuracy, e.g., some fast trials might be inaccurate and some slow trials might be accurate. Figure 1 shows the distribution of the command delivery time for each subject. This figure highlights the need for detecting slow trials, as some of the trials last very long.

As it was mentioned, the main goal was to predict whether a trial is slow or fast. In order to do so, Shannon entropy as a measure of information content of a signal was calculated for fast and slow trials. In order to take into account the temporal structure, we estimated the entropy rate [10] of the signal by considering the conditional entropy of sample x_i , given the previous m samples, $H(x_i | x_{i-1}, x_{i-2}, \dots, x_{i-m})$. For this entropy estimation, an approximation of $m + 1$ -dimensional joint probability, $P(x_i, x_{i-1}, \dots, x_{i-m})$ would be required. For very small m (like in the extreme case of $m = 0$), this can be done. However, for reasonable m , a reliable estimate of this conditional entropy is computationally expensive and requires a large amount of data. These are important issues in our case where we would like to predict if a trial is slow or fast based on samples in a window of few seconds. To alleviate this problem, we employ a linear model and instead estimate the following approximate version of the desired conditional entropy:

$$H(x_i | x_{i-1}, \dots, x_{i-m}) \approx H(x_i | a_1 x_{i-1} + \dots + a_m x_{i-m}) \quad (2)$$

In order to estimate the conditional entropy, an autoregressive model (order = 3) of the original signal ($X = [x_1, x_2, \dots, x_n]$) was built as:

$$x_i = a_1 x_{i-1} + a_2 x_{i-2} + a_3 x_{i-3} + z_{i-3} \quad (3)$$

Based on this estimation, $Z = [z_1, z_2, \dots, z_{n-3}]$ which carries the temporal information of the original signal, X , was formed.

Then, Shannon Entropy of Z was calculated for each trial using Maximum Likelihood estimation [11] as:

$$\hat{H}^{ML} = - \sum_{k=1}^p \hat{\theta}_k^{ML} \log(\hat{\theta}_k^{ML}), \quad (4)$$

With the ML frequency estimates:

$$\hat{\theta}_k^{ML} = \frac{y_k}{n}, \quad (5)$$

Where y_k is the cell count (number of samples in one amplitude bin [11]) and $n = \sum_{k=1}^p y_k$ is the total number of counts. First, the signals were normalized to have maximums and minimums equal to 1. Then, the distribution of the signal was approximated by building histograms using 10 bins. Having equal number of bins and bin widths makes it possible to have comparable values of entropy. For measuring the entropy rate, we considered the first T s of data in each trial, where $T = 2 + 0.5 \times i, 0 \leq i \leq 4$. If the length of trial was less than T s, entropy was estimated for the whole trial. The regression model was recomputed for the first T s of each trial separately. Then, the entropy rates of the signal for the 16 channels were computed and constructed the features for classification of fast vs. slow trials. A Linear Discriminant Analysis (LDA) classifier was used for discrimination between the two classes (slow and fast trials) and five-fold cross-validation was used to assess the performance.

III. RESULTS

Figure 1 depicts the distribution of the command delivery time for different subjects. According to this figure, command delivery time changes across different trials and different sessions due to non-stationarity of the signal. The median time for command delivery is different for the five subjects; for subject 1 to 5 it is equal to 3.31, 2.78, 3.34, 4.25, and 5.9 s, respectively. This highlights the fact that any system designed for differentiating between fast and slow trials should be subject specific.

Figure 2 illustrates the grand averages of entropy rate (slow and fast trials) in all the 16 channels for different subjects, while EEG data of the whole trial is considered for estimating Z . According to this figure, a difference between entropy rates of fast and slow trials is observed in all channels. We used t-test to find the significant differences ($p < 0.05$) between fast and slow trials. For the five subjects, the entropy of fast and slow trials across all sessions are significantly different based on p-values.

Classification performances for the first T s of data ($T = 2 + 0.5 \times j, 0 \leq j \leq 4$) are depicted in Figure 3. This figure illustrates performances in receiver operating characteristic (ROC) space for all subjects. The x axis and the y axis denote the false positive rate (FPR) and the true positive rate (TPR), respectively (slow trials are considered as positive). The performance of an ideal classifier should be located at $(0, 1)$ in ROC space. Each plot in the figure represents the results considering the first T s of data. Following our goal for BCI applications, it is important to capture the slow trials

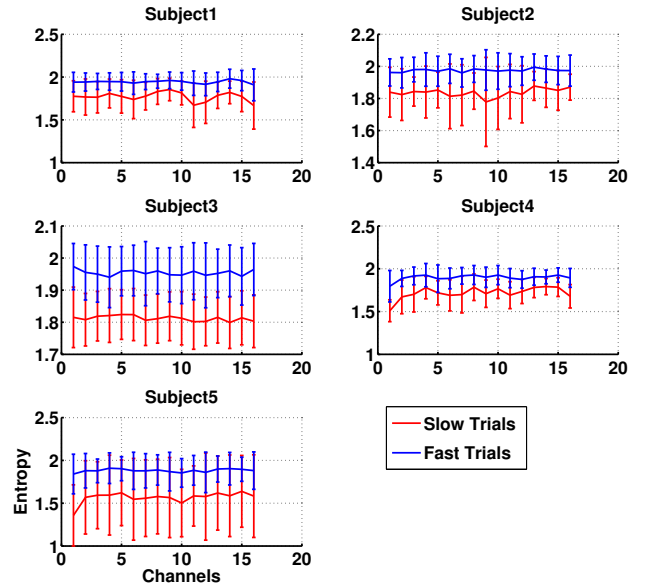


Fig. 2. Average of entropy rate and the standard deviation over trials in all EEG channels. Entropy rate of the fast trials have higher values compared with the slow trials.

precisely in order to adapt the level of assistance. Therefore, misclassification of slow trials as fast ones, reflected in FPR, is an important factor in the classification. Table I shows the area under the ROC curve (AUC) for these plots. The ideal classifier should have the AUC equal to one. The last column in this table shows the AUC when T is equal to median time and it is between 0.79 and 0.9 for all the subjects. Therefore, considering AUC higher than 0.7 is reasonable for our application (highlighted numbers in table I), 3 s of the data for subjects 1 and 5, 2.5 s of data for subjects 3 and 4, and 2 s of data for subject 2 is required to predict whether a trial is slow or fast.

The ROC curve also allows us to determine the optimal threshold for classification, for which the best performance is achieved. This optimal threshold is calculated by maximizing the distance between the ROC curve and the dashed line (random performance) in figure 3. For each subject, the first T s of data, which seems to be sufficient according to the AUCs in table I, was considered. This optimal value and its relevant confusion matrix are shown in table II. These results show that the confusion matrices are quite balanced for all the subjects; that is, the classifier is not biased towards one class.

IV. DISCUSSION

As was mentioned in the introduction, implementing shared control techniques for BCI improves its performance [8]. In order for a BCI system to function efficiently, not only the accuracy but also the speed of delivering a command is an important factor. In this paper, the latter was studied in a motor imagery BCI system, in which the subjects could deliver a command by modulating their brain signals while executing

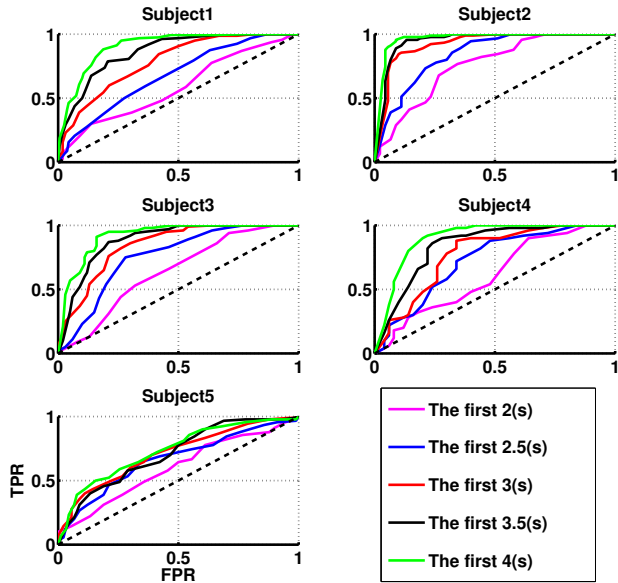


Fig. 3. Performances in ROC space for all the subjects, considering the first T s of signal in each trial.

TABLE I
AREA UNDER THE ROC CURVE WHILE CONSIDERING THE FIRST T s OF TRIALS.

	2 s	2.5 s	3 s	3.5 s	4 s	Median Time
Subject1	0.58	0.67	0.79	0.86	0.91	0.81
Subject2	0.76	0.84	0.92	0.95	0.97	0.85
Subject3	0.64	0.76	0.84	0.88	0.92	0.90
Subject4	0.61	0.72	0.77	0.84	0.9	0.84
Subject5	0.6	0.66	0.71	0.71	0.74	0.79

a motor imagery task. The amount of time it takes for a subject to deliver a command can vary both within and between sessions, most likely due to non-stationarity of EEG or other factors like motivation, fatigue, etc. As depicted in Figure 1, the distribution of command delivery time is different for the five subjects. We referred to the trials as “fast” and “slow” based on being shorter and longer than the median of command delivery time.

In order to model the temporal structure of the signal, an AR model was assumed for the EEG signal in each trial. Entropy rates of the filtered signal over all the trials were compared in Figure 2. As this figure shows, entropy rates of fast trials are higher than those for slow trials. This difference is statistically significant according to p-values of t-test. In general, for two random variables with the same distribution, the one that has more samples will have a higher entropy rate. In this case, on the other hand, the slow trials have lower entropies. This suggests that the difference is not observed only due to different trial lengths.

The goal was to study whether we can find any significant difference between fast and slow trials using only the first few seconds of each trial. LDA classifier was used to differentiate

TABLE II
THE OPTIMAL THRESHOLD FOR CLASSIFICATION AND THE CONFUSION MATRICES.

	Optimal Threshold	Confusion Matrix	
Subject1	0.4	84.4	15.6
		41.9	58.1
Subject2	0.5	76.7	23.3
		34.4	65.6
Subject3	0.5	75.0	25.0
		28.0	72.0
Subject4	0.3	88.0	12.0
		48.0	52.0
Subject5	0.5	60.0	40.0
		30.0	70.0

between fast and slow trials in the first T s of data based on an estimate of entropy rate of EEG signal. As depicted in Figure 3, the longer the data used for calculating entropy, the more accurate the differentiation between the two classes (fast and slow).

The objective was to predict whether a trial was slow or fast as early as possible. According to Table I, considering the first 3 s of data for subject 1 and 5, results in AUC greater than 0.7, which seems to be acceptable. Comparing these cutoff times to the median of command delivery time (depicted in Figure 1) for these two subjects reveals that almost 55% (subject 1) and 73% (subject 5) of trials take more time than that. In other words, we can make a reliable prediction of time efficiency for about 55% (subject 1) and 73% (subject 5) of trials that are longer than the cutoff time. Considering the first 2.5 s of data for subject 3 and subject 4, results in the AUC higher than 0.7 in classification while 58% and 75% of trials are longer. For subject 2, the first 2 s of data attains such an accuracy; almost 63% of trials done by this subjects take more time.

Uncued BCI paradigms, such as driving a wheelchair or a telepresence robot using brain signals, are among the main applications of motor imagery BCI. One of the main issues in such applications would be to adapt the level of assistance provided for the subjects especially when they are not fast enough in delivering a command. In other words, being able to predict the time efficiency of a command within the first few seconds of each trial is important in uncued BCI paradigms, in that the classification algorithm does not have any information about the start and the end of the trials.

V. CONCLUSION

This study aims at real-time differentiation between fast and slow delivery of commands in a motor imagery BCI. The mentioned results reveal that it is possible to predict whether a trial is slow or fast well before a cutoff time (2-3 s), based on measuring the entropy rate of the filtered EEG signal. Estimation of the entropy rate was done by assuming an AR model for the EEG signal. Choosing the optimal order for this model, which might be subject specific, is worth investigating in future. As 55% to 75% of trials take longer than the cutoff time (for the subjects we discussed), this

method seems to be a good predictor of the time efficiency for delivering BCI commands. In fact, the results show that all slow trials can reliably be captured within the first 2-3 s of the trial. This prediction is very important, as it makes it possible to regulate the shared control parameters accordingly. The proposed method can be a factor in the criteria used for adaptation of shared control. This is a point we will strive to address in our future work in a shared control framework to operate a telepresence robot.

ACKNOWLEDGMENT

This work is supported by Swiss National Center of Competence in Research (NCCR) Robotics.

REFERENCES

- [1] J. d. R. Millán, P. W. Ferrez, F. Galán, E. Lew, and R. Chavarriaga, "Non-invasive brain-machine interaction," *Int J Pattern Recognition and Artificial Intelligence*, vol. 22, no. 5, pp. 959–972, 2008.
- [2] J. d. R. Millán, R. Rupp, G. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, C. Neuper, K. Müller, and D. Mattia, "Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges," *Frontiers in Neuroscience*, vol. 4, p. 161, 2010.
- [3] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A brain-actuated wheelchair: Asynchronous and non-invasive brain-computer interfaces for continuous control of robots," *Clin Neurophysiol*, vol. 119, no. 9, pp. 2159–2169, Sep 2008.
- [4] L. Tonin, R. Leeb, M. Tavella, S. Perdakis, and J. d. R. Millán, "The role of shared-control in BCI-based telepresence," in *Proc. of 2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010.
- [5] H. K. Kim, S. J. Biggs, D. W. Schloerb, J. M. Carmenta, M. A. Lebedev, M. A. L. Nicolelis, and M. A. Srinivasan, "Continuous shared control for stabilizing reaching and grasping with brain-machine interfaces," *IEEE Trans Biomed Eng*, vol. 53, no. 6, pp. 1164–1173, Jun 2006. [Online]. Available: <http://dx.doi.org/10.1109/TBME.2006.870235>
- [6] D. Vanhooydonck, E. Demeester, M. Nuttin, and H. V. Brussel, "Shared control for intelligent wheelchairs: an implicit estimation of the user intention," in *Proceedings of the ASER '03 1 st International Workshop on Advances in Service Robotics*, pp. 13–15.
- [7] J. d. R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Trans Biomed Eng*, vol. 51, no. 6, pp. 1026–1033, 2004.
- [8] J. Philips, J. d. R. Millán, G. Vanacker, E. Lew, F. Galán, P. Ferrez, H. V. Brussel, and M. Nuttin, "Adaptive Shared Control of a Brain-Actuated Simulated Wheelchair," in *Proceedings of the 10th IEEE International Conference on Rehabilitation Robotics*, 2007, pp. 408–414.
- [9] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J Neural Eng*, vol. 3, no. 1, pp. R13–R23, Mar 2006.
- [10] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.
- [11] J. Hausser and K. Strimmer, "Entropy inference and the james-stein estimator, with application to nonlinear gene association networks," *Journal of Machine Learning Research*, vol. 10, no. December, p. 18, 2008. [Online]. Available: <http://arxiv.org/abs/0811.3579>