34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Improved Recognition of Error Related Potentials through the use of Brain Connectivity Features

Huaijian Zhang, Ricardo Chavarriaga, Mohit Kumar Goel, Lucian Gheorghe, and José del R. Millán

Abstract— **Brain error processing plays a key role in goal-directed behavior and learning in human brain. Directed transfer function (DTF) on EEG signal brings unique features for discrimination between correct and error cases in brain-computer interface (BCI) system. We describe the first application of brain connectivity features for recognizing error-related signals in non-invasive BCI. EEG signal were recorded from 16 human subjects when they monitored stimuli moving in either correct or erroneous direction. Classification performance using waveform features, brain connectivity features and their combination were compared. The result of combined features yielded highest classification accuracy, 0.85. In addition, we also show that brain connectivity at theta band around $200ms$ after stimuli carry highly discriminant information between error and correct trials. This paper provides evidence that the use of connectivity features improve the performance of an EEG based BCI.**

I. INTRODUCTION

Error related brain activity has been studied in last decade with great interest for its crucial role in goal-directed behavior and learning [1], [2]. Electrophysiological recordings and fMRI studies suggest that error processing involves the dorsal anterior cingulate cortex (ACC) and the medial prefrontal cortex (PFC) [2]. Furthermore, specific brain interaction patterns after presentation of erroneous stimuli have been reported by studies using fMRI and Stereoelectroencephalography (SEEG) signals, in particular, a network comprising the anterior cingulate cortex and other neural sources, including dorsolateral prefrontal cortex, parietal lobe, medial temporal lobe, and thalamus [1], [3], [4].

In particular, it has been reported that error-related potentials can be detected in scalp EEG recordings with human subjects. Typically they consist of an error-related negativity (ERN) located in frontocentral areas, followed by an error positivity (EP) in centroparietal areas. Because of its key role in human brain function, error related potentials have been proposed as input for non-invasive Brain-Computer Interfaces (BCI). Single trial detection of error potential has been applied to monitor erroneous stimulus [5] or during interaction with external devices [6] [7]. So far, all these systems used only waveform or spectral information as discriminant features. We propose that channel interaction and phase differences can also be used as features for classification in BCI systems [8], [9]. We assess the use of directed transfer function (DTF) as a feature extraction method that reflects directional connectivity across multiple channels. This method, which is an extension of Granger causality from pairwise variables [10], allows the estimation of the directed information transfer between multi-variables. It has been applied to compute the brain connectivity in several areas, including localization of epileptic foci [11] and memory information processing [12].

In this study, EEG signal was recorded when human subjects monitored the direction of a moving square. Classification performance between correct and erroneous movement direction using DTF features was evaluated, and the characteristics of connectivity features in temporal, frequency domains and brain regions were assessed.

II. METHODS

A. Experimental protocols

In the experiment, subjects were seated in front of a computer screen, placed at about $50cm$ from their eyes. At the beginning of the experiment, 11 empty white squares arranged horizontally are shown. An orange *target square* appeared randomly to either leftmost or rightmost position. It was followed by a blue *cursor square* in the central location. Then, the cursor moved towards the target square with 80% probability every $2s$. The cursor kept on moving until reaching the target where it changed color to green. If the target was not reached before $40s$, the trial was stopped. During the experiment, subjects were requested to minimize eyes blinking and continuously focus their eyes on the next position of the cursor until $0.5s$ after the cursor reaching the new position. The trials that moved towards the target were considered as correct trials, whereas movements in the opposite direction were considered as error trials. For each subject, more than 100 error trials and more than 400 correct trials were performed. 10 subjects (3 females, age 26 ± 2.40) were included in the experiment. The data from this experiment is referred to as *dataset1* in the following sections. In addition, we also report results on 6 subjects (1 female, age = 27.83 ± 2.23) from a previous study using a similar protocol (here denoted *dataset2*) [5].

B. Directed Transfer Function

We used directed transfer function (DTF) to estimate the brain information interaction between EEG channels. The DTF method is based on multivariate autoregressive (MVAR)

*This study was supported by Nissan Motor Co. Ltd., and carried out under the "Research on Brain Machine Interface for Drivers" project. M.K.Goel is supported by the Swiss National Science Foundation, Project 200021-120293.

H. Zhang, R. Chavarriaga, M. K. Goel, and J. d. R. Millán are with the Defitech Foundation Chair in Non-Invasive Brain-Machine Interface, Center for Neuroprosthetics, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. L. Gheorghe is from Mobility Services Laboratory, Nissan Research Center, Nissan Motor Co., Japan. huaijian.zhang@epfl.ch jose.millan@epfl.ch

model. Multi-channel EEG data X_t at time t is defined as $X_t = [x_{1,t}, x_{2,t}, \cdots, x_{k,t}]^T$, where k denotes the number of channels and T denotes the matrix transposition. The MVAR can be represented by:

$$\sum_{j=0}^{p} A_j X_{t-j} = E_t \tag{1}$$

where E_t is a vector of zero-mean uncorrected white noise process with size $1 \times k$, the coefficients A_1, A_2, $\cdots$, A_p are $k \times k$ matrix with $A_0 = -I$ (I is the identity matrix). In the model, p is the model order of MVAR, which is chosen by the Akaike Information Criteria (AIC) [13], [14]. By multiplying by X_{t-r}^T ($r = 1, 2, \cdots, p$) and taking expectation of both sides of equation (1), we get the Yule-Walker equation:

$$\sum_{j=0}^{p} A_j R(j-r) = 0 \tag{2}$$

In the equation (2), $R(n)$ is the covariance matrix of X_t with lag n. The coefficient matrices A_j can be obtained by calculating the covariance matrix for each lag. After that, we can investigate the relations in the frequency domain by transforming (2) into the frequency domain:

$$A^F X^F = E^F \tag{3}$$

$$X^F = A^{F-1} E^F = H^F E^F \tag{4}$$

where

$$A^F(f) = -\sum_{j=0}^{p} A_j e^{-i2\pi f j} \tag{5}$$

The matrix H^F is the transfer matrix of the model. The MVAR model takes white noise as the input and signals X^F as the output. The system transfer matrix H^F contains information between channels, where $H_{ij}^F(f)$ represents the information transfer from channel j to channel i at a frequency f Hz. The matrix H^F is not symmetric, indicating that the connectivity value from channel j to channel i is different from the connectivity value from channel i to channel j, i.e. there is directionality of information transfer between channels. The values of the matrix are non-zero only when there is phase difference between channels. The non-normalized DTF is defined as:

$$\theta_{ij}^2(f) = |H_{ij}^F(f)|^2 \tag{6}$$

In addition, we can estimate the power spectral matrix of the process:

$$S^F = X^F X^{F*} = H^F V^F H^{*F} \tag{7}$$

where V^F is the covariance matrix of E^F, the asterisk represents the transposition and complex conjugation. Power spectral densities of signals are described in the diagonal elements of the matrix.

C. Signal Processing and Classification

EEG signal was recorded from 64 channels according to the extended 10/20 system using a Biosemi Active Two system. Sampling rate of the recording was $2048Hz$, and downsampled to $512Hz$ afterwards. We used a $4th$ order Butterworth filter to process the raw EEG data with cutoff bands between $0.1Hz$ and $30Hz$. Common average reference (CAR) was used to remove common brain activity. The collected EEG data were segmented in trials for both correct and error cases. The length of each trial was $2s$, including $1s$ before the visual stimulus and $1s$ after. Onsets of stimuli were considered as origin ($t = 0s$).

After preprocessing, we analyzed the brain connectivity patterns within $(-1\ 1)$. To this end, we computed the connectivity features in a sliding windows of size $400ms$ and 90% overlapping ($360ms$), which yielded smoother brain connectivity information. The longer window size, the more stable the MVAR computation, while the shorter the window size, the higher the resolution of the brain connectivity. Here, we used $400ms$ as a trade-off between these two factors. Electrodes Fz, FCz, Cz and CPz were selected for DTF analysis, following early studies on error potentials with scalp EEG [5]. Signal was normalized across time axis for every sliding window (subtracting temporal mean and dividing temporal standard deviation) for all 4 channels separately to meet the zero mean requirement in MVAR [10]. The order of MVAR was 11 based on AIC.

Linear discriminant analysis (LDA) was used to classify correct and error trials. We compare the classification performance when different sets of features are used: *(1)* Waveform features; *(2)* Brain connectivity features; *(3)* Combination of waveform and brain connectivity features.

To extract waveform features, the EEG signals were filtered with $4th$ order Butterworth filter between 1 and $10Hz$ and re-referenced using CAR. Features were extracted from 4 channels (Fz, FCz, Cz and CPz; i.e., the same channels selected for the brain connectivity computation), between $0.2s$ and $0.7s$ after the stimulus, for a total of 20 samples. A total of 52 (13 $samples \times 4$ $channels$) features were selected as waveform features. In the second case, we used $\theta_{ij}^2(f)$ in (6), corresponding to brain connectivity features. For each trial we computed a matrix with size $4 \times 4 \times 30 \times 80$, where 4×4 indicates interaction between 4 channels, and 30×80 indicates frequencies in the [1-30] Hz in the time range between $-0.8s$ and $0.8s$. Since this provides a total of 38400 features, we used Wilcoxon rank sum test to select the 50 most important features; i.e. those features with smallest p-value in the training set. Ten-fold cross-validation was used to test the performance for all feature selections. In the third case, both waveform and connectivity features were computed as above and then feed into the classifier.

III. RESULTS

Figure 1 illustrates the grand average of correct and error trials of channels Fz, FCz, Cz and CPz for all subjects in the dataset1. A positive peak can be observed in all selected channels around $200ms$ after erroneous stimuli (red traces),

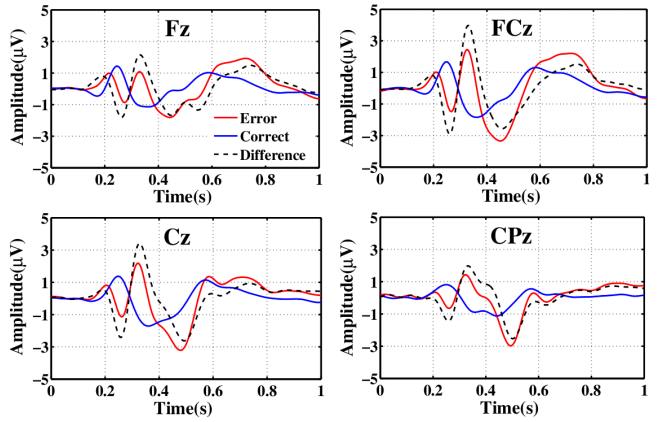

Fig. 1. Grand average and difference of error/correct potentials in Fz, FCz, Cz and CPz channels across 10 subjects. *Red trace*: Error trials. *Blue trace*: Correct trials. *Dashed line*: Error minus correct.

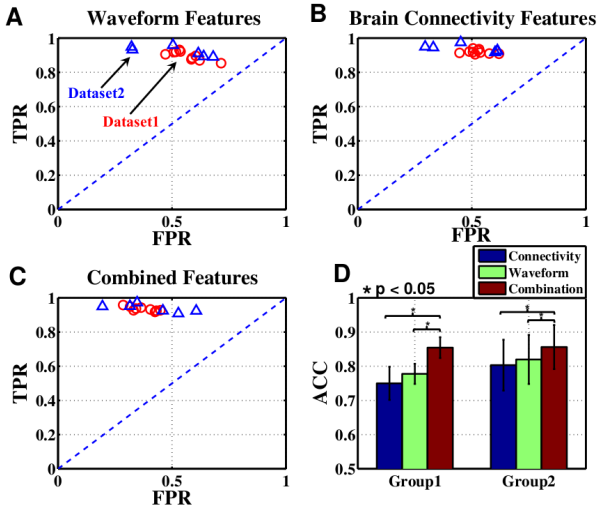

Fig. 2. Classification performances based on different features. A, B and C indicates the classification performance in ROC space for two datasets. D illustrates the average accuracy and standard deviation for two datasets

and a following negative peak happens around $260ms$. After that, a negative peak around $410ms$ can be observed in all selected channels. For the correct condition (blue traces), two peaks can clearly be seen, a positive peak around $230ms$ and a following negative peak near $360ms$. We used Wilcoxon rank sum test to find the significant differences ($p < 0.05$) between correct and error cases: around $180ms$, $250ms$, $350ms$, $455ms$ and $540ms$ in channel Fz; around $240ms$, $320ms$ and $450ms$ in channel FCz; around $250ms$, $320ms$ and $500ms$ in channel Cz; around $500ms$ in channel CPz. Similar waveforms were obtained for the dataset2 [5].

Classification performance is illustrated in Figure 2. Figure 2 A-C denote performances in receiver operating characteristic (ROC) space for all subjects. Here the x axis denotes the false positive rate, and y axis denotes the true positive rate, where we considered correct trials as positive. The performance of an ideal classifier should locate at $(0, 1)$ in ROC space. Each mark in the figure represents one of the subjects for the two datasets (red: dataset1. blue: dataset2). Average performance error across all subjects can be found in Table I. Results of all three methods are above random level. Classification based on both waveform and connectivity features outperforms the other two methods, yielding higher TPR and lower FPR for both datasets. The accuracy of the combination method is significantly higher ($p < 0.05$) than waveform features and connectivity features in both datasets (Figure 2 D).

In the previous results, the combination method uses higher number of features than the waveform. To assess the effect of feature number, we tested a classifier based on waveform using 104 (26×4) features (extracted by taking data per 10 samples). The accuracy is 0.820 ± 0.07 for dataset1 and 0.785 ± 0.05 for dataset2, which is higher than the classifier using 52 waveform features in dataset1, and lower in dataset2. For both datasets, it is not as good as the combination method (see Table I).

The characteristics of the selected connectivity features were assessed. Figure 3 A-B illustrates the feature distribution of all subjects and folds in the frequency domain. It shows that features in the frequency band 7-9Hz are highly discriminant, suggesting that brain connectivity around theta band was more important to discriminate between correct and error conditions than other bands (Figure 3 A-B). In the temporal domain (Figure 3 C-D), the information transfer around $0.2s$ after stimuli carry most essential discriminating information for classification. Figure 3 E-F illustrate the importance of brain connectivity pairs for classification, where the color of the figure denotes the percentage of features selected across all pair of electrodes. The values in the diagonal are zero, since information transfer within one electrode was set to zero in DTF computation. The brain information transfer from FCz to Cz and from frontal to parietal electrodes (Fz to CPz and FCz to CPz) have relatively higher weights than other connections.

TABLE I
CLASSIFICATION PERFORMANCE (MEAN±SD)

		Connectivity	Waveform	Combination
Dataset1	ACC	0.75 ± 0.05	0.78 ± 0.03	0.85 ± 0.03
	TPR	0.90 ± 0.03	0.92 ± 0.01	0.93 ± 0.01
	FPR	0.57 ± 0.07	0.53 ± 0.05	0.38 ± 0.06
Dataset2	ACC	0.80 ± 0.07	0.81 ± 0.07	0.85 ± 0.06
	TPR	0.92 ± 0.02	0.94 ± 0.02	0.94 ± 0.02
	FPR	0.51 ± 0.16	0.48 ± 0.15	0.41 ± 0.15

IV. DISCUSSIONS AND CONCLUSIONS

We show that single trial classification between correct and error can be significantly improved by using connectivity-based features. Because of the signal characteristics of error potential in EEG channels, we chose Fz, FCz, Cz and CPz. These electrodes are located over the cortical areas including ACC, frontal cortex and supplementary motor area (SMA). Functional connectivity between ACC and frontal areas during error processing is supported by previous studies

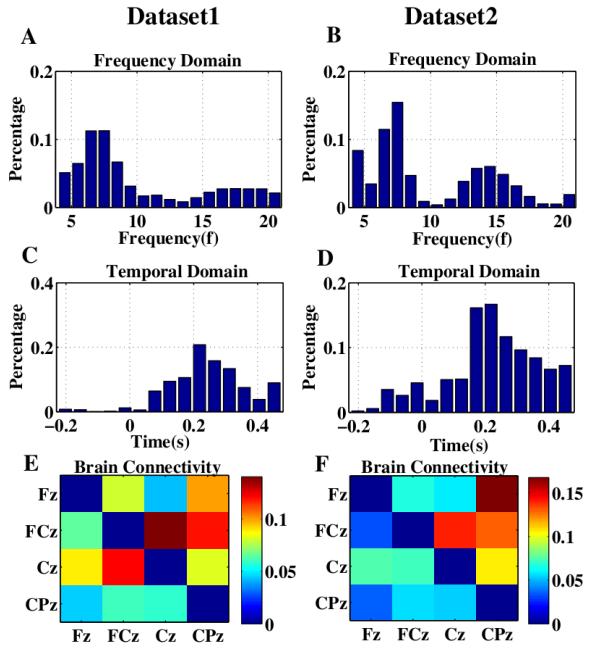

Fig. 3. Characteristics of connectivity features. A and B indicated feature distribution in frequency domain, whereas C and D indicated the distribution in temporal domain. E and F represented the weights for connectivity pairs

[15], [16], as well as coactivation between ACC and SMA [17]. Although the feature selection weights do not reflect the functional brain interaction of cortical areas directly, the selection of connectivity features between FCz/Cz as well as FCz/CPz in this work are consistent with previous studies by fMRI [15] and SEEG [16].

Although DTF uses MVAR model, which is a linear method, the computation of DTF values (i.e. coefficients of MVAR model) from EEG channels is not linear, requiring multiplication between channels (covariance). Hence, it provides classification features by non-linear combination of localized EEG channels. In the future, we would like to compare the use of DTF with spectral coherence, which does not consider directionality between channels (as off-diagonal elements in Equation 7, which is symmetric), to find out the contribution of directionality in classification.

Frequency domain distribution of the connectivity features demonstrates the particular role of theta rhythm in error related processing in agreement with reported results from other studies [16] [18]. Theta rhythm has been considered connecting activity of hippocampal systems and cortical mantle, playing key role in focused attention, working memory and action control [19]. Temporal distribution indicates that the most important features are around $200ms$ to $300ms$ after stimuli. So far, we are not aware of previous studies considering the short time course of brain connectivity during brain error processing. In the future we will assess this issue by a modification of the method, short-time DTF.

Although only 4 channels were included in this study, far from capturing all related brain regions, DTF coefficients provided extra nonlinear features which significantly improved the classification performance. In the next step, we will focus on real time error detection for BCI application using the same approach. In the current implementation, the average time required for DTF computation was around $1.4ms$ per window, supporting the possibility of future online applications.

REFERENCES

[1] C. B. Holroyd and M. G. Coles, "The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity," *Psychol. Revi.*, vol. 109, no. 4, pp. 679–709, 2002.

[2] S. F. Taylor, E. R. Stern, and W. J. Gehring, "Neural systems for error monitoring: recent findings and theoretical perspectives," *Neuroscientist*, vol. 13, no. 2, pp. 160–72, 2007.

[3] M. Brázdil, R. Roman, M. Falkenstein, P. Daniel, P. Jurák, and I. Rektor, "Error processing–evidence from intracerebral ERP recordings," *Exp. Brain Res.*, vol. 146, no. 4, pp. 460–6, 2002.

[4] P. Luu, P. Collins, and D. M. Tucker, "Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring," *J. Exp. Psychol. Gen.*, vol. 129, no. 1, pp. 43–60, 2000.

[5] R. Chavarriaga and J. d. R. Millán, "Learning from EEG error-related potentials in noninvasive brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 4, pp. 381–388, 2010.

[6] X. Perrin, R. Chavarriaga, C. Ray, R. Siegwart, and J. d. R. Millán, "A comparative psychophysical and EEG study of different feedback modalities for human-robot interaction," in *ACM/IEEE Conf on Human-Robot Interaction HRI08*, Amsterdam, Netherlands, 2008.

[7] P. W. Ferrez and J. d. R. Millán, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 923–929, 2008.

[8] H. Zhang, H. L. Benz, A. Bezerianos, S. Acharya, N. E. Crone, A. Maybhate, X. Zheng, and N. V. Thakor, "Connectivity mapping of the human ECoG during a motor task with a time-varying dynamic bayesian network." in *Proceedings of the 32nd EMBS*, pp. 130–133.

[9] B. Hamner, R. Leeb, M. Tavella, and J. d. R. Millán, "Phase-based features for motor imagery brain-computer interfaces," in *Proceedings of the 33rd EMBS*, pp. 2578 –2581.

[10] M. J. Kamiński and K. J. Blinowska, "A new method of the description of the information-flow in the brain structures," *Biol. Cybern.*, vol. 65, no. 3, pp. 203–210, 1991.

[11] P. J. Franaszczuk, G. K. Bergey, and M. J. Kamiński, "Analysis of mesial temporal seizure onset and propagation using the directed transfer function method," *Electroen. Clin. Neuro.*, vol. 91, no. 6, pp. 413–427, 1994.

[12] C. Babiloni, F. Vecchio, S. Cappa, P. Pasqualetti, S. Rossi, C. Miniussi, and P. M. Rossini, "Functional frontoparietal connectivity during encoding and retrieval processes follows HERA model - a high-resolution study," *Brain Res. Bull.*, vol. 68, no. 4, pp. 203–212, 2006.

[13] M. Kamiński, M. Z. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, no. 2, pp. 145–157, 2001.

[14] H. Akaike, "New look at statistical-model identification," *IEEE Trans. Autom. Control*, vol. Ac19, no. 6, pp. 716–723, 1974.

[15] J. Fan, P. R. Hof, K. G. Guise, J. A. Fossella, and M. I. Posner, "The functional integration of the anterior cingulate cortex during conflict processing," *Cerebral Cortex*, vol. 18, no. 4, pp. 796–805, 2008.

[16] M. Brázdil, C. Babiloni, R. Roman, P. Daniel, M. Bares, I. Rektor, F. Eusebi, P. M. Rossini, and F. Vecchio, "Directional functional coupling of cerebral rhythms between anterior cingulate and dorsolateral prefrontal areas during rare stimuli," *Hum. Brain Mapp.*, vol. 30, no. 1, pp. 138–146, 2009.

[17] L. Koski and T. Paus, "Functional connectivity of the anterior cingulate cortex within the human frontal lobe: a brain-mapping meta-analysis," *Exp. Brain Res.*, vol. 134, no. 3, pp. 408–408, 2000.

[18] J. F. Cavanagh, M. X. Cohen, and J. J. Allen, "Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring," *J. Neurosci.*, vol. 29, no. 1, pp. 98–105, 2009.

[19] A. Gevins and M. E. Smith, "Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style," *Cerebral Cortex*, vol. 10, no. 9, pp. 829–839, 2000.