Single Trial Detection of EEG Error Potentials: A Tool for Increasing BCI transmission rates

Benjamin Blankertz[1], Christin Schäfer[1], Guido Dornhege[1], and Gabriel Curio[2]

[1] Fraunhofer-FIRST.IDA, Kekuléstr. 7, D-12 489 Berlin, Germany
{blanker, schaefer, dornhege}@first.fraunhofer.de
[2] Neurophysics Group, Dept. of Neurology, Klinikum Benjamin Franklin, Freie Universitaet Berlin, Hindenburgdamm 30, D-12 200 Berlin, Germany
curio@zedat.fu-berlin.de

Abstract. It is a well-known finding in human psychophysics that a subject's recognition of having committed a response error is accompagnied by specific EEG variations that can easily be observed in averaged event-related potentials (ERP). Here, we present a pattern recognition approach that allows for a robust single trial detection of this error potential from multichannel EEG signals. By designing classifiers that are capable of bounding false positives (FP), which would classify correct responses as errors, we achieve performance characteristics that make this method appealing for response-verification or even response-correction in EEG-based communication, i.e., brain-computer interfacing (BCI). This method provides a substantial improvement over the choice of a simple amplitude threshold criterion, as it had been utilized earlier for single trial detection of error potentials.

1 Introduction

A brain-computer interface (BCI) is a system capable of translating a subject's intention as represented by brain signals into a technical control signal. In EEG-based communication such brain signals are measured non-invasively by means of a multichannel electroencephalogram (EEG). Although research in this area has made great progress in the last years ([1], [2], [3]) there is still a significant lack of accuracy and information transfer speed. Furthermore, all research groups report a considerable intersubject variability: while some subjects learn to operate quite well with the respective BCI system, others have to face error rates that make its usage uncomfortable.

An elegant approach to overcome the problem of low classification accuracy would be a response checking mechanism that is based on the subject's brain signals themselves. It is well-known from the neuroscience literature [4] that a subject's recognition of having committed a response error evokes specific EEG variations, see Section 1.1. The present paper focuses on the ability to pick up this error potential in single trials to provide a tool for response checking in EEG-based communication. Clearly from such a tool those persons would benefit most, who otherwise can only reach a modest BCI control because of a substantial fraction of classification errors.

1.1 Neurophysiological background

The ERP after an error trial is characterized by two components: a negative wave called error negativity (N_E) and a following broader positive peak labeled as error positivity (P_E), [4]. Recent studies revealed that the P_E is more specific to errors while the N_E can also be observed in correct trials, cf. [4]. Although both amplitude and latency depend on the specific task, the N_E occurs delayed and less intense in correct trials than in error trials. The N_E has a fronto-central maximum, the P_E a centro-parietal maximum. At present, there is not yet a final consensus about the underlying cognitive functions. N_E seems to reflect some kind of comparison process. Due to the localization of the origin in the *anterior cingulate cortex* [4] it might be an emotional and/or attentional component. In contrast, P_E seems to be connected to conscious error detection [5].

The one study reporting error potentials in a BCI context [6] is solely based on P_E, but the neurophysiological findings indicate that also the N_E component might be useful to some degree.

2 Aims and Methods

2.1 Response verification for BCIs

Most BCI systems allow the user to select one out of several choices. At present there are often only two classes ([1], [3], [7]), but there are also multi-class BCIs, e.g. [2], [8]. For such BCI systems an error detection algorithm can provide a useful add-on. If there are just two classes, detecting an error allows to correct the BCI classification (response-correction), for more than two classes at least wrong classifications can be rejected (response-verification).

While the idea of correcting BCI misclassifications is tempting, one has to be careful: as the detection method will not work perfectly, some correct BCI classifications can potentially get ›corrected‹ towards a wrong choice. If the proportion of such miscorrections is non-negligible the subject will become irritated. Even if the mechanisms works well enough to theoretically increase the information transfer rate it may be unfavorable in a psychological sense. This implies the need to strictly bound the rate of false positives (FP-rate: the fraction of acual correct trials which is misclassified as an error), where we use the nomenclature that ›positive‹ events are the ones that are to be detected, i.e., trials where the BCI algorithm missed to detect the subject's intention.

2.2 Experiments

At this stage we investigated EEG data from an attention test, while BCI feedback experiments are planned for the next step. Eight healthy subjects took part in one EEG measurement each, in which they had to perform a variant of the ›d2-test‹, [9]. After a computer screen displayed visual stimuli, subjects had to respond to targets by pressing a key with the right index finger and to non-targets with the left index finger. Targets in the d2-test are compound symbols consisting of the letter ›d‹ and exactly two horizontal bars that may occur in four possible positions each. Non-targets either show the letter ›b‹ and an arbitrary number of bars (0–4) or the letter ›d‹ and a number of bars that differs from two, see Fig. 1 for some examples. After the subject's keystroke the reaction time was displayed on the screen, either in green if the response was correct, or in red if it was erroneous. The next trial began 1.5±0.25 s later. A summary of the experiments with reaction times and error rates is given in Table 2.

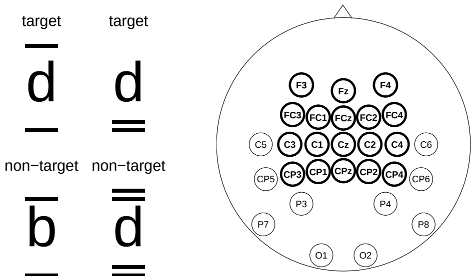

Fig. 1. Examples of targets and non-targets in the d2-test on the left. The electrode montage is shown on the right.

Table 2. Summary of the experiments with ›d2-paradigm‹.

subject [code]	reac. [ms]	trials [#]	err. [#]	err. [%]
aa	539	977	101	10.3
ab	434	893	41	4.6
ac	556	896	8	0.9
ad	477	893	46	5.2
ah	551	884	19	2.1
ae	504	894	39	4.4
af	497	893	35	3.9
ag	529	892	42	4.7

Brain activity was recorded with 28 Ag/AgCl electrodes, cf. Fig. 1, referenced to nasion, with a broad band-pass filter. Besides EEG we recorded a horizontal and vertical electrooculogram (EOG). In an event channel timing and types of stimuli and keystrokes were stored along with the EEG signal.

No (!) trials were rejected due to artifacts, but all trials in which the subject hit two keys (simultaneously or sequentially) were sorted out.

2.3 A Pattern matching method

The two components that are observed in the EEG related to errors are slow cortical potentials (SCPs). In [7] we presented a successful method for classifying single trial EEG based on SCPs which can be used here with some appropriate modifications. In [7] the key for good results was the combination of high-dimensional features and robust learning machines for classification. The features we use in this study are subsampled versions of the relevant channels (marked labels in Fig. 1). Subsampling from 100 Hz to 20 Hz was done by calculating the mean of consecutive 5-tupel of data points.

The advantage of this preprocessing for ERP analysis is that the resulting classification problem has a simple structure, though being high-dimensional. The distributions of the feature vectors of each class can be modelled by a normal distribution, the mean of which is the feature of the ERP of the corresponding condition, cf. [10]. The covariance matrix is determined by non-task related brain activity. As this is approximately the same for both classes, the classification has to separate two normal distributions with equal covariance matrices. The Bayes-optimal classifier for this task is the Fisher Discriminant (FD). Dealing with high noise cases in a high-dimensional space typically requires regularization in order to obtain stable estimates of the covariance matrices.

But in the present situation we are looking for the classifier which is optimal under the constraint that the FP-rate attains a predefined value (on the training set). For linear classifiers in a separating hyperplane formulation ($w^\top x + b = 0$) this can be accomplished by adjusting the threshold b. This procedure is indeed optimal in conjunction with the FD under the forementioned assumptions, which can be seen using the Neyman-Pearson Lemma (we thank Marina Meila for this remark), or by the following direct proof.

Let $(X, Y) \in \mathbb{R}^n \times \{N, P\}$ be random variables such that the conditionals $P(X \mid Y = N)$ and $P(X \mid Y = P)$ are $\mathcal{N}(\mu_N, \Sigma)$ resp. $\mathcal{N}(\mu_P, \Sigma)$ distributed (i.e., normal distribution with mean $\mu_\sim$ and covariance matrix Σ). The problem is to maximize $P(w^\top X + b > 0 \mid Y = P)$ subject to $P(w^\top X + b > 0 \mid Y = N) = \delta$ for some fixed $\delta \in (0, 1)$. Denoting the distribution function of $\mathcal{N}(\mu, \sigma)$ by $F_{\mathcal{N}(\mu,\sigma)}$ we have $\delta = P(w^\top X + b > 0 \mid Y = N) = 1 - F_{\mathcal{N}(w^\top \mu_N + b,\, w^\top \Sigma w)}(0) = 1 - F_{\mathcal{N}(0,1)}(-{}^{(w^\top \mu_N + b)}/_{\sqrt{w^\top \Sigma w}})$. Hence for $\beta := F^{-1}_{\mathcal{N}(0,1)}(1 - \delta)$ we obtain the threshold $b = -\beta\sqrt{w^\top \Sigma w} - w^\top \mu_N$ in dependence from the optimal w. Substituting this term for b one can see that $P(w^\top X + b > 0 \mid Y = P) = \cdots = \int_0^\infty {}^1/_{\sqrt{2\pi}} \exp(-{}^1/_2[t + \beta - w^\top(\mu_P - \mu_N)(w^\top \Sigma w)^{-1/2}]^2)dt$, and this expression is maximized if $w^\top(\mu_P - \mu_N)(w^\top \Sigma w)^{-1/2}$ is maximized. Since the last term is the square root of the Rayleigh coefficient we get the same w as from FD (qed).

2.4 Amplitude threshold criterion.

For comparision we also implemented the absolute amplitude criterion, the only algorithm for the detection of the error potential in a BCI context published so far, [6]. In this method all trials for which the amplitude of the Cz channel averaged over a predefined time period exceeds a predefined threshold are classified as errors. To make the comparison fair we extracted amplitude peaks for all channels that were used in our method, and the optimal hyperplane threshold was determined by the same learning algorithm.

3 Results

In the average difference potential ›miss-minus-hit‹ the two discussed components N_E and P_E emerge very pronounced, cf. the ERP for subject *ad* in Fig. 3. The other ERPs show the same characteristics. For all subjects an early negative and a later positive component is clearly observable. The main intersubject differences concern the latency of the two components and the fall-off after the P_E. For three subjects (*ab*, *ac*, *ad*) the peak of the N_E shows up very early about 10–30 ms, but even in all other subjects it is not later than 60 ms. That peak is too early to be a reaction to the visual feedback at 0 ms. This observation is

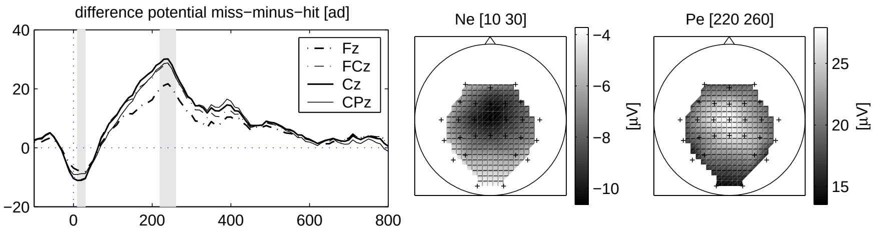

Fig. 3. Average miss-minus-hit EEG-traces at electrodes along the vertex for subject *ad*. Regions of N_E and P_E are shaded and scalp topographies for that regions are shown in the two subplots at the right.

in agreement with the subjects' reports that in erroneous trials they often knew they were going to make a mistake while they initiated the movement but they could not withhold anymore.

As was pointed out in Section 2.1 a special demand on the error detection is the ability to strictly bound false positive classifications. The first question that arises here is: how well does the bound for FPs that we enforce on the training set carry over to the test set. Training a classifier for FP=2% resulted for all 8 data sets in FP-rates between 2.3% and 3.5%. In order to make the comparison between different data sets and parameter choices easier, we used a cross-validation in which (the thresholds of the linear) classifiers were adapted after training to obtain a predefined FP-rate so that the detection performance is reflected by the FN-value only. In Fig. 4 our pattern matching method was evaluated for FP-bounds at 1%, 2% and 3%. Subjects were sorted according to the FN-rate at FP=2%: the performance is ›very good‹ (FN$\leqslant$11%) for 4 subjects, ›good‹ (FN$\leqslant$22%) for 3 subjects and ›not good‹ only for subject *aa*. White bars show the corresponding error rates for the amplitude criteron.

To assess the potential value of the proposed error detection method for improving BCI transmission rates, we take a look at an example. A BCI accuracy of p =0.85 in a two class decision ($N = 2$) has a theoretical information of $\log_2 N + p \log_2 p + (1-p)\log_2(1-p/N-1) = 0.39$ bits per selection. Moderately assuming that the error-correction method works with 20% FN at 3% FP this can be increased by more than 75% to 0.69 bits, where the accuracy of the improved system is calculated by $p \cdot (1 - \mathrm{FP}) + (1 - p) \cdot (1 - \mathrm{FN}) = 0.94$. Obviously the gain gets less the higher the original BCI accuracy is. This trade-off is depicted

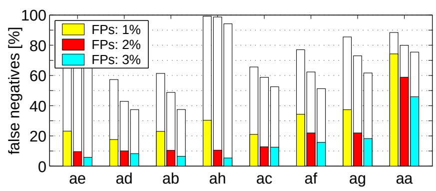

Fig. 4. Rate of FNs for detection at 300 ms with fixed FP-rate. White bars show the corresponding FN-rates for the amplitude criteron, cf. Section 2.4.

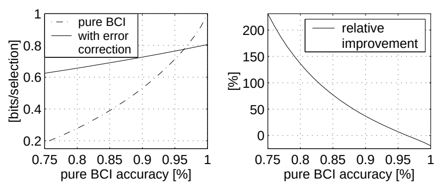

Fig. 5. The plot shows the improvement of the BCI accuracy by the proposed error correction method, when an FP-rate of 3% and an FN-rate of 20% is assumed.

in Fig. 5, note that with the assumed parameters an error correction approach is useful as long as the pure BCI accuracy is lower than 96%.

Ocular artifacts. In [6] it was reported that the end of many trials contained eyeblinks, an effect that is also present in our data. So is has to be made sure that classification success is not based on ocular contamination of the EEG. Therefore we also tried a classification of errors based on the EOG signals in the same way as we did EEG-based classification. The resulting FN-rate was >95% for most subjects, and only for subject *ae* it was 77% which is still more than 8 times higher than in EEG-based detection with 9%.

4 Discussion

Our pattern recognition approach to single trial detection of the error potential provided a substantial improvement in comparison to a simple amplitude threshold criterion, and the expected gain for BCI classification is promising. The important next step is to conduct BCI experiments with real-time feedback to check whether the error potentials are of the same type in such a scenario.

Acknowledgement. We would like to thank Sebastian Mika and Klaus-Robert Müller for valuable discussions. The studies were supported by a grant of the *Bundesministerium für Bildung und Forschung* (BMBF), FKZ 01IBB02A.

References

1. Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., Flor, H.: A spelling device for the paralysed. Nature **398** (1999) 297–298
2. Obermaier, B., Neuper, C., Guger, C., Pfurtscheller, G.: Information transfer rate in a five-classes brain-computer interface. IEEE Trans. Neural Sys. Rehab. Eng. **9** (2001) 283–288
3. McFarland, D.J., Miner, L.A., Vaughan, T.M., Wolpaw, J.R.: Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr. **12** (2000) 177–186
4. Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J.: ERP components on reaction errors and their functional significance: a tutorial. Biol. Psychol. **51** (2000) 87–107
5. Nieuwenhuis, S., Ridderinkhof, K., Blom, J., Band, G., Kok, A.: Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. Psychophysiology **38** (2001) 752–760
6. Schalk, G., Wolpaw, J.R., McFarland, D.J., Pfurtscheller, G.: EEG-based communication: presence of an error potential. Clin. Neurophysiol. **111** (2000) 2138–2144
7. Blankertz, B., Curio, G., Müller, K.R.: Classifying single trial EEG: Towards brain computer interfacing. In Diettrich, T.G., Becker, S., Ghahramani, Z., eds.: Advances in Neural Inf. Proc. Systems (NIPS 01). Volume 14. (2002) to appear.
8. Donchin, E., Spencer, K.M., Wijesinghe, R.: The mental prosthesis: Assessing the speed of a P300-based BCI. IEEE Trans. Rehab. Eng. **8** (2000) 174–179
9. Brickenkamp, R., Zillmer, E.: D2 Test of Attention. Hogrefe & Huber (1998)
10. Kohlmorgen, J., Blankertz, B.: A simple generative model for single-trial EEG classification. In: Artifical Neural Networks – ICANN. (2002)