Classification of error-related potentials from single-trial EEG in association with executed and imagined movements

a feature and classifier investigation

Usama, Nayab; Leerskov, Kasper; Niazi, Imran Khan; Dremstrup, Kim; Jochumsen, Mads

Published in:
Medical & Biological Engineering & Computing

DOI (link to publication from Publisher):
10.1007/s11517-020-02253-2

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Classification of error-related potentials from single-trial EEG in association with executed and imagined movements: a feature and classifier investigation

Nayab Usama[1], Kasper Kunz-Leerskov[1], Imran Khan Niazi[1,2,3], Kim Dremstrup[1], Mads Jochumsen[1§]

[1]Department of Health Sciences and Technology, Aalborg University, Denmark

[2]Health & Rehabilitation Research Institute, Auckland University of Technology, New Zealand

[3]Center for Chiropractic Research, New Zealand College of Chiropractic, New Zealand

Email: nu@hst.aau.dk, kkl@hst.aau.dk, imrankn@hst.aau.dk, kdn@hst.aau.dk, mj@hst.aau.dk

[§] Corresponding author
Mads Jochumsen, PhD
Department of Health Science and Technology
Fredrik Bajers Vej 7D, 9220 Aalborg, Denmark
Telephone: +45 9940 3789
Email: mj@hst.aau.dk

- Total number of words of the manuscript: 7506
- Number of words of the abstract: 199
- The number of figures: 4
- The number of tables: 2

Abstract

Error-related potentials (ErrPs) have been proposed for designing adaptive brain-computer interfaces (BCIs). Therefore, ErrPs must be decoded. The aim of this study was to evaluate ErrP decoding using combinations of different feature types and classifiers in BCI paradigms involving motor execution (ME) and imagination (MI). Fifteen healthy subjects performed 510 (ME) and 390 (MI) trials of right/left wrist extensions and foot dorsiflexions. Sham BCI feedback was delivered with an accuracy of 80% (ME) and 70% (MI). Continuous EEG was recorded and divided into ErrP and NonErrP epochs. Temporal, spectral, discrete wavelet transform (DWT) marginals, and template matching features were extracted, and all combinations of feature types were classified using linear discriminant analysis, support vector machine, and random forest classifiers. ErrPs were elicited for both ME and MI paradigms, and the average classification accuracies were significantly higher than the chance level. The highest average classification accuracy was obtained using temporal features and a combination of temporal+DWT features classified with random forest; 89±9% and 83±9% for ME and MI, respectively. These results generally indicate that temporal features should be used when detecting ErrPs, but there is great inter-subject variability, which means that user-specific features should be derived to maximize the performance.

Keywords Error-related potentials, brain-computer interface

1 Introduction

A brain-computer interface (BCI) system provides a communication pathway between a human brain and a computer by translating the user's brain signals into computer commands [1]. In the past few decades, there have been many advancements in the application of BCIs such as rehabilitation of motor-impaired people [2], induction of neuroplasticity [3], speech synthesis or selection of characters on a screen [4]. Despite these advancements, real-life BCI applications are still limited. One of the main constraints is the constant need for recalibration of the classifier in the BCI to maintain adequate performance, which is caused by e.g. changes in the electrode impedance, amplifier or environmental noise, fatigue or the user's attention level [5,6]. A system that makes an online update of the classifier on its own is termed an adaptive BCI [7]. In an adaptive BCI, the classifier can be updated either after every classification run or after a pre-determined number of trials, but it is imperative that the update is done correctly using the appropriate data. One possible strategy can be to use physiologically elicited signals in response to an error made by a BCI system [6,7]. When a user gets feedback, which differs, from the intended action, the detectable error-related negativity (ERN) or error-related potential (ErrP) is elicited in the fronto-central region of the brain [8]. The ErrP consists of two main components: an error related negativity (ERN or Ne) and an error related positivity (ERP or Pe), which occurs within 50-800 ms after the realization of erroneous feedback [9,10]. It has been shown that various factors can modulate the ErrP; these include the user's attention level [11], feedback modalities [12], the difference in the executed tasks [13, 15], or frequency of erroneous tasks [14]. Previous studies reported the varying amplitudes and latencies of the Ne and Pe peaks of the ErrPs occurring approximately within the first 350 ms [10], 550 ms [12] and 650 ms [15] time interval after the feedback presentation. Since the ErrPs are affected by various factors and are quite variable, single-trial detection of ErrPs is challenging. However, it has been shown that an adaptive BCI can be constructed using ErrPs [16-21]. Successful adaptation of the classifier in the BCI requires that the ErrPs are correctly detected. In the previous studies, several different types of features and classifiers have been employed for the classification of ErrPs [21,23,24]. According to a review on ErrPs classification [11], the two most commonly used features were spectral/frequency band power [26-33] in the theta (4-7 Hz), mu (8-12 Hz), and beta (13-30 Hz) range and temporal/mean amplitude features [27,29,30,34,35]. Generally, the features have been extracted within the first 1000 ms after the presentation of the feedback. The most commonly used classifiers for ErrP classification include linear discriminant analysis (LDA) [10,27,28,35,36], support vector machine (SVM) [22,37,38] and Gaussian filter [14,34,39]. Blankertz et al. classified ErrPs elicited during a motor task for left versus right finger movement using LDA [10]. With a predefined rate of false positives at 2%, more than 85% of the ErrPs were correctly classified. Likewise, Kreilinger et al. [27] also employed an LDA classifier for the ErrP classification during a BCI-driven car game. In [25], Spüler et al. classified the ErrPs elicited during a speller task by implementing the LDA, stepwise LDA, and SVM classifiers, and the results showed that the SVM was best suited for ErrPs detection. Moreover, Chavarriaga et al. [14] and Ferrez et al. [39] used a Gaussian classifier for the classification of ErrPs and reported an average recognition rate of correct and erroneous trials between approximately 65-85%. Hence, several different configurations have been outlined in the literature regarding the choice of features or classifiers for ErrP classification, but there are no clear guidelines about how ErrPs are best detected. In these studies, different control signals have been used to drive the BCI, such as visual evoked potentials [41], P300 [24,40], and motor-based signals [42]. The latter are important in BCI applications concerning neurorehabilitation where neuroplasticity is induced by using movement-related cortical potentials (MRCPs) to trigger neuromuscular electrical stimulation [43,44]. The MRCP can be elicited from either an executed or imagined movement [45,46], but there is a risk that the motor and reafferent potentials of the MRCP would coincide with the ErrP; this can potentially be solved by delaying the feedback to avoid an overlap between the MRCP components and the ErrP [15]. With an MRCP-based BCI in mind, the aims of this study were to: 1) try to isolate the ErrP from the components of the MRCP, 2) quantify potential morphological differences between ErrPs elicited after executed and imagined movements of the upper and lower limbs, and 3) to identify the optimal combination of features and classifier if one such exists. This was tested with a sham BCI (providing pre-fixed mock feedback) that pretended to decode executed and imagined dorsiflexions of the ankle joint and wrist extensions. Temporal, spectral, time-scale, and template matching features were classified using LDA, SVM, and random forest (RF) classifiers. This study contributes to the existing literature with a systematic investigation of features and classifiers for ErrP classification in MRCP-based

BCIs and validates the findings of the ErrP morphology when elicited after the decoding of executed and imaginary movements.

2 Methods

2.1 Subjects

Fifteen healthy subjects (nine males, six females: 25±10 years old) participated in this study. Each subject gave their informed consent before the participation, and the local ethical committee (N-20130081) approved the procedures. None of the subjects had any prior experience with BCI experiments.

2.2 Data recording

64 channels EEG were recorded with a sampling rate of 1200 Hz using active electrodes (g.HIamp G.Tec, Graz, Austria). The electrodes were placed according to the International 10-20 international. A linked ear reference was used, and the ground electrode was located at AFz. The impedance of all electrodes was kept below 5kΩ during the experiments. By using a NI DAQ data acquisition system, an external trigger was sent to the EEG amplifier from custom-made MATLAB (MathWorks, 2015) interface software. The trigger pulses were recorded to synchronize the continuous EEG with the presentation of feedback for dividing it into ErrP ('Incorrect' in Figure 1) and NonErrP ('Correct' in Figure 1) epochs.

2.3 Experimental Details

During the experiment, the subjects were seated in a comfortable chair facing a computer screen. The subjects were asked to avoid eye blinks or any unnecessary movement and keep their gaze at the screen during the movement and feedback-monitoring phase. Figure 1 shows the timeline for the ME and MI experiments, which were scheduled at two separate days, where ME experiments were performed on the first day. Each repetition of a task (either ME or MI) started with an idle phase of five seconds, during which the subjects could blink and relax. Subsequently, a preparation phase was started, which lasted for three seconds during which a text was displayed on the screen: '*Prepare for the movement'*. Next, during the ME or MI phase, a picture of a hand (pointing towards the right or left direction) or a foot was shown in the center of the computer screen to indicate movement of the respective hand or foot. The order of the movement types was randomized for each subject with an equal number of repetitions for each movement type. For the right or left hand, a wrist extension, and for the foot a dorsiflexion movement was performed (ME) or imagined (MI). Instead of recording the movements for both feet, it was restricted to only one side, either the left or the right foot based on the subject's choice and convenience (each foot has approximately the same representation in the motor area [47]). Eight subjects preferred the right foot for ME and MI. Subjects were asked to execute or imagine the movement as soon as they saw the picture on the screen. Contrary to the conventional BCIs [10-12], the feedback was provided with a delay of three seconds in the form of a green tick mark or a red-colored cross sign. The delay of three seconds was chosen to try to better isolate the ErrP from the reafferent potential of the MRCP [48] and the event-related synchronization [49], which are observed up to two seconds after the movement onset, and to obtain a good synchronization of the trials when the subjects were cued to what time the feedback would be presented. Before the experiment, it was conveyed to the subjects that the system was decoding the intended movement from their brain signals, and the feedback type solely depended on their brain signals during the movement. The subjects' brain activity in response to the correct or erroneous feedback was used for discrimination between NonErrPs and ErrPs. To give a notion of the variable performance across the movement types, the ratio of erroneous feedback was set different for ME and MI paradigms. The feedback ratio was set at 80:20 for correct and erroneous feedback in the case of the ME and 70:30 for the MI paradigm based on previous findings of ME and MI detection using MRCPs [24]. To retrieve the approximately same number of ErrPs across both paradigms, 510 movements (170 movements each for the right wrist, left wrist and/foot) were performed for the ME paradigm and 390 (130 movements for each type of the limb) for the MI paradigm. The experiment for the ME paradigm was completed in 17 blocks, where each run consisted of the 30 movement repetitions. For the MI paradigm, the experiment was completed in 13 blocks with the same number of movement repetitions in each trial. After the completion of each run, a break was given to the subjects until they were ready to start the experiment again. The experiments were completed in approximately 180 minutes for the ME experiment and in 120 minutes for the MI experiment.

2.4 Signal Processing

2.4.1 Pre-Processing

Following the experiment, 37 channels of EEG (AF3-4, Fz, F1-6, FCz, FC1-6, Cz, C1-6, CPz, CP1-6, Pz, P1-6) were bandpass filtered between 0.5-30 Hz using an 8th order zero phase-shift Butterworth filter. These channels were chosen, as they have previously been found to be relevant when investigating ErrPs [50] and to reduce the dimensionality of the feature vector. Subsequently, bad channels and epochs were excluded from further analysis. Channels were excluded if the respective channel had a mean amplitude more than three standard deviations above the overall mean amplitude of the used channels. Epochs were excluded if the any samples in the epoch had an absolute value higher than 150 µV. These rejection thresholds were based on manual data exploration. Next, data was separated into 800 ms epochs, starting from the time instant of the feedback presentation (in Figure 1 this was when time = 3 s) balancing the number of ErrPs and NonErrPs. As the number of ErrPs was lower than the NonErrPs, NonErrPs were randomly chosen from those available, to match the number of ErrPs after the rejection of corrupted trials, such that an equal number of uncorrupted ErrPs and NonErrPs were used in the further analyses.

2.4.2 Feature extraction

After the bad channels and epochs were rejected, four types of features were extracted from the remaining channels and epochs. Features included temporal features (Temp), spectral power (Spec), discrete wavelet transform (DWT) marginals, and template matching (Match). Temp, Spec and DWT features were calculated in windows of 100 ms with no overlap. The Temp features were calculated as the arithmetic mean amplitude of the window. Spec features were estimated by integrating a periodogram, constructed using a Hamming window, over given frequencies bins. In this study, frequencies considered were 0-30 Hz in 5 Hz bins, with no overlap. DWT marginals were calculated based on the algorithms introduced in [51], briefly described: DWT coefficients where calculated at N levels of decomposition, where N is given as the binary logarithm of the length of the signal (N = 6, for a signal of l20 samples), using a Daubechies 4 mother wavelet. N DWT marginals were calculated for every window, by dividing the sum of the modulus of DWT coefficients at the nth level by the sum of the modulus of all DWT coefficients. All marginals were subsequently used as features. Match features were calculated as the autocorrelation between a given epoch and the average ErrP template for a given channel. The template was derived for each subject and paradigm individually and consisted of all ErrP trials. A total of 92 features per channel were extracted, giving a total of 3312-3404 features, dependent on whether channels were excluded. To investigate the effect of multiple combinations of features on the classification accuracies of ErrPs, Temp, Spec, DWT, and Match features were investigated in all possible combinations i.e. 15 distinct different combinations.

2.4.3 Feature reduction

Before feature reduction, features were divided into training and test sets, to accommodate a 5-fold cross-validation classification procedure. The feature reduction was then employed on the training set, reducing the features using a PCA, keeping PCA dimensions equivalent to 95 % of the variance in the feature set. The resulting transformation matrix was then applied to the test set. This procedure was done separately for all the feature combinations and all training sets. On average, 13.4 ± 3.8 (mean ± standard deviation) features for the ME paradigm and 14.6 ± 5.1 features for the MI paradigm remained after the feature reduction.

2.4.4 Classification

After feature reduction LDA, SVM, and RF classifiers were trained for the classification of the ErrPs and NonErrPs by using the combinations of the Temp, Spec, DWT, and Match features. A linear SVM was employed, and the RF was trained using 128 trees, as no further improvement in classification accuracy is obtained by adding more trees [52]. To calculate the classification accuracy of each classifier across each set of features, the mean accuracy of the 5-fold cross-validation was reported.

2.5 Morphological Analysis

For morphological analysis of ErrPs and NonErrPs, the average was calculated across epochs from six seconds prior the onset of the presentation of the visual feedback (output of the sham BCI) and 0.8 seconds after the feedback presentation for all subjects at channel FCz [37] for the ME and MI paradigm. The grand averages across subjects are shown in Figure 4. To analyze the morphology of ErrPs and NonErrPs P1, N2 and P3 peaks were

extracted manually based on the peaks' polarity. Then P1-N2 and N2-P3 peak-peak voltage (p-p) and latencies of peaks with respect to the feedback presentation time (time=0 ms) were calculated (see Figure 4).

2.6 Statistical Analysis

Normal distribution of data was checked by using the Shapiro-Wilk test before doing the statistical analysis. All the statistical tests were implemented by using IBM SPSS® software. To analyze the impact of feedback or movement type on the signal morphology two 2-way repeated-measures analysis of variance (ANOVA) were performed on the p-p voltage of the P1-N2 and N2-P3 peaks. The two factors were feedback (2 levels: Error, No-Error) and movement paradigm (2 levels: ME and MI). Moreover, three 2-way repeated measures ANOVA were performed on the latencies of P1, N2 and P3 peaks using the same factors. Furthermore, the effect of feature type, movement paradigm, and choice of the classifier on the classification accuracies was analyzed using a 3-way repeated measure ANOVA. The three factors were feature type (15 levels: all possible combinations of the feature types), classifier (3 levels: LDA, SVM, and RF) and movement paradigm (2 levels: ME and MI). For significant test statistics, a post hoc pairwise comparison with Tukey's correction was employed. The p-value<0.05 was considered statistically significant.

3 Results

On average, 0.8 channels (range: 0-1, maximum one channel was removed) and 54.3 ± 94.8 epochs were excluded from the ME paradigm, and 0.9 channels (range: 0-1) and 5.2 ± 9.6 epochs were excluded from the MI paradigm.

3.1 Classification results

The performance of classifiers was evaluated by calculating the classification accuracy (correct predictions to the total number of samples). The boxplots of the classification accuracies using all feature combinations for LDA, SVM, and RF classifiers across ME and MI paradigms are presented in Table 2 and Figures 2 and 3, respectively. The simple Temp and Temp+DWT features were associated with the highest classification accuracies for all three classifiers. On average, the highest classification accuracies for RF, SVM and LDA classifier across the ME and MI paradigm were observed to be 89±9%, 87±8%, 85±7% and 83±9%, 80±7%, 79±6%, respectively. Generally, there is great inter-subject variability. The results of the 3-way ANOVA showed a significant main effect on classification accuracies based on the choice of features type ($F_{(14,196)}$=20.92; p<0.001), choice of classifiers ($F_{(2,28)}$=52.5; p<0.001) and choice of a paradigm ($F_{(1,14)}$=14.1; p<0.005). The post-hoc test revealed that the Temp and Temp+DWT features were associated with significantly higher classification accuracies compared to the other feature combinations (p<0.01 for all comparisons). The classification accuracies associated with the SVM were significantly lower as compared to LDA and RF (p<0.001), and significantly classification accuracies were obtained for the ME compared to the MI (p=0.02). There was no significant interaction between classifiers*paradigms ($F_{(2,28)}$=0.118; p=0.889), features*paradigms ($F_{(14,196)}$=0.221; p=0.999) or classifiers*features*paradigms ($F_{(28,392)}$=0.184; p=1) but a significant interaction was observed between classifiers*features ($F_{(28,392)}$=42.423, p<0.001).

Table 2 Mean±standard deviation across subjects for the classification of ErrP vs NonErrP in the motor execution and motor imagination paradigm for each feature combination. T: Temporal, S: Spectral, M: Template matching, W: Marginals of the discrete wavelet transform (DWT). The highest classification accuracies were obtained using the temporal and DWT features. Generally, the highest accuraices were obtained using the RF classifier.

	Motor Execution		
Features	*LDA*	*SVM*	*RF*
T	85± 7	87± 8	89± 9
S	72± 9	63± 13	85± 10
M	72± 9	56± 9	82± 12
W	72± 15	74± 15	78± 19
TS	72± 9	64± 13	85± 9
TM	72± 9	57± 9	83± 11
TW	85± 7	87± 8	89± 8

SM	72± 9	56± 9	82± 11
SW	72± 9	63± 13	85± 10
WM	72± 9	57± 9	82± 11
TSM	72± 9	56± 9	82± 11
TSW	72± 9	62± 12	85± 10
TWM	72± 9	57± 9	82± 11
SWM	72± 9	55± 9	82± 11
TSWM	72± 9	55± 9	82± 11
	Motor Imagination		
T	79± 6	80± 7	83± 9
S	70± 6	57± 12	78± 13
M	68± 7	53± 5	77± 12
W	65± 14	66± 15	66± 20
TS	69± 6	55± 13	77± 13
TM	68± 7	53± 5	77± 12
TW	79± 6	80± 7	83± 9
SM	68± 7	54± 6	78± 12
SW	70± 6	58± 11	78± 14
WM	68± 7	53± 5	78± 12
TSM	68± 7	54± 6	77± 12
TSW	69± 6	57± 12	78± 13
TWM	68± 7	53± 5	78± 12
SWM	68± 7	54± 6	77± 12
TSWM	68± 7	54± 6	78± 12

3.2 Morphology results

The grand averages and variability of ErrPs and NonErrPs along with the difference (ErrP minus NonErrP) for ME and MI paradigms are shown in Figure 4a and 4b, respectively. Moreover, the difference plots for ME and MI are plotted in Figure 4c. The mean and standard deviation for the P1-N2, N2-P3 p-p voltage, and P1, N2, P3 peak latencies for the ME and MI paradigm are presented in Table 1. It is evident from Figure 4 and Table 1 that on average P1, N2, and P3 were elicited approximately within the first 100 ms, 200 ms, and 300 ms after the presentation of the feedback. However, there is generally high inter-subject variability. The N1-P2 and N2-P3 p-p voltages for the ErrPs were observed to be approximately in a range of 5-16 µV for the ME and MI paradigms while the range for the NonErrPs was 2-12 µV. For the P1-N2 amplitude, a 2-way ANOVA revealed a significant effect of feedback on the p-p voltage ($F_{(1,14)}$=30.446; $p<0.001$) with higher amplitudes for the ErrPs as compared to NonErrPs ($p<0.001$). There was no effect of movement paradigm on the P1-N2 amplitude ($F_{(1,14)}$=1.104; $p=0.311$). For the N2-P3 amplitude, there was a significant effect of feedback ($F_{(1,14)}$=89.036; $p<0.001$) with higher amplitudes for ErrPs compared to NonErrPs ($p<0.001$). There was no effect of the movement paradigm on the N2-P3 amplitude ($F_{(1,14)}$=4.388; $p=0.055$). There were no significant differences between the latencies of the P1, N2 and P3 peaks for the movement paradigm: ($F_{(1,14)}$=0.559, $p=0.853$), ($F_{(1,14)}$=1.471; $p=0.245$), ($F_{(1,14)}$=0.550; $p=0.471$) or for the feedback type: ($F_{(1,14)}$=0.045; $p=0.449$), ($F_{(1,14)}$=0.512; $p=0.486$), ($F_{(1,14)}$=0.378; $p=0.447$), respectively. In Figure 4c it can be seen that the latencies of the different peaks are similar for the ME and MI paradigms, and the peaks have slightly higher amplitudes for the ME paradigm.

Table 1 Mean±standard deviation of P1-N2, N2-P3 peak-peak (p-p) voltage and P1, N2, and P3 peak latencies across subjects (n=15). The peaks P1, N2 and P3 are elicited within 100ms - 360ms after the presentation of the visual feedback with a p-p voltages between 5-16 µV for ErrPs and 2-12 µV for Non-ErrPs. Similar amplitudes and latencies are obtained for motor execution and motor imagination.

	Motor Execution				
	P1-N2 p-p voltage	N2-P3 p-p voltage	P1 Latency	N2 Latency	P3 Latency

	(μV)	(μV)	(ms)	(ms)	(ms)
ErrP	6±8	16±6	103±16	192±22	283±27
NonErrP	4±4	12±6	132±26	236±28	361±29
	Motor Imagination				
ErrP	5±5	15±9	104±22	178±27	305±23
NonErrP	2±3	10±6	180±29	242±33	352±31
	Mean±std	Mean±std	Mean±std	Mean±std	Mean±std

4 Discussion

A systematic analysis was done to find an optimal set of features and classifier to discriminate between ErrPs and NonErrPs elicited through a sham BCI system designed for movement detection of right/left wrist extension and foot dorsiflexion for the ME and MI paradigms. Overall, in this study, the Temp and Temp+DWT features turned out to be the best choice of features as compared to any other combination of the spectral, DWT marginals or template matching features. While in terms of the classifiers, the RF classifier proved to be the best choice of a classifier as compared to the LDA and SVM classifier.

4.1 Choice of an optimal set of features and classifier

Previous studies that employed temporal features for the classification of ErrPs have reported the accuracies in the range of 72-85% [7,26,27,57]. In this study, the highest average classification accuracies achieved by using the temporal features reached up to 89±9%, 87±% and 85±7% for RF, SVM, and LDA classifier, respectively. The average and median of the classification accuracies are approximately similar to the classification accuracies that have been reported previously [7,19-21,26,27,57]. In this study, the simple mean features, and combination of temporal and DWT marginal features were better for discriminating between ErrPs and NonErrPs as compared to the spectral power, template matching or any other combination of features. Temporal features have generally been reported to be the best feature type for classifying ErrPs [31-33], although in an asynchronous detector spectral features were reported to be better than temporal features [32]. Spectral and template matching features could also be used to classify ErrPs which is in agreement with previous findings [32,58]. The difference between ErrPs and NonErrPs was subtle in the time domain, and the results indicate that it is not possible to discriminate between them in the frequency domain. Moreover, single-trials of ErrPs and NonErrPs may be too similar for template matching to be a viable feature for discriminating the two as suggested by the relatively large overlap of standard deviations of ErrPs and NonErrPs in Figure 4. However, when using RF, fairly high accuracies (above 80% and 70% for ME and MI, respectively) were obtained for all feature combinations. This may indicate that the structure of the feature space based on the current data set is more suited for an ensemble classifier compared to LDA and SVM, which was implemented with a linear kernel. In the literature [7,19,20,59,60], the classification accuracies by employing LDA, SVM, and RF classifiers are reported to be approximately in the range of 70-90%. Although due to differences in experimental paradigms, pre-processing, feature extraction and feature reduction methods, the classification results for this study were similar except for the SVM results. The low performance of the SVM in the current study is not consistent with a recent study [15]; however, the methods differ since they used a radial basis function, which may be more suitable for the classification of ErrPs compared to a linear kernel. It is also possible to optimize the width parameter of the radial basis function, which probably would improve the classification accuracies. It was possible to obtain classification accuracies that were significantly higher than the chance level (α=0.05) for all features types [61]. It should be noted that the epochs containing ErrPs and NonErrPs were extracted with a priori knowledge of where to locate the potential of interest, and that the analysis was performed offline. This information about the location of the ErrP may be available in an online system as well since the potential would be evoked immediately after the output of the BCI. However, if the temporal association between BCI output and the ErrP varies, this may reduce the performance of the ErrP decoder. There was considerable variability in the performance of the ErrP decoding across subjects, and several feature types and combinations could be used (at least for RF). This suggests that user-specific features should be extracted for each user, but it may not be necessary to use all feature types, which would reduce the computational load in an online decoding system. Although the performance of the decoder was significantly higher than the chance level, it can be further improved, potentially using other signal processing approaches. This could be to use other classifiers (e.g.

some that do not require much pre-processing and feature extraction) [62], or signal processing techniques such as blind source separation [63] or spatial filters to improve the signal of interest [64].

4.2 Error-related potential morphology

The BCI system designed in this study varied from the previous studies [10-12,15,53] in terms of the elicitation of the ErrP since the feedback was provided with a delay of three seconds instead of presenting it during or immediately after the task performance [10-12,15,16,53,54]. The reason for the delayed feedback presentation was to avoid an overlap between the reafferent potential of the MRCP [55] associated with the movement and the ErrP/NonErrP. Moreover, with this approach, we aimed at getting good synchronization of ErrP/NonErrP signals across trials to reduce the misalignment of trials with respect to the cue when averaging the epochs. It gave the subjects enough time to be prepared to receive the feedback. However, it appears that there was a negative expectancy potential preceding the presentation of the feedback and the rebound of this potential may overlap with the ErrP/NonErrP [56]. Therefore, this approach for isolating the ErrP/NonErrP was not optimal. Moreover, from a BCI perspective, it will not be realistic to receive feedback with such a delay. There was no difference between the ErrPs associated with ME and MI, this is in agreement with previous findings [15]. The amplitudes of the peaks of the ErrPs and NonErrPs varied considerably across subjects. It is not known if this is a natural variation between the subjects or if the ErrP protocol was not working properly, e.g. the subjects discovered the feedback was sham (this was not tested). Another factor that could contribute to the variability could be fatigue. Although several breaks were included, the experiments were quite long (2-3 hours), so fatigue could cause jitter in the trials if the subjects were not attentive immediately when the feedback was presented; this would affect the averages of the ErrP and NonErrP.

5 Conclusion

It was possible to elicit ErrPs after delayed sham feedback in both an ME and MI paradigm where the ErrPs were isolated from the MRCP. However, when using delayed feedback caution should be taken to avoid an expectancy potential prior the presentation of the BCI output. The ErrPs associated with ME and MI could be classified well-above chance level with all feature types when using the random forest classifier; higher accuracies were obtained for ME. The best features were the temporal and DWT marginal features. It was possible to use a single feature type for the classification, which means that the feature vector does not have to be large. The results suggest that the type of classifier is important. It was shown that it is possible to decode single-trial ErrPs when a negative potential similar to the MRCP precedes them. This implies that ErrPs can be used to correctly identify classified movement trials in MRCP-based BCIs for rehabilitation applications where constant recalibration may be needed to account for shifts in attention during the BCI use.

References

1. Millán J del R, Rupp R, Müller-Putz G, Murray-Smith R, Giugliemma C, Tangermann M, Vidaurre C, Cincotti F, Kubler A, Leeb R, Neuper C, Mueller K.R, Mattia D (2010) Combining Brain–Computer interfaces and assistive technologies: state-of-the-art and challenges. Frontiers in neuroscience 4:161. doi: 10.3389/fnins.2010.00161
2. Ramos-Murguialday A, Broetz D, Rea M, Läer L, Yilmaz Ö, Brasil FL, Liberati G, Curado MR, Garcia-Cossio E, Vyziotis A, Cho W, Agostini M, Soares E, Soekadar S, Caria A, Cohen LG, Birbaumer N (2013) Brain–machine interface in chronic stroke rehabilitation: A controlled study. Ann.Neurol 74:100–108. doi: 10.1002/ana.23879
3. Grosse-Wentrup M, Mattia D, Oweiss K (2011) Using brain–computer interfaces to induce neural plasticity and restore function. J. Neural Engineering 8:025004. doi: 10.1088/1741-2560/8/2/025004
4. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clinical Neurophysiology 113:767–791. doi: 10.1016/S1388-2457(02)00057-3
5. Mohammadi R, Mahloojifar A, Coyle D (2013) Unsupervised short-term covariate shift minimization for self-paced BCI. IEEE symposium series on computational intelligence 101–106. doi:10.1109/CCMB.2013.6609172
6. Yazmir B, Reiner M (2018) Neural correlates of user-initiated motor success and failure–a brain–computer interface perspective. Neuroscience 378:100-112
7. Lopes-Dias C, Sburlea AI, Müller-Putz G (2019) Online asynchronous decoding of error-related potentials during the continuous control of a robot. Sci Rep 9:1-9. doi: 10.1038/s41598-019-54109-x

8. Falkenstein M, Hohnsbein J, Hoormann J, Blanke L (1991) Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Clinical Neurophysiology 78:447–455. doi: 10.1016/0013-4694(91)90062-9
9. Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, Sajda P (2002) Linear spatial integration for single-trial detection in encephalography. IEEE transactions on bio-medical engineering 55(3):223–230. doi: 10.1006/nimg.2002.1212
10. Blankertz B, Dornhege G, Schafer C, Krepki R, Kohlmorgen J, Muller K-R, Kunzmann V, Losch F, Curio G (2003) Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Engineering in Medicine and Biology Society 11(2) :127–131. doi: 10.1109/TNSRE.2003.814456
11. Chavarriaga R, Sobolewski A, Millán JDR (2014) Errare machinale est: the use of error-related potentials in brain-machine interfaces. Frontiers in Neuroscience 8:208. doi: //dx.doi.org/10.3389/fnins.2014.00208
12. Lopes-Dias C, Sburlea AI, Müller-Putz G (2018) Masked and unmasked error-related potentials during continuous control and feedback. Journal of neural engineering 15:036031. doi: 10.1088/1741-2552/aab806
13. Iturrate I, Chavarriaga R, Montesano L, Minguez J, Millán J del R (2014) Latency correction of event-related potentials between different experimental protocols. J. Neural Eng. 11:036005
14. Chavarriaga R, Millán J del R (2010) Learning from EEG error-related potentials in noninvasive brain-computer interfaces 18:381-388. IEEE transactions on neural systems and rehabilitation engineering doi: 10.1109/TNSRE.2010.2053387
15. Omedes J, Schwarz A, Müller-Putz G, Montesano L (2018) Factors that affect error potentials during a grasping task: toward a hybrid natural movement decoding BCI. Journal of neural engineering 15:046023. doi: 10.1088/1741-2552/aac1a1
16. Iturrate I, Montesano L, Minguez J (2010) Robot reinforcement learning using EEG-based reward signals. IEEE International Conference on Robotics and Automation 4822–4829
17. Bhattacharyya S, Konar A, Tibarewala DN, Hayashibe M (2017) A generic transferable EEG decoder for online detection of error potential in target selection. Frontiers in neuroscience 11:226. doi: 10.3389/fnins.2017.00226
18. Parra LC, Spence CD, Gerson AD, Sajda P (2003) Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring. IEEE transactions on neural systems and rehabilitation engineering 11:173-177. doi: 10.1109/TNSRE.2003.814446
19. Ehrlich SK, Cheng G (2018) Human-agent co-adaptation using error-related potentials. Journal of neural engineering.15(6):066014
20. Ehrlich SK, Cheng G (2019) A feasibility study for validating robot actions using eeg-based error-related potentials. International journal of social robotics 11(2):271-283
21. Kim SK, Kirchner EA, Stefes A, Kirchner F (2017) Intrinsic interactive reinforcement learning using error-related potentials for real world human-robot interaction. Scientific reports 7(1):1-16
22. Spüler M, Niethammer C, Rosenstiel W, Bogdan M (2014) Classification of error-related potentials in EEG during continuous feedback. doi: 10.3217/978-3-85125-378-8-6
23. Takahashi H, Yoshikawa T, Furuhashi T (2010) Reliability-based automatic repeat reQuest with error potential-based error correction for improving P300 speller performance. Neural information processing models and applications. Lecture notes in computer science 6444
24. Niazi IK, Jiang N, Tiberghien O, Nielsen JF, Dremstrup K, Farina D (2011) Detection of movement intention from single-trial movement related cortical potentials. Journal of neural engineering 8(6):066009
25. Spüler M, Bensch M, Kleih S, Rosenstiel W, Bogdan M, Kübler A (2012) Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. Clinical neurophysiology 123:1328-1337. doi: 10.1016/j.clinph.2011.11.082
26. Omedes J, Iturrate I, Montesano L, Minguez J (2013) Using frequency-domain features for the generalization of EEG error-related potentials among different tasks:5263-5266. doi: 10.1109/EMBC.2013.6610736
27. Kreilinger A, Hiebel H, Müller-Putz G (2016). Single versus multiple events error potential detection in a BCI-controlled car game with continuous and discrete feedback. IEEE transactions on bio-medical engineering 63(3):519–529
28. Schalk G, Wolpaw JR, McFarland DJ, Pfurtscheller G (2000) EEG-based communication: presence of an error potential. Clinical neurophysiology 111:2138-2144. doi: 10.1016/S1388-2457(00)00457-0
29. Mousavi M, V. R.de Sa (2019) Spatio-temporal analysis of error-related brain activity in active and passive brain–computer interfaces, Brain-computer interfaces, 6:4, 118-127, doi: 10.1080/2326263X.2019.1671040
30. Tong J, Lin Q, Xiao R, Ding L (2016) Combining multiple features for error detection and its application in brain-computer interface. BioMed Eng OnLine 15:n/a. doi: //dx.doi.org/10.1186/s12938-016-0134-9

31. Ehrlich SK, Cheng G (2016) A neuro-based method for detecting context-dependent erroneous robot action. IEEE-RAS international conference on humanoid robots 477-482 doi: 10.1109/HUMANOIDS.2016.7803318
32. Spüler M, Niethammer C (2015) Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity. Frontiers in human neuroscience 9:155 doi: 10.3389/fnhum.2015.00155
33. Omedes J, Iturrate I, Montesano L (2014) Asynchronous detection of error potentials. Proceedings of the brain-computer interface doi: 10.3217/978-3-85125-682-6-11
34. Buttfield A, Ferrez PW, Millan JR (2006) Towards a robust BCI: error potentials and online learning. IEEE transactions on neural systems and rehabilitation engineering 14:164-168. doi: 10.1109/TNSRE.2006.875555
35. Parra LC, Spence CD, Gerson AD, Sajda P (2003) Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring. IEEE transactions on neural systems and rehabilitation engineering 11:173-177. doi: 10.1109/TNSRE.2003.814446
36. Iturrate I, Chavarriaga R, Montesano L, Minguez J, Millán JdR (2012) Latency correction of error potentials between different experiments reduces calibration time for single-trial classification:3288-3291. IEEE engineering in medicine and biology society doi: 10.1109/EMBC.2012.6346667
37. Ventouras EM, Asvestas P, Karanasiou I, Matsopoulos GK (2011) Classification of error-related negativity (ERN) and positivity (Pe) potentials using kNN and support vector machines. computers in biology and medicine 41:98-109. doi: 10.1016/j.compbiomed.2010.12.004
38. Wang S, Lin C, Wu C, Chaovalitwongse WA (2011) Early detection of numerical typing errors using data mining techniques. IEEE Transactions on systems, man, and cybernetics - part A: systems and humans 41:1199-1212. doi: 10.1109/TSMCA.2011.2116006
39. Ferrez PW, Millán J del R (2008) Error-Related EEG potentials generated during simulated brain–computer interaction. IEEE Transactions on Biomedical Engineering 55:923-929. doi: 10.1109/TBME.2007.908083
40. Zeyl T, Yin E, Keightley M, Chau T (2016) Adding real-time bayesian ranks to error-related potential scores improves error detection and auto-correction in a P300 speller. IEEE Transactions on Neural Systems and Rehabilitation Engineering 24(1):46-56, doi: 10.1109/TNSRE.2015.2461495
41. Spüler Martin, Rosenstiel W, Bogdan M (2012) Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning. PloS one 7.12 doi: 10.15496/publikation-4255
42. Ferrez P.W, Millán J del R. (2008b) Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy. Proceedings of the 4th Intl. brain-computer interface workshop and training course
43. Niazi IK, Mrachacz-Kersting N, Jiang N, Dremstrup K, Farina D (2012) Peripheral electrical stimulation triggered by self-paced detection of motor intention enhances motor evoked potentials. IEEE transactions on neural systems and rehabilitation engineering 20(4):595-604 doi: 10.1109/TNSRE.2012.2194309
44. Jochumsen M, Navid MS, Rashid U, Haavik H, Niazi IK (2019) EMG- versus EEG-triggered electrical stimulation for inducing corticospinal plasticity. IEEE transactions on neural systems and rehabilitation engineering 27(9):1901-1908 doi: 10.1109/TNSRE.2019.2932104
45. Jochumsen M, Niazi IK, Mrachacz-Kersting N, Jiang N, Farina D, Dremstrup K (2015) Comparison of spatial filters and features for the detection and classification of movement-related cortical potentials in healthy individuals and stroke patients. Journal of neural engineering 12(5):056003. doi: 10.1088/1741-2560/12/5/056003
46. Jochumsen M, Niazi IK, Taylor D, Farina D, Dremstrup K (2015) Detecting and classifying movement-related cortical potentials associated with hand movements in healthy subjects and stroke patients from single-electrode, single trial EEG. Journal of neural engineering 12(5):056013. doi: 10.1088/1741-2560/12/5/056013
47. Dietrich C, Blume KR, Franz M, Huonker R, Carl M, Preißler S, Hofmann GO, Miltner WHR, Weiss T (2017) Dermatomal organization of SI leg representation in humans: revising the somatosensory homunculus. cereb cortex 27:4564-4569. doi: 10.1093/cercor/bhx007
48. Nascimento O.F.d, Nielsen K.D, Voigt M (2006) Movement-related parameters modulate cortical activity during imaginary isometric plantar-flexions. Exp Brain Res 171, 78–90
49. Pfurtscheller G, Lopes da Silva FH (1991) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical neurophysiology 110: 1842-1857
50. Falkenstein M, Hohnsbein J, Hoormann J, Blanke L (1991) Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Clinical Neurophysiology 78:447–455. doi: 10.1016/0013-4694(91)90062-9
51. Farina D, do Nascimento OF, Lucas MF, Doncarli C (2007) Optimization of wavelets for classification of movement-related cortical potentials generated by variation of force-related parameters. Journal of Neuroscience Methods 162(1-2):357-363 doi: 10.1016/j.jneumeth.2007.01.011
52. Oshiro T M, Perez PS, Baranauskas J.A (2012) How many trees in a random forest? Machine learning and data mining in pattern recognition. Lecture notes in computer science 7376

53. Milekovic T, Ball T, Schulze-Bonhage A, Aertsen A, Mehring C (2012) Error-related electrocorticographic activity in humans during continuous movements. Journal of neural engineering 9(2): 1741-2552 doi: 10.1088/1741-2560/9/2/026007
54. Falkenstein M, Hoormann J, Christ S, Hohnsbein J (2000) ERP components on reaction errors and their functional significance: a tutorial. A tutorial Biol. Psychol 51:87–107. doi: 10.1016/S0301-0511(99)00031-9
55. Shibasaki H, Hallett M (2006) What is the bereitschaftspotential?. Clinical neurophysiology 117:2341-2356
56. Walter WG, Cooper R, Aldridge VJ, McCallum WC, Winter AL (1964) Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. Nature 203:380-384
57. Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, Sajda P (2002) Linear spatial integration for single-trial detection in encephalography. IEEE transactions on bio-medical engineering 55(3):223–230. doi: 10.1006/nimg.2002.1212
58. Salazar-Gomez AF, DelPreto J, Gil S, Guenther FH, Rus D (2017) Correcting robot mistakes in real time using EEG signals. IEEE international conference on robotics and automation 6570-6577 doi: 10.1109/ICRA.2017.7989777
59. Nicolas-Alonso LF, Corralejo R, Gomez-Pilar J, Álvarez D, Hornero R (2015) Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain–computer interfaces. Neurocomputing 159:186-196. doi: 10.1016/j.neucom.2015.02.005
60. Kumar E, Pirogova JQ, Fang (2018) Classification of error-related potentials using linear discriminant analysis. Conference on biomedical engineering and sciences 18-21. doi: 10.1109/IECBES.2018.8626709
61. Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G (2008) Better than random: a closer look on BCI results. International Journal of Bioelektromagnetism 10:52-55
62. Rahimi A, Tchouprina A, Kanerva P, Millán J del R (2017) Hyperdimensional computing for blind and one-Shot classification of EEG error-related potentials. Mobile netw appl https://doi.org/10.1007/s11036-017-0942-6
63. Karimi F, Kofman J, Mrachacz-Kersting N, Farina D, Jiang N (2017) Detection of movement related cortical potentials from EEG using constrained ICA for brain-computer interface applications. Front Neurosci. 11:356 doi:10.3389/fnins.2017.00356
64. Blankertz B, Lemm S, Treder M, Haufe S, Müller KR (2011) Single-trial analysis and classification of ERP components--a tutorial. Neuroimage 56(2):814-25. doi: 10.1016/j.neuroimage.2010.06.048

Captions:

Fig. 1 Time distribution for the experimental protocol. Initially, the subject rests for five seconds (-8 to -3 seconds), after which the subject is instructed to prepare to execute or imagine a movement (-3 to 0 seconds). At 0 seconds, an image is shown indicating what movements should be performed. After a 3-second delay, the output of the sham BCI is displayed to the subject. The feedback was visible for three seconds before a new repetition of the trial started

Fig. 2 Boxplots (mean, median and quartiles) of the classifaction accuracies achieved by using linear discriminant analysis (LDA), support vector machine (SVM) and random forest (RF) classifiers for all combinations of features across the motor execution (ME) paradigm. T: Temporal, S: Spectral, M: Template matching, W: Marginals of the discrete wavelet transform (DWT). The highest classification accuracies were obtained using the temporal and DWT features. Generally, the highest accuraices were obtained using the RF classifier. Note that there is large inter-subject variability

Fig. 3 Boxplots (mean, median and quartiles) corresponding to the classifaction accuracies achieved using LDA, SVM and RF classifiers for all combinations of features across the motor imagination (MI) paradigm. T: Temporal, S: Spectral, M: Template matching, W: Marginals of the discrete wavelet transform. The highest classification accuracies were obtained using the temporal features. Generally, the highest accuraices were obtained using the RF classifier where all four feature types were associated with classification accuracies above 70%. Note that there is large inter-subject variability

Fig. 4 Grand averages across the 15 subjects of the ErrP, NonErrP, and the difference (ErrP - NonErrP) for the Motor Execution paradigm (a) and Motor Imagination paradigm (b). The difference between ErrP and NonErrP for the two paradigms are plotted in the bottom graph (c). The shaded area represents the standard deviation across subjects. 0 seconds is where the feedback was presented to the subject. The ErrPs and NonErrPs are overlapping with slightly higher amplitudes for the ErrPs. There is no apparent difference between the ErrPs and NonErrPs associated with executed and imagined movements for the latencies and there are only slightly higher amplitudes of the peaks for the executed movements. It appears to be an expectancy potential immediately prior the feedback was presented to the subject. It should be noted that there is a considerable amount of inter-subject variability

Author Biographies:

Nayab Usama is an electrical engineer who is enrolled as a Ph.D. student at Aalborg University, Denmark. Her research area includes adaptive brain-computer interface, biomedical signal processing and deep learning.

Kasper Leerskov is a biomedical engineer who is enrolled as a Ph.D. student at Aalborg University, Denmark. His research interests include brain–computer interfacing, neuroplasticity, and neurorehabilitation.

Imran Khan Niazi is a BSc. in Electrical engineering (Riphah University) and a MSc. in Biomedical Engineering (Lübeck University). He did his PhD at Aalborg University in brain-computer interfacing.

Kim Dremstrup received his MSc. in BME and PhD from Aalborg University. He is an Associate Professor and Head of Department of Health Science and Technology at the Medical Faculty, Aalborg University.

Mads Jochumsen is a biomedical engineer with a PhD from Aalborg University, Denmark, where he is an Assistant Professor. His research interests include brain–computer interfacing, signal processing and neurorehabilitation.