Predicting a Subject's Card with the Muse-S headband

Daniel Díaz[1] **Diego Figueroa**[2] **Theodora Gaiceanu**[3]

[1]da5736di-s@student.lu.se [2]di4642fi-s@student.lu.se [3]th6636ga-s@student.lu.se

Abstract: We developed a Brain-Computer Interface system to predict the card a subject is focusing on in real-time with P300 evoked potentials (P300 EPs) recorded with a Muse-S headband. Firstly, we built the 4-card dataset: we told subjects to focus on one out of four cards, and then showed the cards in intervals of 200 ms, labelling the P300 EPs with weather the subject was focusing or not. Secondly, we tested different models to filter and classify the P300 EPs by telling subjects to focus on a card of their choice, and then showed cards in intervals of 200 ms to see which one evoked the P300 EPs that correspond to focusing. The best model for the MOABB dataset and for the 4-card dataset (with average accuracies of 0.925 and 0.865) used Xdawn and ERP covariances for spatial filtering and Logistic Regression for classification. Lastly, we tested this model for the real-time experiment, obtaining promising results that can be improved in the future.

1. Introduction

Our goal is to develop a Brain-Computer Interface (BCI) system that predicts the card a subject is focusing on from brain signals recorded with a Muse-S headband.

BCI systems record brain signals, send them to a computer, and translate them into commands [1]. Their applications are broad. They appear in fields such as medicine, smart environments, education, and neuromarketing. A BCI system has four major components: data recording (invasively or non-invasively), data pre-processing (noise reduction and signal enhancement), feature extraction, and classification [2]. In our case, we have chosen the Muse-S headband, a non-invasive electroencephalogram (EEG), and we have extracted, filtered, and classified the data using machine learning and Riemannian Geometry.

There are many different technologies to record data non-invasively. The most popular are: functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG), and electroencephalograms (EEGs). The Muse-S headband, for example, is a non-invasive EEG that records alpha, beta, theta, delta, and gamma brain waves [3]. In comparison with other devices, it is small, cheap, and has a high temporal resolution. An EEG records brain waves with electrodes placed on the scalp of the user [2] that can analysed using two different approaches: evoked potentials (EP), an approach that studies the effect of the different conditions that trigger the electrical signals [2], and event related desynchronization/synchronization (ERD/ ERS), and approach that detects the brain oscillations that are not necessarily related to external stimuli [2]. Since in our experiments the user is actively focusing on a card, we are interested in EP signals, which are the electrical response that is observable in the recorded brain data after being triggered by a stimulus. In particular, we have focused on the P300 component of event-related potential (ERP) signals, a pattern detectable approximately 300 ms after the presence of a stimulus.

A BCI system poses several challenges. The signals have noise from other brain activity not related to the experiment [2], are nonstationary ([4] and [5]), and change over time for different sessions (at times within the same session) and between subjects. These problems, however, can be solved. The noisiness can be decreased by improving the signal level, decreasing the noise ratio, and applying frequency filters. Transfer Learning techniques can be used to calibrate new data with old data [6]. Furthermore, Riemannian Geometry can be used to extract the features of the signals and to classify them [7].

The remainder of the paper goes as follows: Section 2 describes the equipment used to record the EEG signals; Section 3 describes how the BCI system works; Section 4 describes the experiment; Section 5 describes the datasets; Section 6 describes the models; Section 7 illustrates the results of the proposed models for both offline and real-time classification; and Section 8 summarises the project and presents some future improvements.

2. Equipment

Figure 1. The Muse-S headband [3].

The EEG signals have been recorded with a Muse-S (illustrated in Figure 1). The Muse-S is a low-cost, portable headband broadly used for guided meditation and sleeping that has shown promising results measuring Event-Related

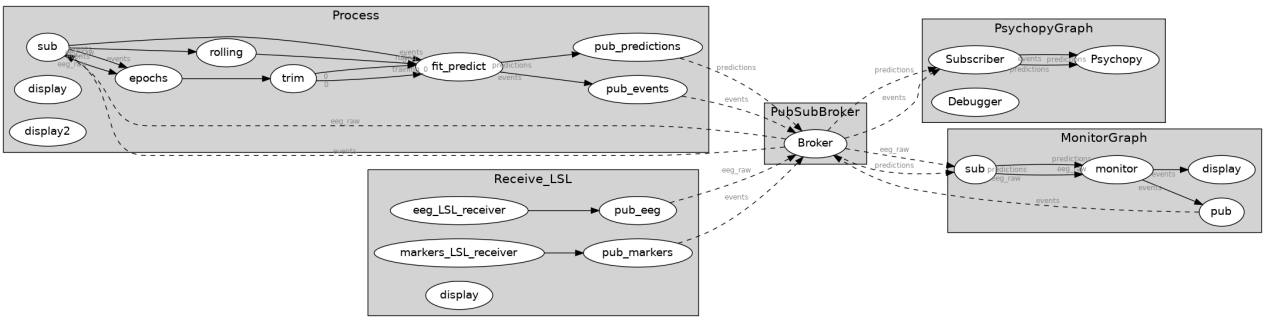

Figure 2. A graph illustrating the structure of the code used for the experiments.

Potentials (ERPs), and, in particular, the P300 component [3]. The Muse-S records alpha, beta, theta, delta, and gamma brain waves with two sensors placed on the forehead and one sensor on top of each ear. They are AF7, AF8, TP9, and TP10. It also uses the Fpz sensor for reference.

3. BCI

The BCI consists mainly of a receiver and an analyser connected by a broker that sends data between them. Figure 2 illustrates the structure of the code used for the experiments.

Receiver

The Muse-S uses the Lab Streaming Layer (LSL) protocol to manage the networking, the time-synchronization, the real time access, the centralized collection, the viewing, and the recording of the EEG signals [8]. Thus, we have used the *muselsl* Python library to acquire the EEG signals broadcasted by the Muse-S, and the *timeflux* Python library to process them in real-time [9].

Figure 2 illustrates the structure of the receiver in the Receive_LSL process. The left nodes (eeg_LSL_receiver and markers_LSL_receiver) receive the data from the LSL stream, and the right nodes publish the raw EEG signals to the broker to deliver them to the other processes.

Analyser

Figure 2 also illustrates the structure of the analyser, the part of the code that trains and tests the models. Its main node is fit_predict. In it, the selected model is trained with the EEG signals broadcasted by the receiver and, once trained, it sends the predictions to the real-time experiment.

4. Experiment

The experiments are performed in a calm room. An assistant places the Muse-S headband and, at times, noise-cancelling headphones on the subject, and instructs him to stay as calm and as static as possible. The assistant then reads the frequency of the signals and only starts the experiment when the frequency of all four signals is stable. Then, the assistant gives the subject a computer and runs the experiment. The experiment is self-explanatory to avoid further interactions between the assistant and the subject.

The experiment consists of showing a random card every 200 ms (with a random standard deviation of 100 ms) and

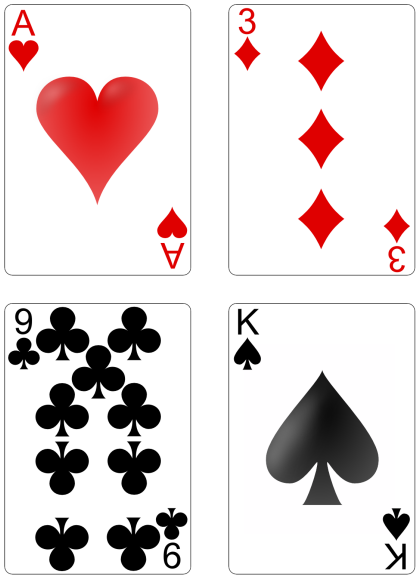

Figure 3. The four cards being used for the experiment.

analysing how the subject responds when his card is shown. The deck consists of four cards (illustrated in Figure 3) that have been chosen to be as different as possible: the Ace of hearts, the 3 of diamonds, the King of spades, and the 9 of clubs.

In the first part of the experiment the program instructs the subject to count the number of times the Ace of hearts appears. The Ace of hearts appears randomly, but to avoid signal overlapping, the program waits for at least three other cards to be shown between each Ace of hearts. The signals are then collected, labelled (with weather the card shown was the Ace of hearts or not), and sent to the analyser as training data. This process is repeated for ten runs, each of them showing thirty cards. After each run the subject is reminded to count the number of times the Ace of hearts appears.

In the second part of the experiment the program presents the four cards in the deck and instructs the subject to choose one. The program then shows cards randomly, sends the signals to the model, and displays the prediction in real time.

The experiment takes approximately ten minutes.

5. Datasets

The project uses two different datasets: the MOABB dataset, used to test the different feature extraction and classification methods, and the 4-card dataset, used to evaluate these methods in an environment similar to the real-time experiments.

MOABB dataset

The MOABB dataset is the BNCI2014009 dataset developed by MOABB, an open-source framework that aims to "build a comprehensive benchmark of popular BCI algorithms applied on an extensive list of freely available EEG datasets" [10]. It consists of P300 evoked potentials recorded in three different sessions from ten healthy subjects with experience in recording EEG data. For simplicity, however, we have reduced the dataset to four subjects.

The subjects were asked to focus on one out of 36 different characters and the P300 evoked potentials were recorded at 256 Hz with a low-pass, high-pass filter (with cutoff frequencies 0.1 Hz and 20 Hz) [11].

4-card dataset

The 4-card dataset consists of the P300 evoked potentials recorded in a session for two different subjects. Each P300 evoked potential is labelled with weather the subject was focusing on the card producing it, label 0, or not, label 1.

Label	Subject 1	Subject 2
Focused	52	44
Unfocused	308	256
Total	360	300

Table 1. Labels of the entries in the 4-card dataset.

The distribution of the labels in the 4-card dataset (illustrated in Table 1) is, roughly, 1:6. It is uneven because the subject is asked to only focus on one out of the four cards shown and, furthermore, the program waits for at least three unfocused cards to be shown after every focused card (always the Ace of hearts).

6. Models

There are many different techniques to extract the feature vector and classify the signals from an EEG. Riemannian Geometry, for example, has shown promising results in recent years [12].

Again, to classify EEG signals, one has to: pre-process the signal, extract the sources with spatial filters, extract the feature vector, and classify the feature vector using a vector-based classifier [7]. All of these steps (but specially the second) involve covariance matrices and they can be simplified with Riemannian Geometry. The steps in a Riemannian Geometry approach are: pre-process the signal, represent it as a covariance matrix, and classify it using Riemannian Geometry-based covariance matrices [7].

Filters

We have used four different filters: ERP covariances, Xdawn, Common Spatial Pattern (CSP) algorithm, and Power Spectral density (PSD).

Firstly, ERP covariances are an estimation of a special form covariance matrix dedicated to ERP processing [13]. They take the mean of trial components X_i to build a prototyped response P, and then compute a super trial $\widetilde{X}_i$ by concatenating P and X. The covariance estimation uses the super trial to take into account the spatial structure of the signal [13]. Secondly, Xdawn is an unsupervised method that enhances "P300 evoked potentials by estimating spatial filters" [14]. It does this by projecting the raw input signal on the estimated evoked space. Lastly, the CSP algorithm maximises the discriminability of two classes by using spatial filters [15], and the PSD estimator "represents the proportion of the total signal power contributed by each frequency component of a voltage signal" [16].

Classifiers

We have used four different classifiers: Logistic Regression, K-nearest neighbours (KNN), Linear Discriminant Analysis (LDA), and Minimum Distance to Mean (MDM).

Logistic Regression and KNN are traditional classifiers broadly used that do not require an explanation. LDA is commonly used in supervised classification tasks. It reduces the dimension of the input space by finding an optimal linear transformation to differentiate distinct classes [17]. In EEG experiments it is normally used to solve the nonstationarity of the EEG signals.

MDM classifies the results by the nearest centroid [18]. It computes a centroid estimation with respect to a certain metric for all known classes and, for the new points, it uses the nearest centroid.

7. Results

MOABB dataset

The models use: (1) Logistic Regression for classification, (2) Xdawn for spatial filtering and K-nearest neighbours (KNN) for classification, (3) Xdawn for spatial filtering and Logistic Regression for classification, (4) Xdawn and ERP covariances for spatial filtering and Logistic Regression for classification, (5) CSP for spatial filtering and LDA for classification, and (6) ERP covarainces for spatial filtering and MDM for classification.

The models were trained on Subject 1 and evaluated on Subjects 1, 2, 3, and 4 (illustrated in Table 2) with data from all three different sessions. The best model is (4), the model that uses Xdawn and ERP covariances for spatial filtering and Logistic Regression for classification. It is the most accurate model for all four subjects, with accuracies of 93%, 96%, 86% and 95%. The second best model is (3), the model that uses Xdawn for spatial filtering and Logistic Regression for classification. The only model that is not suitable for the MOABB dataset is (5), the model that uses CSP for spatial filtering and LDA for classification.

The accuracy of every model is the lowest for Subject 3, suggesting that the data for Subject 3 is more noisy than for the other Subjects.

Model	Filter	Classifier	Subject 1	Subject 2	Subject 3	Subject 4
1	None	LR	0.90	0.95	0.82	0.94
2	Xdawn	KNN	0.80	0.87	0.71	0.89
3	Xdawn	LR	0.87	0.93	0.77	0.92
4	ERP + Xdawn	LR	**0.93**	**0.96**	**0.86**	**0.95**
5	CSP	LDA	0.57	0.61	0.57	0.58
6	ERP	MDM	**0.93**	**0.96**	0.74	0.83

Table 2. Accuracy of the models for four subjects from the MOABB dataset.

4-card dataset

The models were trained on half of the data and evaluated on the other half. The results are illustrated in Table 3. The most accurate models for Theodora are the models: (3), which uses Xdawn for spatial filtering and Logistic Regression for classification; (4), which uses ERP covariances and Xdawn for spatial filtering and Logistic Regression for classification; and (6), which uses Xdawn for spatial filtering and LDA for classification. All of them have an accuracy of 91%. The most accurate models for Martin are the models: (2), which uses Xdawn for spatial filtering and KNN for classification; (4), which uses Xdawn and ERP covariances for spatial filtering and Logistic Regression for classification; (8), which uses ERP covariances for spatial filtering and Logistic Regression for classification; and (9), which uses PSD for spatial filtering and Logistic Regression for classification. All of them have an accuracy of 82%. Taking the different accuracies into account, however, the best model for the 4-card dataset is model (4), which uses Xdawn and ERP covariances for spatial filtering and Logistic Regression for classification. The least accurate model for both subjects is model (5), the model that uses ERP covaraince for spatial filtering and MDM for classification, with an accuracy of 79% for Theodora and of 71% for Martin.

Model	Filter	Classifier	Theodora	Martin
1	None	LR	0.90	0.76
2	Xdawn	KNN	0.90	**0.82**
3	Xdawn	LR	**0.91**	0.77
4	ERP + Xdawn	LR	**0.91**	**0.82**
6	ERP	MDM	0.79	0.71
7	Xdawn	LDA	**0.91**	0.80
8	ERP	LR	0.90	**0.82**
9	PSD	LR	0.84	**0.82**

Table 3. Accuracy of the models for the 4-card dataset.

Every model has a higher accuracy for Theodora than for Martin because the data is affected by the environment, the conditions, and the subject's overall mental state. Furthermore, as mentioned above, the 4-card dataset is unbalanced: there are roughly six unfocused entries for every focused entry. To ensure that model (4) is not over-fitted, we evaluate the confusion matrices for both subjects (illustrated in Figures 4 and 5). Theodora does not have false negatives and has 8 false positives, so 30% of the focused labels were misclassified; Martin does not have false negatives and has 11 false positives, so 40% of the focused labels were misclassified. Thus, model (4) is slightly over-fitted.

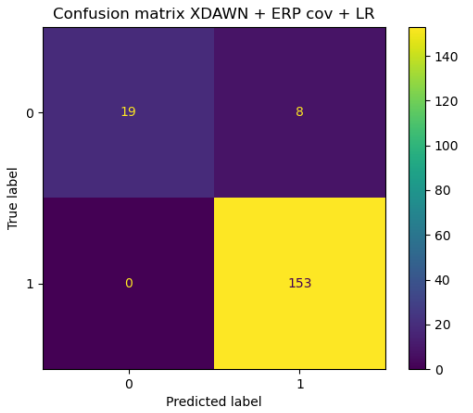

Figure 4. Confusion matrix of model (4) for Theodora.

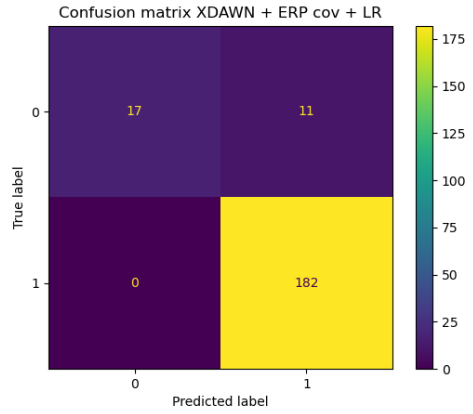

Figure 5. Confusion matrix of model (4) for Martin.

Real-time

In the real-time experiment we only used model (4), the model that uses ERP covariances and Xdawn for spatial filtering and Logistic Regression for classification, and that achieves the highest accuracies in both the MOABB dataset and the 4-card dataset.

We ran the real-time experiment in six different sessions for five different subjects (Subject 3 took part in the experiment twice). At the second session for Subject 3 and at the session for Subject 5, the BCI system did not make a prediction. At the rest of the sessions the BCI system made the predictions

illustrated in the Table 4. The fractions represent the number of focused predictions over the number of total appearances. The results differ greatly between subjects. The BCI system is confident about its prediction for Subject 1 and uncertain for Subjects 2, 3, and 4.

Explanation on why the results in real-time are so much worse.

Card	Subject 1	Subject 2	Subject 3	Subject 4
A.	1/85	0/40	1/85	1/85
3	0/76	**3/41**	0/76	0/69
K.	**15/71**	0/38	0/71	0/79
9	0/69	1/31	**2/68**	**1/67**
Total	16/300	4/150	3/300	2/300

Table 4. Real-time focused predictions over total appearances.

We also ran a cross-session experiment for Subject 1. We used data collected six days prior and, after adding only 60 new data-points, the program attempted to predict Subject's 1 card of choice. The results are illustrated in Table 5. They are quite similar to the ones obtained in a single session (illustrated in Table 5), meaning that Subject's 1 data can be reused without significantly decreasing the accuracy of the predictions.

Prediction	A.	3	K.	9	Total
Focused	2	**10**	0	0	12
Unfocused	83	**68**	73	64	288

Table 5. Real-time cross-session predictions for Subject 1.

8. Conclusion

We developed a BCI system to predict the card a subject is focusing on in real-time with the signals recorded by a Muse-S headband. We first tested different models with the MOABB dataset and achieved promising results. Then, we tested more models with the 4-card dataset, a dataset developed by us that consists of P300 evoked potentials, recorded with the Muse-S, and labelled with weather the subject was focusing on the card evoking them or not, and once again achieved promising results. The best model for both datasets used Xdawn and ERP covariances for spatial filtering and Logistic Regression for classification, and, therefore, it was the only model we used for the real-time experiment. The accuracy of the models was high in the MOABB dataset and in the 4-card dataset, and significantly lower in the real-time experiment.

The BCI system can be improved by switching the card a subject is asked to focus on at every run, by adding more cards, and by performing more real-time experiments. Furthermore, more spatial filtering techniques and classification methods could be used and the 4-card dataset could be extended to assist in choosing a better model. Finally, Transfer Learning techniques could be used to reduce the amount of data the BCI system records in the real-time experiment, and to make the data more robust to noise, movements, etc.

References

[1] D. J. K. Jerry J. Shih and J. R. Wolpawc. "Brain computer interfaces in medicine". *Mayo Clinic Proceedings* **87**:3 (2012), pp. 268–279. DOI: 10.1016/j.mayocp.2011.12.008. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497935/.

[2] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa. "Brain computer interfacing: applications and challenges". *Egyptian Informatics Journal* **16**:2 (2015), pp. 213–230. ISSN: 1110-8665. DOI: https://doi.org/10.1016/j.eij.2015.06.002. URL: https://www.sciencedirect.com/science/article/pii/S1110866515000237.

[3] *Muse headbands*. URL: https://choosemuse.com/how-it-works/?utm_source=Google&utm_medium=PaidSearch&utm_campaign=Non-Brand&gclid=CjwKCAjwjZmTBhB4EiwAynRmD1WYqQNv7PwqZBayIy8Kw%20s6y8Hepfn9sStjcpadZ8RMN0MYY9_MdqxoCE-0QAvD_BwE (visited on 2022-04-25).

[4] G. S. Matthias Kreuzer Eberhard F Kochs and D. Jordan. "Non-stationarity of eeg during wakefulness and anaesthesia: advantages of eeg permutation entropy monitoring". *Journal of clinical monitoring and computing* (2014). DOI: 10.1007/s10877-014-9553-y.

[5] R. W. T. Dikanev D. Smirnov, J. L. P. Velázquez, and B. Bezruchko. "Eeg nonstationarity during intracranially recorded seizures: statistical and dynamical analysis." *Clinical Neurophysiology* (2005). DOI: 10.1016/j.clinph.2005.04.013.

[6] C. J. Paolo Zanini Marco Congedo, S. Said, and Y. Berthoumieu. "Transfer learning: a riemannian geometry framework with applications to brain-computer interfaces". *IEEE Transactions on Biomedical Engineering* **65**:5 (2018), pp. 1107–1116. DOI: 10.1109/TBME.2017.2742541.

[7] F. L. Florian Yger Maxime Berar. "Riemannian approaches in brain-computer interfaces: a review". *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2017).

[8] *Lab streaming layer*. URL: https://labstreaminglayer.readthedocs.io/info/intro.html (visited on 2022-04-25).

[9] *Muse lsl*. URL: https://github.com/alexandrebarachant/muse-lsl (visited on 2022-04-25).

[10] A. Barachant and S. Chevallier. *Mother of all bci benchmarks*. URL: http://moabb.neurotechx.com/docs/index.html.

[11] *Moabb dataset bnci2014009*. URL: http://moabb.neurotechx.com/docs/generated/moabb.datasets.BNCI2014009.html (visited on 2022-04-25).

[12] R. B. Marco Congedo Alexandre Barachant. "Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review". *Brain-Computer Interface* (2017). DOI: 0.1080/2326263X.2017.1297192.

[13] *Erp covariances - pyriemann*. URL: https://pyriemann.readthedocs.io/en/latest/generated/pyriemann.estimation.ERPCovariances.html (visited on 2022-04-25).

[14] B. Rivet*, A. Souloumiac, V. Attina, and G. Gibert. "Xdawn algorithm to enhance evoked potentials: application to brain–computer interface". *IEEE Transactions on Biomedical Engineering* **56**:8 (2009), pp. 2035–2043. DOI: 10.1109/TBME.2009.2012869.

[15] Q. Ai, Q. Liu, W. Meng, and S. Q. Xie. "Chapter 6 - eeg-based brain intention recognition". In: Q. Ai et al. (Eds.). *Advanced Rehabilitative Technology*. Academic Press, 2018, pp. 135–166. ISBN: 978-0-12-814597-5. DOI: https://doi.org/10.1016/B978-0-12-814597-5.00006-0. URL: https://www.sciencedirect.com/science/article/pii/B9780128145975000060.

[16] J. DEMPSTER. "Chapter six - signal analysis and measurement". In: J. DEMPSTER (Ed.). *The Laboratory Computer*. Biological Techniques Series. Academic Press, London, 2001, pp. 136–171. DOI: https://doi.org/10.1016/B978-012209551-1/50039-8. URL: https://www.sciencedirect.com/science/article/pii/B9780122095511500398.

[17] E. Neto, F. Biessmann, H. Aurlien, H. Nordby, and T. Eichele. "Regularized linear discriminant analysis of eeg features in dementia patients". *Frontiers in Aging Neuroscience* **8** (2016). ISSN: 1663-4365. DOI: 10.3389/fnagi.2016.00273. URL: https://www.frontiersin.org/article/10.3389/fnagi.2016.00273.

[18] C. J. Marco Congedo Pedro Luiz Coelho Rodrigues. "The riemannian minimum distance to means field classifier". *BCI 2019 - 8th International Brain-Computer Interface Conference* (2019).