DECODING OF ERROR RELATED POTENTIALS

BCI project 3

June 7, 2019

Kaleem Corbin, Jan Frogg, Aindrias Lefévère-Laoide, Philipp Spiess

Supervisor: Fumiake Iwane

Contents

1 INTRODUCTION

Research on Brain-Computer Interaction (BCI) started in the mid-1970s and is making quick progress thanks to its increasing popularity, in particular for its exciting prospects and the world of possibilities it opens up in terms of translational applications regarding computed-aided human actions. Such systems have especially valuable applications for people suffering from severe neuromuscular disorders, such as amyotrophic lateral sclerosis (ALS), brain-stem stroke, and spinal cord injury. To give a brief definition, a BCI may be described as a direct communication and control system that does not depend on the brain's normal output pathways of peripheral nerves and muscles [1]. In other words, it is a direct communication interface between an external device and the brain, which bypasses the need for an embodiment. It is a closed loop which allows the user to get extended feedback and to learn from the way the computer reacts while the decoder learns to classify in real time with better accuracy with time and data. One common way of implementing a BCI is to use electro-encephalography (EEG), which is a non-invasive technique to record brain activity with an electrode-carrying cap at the surface of the scalp. The brain waves are analyzed in order to extract features of interest. One particular feature researchers have been working on lately is the detection of Error-Related Potentials (ErrPs), which occurs when a person perceives an erroneous behavior in any kind of normal situation. An example of this would be the detection of a wrong letter when spelling a word.

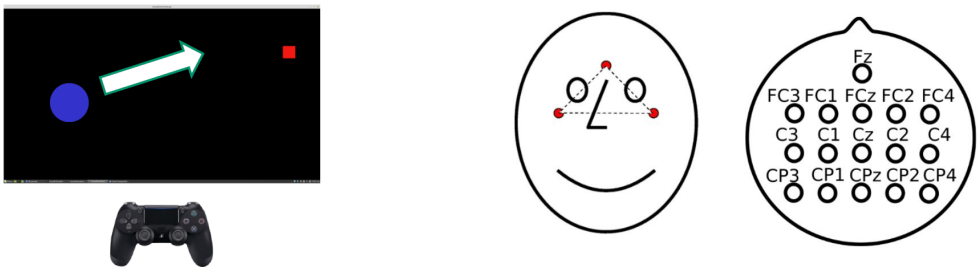
Our project was based on the detection of ErrPs when playing a simple video game, where upon an erroneous behavior was induced and had to be detected by the subject. EEG, as well as electro-oculographic (EOG) recordings were done for all of the group's members at the Biotech Campus in Geneva and the ultimate aim of the project was to build a decoder capable of detecting and classifying these ErrPs. The programming environment that was used for this project was MATLAB. The following sections will be dedicated to describing each step in building the decoder, as well as the results and performance it yielded for the EEG recordings across all subjects.

2 METHODS

2.1 Experimental Protocol

The experimental protocol was designed to elicit ErrPs when subjects perceived erroneous behavior. The setup consisted of a monitor, an EEG device with EOG electrodes and a joystick controller. The device consisted of 16 EEG electrodes following the configuration shown in

figure 1b. Furthermore, 3 EOG electrodes were placed in a triangular formation, as seen in figure 1b, to allow for EOG calibration and eventual removal of EOG artifacts. EEG activity was recorded at a sampling rate of 512 Hz. The monitor displayed a blue circle (cursor) that was controlled by the controller and a red square that served as the target. The task consisted of bringing the cursor to the target as quickly as possible using the joystick as seen in figure 1a. An erroneous trial was created by changing the joystick-to-cursor mapping by a rotation of ± 20°, 40°or 60°, whereas during a normal trial, the mapping remained unchanged. The change in mapping should be sufficient for the subject to perceive an error and thus should yield an error related potential. The magnitude of the change of mapping in a given erroneous trial was arbitrary, thus preventing the subject from guessing or anticipating behavior. During the online decoding trials, the aim was for the classifier to detect in real time if the ongoing trial was a "rotation" trial. If there was detection, the circle turned green and the subject gained a 1 second bonus if this proved to be a true positive. The entire duration of a trial (start until the target was reached) would take on average 3s for a normal (correct) trial, whereas an erroneous trial would last slightly longer (0.5-1s more). Each block consisted of 40 trials, whereby approximately 30% were error trials. The percentage of error trials in each block was selected because a 50% split between error and correct would not allow the subject to distinguish between erroneous and normal behavior. During each session, the subject would undergo 10 blocks of trials, therefore a total of 400 trials were recorded of which 30% of these trials were erroneous.



(a) Simulation task **(b)** Electrode placement

Figure 1: Experimental protocol

2.2 Pre-processing

2.2.1 EOG Filtering

One of the first steps in preprocessing the signal was to remove possible artifacts, more specifically EOG, as it can generally be a major noise source in EEG recordings. Eye move-

ments are difficult to suppress over a sustained period of time, therefore it is valid to assume that every EEG recording is contaminated with EOG artifacts. In order to try and remove EOG-based artifacts, we used a linear model described in [2] which considers three spatial EOG components and is described as follows:

$$Y(t,ch) = S(t,ch) + [EOG1(t), EOG2(t), EOG3(t)] \times [b1(ch), b2(ch), b3(ch)]^T,$$

where Y is the EOG contaminated signal, S is the signal we wish to extract (contamination-free), EOG123 represent the noise source vector U of the three spatial EOG components and lastly, the b vector indicates the weights of the EOG artifacts at the EEG channel ch. These weights are computed using a covariance analysis between the EOG and EEG channels. The non-contaminated signal can therefore be expressed as:

$$S = Y - U \cdot b,$$

which is what we apply on the signals we recorded.

2.2.2 Spectral Filtering

A spectral filter amplifies the desired frequency band and attenuates unwanted frequencies. A primary source of noise originates from the electromagnetic interference caused by power lines that are set at 50 Hz. For analysis of ErrPs, the desired frequency band has been shown to be between 1-10 Hz [3]. Therefore, we applied a Butterworth band-pass filter with the following parameters: low cutoff frequency set to 1 Hz, high cutoff frequency set to 10 Hz and filter order set to 2. This band-pass filter additionally dealt with the electromagnetic interference issue.

2.2.3 Spatial Filtering

Spatial analysis allows for a more accurate representation of EEG activity at each electrode. Three spatial filters were considered for our project: Common Average Reference (CAR), Laplacian, and Canonical Correlation Analysis (CCA). The combination of both spatial and spectral filtering can increase the Signal-to-Noise Ratio (SNR) of the signal by improving the signal of interest and by reducing the noise, which further leads to an improvement in classification accuracy.

Firstly, CAR-based spatial filter consists of computing the mean activity from all electrodes (the common average) and using it as a reference that will then be subtracted from the activity observed at each electrode. With the assumptions that the head is covered by equally-spaced and that the potential on the head is generated by point sources, this method of

filtering should, in theory, lead to spatial voltage distribution with mean zero. Although these assumptions are not always met, the CAR filter is still widely used and generally yields EEG recordings which are nearly reference-free.

Secondly, we tried to implement a Laplacian filter, which essentially consists of computing the second derivative of the voltage distribution at each electrode, which therefore emphasizes EEG activity originating in radial sources immediately below each electrode. Thus, this spatial filter acts as a high-pass filter that also improves the EEG's SNR by emphasizing localized activity and reducing more diffuse activity.

While the CAR and Laplacian are the most widely used methods for spatial filtering and are fairly easy to implement, we found it interesting to explore the more specific CCA method to compute a spatial filter, based on the description given in Spüler et al, 2014. This method is based on the assumed existence of an underlying correlation between two datasets, the ultimate goal being to maximize this correlation after transformation of these two datasets. In order to understand CCA computation for spatial filtering, let us consider two datasets X and Y which represent the raw EEG data and by the waveform of the average evoked response, respectively. The aim of the method is to find the following transformations:

$U = W_X^T X$ and $V = W_Y^T Y$

The idea in the end is to find the transformation matrix W_X^T, which will be used as a spatial filter. Moreover, it is important to note that single-trial data of EEG recording generally have a bad SNR and therefore, we average over multiple trials in order to improve the SNR. The main principle when using CCA for spatial filtering is to find a filter that, as stated earlier, maximizes the correlation between the transformed signal and the average evoked response (apparition of an ErrP). By doing this, the part of the average evoked response is enhanced, while the noise is attenuated. This ends up in an improvement in the SNR, once again. Lastly, one advantage of using CCA over the previously mentioned other methods is that it allows for dimensionality reduction by selecting only the most relevant CCA coefficients, as the resulting matrix W_X^T ranks them from most explanatory to less explanatory. Although this might be where the actual "spatial" component of the filter comes in, as it seems to perform a channel selection when choosing the most important components, it is still unclear to us how this is done and exactly what channels are selected when retaining certain coefficients.

2.3 EEG correlates

Visualizing ErrPs was an important step in understanding our data and allowed us to make initial hypotheses on how to develop our decoder. EEG correlates were derived by taking

the average activity across all trials (grand average) for each type (correct and error) for a given electrode. This was applied to the original signal, spatially filtered signal using CAR, spectrally filtered signal and a signal using both spectral and spatial filters on electrode FCz for the competition data as seen in figure 2. The first observation is that the addition of these filters progressively made the signal cleaner and easier to distinguish the ErrPs from baseline activity. Furthermore, it can be observed from the topological maps that activity at the peaks is more prominent at certain electrodes compared to others. This observation served as our initial channel selection mechanism. Seven central electrodes (Fz, FC1, FCz, FC2, C1, CZ, C2) were identified to consistently yield higher amplitudes at the peaks and troughs of the ErrPs as seen in figure 4. The selection of these electrodes is consistent with the choice of electrodes used in current literature, whereby the eight fronto-central electrodes are generally selected: Fz, FC1, FCz, FC2, C1, CZ, C2, CPz [3]. Furthermore, it can be noted that the ErrP does not occur instantly after the onset of rotation, but generally appears after 0.2 seconds. EEG correlates were also observed as a function of their magnitude. The grand average was also performed on all trials according to their magnitude as seen in figure 3. With regards to the competition data (subject B4), it can be observed that ErrPs with ± 20°rotation resulted in a weaker response than ± 40°and 60°respectively. Therefore, we hypothesized that our decoder would have more difficulty decoding trials with ± 20°rotation compared to the latter two.

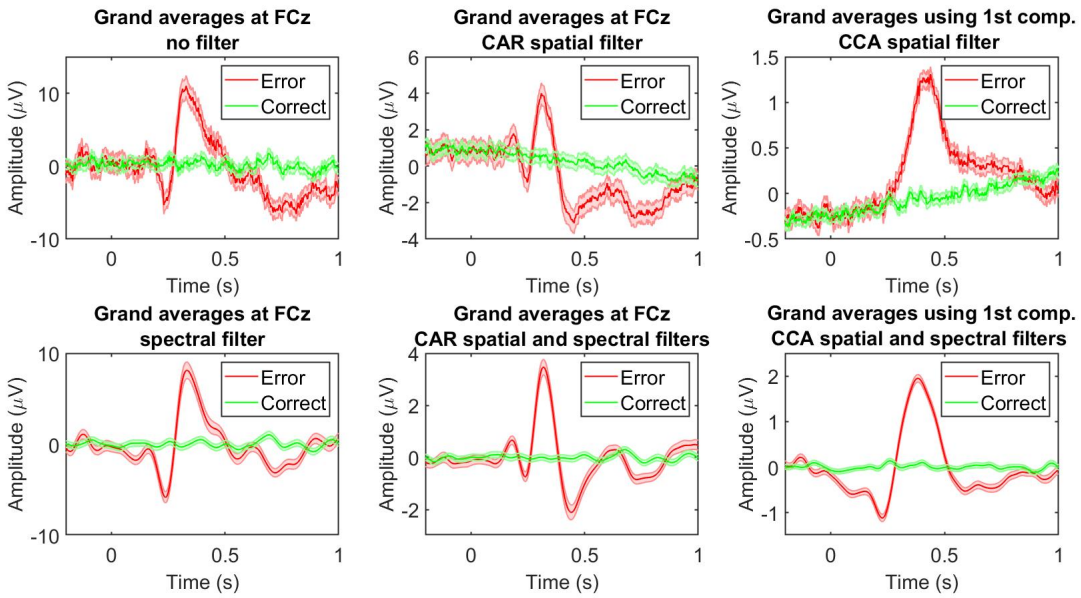


Figure 2: Grand averages of the competition data using different sets of filters. Thick line represents the mean and shaded area represents the standard error across trials.

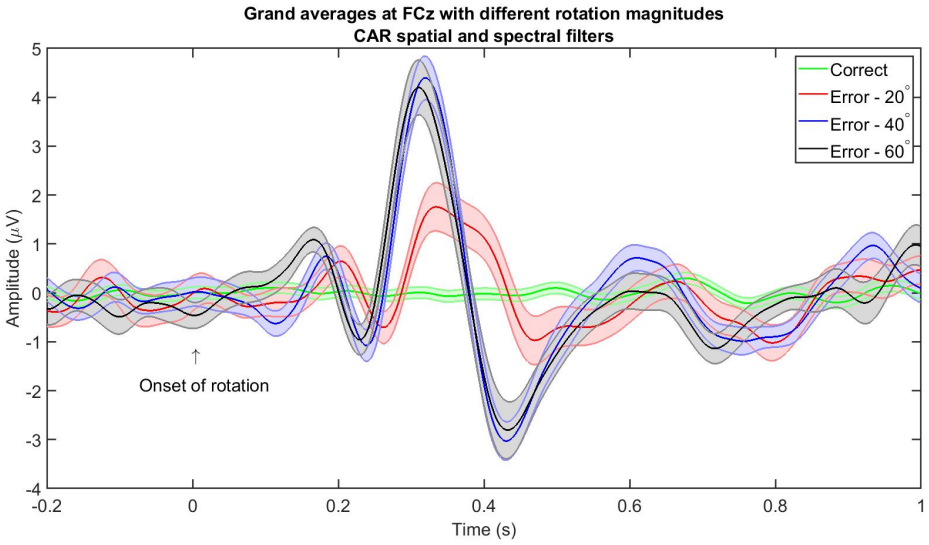


Figure 3: Grand averages of the competition data with respect to different magnitudes of rotation using a CAR spatial filter and a spectral filter. Thick line represents the mean and shaded area represents the standard error across trials.

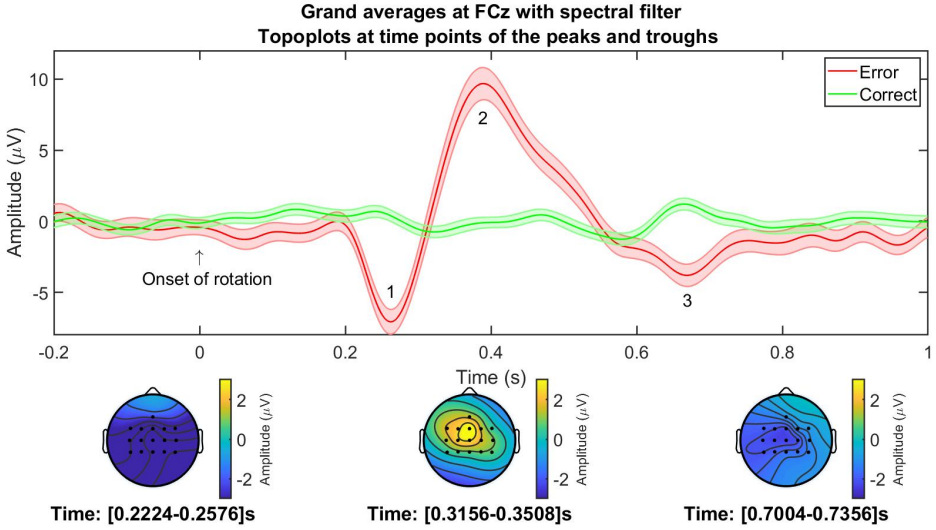


Figure 4: Grand averages of the competition data using a spectral filter. Associated topoplot at the time points of the peaks and troughs of the ErrP. The topographical graphs were obtained by taking the average over a time window of 20 samples (0.0352s).

2.4 Decoder

To find the best way to decode the error related potentials, several classification methods were used combining unsupervised and supervised learning. All the following methods

were trained on only the training sets generated by the 10-fold cross-validation to avoid overfitting. First, the CCA, as stated before, was used as the spatial filter which allows to combine the 16 EEG signals into 4 channels in a way to maximize the correlation and reduce the dimensionality. The first method used for feature selection is the fisher score [4] , which allows us to evaluate labeled datapoints (see figure 5). We decided to to apply a threshold which removes all Z-values under 0.8.

Principal Component Analysis (PCA) was then used in addition, as an unsupervised method to select only features which describe the variance in the principal components space the most. [5] suggests in fact that combining both of these supervised and unsupervised feature selection methods can be beneficial. On figure 6, we plot the Matthews Correlation Coefficient score (MCC), which is a good way to evaluate binary classification performance for imbalanced data. The MCC score is ploted as a function of the % of variance explained for each of the 4 data-sets of our group. As a result, we decided to choose 75% explained variance is an optimal solution to be used henceforth. We then compared 4 supervised learning methods, (Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), naive bayes and Support Vector Machine (SVM)) to train a classifier which will then classify the test set. We chose to retain QDA given that it gaves us on average the best MCC score as it will be shown in the results section and is in accordance with [6].
The pipeline of the whole feature extraction process is shown in figure 7. The MCC score formula is given below with the corresponding variables defined in the following confusion matrix.

		Predicted	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	TN	TP

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

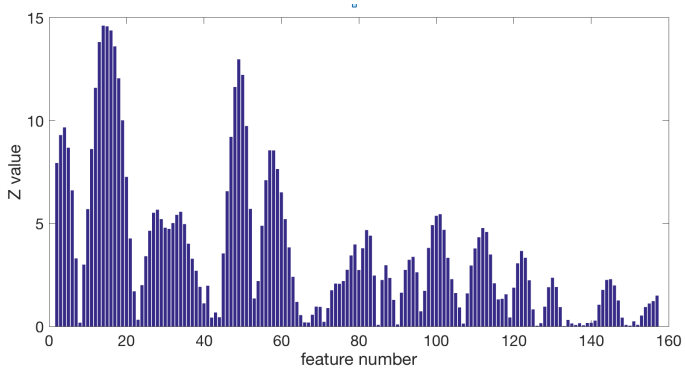


Figure 5: Fisher score of every features explaining the distance to a hypothetical linear discriminant hyper-plane

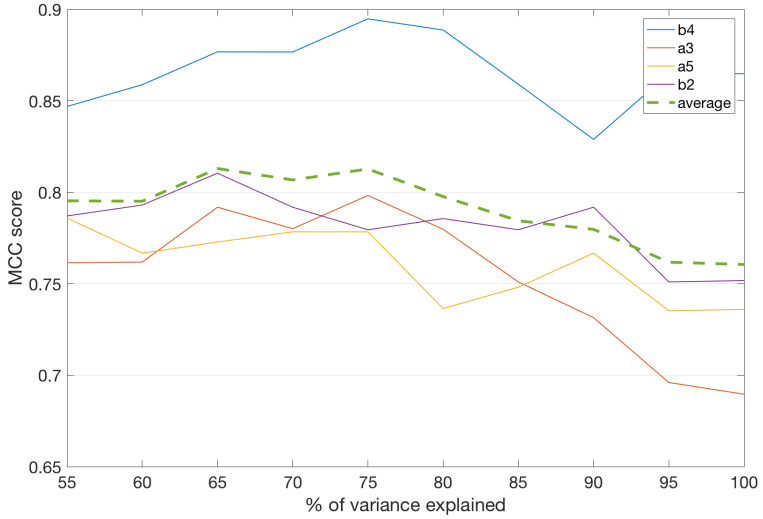


Figure 6: MCC score as function of the % of variance explained by n number of principal components.

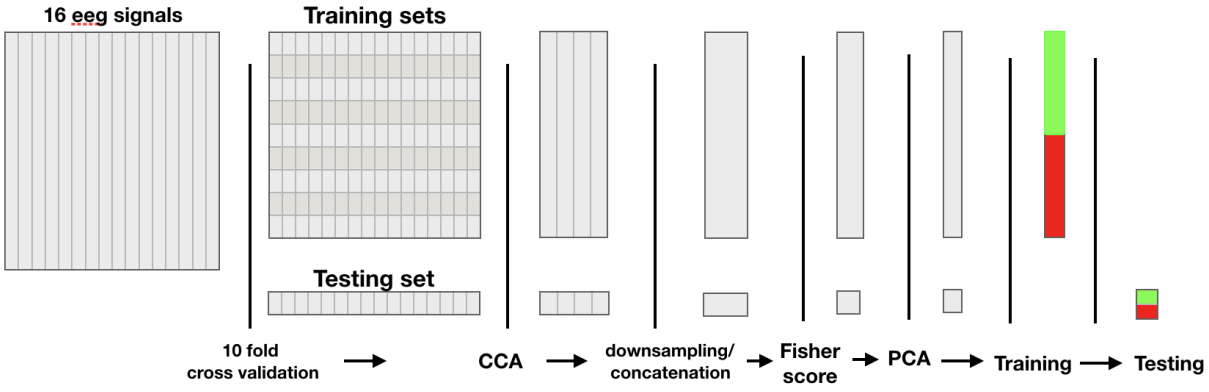


Figure 7: Pipeline of features extraction and classification.

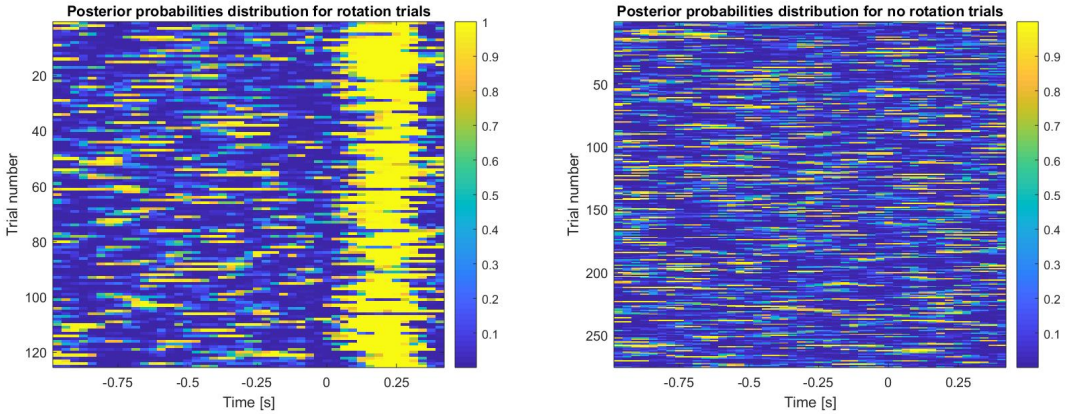


Figure 8: Posterior probabilities for rotation and no rotation trials and for a quadratic classifier, aligned on onset of the rotation

2.5 Continuous classification

Continuous classification of successive epochs for a unique trial allows the decoder to work online. Training the decoder for online classification necessitates the optimization of two hyperparameters: the threshold value for posterior probability classification and a second hyperparameter which can be either the number of consecutive values above that threshold or the number of samples in the moving average which has to be above the threshold.

The classification was performed on a sliding window of size 0.6 seconds. Since our data was downsampled from 512 Hz to 64 Hz, this amounted to 39 samples per window. The online data is collected for classification at a frequency of 32 Hz, which means that every new window had two new samples compared to the previous one. The performance of the classifier was assessed on the whole length of the trial. We used a quadratic classifier because it yielded the best performance in the simple offline decoding.

We plotted the posterior probabilities using the MATLAB predict function with a 10-fold cross-validation (see Figure 8). We did this to get an initial idea of whether our classifier could discriminate ErrPs. We observe that there is a clear peak in posterior probability for the window ranging from 0.2 to 0.8 seconds after onset for the rotation trial. This seems to indicate that we should be able to build a performing classifier. We also observed a peak in posterior probabilities just after the onset of the trial, not shown on this data, probably corresponding to an event-related potential linked to the onset. We therefore decided to discard the 0.25s at the start of the trial so as to obtain better performances. We tested values for threshold between 0.5 and 0.99. We decided to classify positive predictions (i.e rotations) as false positives if the time window started more than 0.3s before the onset of the trial.

To assess the performance of our online decoder, we wanted to use unseen data as

test data. Therefore, we divided our dataset into 10 blocks and kept one block of 40 trials apart as testing set. We computed a 9-fold cross-validation on the remaining data to perform optimization of the hyperparameters. We then trained the classifier on the 9 blocks of training data and computed the performance on the unseen data corresponding to the tenth block, with the optimal combination of hyperparameters which had been found in the previous step.

3 RESULTS

3.1 Artifacts-related results

As outlined in the prepocessing section, EOG calibration data was made available in order to account for EOG-based artifacts. However, our implementation of the EOG decontamination algorithm was determined not to be effective. Therefore, our decoder was still able to yield relevant results even without removing EOG-based artifacts. Another source of artifacts that we looked for was the identification of bad channels (electrodes). To identify these poor channels, we plotted the time evolution of the topographical maps of the grand averages of correct and erroneous trials using a moving average of 20 time points (available in source code). A bad channel was identified through visual inspection of these topoplots and observing abnormal behavior at a specific electrode. Examples of abnormal behavior would be if an electrode displayed a fixed constant value throughout the trial or if an electrode had highly fluctuating activity relative to the activity of the other electrodes. Through visual inspection, no channels were determined to contaminate our data for any of our subjects.

3.2 EEG correlates

EEG correlates were analyzed across all four of our team members' EEG signals using the same process detailed in the previous section. It can be observed that applying a spatial and spectral filter on the data yielded an ErrP in all subjects. However, the shape and magnitude differed across subjects. The expected N200 and P300 phenomena were observed in Subjects A3, A5 and B4 using the CAR spatial filter and spectral filter as seen in figure 9. In contrast, subject B2 had an unexpected negative deflection at 300-400ms, which was later determined to be due to the CAR filter (available in source code). In all subjects, it can be observed that significant discriminability exists between the error and correct trials, which allowed our decoder to distinguish between both states. A similar analysis was done using the CCA spatial filter (only the 1st component) and the grand averages of all subjects were plotted as seen

in figure 10. Given that the CCA filter transforms the feature space, it is more difficult to interpret the spatial location of the grand average. In all subjects, it can be observed that there are significant peaks and troughs at approximately 250ms and 350ms, which can be associated with the N200 and P300 phenomena respectively.

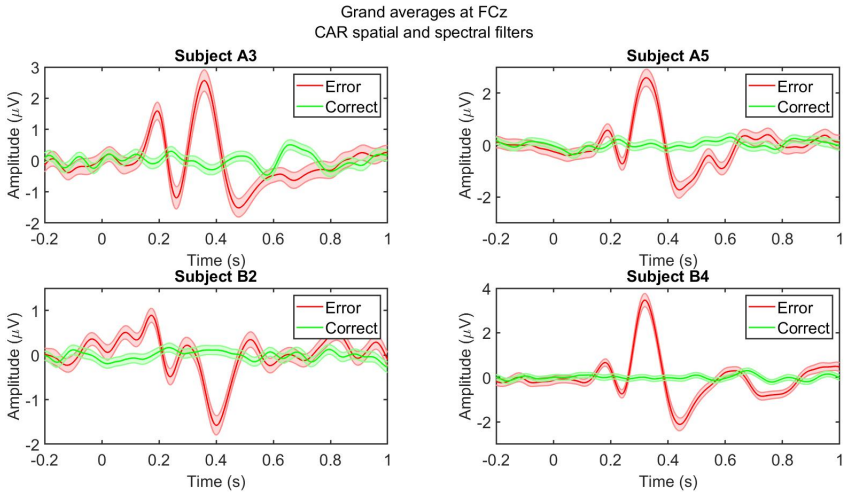


Figure 9: Comparison of the grand averages at FCz using CAR spatial filter and spectral filter across all subjects. Thick line represents the mean and shaded area represents the standard error across trials.

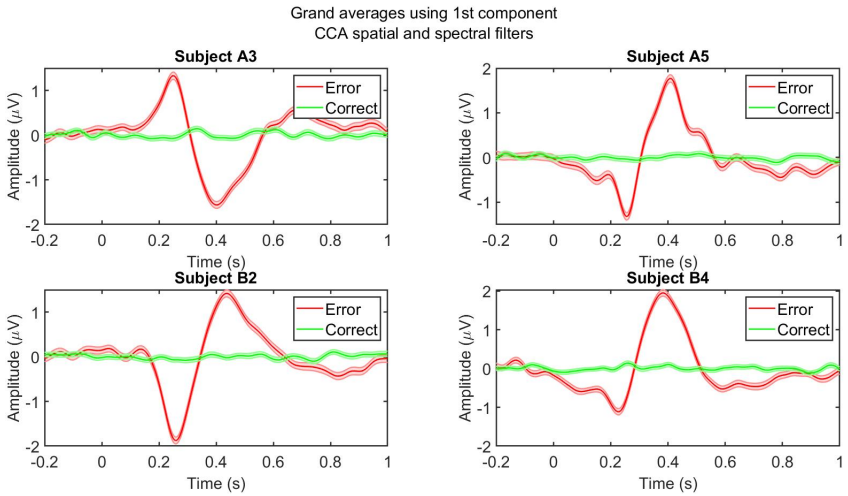


Figure 10: Comparison of the grand averages with a spectral filter and using the first component of the CCA spatial filter across all subjects. Thick line represents the mean and shaded area represents the standard error across trials.

3.3 Decoding results

3.3.1 Offline

To measure the performance of the classification, a boxplot presents on figure 11 the MCC score obtained for the 4 classifiers tried on all subjects. It allowed us to confirm that QDA was the optimal classifier given that it gave the best results for all four sets of data.

Figure 11: Comparison of different classifiers performance in offline classification through the MCC score over all datasets of our group (b4, a3, a5, b2)

Throughout the project, we were also interested in the fact that the magnitude of the rotation errors has a clear impact on the resulting cognitive state of the subject observing the error. On figure 12, we compare the classification accuracy when using only specific onsets of rotations. It first appears that a 20°rotation is difficult to detect. Also interestingly, the results show that above a certain value (40°), a bigger rotation error (60°) does not increase the classification score by a lot.

3.3.2 Online

We obtained an MCC score of 0.72 on unseen data using PCA as the feature selection method for the competition dataset. This was obtained with the hyperparameters tuned to the values which can be observed in table 1.

The performance varies significantly between the subjects, with a mean MCC of 0.668 and a standard deviation of 0.133. The posterior probability threshold also varies between 0.88 and 0.95, while the number of consecutive samples remains quite constant around 5.

Figure 12: Comparison of classification performance for specific types of rotation errors magnitudes (averaged over the 4 datasets of our group: b4, a3, a5, b2)

	Subject a3	Subject a5	Subject b2	Subject b4
MCC score	0.529	0.594	0.827	0.722
Posterior probability threshold	0.88	0.92	0.95	0.88
Number of consecutive samples	5	5	5	6

Table 1: Online performances and hyperparameters associated for the four subjects

3.4 Competition results

The competition results were obtain using the 10-fold cross-validation in offline classification and are based on competition dataset containing 400 trials. (the eog filter did not show an increase in performance and was thus not used for the classification of the competition dataset: b4):

$$MCCscore = 0.895$$

$$Accuracy = 0.924$$

Confusion Matrix	Predicted: No	Predicted: Yes
Actual: No	271	3
Actual: Yes	15	110

4 DISCUSSION

Regarding the spatial filter implementation, while the CAR and the Laplacian spatial filters were quite easy to implement and yielded good results as spatial filters for us, we chose to stick with the CCA approach, as it was the one with the most tuning possibilities and eventually gave us the best spatial filtering, even though it did take us some time to understand and implement correctly.

Computing grand averages allowed us to successfully identify ErrPs in all of our subjects. The addition of spectral and spatial filters served to clean up our signal and made it easier to identify the N200 and P300 phenomena. Applying a topographical mapping at these time points then allowed us to select channels that were more a contributing factor. The identified channels turned out to be consistent with those used in recent literature [3]. Furthermore, we observed that there was a correlation between the amplitude of the ErrP and the magnitude of rotation. This led to an early hypothesis that higher magnitude rotations would be easier to decode and was later shown to be the case. With regards to artifact removal, our current implementation of the EOG-based artifact removal algorithm was shown to be ineffective. Therefore, it remains a feature that can be further studied and improved upon in future iterations. In this project, bad channels were identified visually based on a certain criteria, however it would be more objective if we used a quantitative metric. In future developments, it would be necessary to develop a method to quantify how certain activity can fall in to the category as bad channel. We would then proceed to automatically remove said channel from our data. Although, an ErrP can be easily identified using grand averages, it remains a challenge to decode them in single trials and thus additional parameters were used.

Concerning the online decoding, it seems logical that we should have variation in performance between the individuals since we can see from the grand averages that the waveforms are very different for each individual. The performance is overall satisfying, with very good performance for the subject b2. We must note however that since the test data consists of only 40 trials, this performance is most certainly biased and we would need more data to be able to assess the online performance of our classifier. The fact that the optimal threshold is always at 0.88 or higher is a good indicator that this hyperparameter is optimized to limit the number of false positives mostly before the onset of the rotation. Indeed, observing in detail the performance of the classifier showed that when working with lower threshold there were a huge amount of false positives due to early detection (before the onset). Concerning the consecutive number of samples, the number remains constant at 5 or 6. This is consistent with what we can expect: since the analysis frequency is set at 32Hz, 5 samples corresponds

to a shift in time window of 0.125, which implies that most of the data in the time window (at least 0.5 seconds) is important for classification. Concerning the computation time, the processing for each window in real time takes 3 ms, which is approximately 10 times faster than the new data input frequency of 32Hz. Processing of each fold takes approximately 6 seconds which is very reasonable. Finally, we also tried the moving average as an alternative method for hyperparameter optimization in complement to optimal threshold. However, the performances obtained were not as good (around 0.3) as those for the consecutive samples above threshold method, therefore we decided not to continue with that method.

5 CONCLUSION

We have successfully constructed an online classifier capable of detecting error related potentials in real time from the input of 16 EEG electrodes. For future improvements, and in case more data is available to train a classifier, a deep neural network could be an interesting option to reach good classification accuracy [7]. We tried a neural network on our data but it did not yield good performance (MCC around 0.55). If we were to choose that type of classifier, we would need a performance much higher than that obtained with the quadratic classifier, to "compensate"the longer computing time. The performance should not just be based on the MCC score but also on the generalization capacity of the classifier (how stable is the performance?) and the computing power needed. When choosing between classifier with similar performances, we would follow the law of parcimony and choose the simplest model.

Detection of ErrPs is a major challenge for BCI development. However, its utility is valuable as it could be used in real world applications so as to complement the main function of a BCI. One example of its use case is in detecting when a subject is surprised by the incorrect output of the BCI (incorrect movement of the robotic arm or incorrect direction taken by the wheelchair for example) and correct it automatically. This process still requires significant fine-tuning if it is to be used in the real world since we elicit ErrPs, or event-related potentials of similar form, quite frequently in our everyday life in reaction to something we do not expect.

Bibliography

1. Wolpaw, J. R. *et al.* Brain-computer interface technology: a review of the first international meeting. eng. *IEEE transactions on rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society* **8,** 164–173. ISSN: 1063-6528 (June 2000).
2. Schlögl, A. *et al.* A fully automated correction method of EOG artifacts in EEG recordings. eng. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* **118,** 98–104. ISSN: 1388-2457 (Jan. 2007).
3. Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J. & Millan, J. d. R. Latency correction of error potentials between different experiments reduces calibration time for single-trial classification. eng. *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* **2012,** 3288–3291. ISSN: 1557-170X (2012).
4. K, A., KamatchiPriya.L & Askerunisa.A. Fisher Score Dimensionality Reduction for Svm Classification. en. *International Journal of Innovative Research in Science, Engineering and Technology* **3.** ISSN: ISSN ONLINE(2319-8753)PRINT(2347-6710). http://www.rroij.com/peer-reviewed/fisher-score-dimensionality-reduction-for-svmclassification-50579.html (2019) (Jan. 1970).
5. Pechenizkiy, M. & Tsymbal, A. *On Combining Principal Components with Fisher's Linear Discriminants for Supervised Learning* ().
6. Eva, O. D. & Lazar, A. M. Comparison of Classifiers and Statistical Analysis for EEG Signals Used in Brain Computer Interface Motor Task Paradigm. en. *International Journal of Advanced Research in Artificial Intelligence (IJARAI)* **4.** doi:10.14569/IJARAI.2015.040102. http://thesai.org/Publications/ViewPaper?Volume=4&Issue=1&Code=IJARAI&SerialNo=2 (2019) (2015).
7. Craik, A., He, Y. & Contreras-Vidal, J. L. Deep learning for electroencephalogram (EEG) classification tasks: a review. eng. *Journal of Neural Engineering* **16,** 031001. ISSN: 1741-2552 (June 2019).