2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Classification of erroneous actions using EEG frequency features: implications for BCI performance*

Camila Dias, Diana M. Costa, Teresa Sousa, João Castelhano, Verónica Figueiredo, Andreia C. Pereira, Miguel Castelo-Branco

Abstract—**Several studies have demonstrated that error-related neuronal signatures can be successfully detected and used to improve the performance of brain-computer interfaces. However, this has been tested mainly in well-controlled environments and based on temporal features, such as the amplitude of event-related potentials. In this study, we propose a classification algorithm combining frequency features and a weighted SVM to detect the neuronal signatures of errors committed in a complex saccadic go/no-go task. We follow the hypothesis that frequency features yield better discrimination performance in complex tasks, generalize better, and require fewer pre-processing steps. When combining temporal and frequency features, we achieved a balanced classification accuracy of 75% - almost the same as using only frequency features. On the other hand, when using only temporal features, the balanced accuracy decreased to 66%. These findings show that subjects' performance can be automatically detected based on frequency features of error-related neuronal signatures. Additionally, our results revealed that features computed in the pre-response time contribute to the discrimination between correct and erroneous responses, which suggests the existence of error-related patterns even before response execution.**

I. Introduction

Brain-computer interfaces (BCIs) are systems that translate neuronal activity patterns into commands for an interactive application, thus decoding user's intentions from brain activity, which is typically measured using electroencephalography (EEG) [1, 2]. Therefore, BCIs enable computer control without any physical activity, allowing, e.g., motor-impaired users to control assistive tools [1, 3].

There are some difficulties in the classification for EEG-based BCIs, namely the low signal-to-noise ratio (SNR), the limited amount of data, and data variability over time, between users and tasks. One solution to overcome these limitations is to use error-related potentials (ErrPs) to deduce which user's intentions are incorrectly decoded by the BCI. This allows BCIs to learn online (requiring fewer offline training data) and to continuously update their parameters. When the output is not the one intended by the user, it elicits ErrPs that can be automatically detected. Hence, the label is estimated as the one predicted by the BCI if no ErrP is detected, and as the opposite otherwise (for binary classification) [1]. Then, ErrPs can be decoded to revert the outcome of erroneous commands, derived from machine [4] or subject errors [5].

It is hence of extreme importance to properly classify errors in this context. It has been shown that the ErrPs can be recognized on a single-trial basis [5, 6]. However, there are still improvements to make, because ErrPs-based studies are usually performed in controlled laboratory conditions using simple tasks, which leads to high SNR and avoids typical confounds of realistic scenarios [2]. Moreover, the large majority of ErrP classification studies are based on temporal features [2, 6], which might not be ideal during complex tasks [7]. Therefore, it would be beneficial to find other types of features with error-related discriminative power.

The extensive use of temporal features to classify ErrPs is related to the appearance of a negative deflection at 50-100 ms after the subject's errors at mid-frontal electrodes, centered at FCz, named error-related negativity (ERN). A similar peak, the feedback-related negativity (FRN), arises 200-300 ms after the perception of an error [2], which may be committed by another subject or a machine [8]. Moreover, the ERN may be followed by an error-related positivity (Pe), a mid-parietal positive peak centered at the P_Z electrode, which arises 200-400 ms after the erroneous action [9]. Nevertheless, these are not the only correlates of error monitoring. There is large evidence that mid-frontal theta oscillations are associated with error monitoring, which is reflected by an increase in theta power during errors peaking on the FCZ electrode [2, 9, 10]. Moreover, parieto-occipital alpha activity has been related to errors caused by attentional lapses [9, 11] and the delta band has also been associated with error commission at central [10], mid-frontal and parietal channels [12].

Few studies have explored the use of frequency features to detect errors [10, 13], but have brought promising results. Omedes *et al*. [13] demonstrated that frequency features generalize better across tasks due to their lower sensitivity to latency shifts [2]. Furthermore, Spüler *et al*. [10] showed that frequency features yield much higher classification performance in an asynchronous context [14]. Also, frequency features require less pre-processing than temporal features [1], which is advantageous in the BCI context.

*Research supported by FCT – Portuguese national funding agency for science, research and technology [grants CONNECT-BCI: POCI-01-0145-FEDER-30852, FCT/UID/4950/2020, DSAIPA/DS/0041/2020, B-RELIABLE: PTDC/EEI-AUT/30935/2017, FCT/UID/150832/2021] and Bial Foundation [grant 306/18].

C. Dias, D. M. Costa, T. Sousa, J. Castelhano, V. Figueiredo, A. C. Pereira and M. Castelo-Branco are with the Institute of Nuclear Sciences Applied to Health ICNAS, Coimbra Institute for Biomedical Imaging and Translational Research CIBIT, University of Coimbra, Portugal. M. Castelo-Branco is also with Faculty of Medicine, University of Coimbra, Portugal (e-mail: camilageraldodias@gmail.com).

Here, we tested a set of frequency features and compare their discriminative power of erroneous *vs*. correct actions to the temporal features. These were used to build a classifier to detect errors committed by the user in a complex saccadic go/no-go task. We computed not only theta [13] but also delta and alpha power features. In addition to the features computed from EEG data in the post-response time [10, 13], we assessed the pre-response time.

II. METHODS

A. Data Acquisition

Data from 20 healthy participants (nine female, mean age 26.80 ± 4.51 y) were used in this study. All provided written consent in accordance with the Declaration of Helsinki, and the study followed the safety guidelines for research on humans. The work was approved by the Ethics Committee of the Faculty of Medicine of the University of Coimbra.

The experiment was based on a go/no-go saccadic task with facial expression cues. During "go" trials, a happy averted face was the cue to look in the same direction of the face shown (pro-saccade), while a sad averted face informed the participants to look in the opposite direction of the face shown (anti-saccade). The "no-go" trials were signalized by a face (happy/sad) looking straight ahead. There were six distinct instructions (Fig. 1). The facial expression images of a white young-adult male were obtained from the Radboud Faces Database [15].

The paradigm included six stages (Fig. 2). Firstly, a Neutral face (preparatory cue) was exhibited and followed by a Gap period. Then, the Instruction (randomly selected) was given and followed by the Fixation cross. In the Target phase, a square (distractor) appeared on either the right or left part of the screen, coherent with the gaze direction (when the face was looking forward, the target was randomly placed right or left). The last stage was destinated to the execution of the Response (saccade or no-go). However, many participants performed some saccades in other stages (mainly anticipatory reflexive saccades). The task comprised four runs of 84 trials each (28 pro-saccade, 28 anti-saccade, and 28 no-go trials). Happy and sad, right and left trials were counterbalanced. The task was programmed in Presentation software (v. 12.0, Neurobehavioral Systems Inc.).

EEG and electrooculogram (EOG) were recorded from 64 channels (QuickCap, NeuroScan) with an extended international 10-20 system at 1000 Hz. The online reference was a channel close to Cz. EEG signals were amplified using a SynAmps 2 amplifier (Compumedics NeuroScan) and recorded using Curry Neuroimage 7.08 (NeuroScan). Eye tracking (ET) data were acquired simultaneously to record saccades at 120 Hz (iView XTM Hi-Speed, SMI). Data from one participant was excluded due to synchronization problems between EEG and ET data.

B. EEG Data Pre-processing

EEG data were down-sampled to 500 Hz, filtered (0.5 - 45 Hz), and the noisy channels were removed. The electrodes were re-referenced to the average of all EEG (excluding EOG) channels. To minimize artifacts, Independent Component Analysis [16] was used. Afterwards, the previously removed noisy channels were interpolated (spherical interpolation).

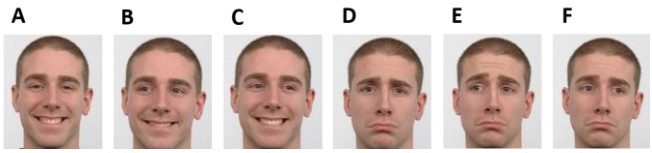

Figure 1. Facial instructions: A) happy no-go, B) left pro-saccade, C) right pro-saccade, D) sad no-go, E) left anti-saccade, and F) right anti-saccade.

The data were segmented into epochs of 1000 ms locked to the onset of the saccade and starting 500 ms before it (trials without saccades were not used). These were corrected for a baseline computed from the mean activity of the entire epoch.

Trials were defined as correct (and used for classification with label = 0), if their first saccade was executed according to the given instruction during the stage Response (section II-A), to avoid error-related signals due to anticipation of responses. In opposition, trials were defined as errors (label = 1), if the first saccade was not according to the given instruction (e.g., pro-saccades instead of no-go), regardless of the response timing. Given that all correct but anticipated responses were removed and that trials without any saccade were not used, the number of observations were 1283 (1056 correct and 227 erroneous instances).

C. Feature Extraction

To build the classification model, several features were computed per trial (table 1). Frequency band power (e.g., theta) "before" and "after" is the mean power spectral density (PSD) - computed using the Welch's method - in that band at [-500, 0] ms and [0, 500] ms, respectively, where $t = 0$ is the response onset. Most of the features were computed for each individual channel, except for the features regarding the mean theta activity in the channel cluster around FCz. The number of initially extracted features was 1559, with 1307 frequency features - 587 of which from the theta band -, 250 temporal features, gender and age. EEG data were pre-processed, and features were computed using EEGLAB functions (v. 2) in a MATLAB script (v. R2018b).

D. Feature Selection

Firstly, all features with a coefficient of variance lower than 0.2 were removed. The remaining steps of feature selection were done using cross-validation [17] and, accordingly, different features were selected in each iteration. When the absolute value of the Pearson correlation between two features was higher than 0.9, one of them was nominated for elimination in that iteration (the one with the highest *p*-value obtained from *t*-tests between classes). After all iterations, the features designated for removal in less than half of the iterations were selected. Then, we selected the 40 most discriminative from the remaining ones - the *p*-values obtained from the *t*-tests between classes on each iteration were sorted and the features received from one (lowest *p*-value) to *N* (highest *p*-value) votes to be removed. In the end, the 40 features with less votes were selected.

E. Classification

The Monte-Carlo cross-validation was applied, in which the data were randomly partitioned in training (70%) and test (30%) sets in 100 iterations [18]. The features were standardized (z-score transformation) in each training set and the same transformation was then applied to the test set. The

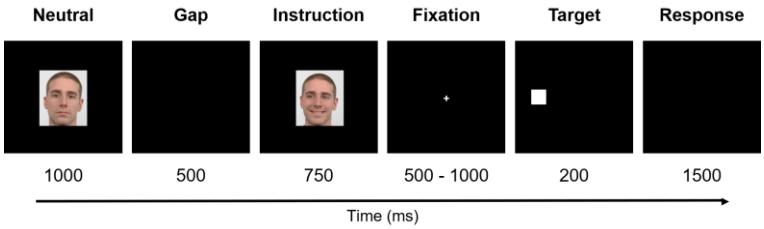

Figure 2. Experimental design.

classification was performed using a support vector machine (SVM) with a radial basis function (RBF) kernel. Given that our data were imbalanced, we used a weighted SVM [10] that adjusts weights inversely proportional to class frequencies [19]. Feature selection and classification was performed in Python (v. 3.8) using Scikit-learn package.

III. Results

A. Selected Features

The 40 most discriminative features are flagged in table 1 with a number. Features 1-4 measure the theta activity of the cluster around FC_Z, but the remaining selected features assess the activity of single channels. Feature 5 was selected for FC1, FCZ, FC2; feature 6 for FCZ, FC2; feature 7 for F3, FZ, FC1, FCZ, FC2, P4; feature 8 for FCZ; feature 9 for FC3, FC1, FCZ, FC2, CZ; feature 10 for FP1; feature 11 for C2, P6, PO4, PO6, PO8, O2; feature 12 for FC2, T8, CP5, P5; feature 13 for CP4; feature 14 for C5, T8, CP5, TP8, P5; and feature 15 for F8.

We verified that 95% of the selected features were based on frequency metrics, 55% of which were theta-related, 24% were low frequency (average between delta, theta and alpha) features, 16% were alpha features and the remaining 5% were delta features. Moreover, 95% of the selected theta features were related to mid-frontal channels. On the other hand, 33% of the selected features were based on data prior to the response execution. The 10 most discriminative features (sorted) were feature 9 for FCZ, feature 3, feature 14 for T8, feature 4, feature 14 for P5, feature 9 for FC1, feature 14 for CP5, feature 11 for PO8, and feature 7 for FC1 and FCZ.

Table 1. Features extracted.

[I] "All freq." represent the mean PSD at [1.5, 45] Hz for a certain channel/cluster of channels. [II] "High freq." is the mean PSD at [35, 45] Hz.
[III] "Other channels" is the cluster of all channels not included in the set of mid-frontal channels.

Frequency features			
Theta features (4-8 Hz)			
Channel cluster around FC_Z (Fz, FC1, FCz, FC2, Cz)			
$\frac{Theta\ after_{around\ FCz}}{All\ freq.^{I}\ after_{around\ FCz}}$ (1)	$\frac{Theta\ after_{around\ FCz}}{High\ freq.^{II}\ after_{around\ FCz}}$ (2)	$\frac{Theta\ before_{around\ FCz}}{All\ freq.^{I}\ before_{around\ FCz}}$	$\frac{Theta\ before_{around\ FCz}}{High\ freq.^{II}\ before_{around\ FCz}}$
$\frac{Theta\ after_{around\ FCz}}{Theta\ after_{other\ channels^{III}}}$ (3)	$\frac{Theta\ before_{around\ FCz}}{Theta\ before_{other\ channels^{III}}}$	$\frac{Theta\ after_{around\ FCz}}{Theta\ before_{around\ FCz}}$ (4)	
All channels			
$\frac{Theta\ after_{Ch}}{All\ freq.^{I}\ after_{Ch}}$ (5)	$\frac{Theta\ after_{Ch}}{High\ freq.^{II}\ after_{Ch}}$ (6)	$\frac{Theta\ before_{Ch}}{All\ freq.^{I}\ before_{Ch}}$	$\frac{Theta\ before_{Ch}}{High\ freq.^{II}\ before_{Ch}}$
$\frac{Theta\ after_{Ch}}{Theta\ before_{Ch}}$ (7)	$\frac{Theta\ after_{Ch}}{Delta\ after_{Ch}}$	$\frac{Theta\ after_{Ch}}{Alpha\ after_{Ch}}$ (8)	$\frac{Theta\ before_{Ch}}{Delta\ before_{Ch}}$
$\frac{Theta\ before_{Ch}}{Alpha\ before_{Ch}}$			
Mid-frontal channels Single channels: Ch = {FP1, FPz, FP2, AF3, AF4, F5, F3, F1, Fz, F2, F4, F6, FC3, FC1, FCz, FC2, FC4, C1, Cz, C2}			
$\frac{Theta\ after_{Ch}}{Theta\ after_{other\ channels^{III}}}$ (9)		$\frac{Theta\ before_{Ch}}{Theta\ before_{other\ channels^{III}}}$	
Delta (1.5-3.5 Hz)**, alpha** (8.5-12 Hz) **and low frequency** (1.5-12 Hz) **features** (all channels)			
$\frac{Power^{*}\ after_{Ch}}{All\ freq.^{I}\ after_{Ch}}$ *Delta (10), alpha (11), low freq. (12)	$\frac{Power^{*}\ after_{Ch}}{High\ freq.^{II}\ after_{Ch}}$ *Delta, alpha, low freq.	$\frac{Power^{*}\ before_{Ch}}{All\ freq.^{I}\ before_{Ch}}$ *Delta (13), alpha, low freq. (14)	$\frac{Power^{*}\ before_{Ch}}{High\ freq.^{II}\ before_{Ch}}$ *Delta, alpha, low freq.
Temporal features			
ERN features (around FC_Z) Single channels: Ch = {Fz, FC1, FCz, FC2, Cz}		**Pe features** (around P_Z) Single channels: Ch = {CPz, P1, Pz, P2, POz}	
Mean potential at [70, 160] ms		Mean potential at [200, 500] ms	
Other temporal features (all channels)			
Mean potential at [-500, -250] ms (15)	Mean potential at [-250, 0] ms	Mean potential at [0, 250] ms	Mean potential at [250, 500] ms
Demographic features			
Gender (16)		Age	

B. Performance Evaluation

The classification performance metrics are shown in table 2. We tested four models using different features sets: all features, and with temporal, frequency, and theta features individually. All classifiers were tested with 40 features.

Table 2. Classification performance using different sets of features.

Features	Sensitivity (%)	Specificity (%)	Bal. accuracy (%)
Temporal	54.35 ± 5.15	78.45 ± 2.26	66.40 ± 2.73
Theta	57.35 ± 7.01	81.52 ± 2.52	69.44 ± 3.24
Frequency	59.46 ± 5.32	87.53 ± 2.38	73.49 ± 2.46
All	62.59 ± 5.56	87.95 ± 2.15	75.27 ± 2.74

IV. Discussion

With this study, we aimed to test whether frequency features add discriminative power to the automatic recognition of error-related EEG patterns. We intended to compare the use of frequency and temporal features in complex tasks. We built SVM classifiers and achieved a balanced accuracy of 75% using temporal, frequency and demographic features, similar to the one obtained using only frequency features (73%). Using only theta features, the performance decreased to 69% and, only with temporal features, to 66%. Moreover, 95% of the most discriminative features were frequency-related (theta, alpha and delta). On the one hand, these results demonstrate that, in this context, the frequency features discriminate better between correct and erroneous responses than temporal features. On the other, we revealed that it is beneficial to explore not only theta, but also alpha and delta features. As expected, the great majority of selected theta features were computed from mid-frontal channels [9, 10]. Regarding the alpha band, features were selected mainly from parieto-occipital channels, which is in line with the link between attentional errors and parieto-occipital alpha [9].

Although some previous studies report classification accuracies above 80% [4] or even 90% [20], a great part of literature describes accuracies between 70% and 80% [7, 10, 13]. Moreover, most of these studies were tested in well-controlled conditions with time-locked events. When trying to detect ErrPs asynchronously, Spüler *et al*. [10] reported only 66% of accuracy, revealing the challenge of recognizing errors in continuous events. Our experiment was based on a go/no-go saccadic task and, given that saccades are semi-automatic oculomotor responses to visual stimuli, there was a high variability of response timing. This context, with asynchronous saccades, is hence more realistic and challenging than the usual simple tasks [2]. The better performance derived from the use of frequency features found here supports the notion that oscillatory activity allows asynchronous recognition of erroneous actions [2, 10].

In addition to the usual features computed in the post-response time, we assessed the pre-response time as well. Interestingly, one third of the selected features were computed from the pre-response time, which suggests the existence of error-related patterns prior to response execution, namely regarding theta and delta power. A future study comparing correct and erroneous trials regarding delta and theta power in the pre-action time is needed to confirm it.

Summing up, we have found that frequency features add discriminative power to the automatic detection of errors. Our results contribute to improve current classification algorithms; particularly the ones based on temporal features and that are only robust in synchronous contexts [2, 10]. The classification algorithm here proposed still needs to be tested for online applications. Nevertheless, the presented features might bring an important contribution to BCI systems optimization.

References

[1] F. Lotte *et al*., "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update," *J. Neural Eng*., 2018.

[2] R. Chavarriaga, A. Sobolewski, and J. del R. Millán, "Errare machinale est: the use of error-related potentials in brain-machine interfaces," *Front. Neurosci*., vol. 8, no. 208, 2014.

[3] G. Pires, U. Nunes, and M. Castelo-Branco, "Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis," *J. Neurosci. Methods*, 2011.

[4] A. Cruz, G. Pires, and U. J. Nunes, "Double ErrP Detection for Automatic Error Correction in an ERP-Based BCI Speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 26–36, 2017.

[5] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Response Error Correction—A Demonstration of Improved Human-Machine Performance Using Real-Time EEG Monitoring," in *IEEE Trans. on Neural Syst. and Rehabil. Eng.*, 2003, vol. 11, no. 2, pp. 173–177.

[6] T. Plewan, E. Wascher, M. Falkenstein, and S. Hoffmann, "Classifying response correctness across different task sets: A machine learning approach," *PLoS One*, vol. 11, no. 3, 2016.

[7] H. Zhang, R. Chavarriaga, Z. Khaliliardali, L. Gheorghe, I. Iturrate, and J. R. Millán, "EEG-based decoding of error-related brain activity in a real-world driving task," *J. Neural Eng*., vol. 12, 2015.

[8] A. Paas, G. Novembre, C. Lappe, and P. E. Keller, "Not all errors are alike : Modulation of error-related neural responses in musical joint action," *Soc. Cogn. Affect. Neurosci*., 2021.

[9] E. F. Pavone, G. Tieri, G. Rizza, E. Tidoni, L. Grisoni, and S. M. Aglioti, "Embodying Others in Immersive Virtual Reality: Electro-Cortical Signatures of Monitoring the Errors in the Actions of an Avatar Seen from a First-Person Perspective," *J. Neurosci*., vol. 36, 2016.

[10] M. Spüler and C. Niethammer, "Error-related potentials during continuous feedback: Using EEG to detect errors of different type and severity," *Front. Hum. Neurosci*., vol. 9, pp. 1–10, 2015.

[11] J. van Driel, K. Richard Ridderinkhof, and M. X. Cohen, "Not all errors are alike: Theta and alpha EEG dynamics relate to differences in error-processing dynamics," *J. Neurosci*., vol. 32, 2012.

[12] M. Völker et al., "The dynamics of error processing in the human brain as reflected by high-gamma activity in noninvasive and intracranial EEG," *Neuroimage*, vol. 173, pp. 564–579, 2018.

[13] J. Omedes, I. Iturrate, L. Montesano, and J. Minguez, "Using frequency-domain features for the generalization of EEG error-related potentials among different tasks," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 5263–5266, 2013.

[14] M. Simões *et al.*, "A novel biomarker of compensatory recruitment of face emotional imagery networks in autism spectrum disorder," *Front. Neurosci.*, vol. 12, no. 791, 2018.

[15] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cogn. Emot*., vol. 24, no. 8, pp. 1377–1388, 2010.

[16] O. Dimigen, "Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments," *Neuroimage*, 2019.

[17] J. Li et al., "Feature Selection: A Data Perspective," *ACM Comput. Surv.*, vol. 50, no. 6, p. 94:1-94:45, 2017.

[18] Q. S. Xu and Y. Z. Liang, "Monte Carlo cross validation," *Chemom. Intell. Lab. Syst*., vol. 56, no. 1, pp. 1–11, 2001.

[19] Y. M. Huang and S. X. Du, "Weighted support vector machine for classification with uneven training class sizes," *Int. Conf. Mach. Learn. Cybern. ICMLC 2005*, vol. 18, no. 21, pp. 4365–4369, 2005.

[20] S. K. Kim, E. A. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning-Using error-related potentials for real world human-robot interaction," *Sci. Rep*., vol. 7, 2017.