OCULAR ARTIFACT REMOVAL FROM EEG: A COMPARISON OF SUBSPACE PROJECTION AND ADAPTIVE FILTERING METHODS.

Eleni Kroupi[1], *Ashkan Yazdani*[1], *Jean-Marc Vesin*[2], *Touradj Ebrahimi*[1]

[1]Multimedia Signal Processing Group, École Polythechnique Fédérale de Lausanne-EPFL
[2]Applied Signal Processing Group, École Polythechnique Fédérale de Lausanne-EPFL
CH-1015 Lausanne, Switzerland

ABSTRACT

One of the fundamental challenges in EEG signal processing is the selection of a proper method to correct ocular artifacts in the recorded electroencephalogram (EEG). Several methods have been proposed for this task. Among these methods, two main categories, namely subspace projection and adaptive filtering, have gained more popularity and are widely used in EEG processing applications. The main objective of this paper is to perform a comparative study of the performances of these methods using two measures, namely the mean square error (MSE) and the computational time of each algorithm. According to this study, ICA (independent component analysis) methods appear to be the most robust but not the fastest ones. Hence, they could be easily used for off-line applications. Moreover, PCA (principal component analysis) is very fast, but less accurate, so it could be used for real-time applications. Finally, adaptive filtering appears to have the worst performance in terms of accuracy, but it is very fast. Therefore, it could be also used for real-time applications, in which speed matters more than accuracy.

1. INTRODUCTION

Electroencephalographic (EEG) signal processing is becoming more and more important in studying brain functionality, especially for clinical purposes such as the diagnosis of brain disorders. Furthermore, EEG based brain-computer interface (BCI) systems, which aim at establishing an alternative communication channel between the brain and machines or peripheral devices, have attracted an increasing interest in the recent years. Thanks to the reasonable price and the decent time resolution of EEG acquisition devices, this signal is now the main tool in brain research. Thus, developing and implementing advanced signal processing tools and denoising algorithms for the analysis of the EEG signals is crucial.

The principal bottleneck in EEG signal processing stands in the identification and removal of the undesirable non-cerebral artifacts that interfere with the neuronal activity of the brain and can be mistakenly taken as brain activity patterns. The most common examples of such artifacts are the electro-oculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) emanating from ocular, muscular, and cardiac activities, respectively, as well as other external artifactual distortions induced by equipments, environment, etc. Among these artifacts, the most frequent and intrusive is the EOG, generated by the changes of the electric field of the scalp during blinking and eye movements. These involuntary eye movements and blinking often cause significant EOG artifacts especially in the frontal and central regions of the cortex and subsequently result in the loss of recorded data. EMG artifacts can be minimized by training the subjects to avoid any head movement and facial expression during the test, however, it is infeasible for most people to control their eye movement and avoid blinking. Therefore, the development of appropriate EOG artifact removal methods is a necessity in EEG studies.

Basically, the approaches proposed for dealing with EOG-contaminated EEG signals fall into two main categories, namely rejection or correction of the contaminated signal segments. In rejection methods, the recorded data is often scanned by an expert and artifactual EEG segments are excluded from the data. This approach is frequently used in medical research due to its simplicity. However, it can be a cumbersome procedure when analyzing long data and many non-obvious artifacts can be neglected. The other drawback of the rejection method is data reduction and loss, which restrains the use of this method.

Conventional EOG artifact correction methods include subspace projection, such as independent component analysis (ICA) and principal component analysis (PCA), which decompose the signal into independent and uncorrelated components, respectively. Moreover, regression-based methods have been proposed for artifact correction. For instance, adaptive filtering techniques [1] remove the contaminated parts of the signal by producing an optimized estimate of the original source. Finally, methods based on wavelet decomposition have been developed and used to correct ocular artifacts [2].

Since different artifact removal algorithms are required for each application, it is essential to chose the most appropriate one depending on the problem. For instance, in real time applications of EEG signals such as BCI systems, epileptic seizure detection etc., fast online denoising techniques are required. On the contrary, in off-line analysis of EEG signals, algorithms with better performance are often preferred regardless of how computationally complex and time-consuming they are. Therefore, comparing the efficiency and speed of the various proposed methods for artifact removal of the EEG signal could be of great interest. Nevertheless, so far only few research studies have been conducted in this regard [3, 4, 5]. In [3], time domain regression methods are compared with frequency domain regression methods, while in [4] the comparison is limited to time regression-based techniques. More precisely, an adaptive filtering method is compared with a simple time regression method. In [5] the compared techniques include two methods based on regression analysis (with and without adaptive filtering), an automated PCA, a manual PCA, and a method based on manual selection of the artifacts using ICA. The emphasis is given, however, to the implications of these methods on the spectral domain of the EEG signal.

In the current study, we compare the most conventional regression-based and the most conventional decomposition procedures using simulated EEG data. More specifically, the decomposition techniques include the commonly used ICA methods (SOBI [6], Infomax [7] and FastICA [8]) and an ICA method based on Bayesian learning and variational approximation [9]. The latter is used in order to investigate the performance of an ICA method that takes into consideration a possible additional Gaussian noise. The performance of these techniques is then compared with that of PCA and of the recursive least square (RLS) adaptive filtering that has a better convergence than the simple least mean square (LMS) adaptive filtering. Our aim is to present and compare the performance of these algorithms both in terms of accuracy and of computational time so as to give an insight on the methods that can be accurately used for real-time applications, off-line analysis or both.

The paper is organized as follows. Section 2 reviews the methods that will be used. Section 3 describes the data used in this study. The results and the further discussion are detailed in Section 4. Finally, the conclusions are presented in Section 5.

2. METHODS

2.1 Independent Component Analysis (ICA)

Blind source separation (BSS) [10, 11] has received considerable attention due to the fact that it addresses the significant problem of finding a suitable representation of multivariate data. ICA is a popular method for BSS using the assumption that the original sources are non-Gaussian, mutually independent, and the measurements are a linear transformation of the original sources. Under these assumptions the ICA problem can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{u}, \tag{1}$$

where the observed variables ($\mathbf{y}$) of dimension S are modeled as a linear combination of the statistically independent sources ($\mathbf{x}$) of dimension L with added S-dimensional Gaussian noise $\mathbf{u}$. $\mathbf{A}$ is an SxL mixing matrix. If the noise is not taken into consideration, the BSS problem represents the simplest case for ICA.

2.1.1 FastICA

FastICA is a fast algorithm for ICA, which can be used for BSS and feature extraction. It was initially introduced in [8]. This algorithm is based on a fixed-point iteration scheme and maximizes the non-Gaussianity as a measure of statistical independence. This algorithm is publically available in the FastICA package[1].

2.1.2 SOBI

Introduced by Belouchrani *et. al* [6], the second-order blind identification (SOBI) algorithm is a blind source separation technique for temporally correlated sources. More specifically, it is based on the joint diagonalization of an arbitrary set of covariance matrices. Therefore, it relies only on second-order statistics of the processed signals, in contrast to FastICA, which uses high order cumulant techniques. The algorithm is implemented in the publically available EEGLAB toolbox[2].

2.1.3 RunICA

Bell and Sejnowski proposed the Infomax ICA algorithm [7] based on a maximization of the mutual information between the sources and the sensors. Cardoso [12] showed that information-maximization is actually identical to minimization of the KL-divergence between the distribution of the output vector and the sources vector. Finally, in the same research study, the Infomax principle was shown to coincide with the maximum-likelihood principle in the case of source separation. RunICA algorithm, which is implemented in EEGLAB, performs ICA decomposition of input data using the Infomax ICA algorithm.

For these algorithms the sensors depend deterministically on the independent sources. In other words, once the mixing matrix $\mathbf{A}$ is found, the sources can be recovered exactly from the observed data, using the inverse or the pseudo-inverse of $\mathbf{A}$ [10].

2.1.4 Variational Bayesian Independent Component Analysis

In variational Bayesian independent component analysis (VbICA) [9], the noise is assumed to be Gaussian, with zero mean and diagonal precision matrix $\mathbf{\Lambda}$ and the distributions of the sources are represented by Mixtures of Gaussians (MoGs) [9]. A MoG is in general of the form:

$$p(x) = \sum_k \pi_k \mathcal{N}(x|\mu_k, \beta_k^{-1}), \tag{2}$$

where π_k, μ_k and β_k are respectively the weights, mean vectors and precision matrices of the Gaussian components. The variational Bayesian methods in general aim at approximating intractable posterior distributions by finding an appropriate distribution over the parameters and the latent variables, such that it factorizes [9, 10]. This factorization consists of the priors over the variables and the parameters. The variational approach of the algorithm lies on the fact that each factor of the approximate distribution iterates individually until convergence, in order to minimize the Kullback-Leibler (KL) divergence between the true and the approximate distributions. The VbICA algorithm used in this study is publicly available [3].

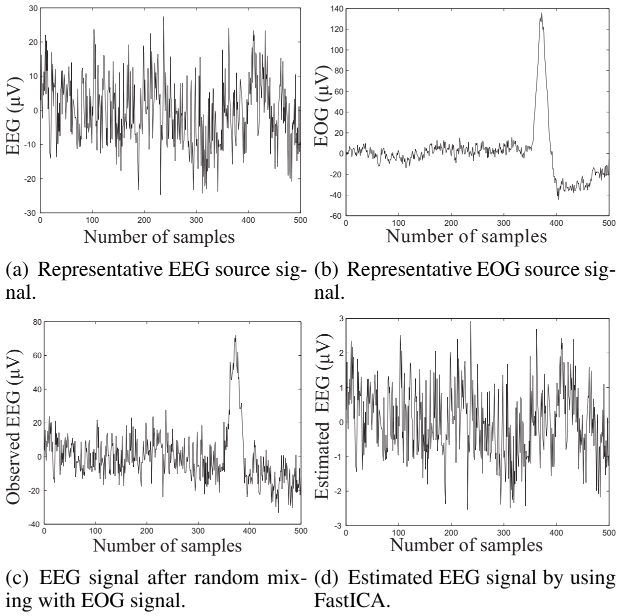
2.2 Principal Component Analysis (PCA)

PCA is another approach to construct the mixing matrix $\mathbf{A}$ and decompose the data into spatial components. In this technique, the coefficient vectors are the normalized eigenvectors of the covariance matrix. They are sorted according to the variance, with the first component having the largest variance. The coefficient vectors constructed in this manner are orthogonal. The main difference between PCA and ICA is that the sources are assumed to be uncorrelated and not statistically independent.

[1] http://www.cis.hut.fi/projects/ica/fastica/

[2] http://sccn.ucsd.edu/eeglab/

[3] http://www.robots.ox.ac.uk/parg/projects/ica/riz/code.html



(a) Representative EEG source signal.

(b) Representative EOG source signal.

(c) EEG signal after random mixing with EOG signal.

(d) Estimated EEG signal by using FastICA.

Figure 1: Example of two signals selected to be the sources, one possible observable EEG (which is a mixture of the sources) and the illustration of the initial clean EEG with the estimate of the FastICA .

2.3 Adaptive Filtering

In this study the RLS algorithm was used to determine the coefficients of the adaptive filter. Briefly, the RLS algorithm takes into account information from all the past input samples to estimate the autocorrelation matrix of the input vector. The goal of this algorithm is to produce an estimated signal which is as similar to the original source as possible, by adjusting the filter coefficients. The merit of this algorithm (against other adaptive filtering techniques) is the fact that it estimates recursively the filter coefficients that minimize a deterministic objective function [1, 13].

3. DATA

The data set used in this study was obtained from the BCI Competition of 2008 [14]. It consists of EEG data from nine subjects, who performed a paradigm of four different motor imagery tasks, namely the imagination of movement of the left hand, right hand, both feet and tongue. This publicly available dataset is already preprocessed (band-pass filtering and line noise suppression). Three monopolar EOG channels were also recorded and preprocessed with the same techniques.

In order to compare the different methods of artifact correction, the signals were separated into different trials and 40 EEG segments as well as 40 representative EOG segments were carefully selected. The duration of these segments was two seconds. The signals were then mixed one by one using a random mixing matrix of the form

$$\mathbf{W} = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}, \tag{3}$$

where α is the random attenuation constant of the EOG (different for each mixing procedure), taking values between 0 and 1. Since α should avoid taking values very close to 0 or to 1 (in order for the mixture to resemble as much as possible the real contaminated EEG), ten different random α values were created for each mixing procedure and their mean value was used for each mixing. The values were selected to be different for each mixing case in order to take into consideration the attenuation of the EOG effect with the distance to the eye-dipole. The EEG segments were selected from the Pz electrode, according to the 10-20 International System. The experiment was repeated with signals from other electrodes, but since clean EEG signals were manually selected, the results were not significantly different. A source EEG signal is shown in Figure 1(a) and a source EOG signal in Figure 1(b). A random mixing matrix was then created following the above-mentioned procedure, to enable the construction of the contaminated (observed) EEG signals;

$$\begin{bmatrix} \text{EEGobs} \\ \text{EOG} \end{bmatrix} = W \begin{bmatrix} \text{EEG} \\ \text{EOG} \end{bmatrix}. \tag{4}$$

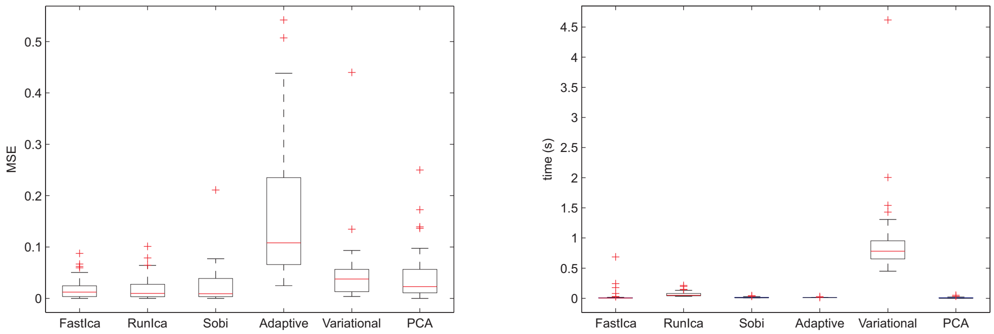
In (4), EEGobs is the observed EEG segment. A typical observed EEG signal created using equation 4 is shown in Figure 1(c). Data were artificially mixed for this study because the whole concept lies on the evaluation of the results that are generated using the different artifact removal methods. Therefore, real EEG data could not be taken into consideration due to the fact that they are always contaminated with EOG artifacts, and thus, there is no clean source that can be used for evaluating the denoised signals.

Finally, regarding the decomposition methods, the final selection of the estimated EEG signal was performed by comparing the correlation coefficient of each decomposed signal with the corresponding source (clean) EEG signal. The signal with the highest absolute correlation was then selected. A normalized (mean = 0 and standard deviation = 1) estimated EEG signal using FastICA is represented in Figure 1(d). The normalization was done due to the fact that ICA methods scale the data randomly. For the comparison of the true signals with the estimated ones, all of them were normalized in the same way.

4. RESULTS AND DISCUSSION

In order to compare the algorithms in terms of accuracy and speed, the mean square error (MSE) metric, as well as the computational time of each algorithm for one decomposition were used. The implementations were done on a Duo Processor T7500 with 2048MB RAM. The Bootstrap hypothesis test was applied to estimate the significant differences among the methods. This test was chosen due to the fact that it is a non-parametric significance test and thus it does not assume a Gaussian distribution for the data. Furthermore, the number of MoGs used for the variational Bayesian ICA was fixed to 5 mixtures, although the experiment was also performed using 3, 4, and 7 mixtures. Due to the fact that there were no significant differences, the results will be presented for the case of 5 MoGs. Finally, regarding the RLS adaptive filter, the parameters were set up in accordance with [1].

In Figure 2, the MSE and the elapsed time values of the six methods (FastICA, SOBI, RunICA, VbICA, adaptive filtering and PCA) are illustrated with box plots. Also, Table 1 presents the mean ± SE (standard error) of each method. According to the Bootstrap hypothesis test, there



(a) MSE of the different algorithms.

(b) Elapsed time of the different algorithms for one decomposition.

Figure 2: MSE and elapsed time to compare the algorithms.

is no significant difference among the three basic ICA methods (FastICA, SOBI and RunICA) in terms of their MSE, with $p > 0.05$. However, there is a significant difference (with $p < 0.05$) between FastICA and VbICA, which indicates that there is evidence against the null hypothesis that the MSE distributions of these two methods belong to the same population. FastICA outperforms VbICA in terms of MSE. Furthermore, there is an even more significant difference between RunICA and VbICA, with $p < 0.01$, whereas there is no difference between SOBI and VbICA ($p > 0.05$). According to the same hypothesis test, there is very significant difference between all the methods and adaptive filtering, with $p < 0.001$, but there is no significant difference between PCA and VbICA ($p > 0.05$). Regarding the performance of the methods in terms of the computational time, there are significant differences among all the methods, with $p < 0.001$ for every combination, apart from FastICA and SOBI ($p > 0.05$).

Methods	MSE	time (s)
FastICA	0.02 ± 0.003	0.02 ± 0.006
RunICA	0.02 ± 0.003	0.05 ± 0.008
SOBI	0.02 ± 0.006	0.02 ± 0.001
Adaptive RLS	0.2 ± 0.02	0.012 ± 0.0004
VbICA	0.03 ± 0.004	0.9 ± 0.05
PCA	0.04 ± 0.008	0.003 ± 0.0004

Table 1: Each value is the mean $\pm$ SE for both the MSE and the elapsed time (in sec).

Finally, the same experiment was repeated by adding zero-mean Gaussian noise with standard deviation SD = 0.5, 2, and 5 (signal-to-noise ratio, SNR = 20, 5 and 2 respectively). According to the Bootstrap hypothesis test, for any value of SD, the RLS adaptive filtering and the VbICA were significantly different from the other methods. However, there was no significant difference among the rest of the methods, but they all deteriorated with an increase of SD (see Figure 3). A difference was also observed regarding the computational time of RunICA, which deteriorated as well.

Therefore, taking into account both the significant differences (indicated by the $p-$values) and Table 1, common ICA techniques (FastICA, RunICA and SOBI) appear to be the most robust methods, VbICA and PCA are following in robustness, whereas adaptive filtering showed the poorest behavior in terms of accuracy. Improved performance could probably be obtained if two EOG channels were used instead of one to better describe the EOG artifacts. Nevertheless, in terms of computational time, adaptive filtering and PCA seem to be the fastest, while VbICA appears to have the slowest behavior. Regarding the conventional ICA methods, both FastICA and SOBI seem to be faster than RunICA. In addition, SOBI seems to be the most consistent among ICA methods since it differs significantly from the rest with a lower value of SE.

The performance of VbICA is quite poor in comparison with the other methods due to its large computational complexity. A possible explanation for diverging a lot from the other techniques, especially the other ICA techniques, is the fact that VbICA approaches the problem from a probabilistic point of view (since additional noise is assumed) and it integrates the decomposition into two parts. More specifically, the source densities, mixing matrix and noise covariance are estimated from the observed data and then the unmixing procedure is inferred. However, the rest of the ICA algorithms use a maximum likelihood approach to estimate the sources, but they reconstruct them deterministically, using directly the pseudo-inverse of the mixing matrix.

A noteworthy limitation of ICA algorithms is that they perform unsupervised learning and hence, the order of the estimated sources is random. Therefore, if the order of the sources has to be retained, another automated routine has to be implemented, in order to map the sources with their corresponding sensors. Obviously, this makes the use of ICA difficult for real-time applications. Moreover, the artifactual components have to be selected either manually or by implementing an automated routine. This issue can be solved by sorting the components in terms of variance maximization and reject the ones with the highest variance. Another automated way of artifact detection after ICA decomposition, is based on sorting the components in terms of power spectral density (PSD) maximization of the lowest frequencies. According to this technique, components with the highest PSD in the lower frequencies are excluded. However, the latter technique is time consuming and can be used only for off-line applications. Since PCA automatically ranges the

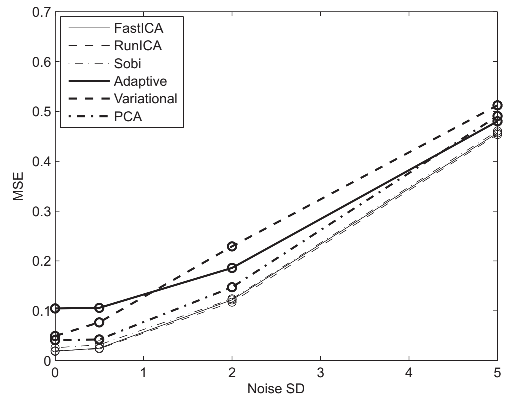


Figure 3: MSE of the different algorithms after adding zero-mean Gaussian noise with SD = 0, 0.5, 2, 5, respectively.

components based on variance maximization, it is easier to detect the artifactual components. In particular, the latter are observed in the main principal components, since they correspond to a greater variance. Obviously, this is not an issue for adaptive filtering, due to the fact that the denoising is performed individually for each sensor. Finally, ICA techniques scale the data randomly and hence extra information is needed in order to identify changes in the amplitude and sign of the components.

5. CONCLUSIONS

In this paper, subspace projection and adaptive filtering methods were used to eliminate EOG artifacts in recorded EEG signals. It was shown that conventional ICA methods do not seem to differ much in terms of accuracy, whereas FastICA and SOBI seem to be faster than RunICA, SOBI being the most consistent one. Since they are the most robust and quite fast, with SOBI having the lowest computational complexity, they could be used in real-time applications if the order of the sources is not important, or in off-line analysis otherwise. The same accounts for PCA which is quite fast and robust, so it could constitute an effective solution for both off-line and real-time applications. Finally, adaptive filtering could be also used for real-time applications, such as BCIs, although it has poorer performance in comparison with the others, and it would probably behave better by acquiring both vertical and horizontal EOG signals.

A possible extension to the current study could be to compare the performance of no-reference artifact correction algorithms in order to explore their behavior without the EOG channel. Due to the fact that in many research studies the EOG reference channel is not acquired, it is quite challenging to develop algorithms that can identify and remove the EOG artifacts. Moreover, the current study could be further extended in order to include signal decomposition methods for artifact manipulation, such as wavelet decomposition and empirical mode decomposition. According to these techniques, the decomposed signal has to be further processed in order to manipulate the artifactual components, by using either fixed thresholding methods or adaptive ones. The former do not perform well, due to the fact that they do not consider the properties of the different kinds of artifacts, since they are based on predefined thresholds. However, the latter seem quite promising, and it is very challenging to explore the behavior of adaptive algorithms considering both time and frequency characteristics of the signal.

REFERENCES

[1] P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering." *Med. Biol. Eng. Comput.*, vol. 42, pp. 407–412, 2004.

[2] V. Krishnaveni, S. Jayaraman, L. Anitha, and K. Ramadoss, "Removal of ocular artifacts from eeg using adaptive thresholding of wavelet coefficients." *J. Neural Eng.*, vol. 3, pp. 338–346, 2006.

[3] J. L. Kenemans, P. C. Molenaar, M. N. Verbaten, and J. L. Slangen, "Removal of the ocular artifact from the eeg: A comparison of time and frequency domain methods with simulated and real data," *Psychophysiology.*, vol. 28, pp. 114–121, 1991.

[4] P. He, G. Wilson, C. Russell, and M. Gerschutz, "Removal of ocular artifacts from the eeg: a comparison between time-domain regression method and adaptive filtering method using simulated data." *Med. Biol. Eng. Comput.*, vol. 45, pp. 495–503, 2007.

[5] G. L. Wallstrom, R. E. Kassb, A. Millerc, J. F. Cohnd, and N. A. Foxe, "Automatic correction of ocular artifacts in the eeg: a comparison of regression-based and component-based methods." *International Journal of Psychophysiology*, vol. 53, pp. 105–119, 2004.

[6] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics." *IEEE Transactions on Signal Processing*, vol. 45, pp. 434–444, 1997.

[7] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[8] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis." *Neural Computation*, vol. 9, pp. 1483–1492, 1997.

[9] R. A. Choudrey and S. J. Roberts, "Flexible bayesian independent component analysis for blind source separation." *In proceedings of ICA -2001, San Diego*, dec 2001.

[10] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, pp. 803–851, 1999.

[11] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications." *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, 2000.

[12] J. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Processing Letters*, vol. 4, pp. 112–114, 1997.

[13] S. Vaseghi, *Advanced signal processing and noise reduction.* 2nd edition, LinkChichester: Wiley, 2000.

[14] R. Leeb, C. Brunner, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, "Bci competition 2008-graz data set a." 2008.