Classification of Imaginary Motor Task from Electroencephalographic Signals: A Comparison of Feature Selection Methods and Classification Algorithms

Clasificación de una Motricidad Imaginaria por medio de Señales Electroencefalograficas: Una Comparación de Métodos de Selección de Características y Algoritmos de Clasificación

H. J. Vélez-Lora, D. J. Méndez-Vásquez, J. F. Delgado-Saa

Fundación Universidad del Norte

ABSTRACT

In this work, a Brain Computer interface able to decode imagery motor task from EEG is presented. The method uses time-frequency representation of the brain signal recorded in different regions of the brain to extract important features. Principal Component Analysis and Sequential Forward Selection methods are compared in their ability to represent the feature set in a compact form, removing at the same time unnecessary information. Finally, two method based on machine learning are implemented for the task of classification. Results show that it is possible to decode the mental activity of the subjects with accuracy above 80%. Furthermore, visualization of the main components extracted from the brain signal allow for physiological insights on the activity that take place in the sensorimotor cortex during execution of imaginary movement of different parts of the body.

KEYWORDS: Sensorimotor Rhythms; BCI; Spectral Decomposition; Feature Selection; PCA; SFS; SVM; LDA.

RESUMEN

En este trabajo es presentada una Interfaz Cerebro Computadora que tiene la capacidad de decodificar actividades motrices. El método utiliza representación en el dominio de la frecuencia y el tiempo de las señales del cerebro grabadas en distintas regiones de este mismo, con el fin de extraer características importantes. Los métodos: Análisis de Componentes Principales y Selección Secuencial, son comparados en términos de su capacidad para representar características de la señal de una forma compacta, removiendo de esta forma, información innecesaria. Finalmente, dos métodos basados en aprendizaje de máquinas fueron implementados para la clasificación de actividades motrices utilizando solo las señales cerebrales. Los resultados muestran que es posible decodificar la actividad mental en los sujetos con una precisión superior al 80%. Además, la visualización de las componentes principales extraídas de las señales del cerebro permite un analísis de la actividad que toma lugar en la corteza cerebral sensorimotora durante la ejecución de la imaginación de movimientos de distintas partes del cuerpo.

PALABRAS CLAVE: Ritmos Sensorimotores, BCI, Descomposición Espectral, Selección de Características, PCA, SFS, SVM, LDA.

Correspondencia

DESTINATARIO: **Jaime Fernando Delgado Saa**
INSTITUCIÓN: **Fundación Universidad del Norte**
DIRECCIÓN: **Km. 5 Vía Puerto Colombia, Barranquilla, Atlántico, Colombia**
CORREO ELECTRÓNICO: **jadelgado@uninorte.edu.co**

INTRODUCTION

Brain computer interfaces (BCI) are systems that aim to establish a non muscular communication path between a person and the environment. BCI systems are aimed to population who suffer from motor disabilities, making use of different phenomena present in the brain activity [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. However, the application of BCI is not restricted only to rehabilitation. Recently, BCI has been used for entertainment in multimedia games with healthy people [1, 11]. Different methods for acquiring the brain signals can be used in BCIs, however for practical considerations the electroencephalogram (EEG) is most used. EEG has the advantage of being a non-invasive technique, of easy use, low risk for the subject and relatively cheap compared to other techniques as electrocorticography (ECoG) which is an invasive method that need of surgery for correct placement of the recording electrodes over the brain cortex.

Depending on the activity executed by a person, different potentials are elicited in the brain signal, such as P300, steady-state evoked potentials (SSVEP), slow cortical potentials (SCP) and sensorimotor rhythms (SMR). Those potentials can be used as features for decoding subjects intention and provide mechanism for interaction with the environment [12].

Of great interest is the SMR, these rhythms are observed as a decrease in the activity in the Alpha band (8-12 Hz) and Beta band (16-26 Hz) when the subject executes motor task. Interestingly the decrease in the power in the frequencies mentioned before is also observed when the subject imagines the movement, which is of especial interest in applications for people with degenerative diseases. Topographically, the decrease in power occurs over the brain areas known as the primary motor cortex and the sensorimotor cortex. Several methods have been proposed for extraction of the features of the SMR such as adaptive auto-regressive models (AAR) [13, 14],

common spatial pattern (CSP) [15], filter bank CSP (FBCSP) [16], Common Spatial Subspace Decomposition (CSSD) [17, 18] among others. After feature extraction, classification is performed using machine learning techniques. Of particular interest are the Linear Discriminant Analysis (LDA) due to its simplicity and the Support Vector Machines for its optimal performance. These two machine learning techniques dominate the BCI literature [19, 20, 12, 13, 5, 21, 14].
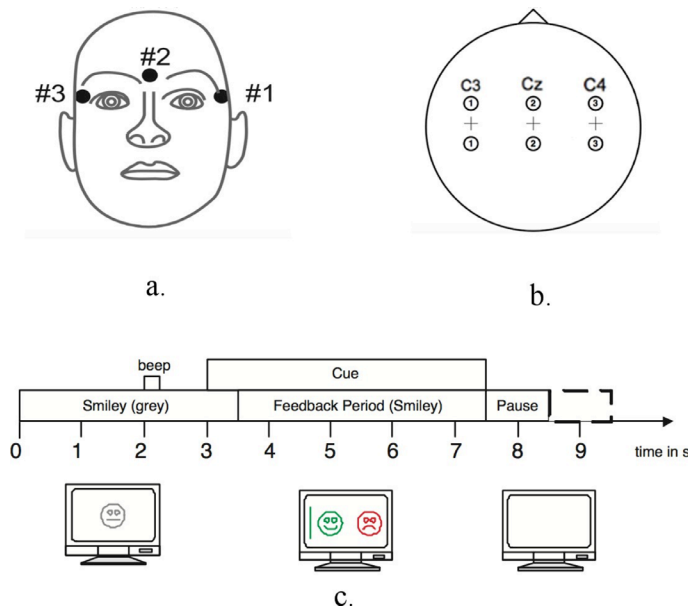
In this work a method based on short time Fourier Transform (STFT) is presented for feature extraction. Algorithms for feature selection based on Sequential Forward Selection (SFS) and Principal Component Analysis (PCA) are compared. Finally, classifiers based on LDA and SVM were implemented and compared in terms of performance.

METHODS

Data-set Description Data-set 2b from BCI competition IV was used for this analysis. The experimental paradigm was applied to nine subjects. Each subject was sitting in an armchair in front of a screen at the eye level from one meter of distance. There were five sessions, whereby the first two sessions were for training purpose. The other were feedback sessions, consequently the subject knew the immediate results of the imagery tasks. The imagery task concern to two classes, which were the motor imagery of left hand (class 1) and right hand (class 2). Given that the sessions were recorded in different days, only one of the sessions was used for each subjects. This was done due to the session transfer problem that is present given that the conditions of the experiment may change across sessions.

The description of the experiment can be seem in Figure 1. For each session, an electro-oculogram (EOG) was realized before starting the experiment in order to estimate its influence on the electroencephalogram signals (EEG). Each session is composed of six runs

with ten trials each per class, for a total of 120 trial for each class in the session used. Each trial began with a fixation cross and an acoustic sound that indicated the start of the trial. Then, a cue indicating the motor imagery task to be executed by the subject appeared. In parallel, the EOG signal was recorded together with the EEG, in order to use for removal of the artifacts caused by subject's eye movements.

The EEG signal was recorded through three bipolar electrodes centered on C3, Cz and C4 (see Figure 1) with a sample frequency of 250 *Hz*. The EEG recordings were bandpass filtered between 0.5 *Hz* and 100 Hz and the line component of 50 *Hz* was removed using a notch filter.

The electrode Fz was used as a signal reference. The EOG signal was recorded through three monopolar electrodes as displayed on Figure 1 (left mastoid serving as a reference for the signal). These channels were recorded in parallel with the EEG signal, with the same amplifier settings.

Feature Extraction

For feature extraction, the STFT was computed for each trial obtaining a representation of the changes of power with time for each frequency value. Since EEG signals are of non-stationary nature, the use of the STFT is required instead of the Fourier transform. The Fourier Transform (FT) is defined as:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$$

where ω means the angular frequency and $X(\omega)$ is the FT of $x(t)$. The STFT [22] computes the Fourier transform at different time intervals of the signal using a window function. In this way the new domain, which is the frequency domain, change over time. The STFT is defined as:

$$X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)g(t-\tau)e^{-j\omega t}dt$$

in this case τ refers to the corresponding time of the FT computed and $g(t)$ concern to the time window function. In this work a Hamming window was used. The time window function define the time resolution and the frequency resolution [20]. However there is a trade of between frequency resolution and time resolution. The best of compromise can be obtained using overlapping windows.

Feature Selection

After feature extraction using the STFT, two approaches of feature selection were applied to data.

Principal Component Analysis (PCA)

PCA is a dimensionality reduction method. Given a set of signals, PCA find a projection to a new space where the new total variance of the original data is mapped to a reduced number of components. The components are ranked according to the amount of variance retained. The component with the maximum

variance is termed the principal component. In this fashion, the data is represented in a compact form, reducing the number of features to be input to the classifier in subsequent stages. The k-th principal component can be computed as:

$$\hat{X}_k = X - \sum_{s=1}^{k-1} X w_{(s)} w_{(s)}^T$$

where X is the data matrix and w is the projection vector. The projection vector which extracts the maximum variance must fulfill the following expression:

$$w_k = argmax \left\{ \frac{w^T \hat{X}_k^T \hat{X}_k w}{w^T w} \right\}$$

It can be shown that the solution for this optimization problem is that the set of filters w that extract the principal components is in fact the eigenvectors of $X^T X$ [21].

Sequential Forward Selection (SFS)

This method uses an iterative approach for selection of relevant features. It initializes with an empty feature vector. Each feature is tested individually and the feature providing the higher improvement in classification is selected. Once the first feature is fixed, the process is repeated with the remaining features. This iterative process is repeated until when further inclusion of features does not incur in a increased performance [23, 20]. We selected a threshold of 1% as minimum accuracy increase in order to include a feature in the set of selected features.

Classification

Linear Discriminant Analysis (LDA)

Linear discriminant analysis is a method for classification based on linear separation of two or more classes. The operation principle is to find a straight line that divides two regions (in a two classes problem), such that depending on the location of a new test

sample the algorithm determine to which class the new sample belongs.

The straight line is ruled by the following equation

$$y(x) = w^T \cdot x + w_0$$

where w^T are denominated the weight vector, w_0 is a bias and x is the sample vector [19]. Basically, when a sample x_+ is located over the straight line $y(x_+) = 0$. Hence, $y(x)$ represents the distance r from the sample to the decision surface. Accordingly, r and $y(x)$ are related by

$$r = \frac{y(x)}{\|w\|}$$

r determine the class assigned to the sample. The parameters w^T and w_0 are learned from training data such that the error in classification is minimized.

Support Vector Machines (SVM)

This classification algorithm is based on projections onto a vector that is composed of the sum of particulars vectors denominated support vectors. It follows that the projection of any sample vector $\underline{u}$ onto the decision vector $\underline{w}$ has to be greater or lower than a constraint b.

$$f(u) = \vec{u} \cdot \vec{w} + b > 0$$

The decision vector must be orthogonal to the hyperplane which separates the classes. In order to achieve a good performance with this classifier, the width of the gap that separates classes must be maximized. After some algebra [21], the width of the gap to maximize is as follows:

$$Gap = \frac{2}{\|\vec{w}\|}$$

as a matter of fact, to maximize the width of the gap it is necessary to minimize the magnitude of _w_. Applying the Lagrange's method for optimization problems it is found that the decision vector has to be:

$$\vec{w} = \sum \alpha_i y_i \vec{x}_i$$

in this equation, the vector x_i is the group of samples vectors used for creating the decision vector _w_, they are called support vectors. The others are parameters, such as α_i which is the Lagrange multiplier and y_i is a parameter that indicate the class at which the support vector belongs. This classifier was used with a kernel function of type linear. The cost parameter (C) was trained using cross-validation. The formulation of C-Support vector classification (C-SVC) was used. Prior to the classification, training data were centered and scaled between [-1, 1]. For a fair validation, test data was also centered using the parameters learned over the training data [24].

RESULTS AND DISCUSSION

Average of the spectrogram of single trials, for all electrodes and each classes in subject 4 is shown in Figure 2 as an example of the features extracted using STFT. It is observed that the magnitude of the spectrum decreases at the electrode C3 around 10 _Hz_ and 20 _Hz_ when the subject execute the motor imagery task corresponding to the movement of the right hand. Similar activity is observed at electrode C4 in the same frequency bands when the subject executes the imagination activity referring to the left hand movement. This phenomena is observed across subjects and is termed in the literature Event Related De-synchronization (ERD). Based on this, the feature vector is composed of the average activity across time (between 3.5 to 5.5 _secs_) for each individual frequency. That is, the feature vector represent the power in each frequency band during 2 _seconds_ after the beginning of the motor task. The size of the feature vector is

defined by the frequency resolution used in the calculation of STFT. In this work a frequency resolution of 1 _Hz_ was used and the frequencies of interest are in the range of 2 _Hz_ to 30 _Hz_. This gives a features vector of 29 elements per electrode, which sum to a total of 87 features. Following the feature extraction, PCA and SFS (see Methods section) were tested for dimensionality reduction. These method aim to retain most of the information present in the feature vector but in a compact way. This is important because reducing the number of features avoid over fitting of the classifier.

Training and testing were implemented using a five-fold cross-validation procedure. The data was segmented in five folds. Four folds were used for training and one for testing. This is repeated five times giving statistical robustness to the results.

In the case of PCA, the projection to the new space is learned in the training set and used in the testing stage. In the case of SFS, the training set is further divided in 3 sub-folds to learn which features are relevant. Once this is learned, the selection is applied to the testing data.
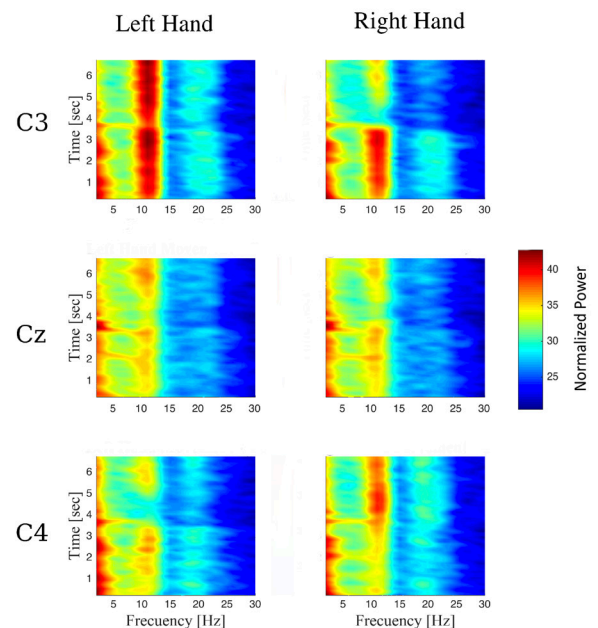

FIGURE 2. Spectrogram over time and frequency from S4.

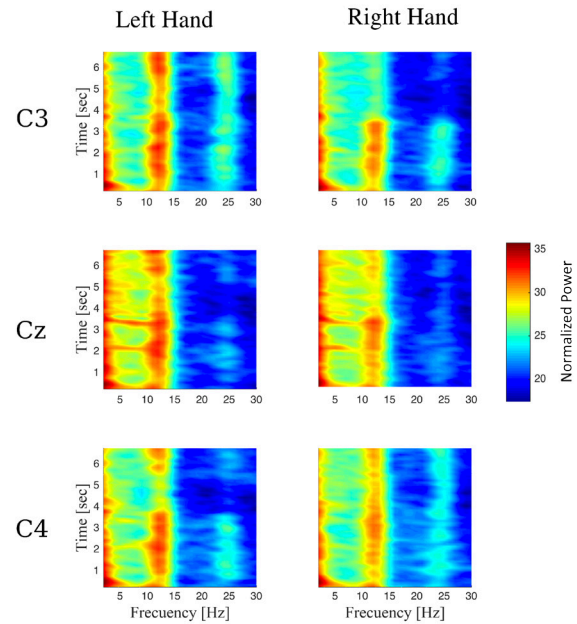**TABLE 1. Results for Cross-Validation Accuracy with and without PCA approach.**

	LDA (%)	LDA after PCA (%)	SVM (%)
S01	67.50	86.87	81.87
S04	82.50	91.87	90.62
S05	63.12	64.37	80.00
S06	68.12	79.37	72.50
S07	72.50	86.87	90.00
S08	73.75	83.75	84.37
S09	58.12	75.56	78.75
Average	69.37	81.23	82.58

Classification results using PCA are shown in Table 1. Using the original feature vector results in poor classification performance while using PCA produces a significant increase in the performance. In the case of LDA this can be explained by the nature of the classifier. LDA works under the assumption that the distribution of the features is Gaussian and that the covariance matrix of the feature vector is diagonal. This last condition imply uncorrelated features, a condition that is difficult to fulfill in practical problems. However, PCA projects the data to another space where the components are uncorrelated, that is, the covariance matrix is diagonal.

This makes the data more fitted to the assumptions of LDA and therefore better performance is achieved. In Table 1 the results obtained using SVM are also shown. As SVM is a discriminative method, no assumption is made about the distribution of the data. More important, the classification depends only on the support vectors (see methods) making the algorithm very robust, therefore PCA is not required for achieving good results with SVM.

SFS was also tested for feature selection. Results in Table 2 show the performance of LDA and SVM when SFS is used. Similar performances are observed comparing SFS to PCA in the case of LDA, and SVM does not seem to benefit from feature selection. To test whether the differences using PCA or SFS are significant, a paired t-test was performed. The results show no significant differences (at 0.05 level) between SFS and PCA in LDA or SVM.

FIGURE 3. Spectrogram over time and frequency from S9.

TABLE 2. Results for Cross-Validation Accuracy with SFS approach.

|	LDA (%)	SVM (%)
S01	78.12	80.00
S04	96.25	93.12
S05	83.75	80.00
S06	77.50	71.87
S07	87.50	86.25
S08	81.25	82.50
S09	83.12	83.75
Average	83.92	82.50

While no statistical differences are observed in between methods, SFS has the possibility of provide qualitative information about what features are more important. For instance, the average of the spectrogram for Subject 9 are shown in Figure 3. Note that the frequency band with strongest variation related to the

task is outside of the standard definition of Alpha band. Therefore fixing the features would have produced poor results for this subject. In order to visualize the spatial characteristic of the signals, the best feature selected using SFS was plotted in the scalp for each electrode. This is shown in Figure 4. The left and right images for each subject correspond to the left and right hand imagination respectively. These topographies are indicative of which electrode is more significant discriminant between classes. To illustrate this in a more particular way, lets take Subject 1 and 4 as examples. In Subject 1, the electrode Cz has more power or energy in relation to other electrodes, which means that has discriminant information for class one, meanwhile, class two is discriminated by electrode C4. On the other hand, in Subject 4 class one is discriminated by looking at the electrode C3 and the class two conserve the same discriminant electrode as subject one. Moreover, the SFS method selected the frequency bands around the Alpha band as main feature for all subjects. It is worth noting the inverse relationship between class and the hemisphere with strongest activity. The hemisphere with strongest activity is the one with lower power due to the nature of the Alpha rhythm. This inverse relationship is related to the physiology of the neural system where structures in the left hemisphere control the motor functions in the right part of the body and vice-versa. In addition, according to [25], the motor output leads to a continuous rhythmic sub-cortical discharge of neurons. Therefore, depending on the cortex activation, cortical cells would be affected doing the same activity. In this sense, Figure 4 also shows a pattern in all subjects about the location in the scalp of these discharges depending on the current imagination of the hand movement.

It is necessary to note that PCA will also provide a way to determine what is being extracted as main feature. The projection matrix w (see methods section) acts a spatial filter across features. However, given that the method has as priority to accommodate the highest amount of variance in few components, its operation is highly influenced by artifacts occurring in particular frequency bands, if such artifacts contain high amplitude.
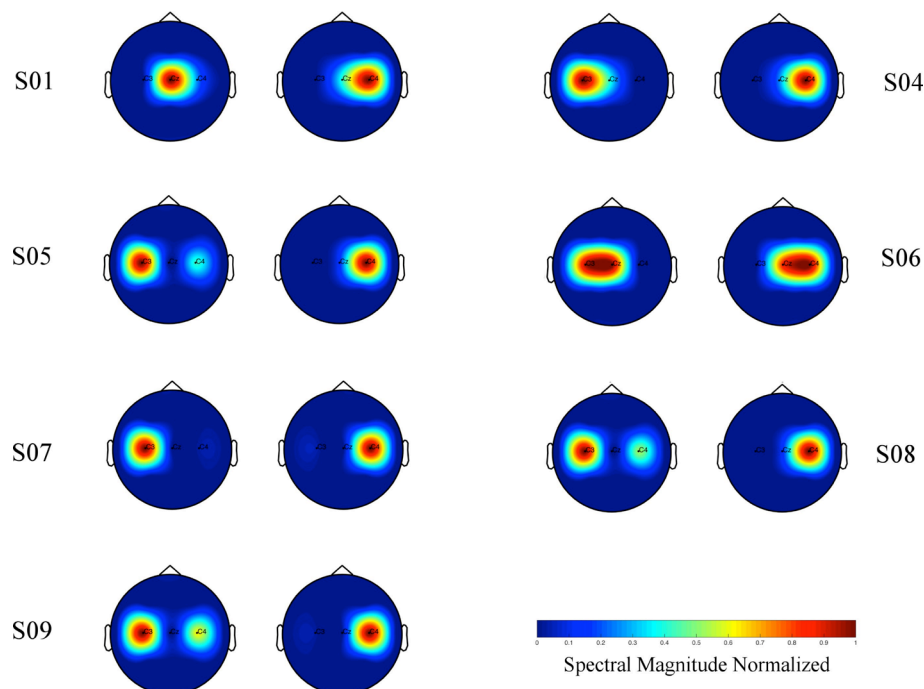
FIGURE 4. **Topographical distribution of energy.**

CONCLUSIONS

In this work, a BCI system able to classify two different types of imaginary motor movement is presented. Feature extraction is performed by Short Time Fourier Transform, providing a representation of the brain signals in the time-frequency domain. Feature selection methods based on Principal Component Analysis and Sequential Forward Selection are implemented and compared as methods for a compact representation of the information available in the EEG signals. After feature selection, machine learning algorithms are used for classification. LDA and SVM were compared as their use in the BCI community is extensive. Results show that it is possible classify the mental imagery of subjects with accuracy above 80%, these results as good as the winner algorithm of the BCI competition IV [26]. Also LDA obtain more benefits from the feature selection than SVM. This is mainly due to the nature of the classifiers. LDA makes assumptions that are not usually fulfilled in real data (Gaussianity and uncorrelated features) which affect the performance of the system. SVM on the other side can deal with non-Gaussian data and is less affected the size of the feature vector that are input to the classifiers. Beyond the performance obtained, information about the spatial characteristics of the brain activity were obtained by plotting the topographical distribution of the more important features detected by SFS.

Our future work will involve adding new features to the system that could improve the performance. Some features to consider would be the difference of the instantaneous phase of signals and the relationships among different frequencies and electrodes.

ACKNOWLEDGMENT

REFERENCES

[1] Laurent Bonnet, Fabien Lotte, and Anatole Lécuyer. Two brains, one game: design and evaluation of a multiuser bci video game based on motor imagery. IEEE Transactions on Computational Intelligence and AI in games, 5(2):185-198, 2013.

[2] J. F. Delgado Saa and M. Çetin. Discriminative methods for classification of asynchronous imaginary motor tasks from eeg data. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 21(5):716-724, Sept 2013.

[3] Jaime F Delgado Saa and Müjdat Çetin. A latent discriminative model-based approach for classification of imaginary motor task from eeg data. Journal of Neural Engineering, 2012.

[4] Jaime F. Delgado Saa, Adriana De Pesters, and Mujdat Çetin. Asynchronous decoding of finger movements from ecog signals using long-range dependencies conditional random fields. Journal of Neural Engineering, 13(3), 2016.

[5] Fabien Lotte. Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications. PhD thesis, INSA de Rennes, 2008.

[6] Fabien Lotte, Josef Faller, Christoph Guger, Yann Renard, Gert Pfurtscheller, Anatole Lécuyer, and Robert Leeb. Combining bci with virtual reality: towards new applications and improved bci. In Towards Practical Brain-Computer Interfaces, pages 197-220. Springer Berlin Heidelberg, 2012.

[7] D. McFarland, M. Parvaz, W. Sarnacki, R. Goldstein, and J. Wolpaw. Prediction of subjective ratings of emotional pictures by eeg features. Journal of Neural Engineering, 14:016009, 2017.

[8] Dennis J. McFarland, Janis Daly, Chadwick Boulay, and Muhammad A. Parvaz. Therapeutic applications of bci technologies. Brain-Computer Interfaces, 4(1-2):37-52, 2017.

[9] D.J. McFarland and T.M. Vaughan. Bci in practice. Progress in Brain Research, 228:389 - 404, 2016. Brain-Computer Interfaces: Lab Experiments to Real-World Applications.

[10] G Pfurtscheller and C Neuper. Motor imagery and direct brain-computer communication. Proceedings of the IEEE, 89(7):1123-1134, jul 2001.

[11] D Coyle, J Garcia, A R Satti, and T M McGinnity. EEG-based continuous control of a game using a 3 channel motor imagery BCI: BCI game. In 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), pages 1-7, apr 2011.

[12] G. Dornhege. Toward Brain-computer Interfacing. A Bradford book. MIT Press, 2007.

[13] S Lin, S Guo, and Z Huang. Determining AR order for BCI based on motor imagery. In 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), pages 174-178, oct 2015.

[14] Alois Schlögl, Felix Lee, Horst Bischof, and Gert Pfurtscheller. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. Journal of Neural Engineering, 2(4):L14, 2005.

[15] R Ortner, J Scharinger, A Lechner, and C Guger. How many people can control a motor imagery based BCI using common spatial patterns? In 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), pages 202-205, apr 2015.

[16] Chin Zheng Yang, Ang Kai Keng, Wang Chuanchu, Guan Cuntai, Zhang Haihong, Phua Kok Soon, Hamadicharef Brahim, and Tee Keng Peng. BCI Competition IV Dataset 2b, 2008.

[17] Huang Gan, Liu Guangquan, and Zhu Xiangyang. BCI Compition 2008, Dataset IIb, 2008.

[18] Damien Coyle, Abdul Satti, and Martin Mcginnity. BCI-competition IV - Data set IIa and IIb Algorithm Description, 2008.

[19] C Bishop. Pattern recognition and machine learning. Springer Verlag, 2006.

[20] Jaime F Delgado Saa. Probabilistic Graphical Models for Brain Computer Interfaces. PhD thesis, 2014.

[21] S Sanei and J A Chambers. EEG Signal Processing. Wiley, 2008.

[22] M.X. Cohen. Analyzing Neural Time Series Data: Theory and Practice. MIT Press, 2014.

[23] E. Alpaydin. Introduction to Machine Learning. Adaptive computation and machine learning. MIT Press, 2004.

[24] Chih-wei Hsu, Chih-chung Chang, and Chih-jen Lin. A Practical Guide to Support Vector Classification. 1(1):1-16, 2016.

[25] William Mackay. 7 wheels of motion: Oscillatory potentials in the motor cortex. 01 2005.

[26] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b. Frontiers in Neuroscience, 6:39, mar 2012.