Detection and Removal of Muscle artifacts from Scalp EEG Recordings in Patients with Epilepsy

Maria Anastasiadou[1], Avgis Hadjipapas[2], Manolis Christodoulakis[1], Eleftherios S. Papathanasiou[3], Savvas S. Papacostas[3] and Georgios D. Mitsis[1,4]

[1]Department of Electrical and Computer Engineering, University of Cyprus Nicosia, Cyprus (email: manast09@ ucy.ac.cy, manolisc@ucy.ac.cy)
[2]St George's University of London Medical Programme, Medical School, University of Nicosia, Nicosia, Cyprus (email: hadjipapas.a@unic.ac.cy)
[3]The Cyprus Institute for Neurology and Genetics, Nicosia, Cyprus (email: neurophy@cing.ac.cy, savvas@cing.ac.cy)
[4]Department of Bioengineering, McGill University, Montreal QC, Canada (email: gmitsis@ucy.ac.cy)

***Abstract*—The Electroencephalogram (EEG) is often contaminated by muscle artifacts. EEG is a widely used recording technique for the study of many brain related diseases such as epilepsy. The detection and removal of muscle artifacts from the EEG signal poses a real challenge and is crucial for the reliable interpretation of EEG-based quantitative measures. In this paper, an automatic method for detection and removal of muscle artifacts from scalp EEG recordings, based on canonical correlation analysis (CCA), is introduced. To this end we exploit the fact that the EEG signal may exhibit altered autocorrelation structure and spectral characteristics during periods when it is contaminated by muscle activity. Therefore, we design classifiers in order to automatically discriminate between contaminated and non-contaminated EEG epochs using features based on the aforementioned quantities and examine their performance on simulated data and in scalp EEG recordings obtained from patients with epilepsy.**

Keywords—Epilepsy; EEG; Muscle Artifacts; Blind Source Separation; Canonical Correlation Analysis

I. INTRODUCTION

The Electroencephalogram (EEG) is one of the most common recording techniques of brain activity. However, EEG does not only record brain activity but is frequently contaminated by non-cerebral electrical activity such as line noise, cardiac signals, eye blinks and movements and muscle contraction due to biting, chewing and frowning. There are several filters and other methods, for removing line noise, cardiac signals and eye blinks and movements [1]. Muscle artifacts, also known as Electromyographic (EMG) artifacts, have presented a real challenge over the years, since they overlap with brain activity over a wide frequency range. EMG contamination is particularly problematic in epilepsy where muscle artifacts are prevalent and also overlap the EEG signal and complicate its interpretation, often making this interpretation infeasible. Thus, EMG artifacts are a major obstacle in the characterization of electrophysiological properties [2]. Muscle artifact removal or reduction is an important problem that needs to be considered. Therefore, automatic detection of muscle activity and its subsequent removal is important, particularly for long duration EEG recordings.

Currently, there are a number of algorithms to remove muscle artifacts from physiological signals, including independent analysis (ICA) and canonical correlation analysis (CCA) [3-4]. ICA [5], which separates the EEG into statistically independent components, has been applied, with good results, to ocular artifact removal [6-11]. However, in the case of muscle artifacts the removal was in many cases suboptimal [12].

CCA [13] has been recently used as a tool for removing EMG-related contamination from the EEG [4, 12]. The CCA method, as well as ICA, belongs to a group of data-driven techniques for solving the Blind Source Separation (BSS) problem. According to the CCA method, the BSS problem is solved by forcing sources to be mutually uncorrelated and maximally autocorrelated. Moreover, the CCA method is computationally much less intensive than ICA [12]. Muscle artifacts are characterized by high amplitude, a broad frequency spectrum and low autocorrelation values (resembling random noise more than the brain-related EEG signals). Therefore, they tend to exhibit lower autocorrelation values compared to cerebral EEG signals [12].

The EMG signal also exhibits topographical features that can be used in order to automatically detect muscle activity. The most common sources of EMG signal and therefore EEG muscle artifacts are the frontalis and temporalis muscles in the, frontal and central head regions [14]. The peak frequency of frontalis muscle contraction varies from 16 to 38 Hz, while for temporalis muscle contractions it ranges from 13 to 34 Hz [14]. Therefore, the spectral signal power fraction that resides in this range may be an important feature for detection of muscle activity in the EEG signal. In this paper, we present an automatic method for detection and removal of muscle artifacts from scalp EEG recordings in patients with epilepsy, by taking into account the above observations. Specifically, we exploit the fact that muscle artifacts exhibit lower autocorrelation values and higher relative power within the aforementioned range (13 to 38 Hz), compared to normal EEG signals. Therefore, by using these two quantities as features,

This work was partially supported by the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation (Project ΥΓΕΙΑ/ΔΥΓΕΙΑ/0609(ΒΕ)/11).

978-1-4799-7502-0/14 $31.00 © 2014 IEEE
DOI 10.1109/BIBE.2014.52

we design classifiers that automatically detect muscle activity and subsequently remove this signal from the EEG recordings.

II. METHODS

A. EEG Recordings and Preprocessing

The proposed method was tested on 30min epochs recorded from three patients with epilepsy, extracted from longer duration EEG recordings. The experimental data were collected at the Neurology Ward of the Cyprus Institute for Neurology and Genetics. Twenty-one electrodes with two additional anterotemporal electrodes were placed according to the 10-20 system. The data were recorded using an XLTEK video-EEG recording system with a sampling frequency of 200Hz. The data was band-pass filtered between 1 and 45Hz and the Temporal Decorrelation source SEParation (TDSEP) algorithm was subsequently applied to remove ocular artifacts using simultaneously recorded Electroocculogram (EOG) recordings (2 channels), as reference signals [3, 15].

TDSEP is based on the 'simultaneous diagonalisation of several time-delayed correlation matrices' [15]. Therefore because separation is based on the correlation of the sources, one of the advantages of TDSEP is that, it can separate signals whose amplitude distribution is Gaussian [3].

For this work, the selected 30 minutes segments for each patient (in which a seizure is included), were then segmented into 5-second non-overlapping windows. Each window was marked as 'noisy' if muscle artifacts were identified within its duration otherwise was marked as 'noise-free' (Fig. 2). The muscle artifacts were identified and marked by expert neurophysiologists (coauthors ESP and SSP). The labeling is used in order to assess whether the algorithm is able to detect the 'noisy and noise-free' windows in data segments used for testing (cross validation); however, in the future our goal is to reduce or remove the need for expert marking by performing unsupervised clustering.

B. Canonical Correlation Analysis

CCA [13] provides a way of measuring the linear relationship between two multidimensional variables. This technique solves the BSS problem by forcing the sources to be maximally autocorrelated and mutually uncorrelated [12]. Consider two multi-dimensional random variables **x** and **y**. The input signal is x and let y be a temporally delayed version of the input **x** (i.e., **y** (t)= **x** (t – 1)). Consider the linear combination of the two variables, $\mathbf{x} = \mathbf{w}_x^T (\mathbf{x} - \mathbf{x}')$ and $\mathbf{y} = \mathbf{w}_y^T (\mathbf{y} - \mathbf{y}')$, respectively. The correlation between x and y is given by:

$$\rho(x, y) = \frac{w_x^T \, C_{xy} \, w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}}, \quad (1)$$

where $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ are the within-set covariance matrices and $\mathbf{C}_{xy}$ is the between-sets covariance matrix. The largest canonical correlation is the maximum of ρ with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$. Optimization of (1) with respect to respect to $\mathbf{w}_x$ and $\mathbf{w}_y$, i.e., $\max_{w_x, w_y} \rho(\mathbf{x}, \mathbf{y})$, results in the following two eigenvalue problems [12, 16]:

$$\begin{cases} C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \widehat{w}_x = \rho^2 \widehat{w}_x \\ C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} \widehat{w}_y = \rho^2 \widehat{w}_y \end{cases} \quad (2)$$

Solving (2) gives N solutions $\{\boldsymbol{\rho}_n, \mathbf{w}_{xn}, \mathbf{w}_{yn}\}$, n={1…N}. N is the minimum of the dimensionalities of **x** and y and ρ_n are the canonical correlations. More details can be found in [16]. The N estimates of the sources are then given by $\mathbf{z_i(t)} = \widehat{\mathbf{w}}_{\mathbf{xi}}^{\mathbf{T}} \mathbf{x(t)}$. The EEG is reconstructed after the exclusion of components containing the artifacts, by projecting the selected components back onto the scalp. The enhanced EEG $\mathbf{x_{clean}(t)}$ is then:

$$\mathrm{x_{clean}}(t) = A_{clean} \mathrm{z}(t), \quad (3)$$

with **z(t)** the sources obtained by CCA and $\mathbf{A_{clean}}$ the mixing matrix with the columns representing the artifacts sources, set to zero [12].

C. Classifier design

In this work we aim to detect and remove muscle activity from long duration EEG recordings by using the autocorrelation value at time lag one (we also examined the sum of the component autocorrelation values over the first 5-10 lags but the results did not improve) and the relative power of each component in the range between 13-38 Hz as features. Brain activity corresponds to a structured signal having high autocorrelation values, in contrast to muscle activity which is less structured and exhibits properties that more closely resemble random (white) noise compared to normal EEG [12]. Furthermore, the relative power is fraction within the frequency range 13-38Hz is expected to be higher in muscle contaminated EEG signal and lower in artifact-free EEG signal [14]. Therefore, we computed the relative power (RP) as:

$$\mathrm{RP} = \frac{P_{MA}}{P_{total}}, \quad (4)$$

for each canonical component, where $\mathbf{P_{MA}}$ is the component power within 13-38 Hz, calculated as the integral of the power spectral density (PSD) within this range and $\mathbf{P_{total}}$ is the total component power, calculated as the PSD integral between 1 and 45 Hz. CCA was applied in 30min epochs containing muscle artifacts and artifact-free periods and 23 components were extracted. As described in Methods, 5-second non-overlapping windows were labeled as 'noisy' or 'noise-free' by expert neurophysiologist within the corresponding window or 'noise-free' otherwise (Fig. 3). For each 5s window, the autocorrelation at time lag one and the RP were obtained for all 23 components. In total, we selected 46 features, of which half corresponded to autocorrelation values and half, corresponded to the RP of all 23 canonical components as identified by CCA. Therefore, a two-class classification problem was subsequently solved in order to correctly classify windows as 'noisy' or 'noise-free'. We also examined the spatial components maps

for each patient. Visual inspection revealed that the spatial map of the last 15 components was above possible sources of muscle activity (Fig. 4), which corroborates the results yielded by the classifier (see below). After feature extraction, we ended up with a high dimensional feature set (46 dimensions). Note that we decided to keep all components in order to remove any need for visual inspection of the spatial maps. Instead, we used a feature selection algorithm in order to identify the most important feature subset of the original features, which can be used for clustering, classification and artifact removal. To this end, several feature selection algorithms [17-18] were used and compared using a measure of the quality of the binary (two-class) classification performance termed the Matthews correlation coefficient. MCC [19] is regarded as one of the most balanced measures which can be used even if the classes have considerably different sizes. MCC values range between −1 and +1. A value of +1 indicates perfect prediction, 0 indicates random prediction and −1 indicates total disagreement between prediction and observation. Among the feature selection algorithms used, random forests yielded the largest MCC values; therefore, we present results based on these.

Random Forests [18] are an ensemble learning method for classification. According to the method, many classification trees are grown, where by each tree depends on the values of an input vector sampled independently and with the same distribution for all trees in the forest. Each tree is generated using different bootstrap samples from the samples from the original data. About one third of the cases were left out of the bootstrap sample and used as cross-validation in order to compute the classification error. The selected features are placed in order of most important based on their importance score, which is computed by averaging the difference in the cross-validation error before and after the permutation over all trees. Therefore, the important features are those that yield large values for these scores [18]. Therefore, we obtained a vector of forty-six features, of which the first one was the most important and the last one the least important. We repeated this method forty-six times and in each run we removed the last feature. The MCC was calculated in each run and was used to select the number of features (< 46) needed to achieve the best classification accuracy.

D. Simulation

The aim of the simulation study was to evaluate the performance of the proposed method in order to detect and remove muscle artifacts. A simulation study with different signal to noise ratios in different channels and periods was implemented. Specifically an EEG epoch of 10min containing only brain activity was selected. This epoch was without muscle artifact according to the visual inspection by expert neurophysiologists. The simulated muscle activity was White Gaussian Noise (WGN) of different signal to noise ratios (SNR, -5, -10, -15, -20 dB) band-pass filtered in the frequency range 13-38Hz. The muscle activity was added in several randomly selected channels (1, 2, 14, 15 and 18) and only in within a 5min window in the epoch. The simulated epoch was segmented into 5-second non-overlapping windows of which half windows contained muscle activity and the rest did not.

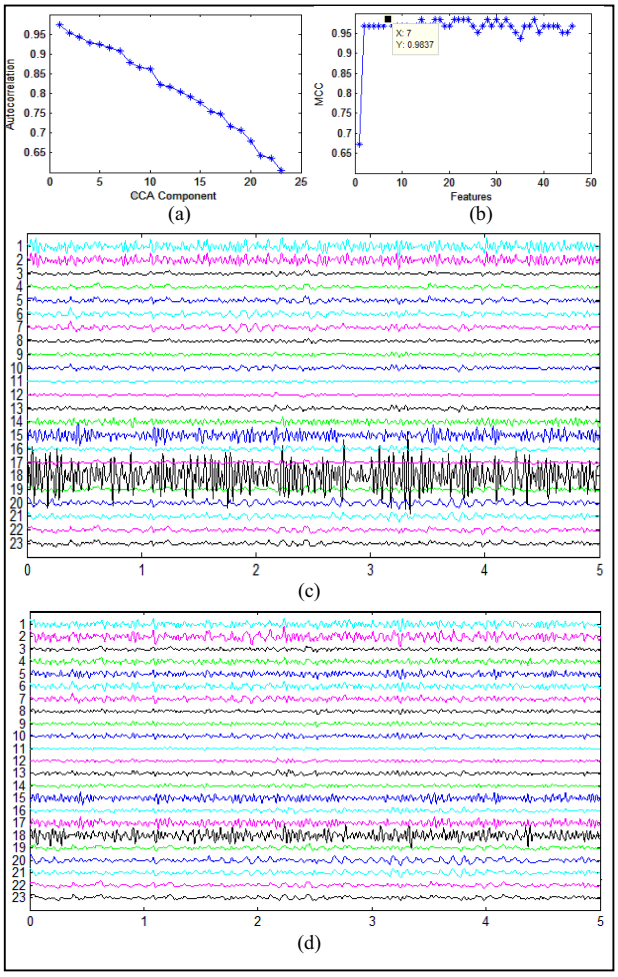

Figure 1. a) The autocorrelation of the Canonical Correlation Analysis (CCA) components, b) MCC values for different number of features used for classification, c) EEG signal (5s window) with muscle artifacts in channels 1, 2, 14, 15 and 18. (b) EEG signal after removal of components (17, 18, 19, 20, 21, 22 and 23). The EMG artifacts that are apparent in channels 1, 2, 14, 15 and 18 are completely removed from the reconstructed EEG signal.

We repeated the procedure described above in order to detect and exclude muscle artifacts from the EEG signal.

III. RESULTS

A. Simulation

Fig. 1a shows the autocorrelation of the CCA components and Fig. 1b shows the MCC values as a function of the selected features. The components accounting for the muscle artifact are present in the components with lower autocorrelation values. Fig. 1c and 1d show a 'noisy' 5s window before and after EEG reconstruction, respectively. The muscle artifacts that are apparent in channels 1 (SNR=-5dB), 2 (SNR=-5dB), 14 (SNR=-10dB), 15 (SNR=-15dB) and 18 (SNR=-20dB) are completely removed from the reconstructed EEG signal. The 'noise-free' windows are largely unaffected. Table I provides the goodness of fit between channels of a 'noisy' and 'noise-free' window before and after EEG reconstruction. The cost

function to determine goodness of fit is one minus the Normalized Mean Square Error (NMSE) which varies between minus infinity (bad fit) to 1 (perfect fit).

TABLE I.

'Noisy' Window Channels	NMSE After EEG reconstruction	'Noise-free' Window Channels	NMSE After EEG reconstruction
1	0.0840	1	0.8574
2	0.0983	2	0.8152
14	0.1119	14	0.9855
15	0.0355	15	0.9200
18	0.3140	18	0.9755

In the 'noisy' window, the channels with muscle activity are affected significantly (except channel 18, Table 1) since the EMG artifacts are completely removed from the reconstructed EEG signal. . The noise added in channel 18 was with the largest SNR (-20dB), thus there is a possibility that was not completely removed after EEG reconstruction In contrast, in the 'noise-free window the same channels are largely unaffected.

B. Experimental data

- *Canonical Correlation Analysis*

Figs. 2a and 2b show the EEG signal of a 'noisy' and a 'noise-free' 5s window from patient 1, identified and marked by expert neurophysiologists, respectively. Figs. 3a and 3b show the components of a 'noisy' and a 'noise-free' 5s window from patient 1, respectively. We observe that there are visible differences between in the components obtained from the two windows that could be used to detect and remove muscle artifacts from the EEG signal. Also, the muscle activity in the 'noisy' window does not seem to reside in the components with high autocorrelation values (e.g. components 1 and 2 in Fig. 2a), which are more likely related to normal brain activity. Rather, muscle activity is present in the CCA components exhibiting lower autocorrelation values (i.e. those that correspond to the lower traces of Fig. 1a). Fig. 4a presents the spatial map of the retro-projected components weights extracted from the 30 min epoch of patient 1. Fig 4b represents the autocorrelation of the CCA components. The selected features from the random forest algorithm (see below) were different for each patient; however, in all cases they corresponded to the last 15 canonical components (components 9 to 23 - Fig. 6a), as implied from the visually inspected spatial maps (e.g. Fig. 4).

- *Classifier*

Table II provides representative MCC values for all three patients after feature selection was performed. The results varied between the three patients. Patient 1 yielded the best MCC values. Figs. 4a and 4b shows the MCC values as a function of the selected features for patients 1 and 2, respectively. The selection of CCA components was done by keeping a trade-off between the MCC values (as high as possible) and the number of features (as low as possible). In order to remove the artifacts from the EEG signal we selected different number of CCA components to exclude for each patient. The selected components were excluded from the 30 min epoch for each patient in order to keep continuity of the EEG signal. Figs. 5a show the reconstructed artifact free EEG signal of a representative 5s window marked as 'noisy', of patient 1. We observed that the EMG artifacts were removed or reduced from the reconstructed EEG signal.

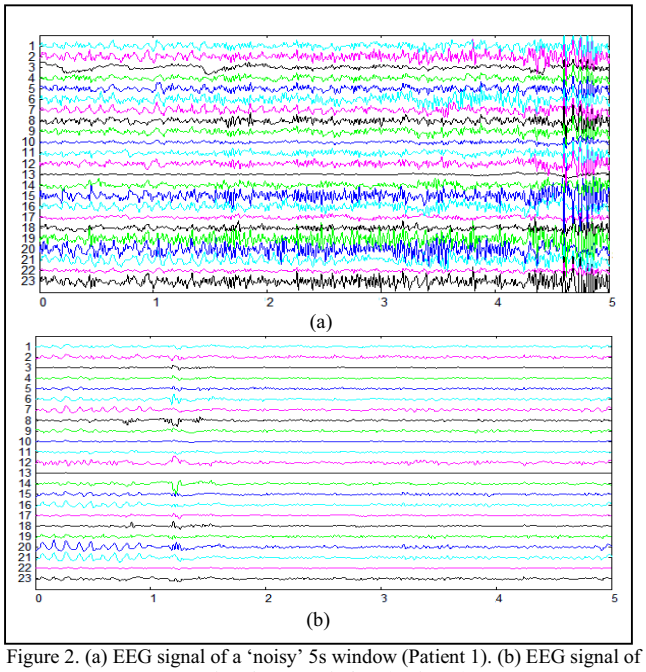

Figure 2. (a) EEG signal of a 'noisy' 5s window (Patient 1). (b) EEG signal of a 'noise- free' 5s window (Patient 1).

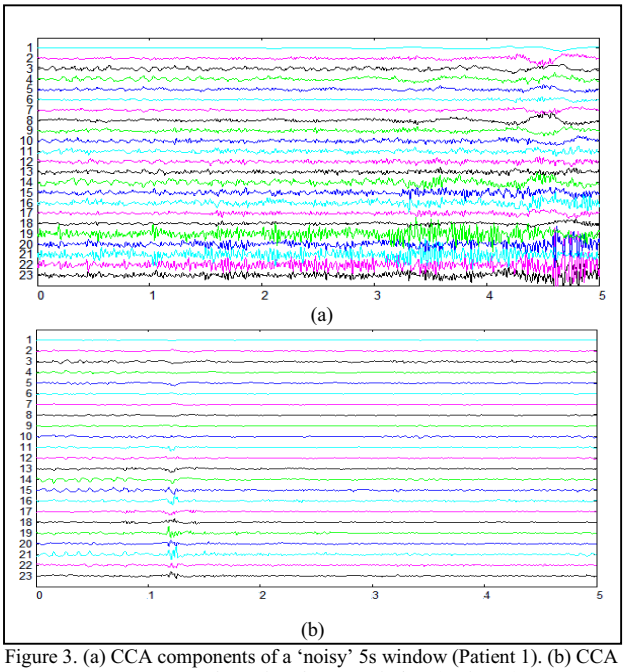

Figure 3. (a) CCA components of a 'noisy' 5s window (Patient 1). (b) CCA components of a 'noise- free' 5s window (Patient 1).

More specifically, the artifacts from the 2nd, 8th, 12th, 15th 19th and 23rd channels are completely removed, while the artifacts from the 15th and completely removed, while the artifacts from the 15th and 20th channels are reduced. Table III provides the goodness of fit between channels of a 'noisy' and 'noise-free' window (Fig. 1a, 5a) before and after EEG reconstruction, of patient 1. In the 'noisy' window, the channels with muscle activity, are affected significantly (except channel 1, Table III) since the EMG artifacts are removed almost completely from the reconstructed EEG signal. In contrast, in the 'noise-free window the same channels are largely unaffected. Figs. 5b show the EEG signal of a representative 5s window for patient 1 marked as noise- free after EEG reconstruction, of patient 1. We observed that after exclusion of components from the 30 min epoch only the 'noisy' window was affected (Fig. 6a) unlike the 'noise-free' window which was largely unaffected (Fig. 6b).

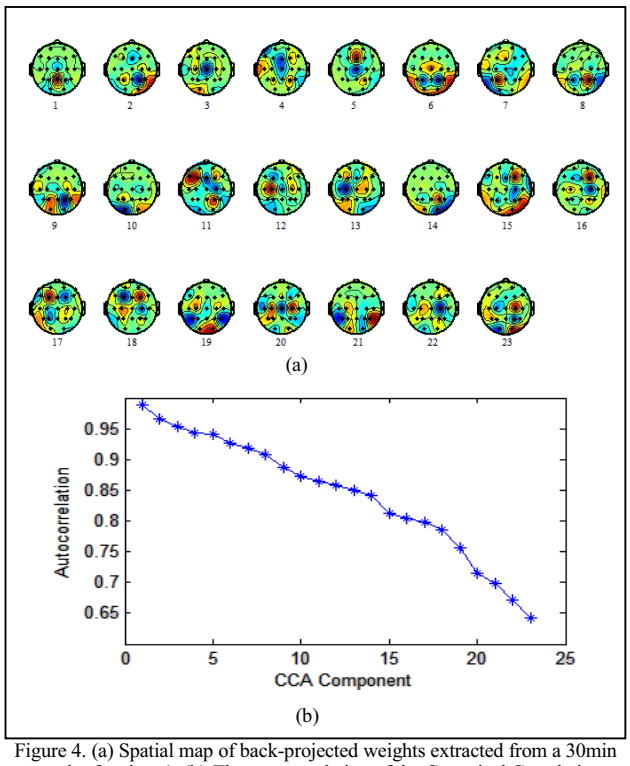

Figure 4. (a) Spatial map of back-projected weights extracted from a 30min epoch of patient 1. (b) The autocorrelation of the Canonical Correlation Analysis (CCA) components of Patient 1.

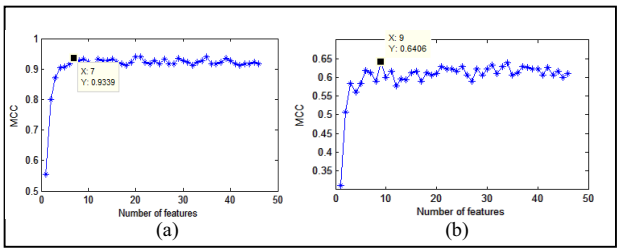

Figure 5. (a) MCC values for different number of features used for classification (Patient 1). (b) MCC values for different number of features used for classification (Patient 2).

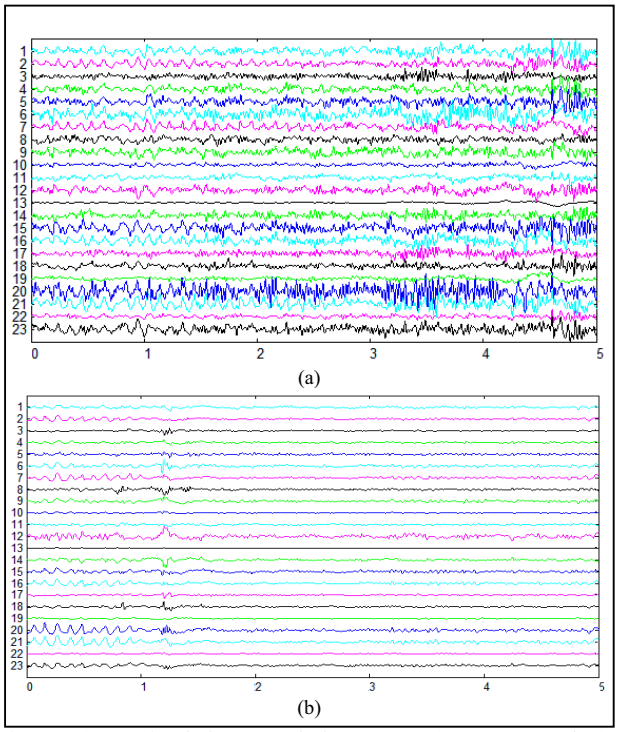

Figure 6. a) EEG signal after removal of components (9, 15, 18, 22 and 23). The EMG artifacts that are apparent in channels 2, 8, 11, 12, 15, 17, 19 and 23 are completely removed from the reconstructed EEG signal. (b) EEG signal (5s window) with no muscle artifacts marked as 'noise-free, after exclusion of the above components, of patient 1. The 'noise-free' window is largely unaffected.

TABLE II.

Number of features used for classification	MCC		
	Patient 1	*Patient 2*	*Patient 3*
46	0,916	0.611	0.508
40	0,917	0,612	0518
30	0,923	0.584	0.523
20	0.934	0.617	0.510
10	0.922	0.618	0.502
5	0.906	0,584	0,423
1	0,556	0,312	0,419

TABLE III.

'Noisy' Window Channels	NMSE After EEG reconstruction	'Noise-free' Window Channels	NMSE After EEG reconstruction
1	0.8349	1	0.9469
2	0.1374	2	0.7797
8	0.2423	8	0.8893
12	0.1807	12	0.9604
15	0.1023	15	0.8694
19	0.0546	19	0.9775
23	0.1604	23	0.8280

This was observed in almost all cases. It is worth mentioning that despite the relatively low MCC classification scores for patients 2 and 3, the results of the reconstructed artifact-free

EEG signal were overall satisfactory for all subjects (e.g. Table III for patient 1).

IV. DISCUSSION

In this paper an automatic detection and removal of muscle artifacts from EEG recordings based on CCA was presented. The performance of the method was tested by using 'noise-free' and 'noisy' windows marked by expert neurophysiologists, on both synthetic data and ictal EEG. The results from the simulation study were similar to that of ictal EEG. Most papers obtain CCA components by applying BSS-CCA to 5-10s epochs [1, 9]. In this study we defined the CCA components from a long EEG epoch (30min) in order to be able to use these components in order to automatically classify shorter (5s) windows as 'noisy' or 'noise-free' (Figs. 1a and 1b). The exclusion of the selected components from a 30min epoch for all patients affected only the 'noisy' windows (Fig. 1a, 5a) and not the 'noise-free' windows (Fig. 1b, 5b). The autocorrelation value at time lag one and the relative power of the CCA components in the frequency domain were found to be different between 'noise-free' and 'noisy' windows. In 'noisy' windows the autocorrelation values of the components corresponding possibly to muscle artifacts, were lower than the 'noise-free' windows reflecting the fact that these artifacts exhibit more random characteristics. In contrast, the relative power in the selected range of frequencies (13-38 Hz) was higher in 'noisy' windows and lower in 'noise-free' windows. The selection of the CCA components for exclusion was based on a trade-off between the obtained MCC values and the number of features (Figs. 4a and 4b). The proposed method succeeded in removing muscle artifacts from brain signals in 'noisy' windows for all patients. However, the classification performance varied substantially across patients. Patient 1 yielded the best MCC values and lowest classification error compared to the other two patients. Although, for patients 2 and 3 classification MCC values were relatively low, the results of the reconstructed artifact-free EEG signal were satisfactory for all patients. Furthermore, a possible explanation for these observations could be found in the origin of the muscle artifact signal. All the muscle artifacts in patient 3 were marked after the seizure (after the first 15 min mark in a 30 min epoch). Thus, there is a possibility that in this case the muscle activity originated from different muscles than the frontalis and temporalis. Activation of other muscles may lead to an artifact signal with somewhat different spectra and frequency characteristics. Another possible explanation is the type of epilepsy. Patient 1 had generalized epilepsy, patient 2 had intractable generalized epilepsy and patient 3 had simple focal onset epilepsy. Further investigation is needed in a larger set of patients with different types of epileptic seizures and muscle artifacts originating from different muscle groups. Also in the future, the method will be further validated on long duration EEG recordings with simultaneous EMG recordings used as reference.

REFERENCES

[1] M. Mennes, H. Wouters, B. Vanrumste, L. Lagae and P. Stiers, "Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP", Psychophysiology, Vol. 47, pp. 1141-1150, 2010.

[2] W. De Clercq, P. Lemmerling, S. Van Huffel, and W. Van Paesschen, "Correspondence : Anticipation of epileptic seizures from standard EEG recordings," The Lancet, vol. 361, pp. 970–971, 2003.

[3] N. Nicolaou and S. J. Nasuto, "Comparison of Temporal and Standard Independent Component Analysis (ICA) Algorithms for EEG Analysis", ICANN/ICONIP,Vol. 3, 2003.

[4] K. T. Sweenen, S. F. McLoone and T. E. Ward, "The Use of Ensemble Empirical Mode Decomposition With Canonical Correlation Analysis as a Novel Artifact Removal Technique ", IEEE Transactions on biomedical engineering, vol. 60, no. 1, 2013.

[5] P. Comon, "Independent component analysis, A new concept?", Signal processing, vol. 36, pp. 287-314, 1994.

[6] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, et al., "Removing electroencephalographic artifacts: comparison between ICA and PCA," in Proc. IEEE Signal Processing Society Workshop Neural Networks for Signal Processing VIII, ed, 1998, pp. 63-72.

[7] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, et al., "Removing electroencephalographic artifacts by blind source separation.," Psychophysiology, vol. 37, pp. 163-178, 2000.

[8] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods.," Int J Psychophysiol, vol. 53, pp. 105-119, 2004.

[9] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis," J Neurosci Methods, vol. 158, pp. 300-312, 2006.

[10] S. Makeig and T. Jung, "Tonic phasic, and transient EEG correlates of auditory awareness in drowsiness," cognitive brain research, vol. 4, pp. 15-25, 1996.

[11] R. N. Vigario, "Extraction of ocular artefacts from EEG using independent component analysis," Electroencephalography and Clinical Neurophysiology, vol. 103, pp. 395-404, 1997.

[12] W. D. Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," IEEE Trans. Biomed. Eng., vol. 53, no. 12, pp. 2583–2587, Nov. 2006.

[13] H. Hotelling, "Relations between two sets of variates", Biometrika, vol. 28, no. 3/4, pp. 321–377, 1936.

[14] I.I. Goncharova, D.J. McFarland, T.M. Vaughan, J.R. Wolpaw, "EMG contamination of EEG: spectral and topographical characteristics", Clinical Neurophysiology, Vol. 114, pp. 1580–1593, 2003.

[15] A. Ziehe and K. R. Muller, "TDSEP - an efficient algorithm for blind separation using time structure", in Proceedings of ICANN '98, pp. 675-680, December, 1998.

[16] T. W. Anderson. "An Introduction to Multivariate Statistical Analysis", John Wiley & Sons, second edition, 1984.

[17] D. Cai, C. Zhang, X. He. "Unsupervised Feature Selection for Multi-Cluster Data", Proceedings of International Conference on Knowledge and discovery and data mining, pp. 333-342, 2011.

[18] L. Breiman. "Random Forests", Machine Learning Journal, Vol. 45, no.1, pp. 5–32, 2001.

[19] P. Perruchet, R. PeeremaN. "The exploitation of distributional information in syllable processing". J. Neurolinguistics, Vol. 17, pp. 97−119, 2004.