World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:5, No:9, 2011

Wavelet Enhanced CCA for Minimization of Ocular and Muscle Artifacts in EEG

B. S. Raghavendra, and D. Narayana Dutt

***Abstract*—**Electroencephalogram (EEG) recordings are often contaminated with ocular and muscle artifacts. In this paper, the canonical correlation analysis (CCA) is used as blind source separation (BSS) technique (BSS-CCA) to decompose the artifact contaminated EEG into component signals. We combine the BSS-CCA technique with wavelet filtering approach for minimizing both ocular and muscle artifacts simultaneously, and refer the proposed method as wavelet enhanced BSS-CCA. In this approach, after careful visual inspection, the muscle artifact components are discarded and ocular artifact components are subjected to wavelet filtering to retain high frequency cerebral information, and then clean EEG is reconstructed. The performance of the proposed wavelet enhanced BSS-CCA method is tested on real EEG recordings contaminated with ocular and muscle artifacts, for which power spectral density is used as a quantitative measure. Our results suggest that the proposed hybrid approach minimizes ocular and muscle artifacts effectively, minimally affecting underlying cerebral activity in EEG recordings.

***Keywords*—**Blind source separation, Canonical correlation analysis, Electroencephalogram, Muscle artifact, Ocular artifact, Power spectrum, Wavelet threshold.

I. INTRODUCTION

ELECTROENCEPHALOGRAM (EEG) measures electrical potentials on the scalp and provides a continuous measure of cortical functions. These brain signals typically are of very low amplitude and hence they are prone to noise [1]. During signal acquisition, the EEG recordings are often contaminated with muscle, head movement and ocular artifacts. These artifacts may originate due to many reasons such as eye blinks and eye ball rotations, head movements, biting, chewing etc. The artifacts produce large electrical potentials and spread across the scalp to contaminate the EEG signal, and ultimately affect the assessment of neurological phenomenon. The ocular artifacts are generally high amplitude and low frequency signals affecting delta and theta band, and the muscle artifacts affect high frequency (alpha, beta and gamma) bands of EEG power spectrum [2], thus in turn affecting extraction of valid information from the EEG recordings and its interpretation. Hence minimization of these artifacts forms an important preprocessing step before making further quantitative analysis of the EEG data.

B. S. Raghavendra and D. Narayana Dutt are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore – 560 012, India (corresponding author to provide phone: +91 80 2293 2742; fax: +91 80 2360 0563; e-mail: r.bobbi@ece.iisc.ernet.in).

Several methods have been proposed in the literature for minimization of artifacts in EEG signals. The main ocular artifact correction methods are as follows. Artifact avoiding or eye fixation, artifact rejection, linear filtering, regression based methods [3]-[5], blind source separation (BSS) based methods (including principal component analysis (PCA) and different versions of independent component analysis (ICA)) [5]-[8], singular value decomposition (SVD) based techniques [9], adaptive noise cancellation based methods [10], and methods that make use of some special properties of artifacts [11], [12], etc. However, the ICA based methods have become one of the most successful methods applied to multichannel EEG recordings. The ICA is also used in minimization of muscle artifacts. In addition to this, BSS and SVD based methods are also found useful for muscle noise correction. Recently, BSS based method using canonical correlation analysis (CCA) technique has been introduced to remove muscle artifacts, and has produced better performance [13].

However, each of the methods is having its own advantages and disadvantages, and improvements in the performance of the methods have been suggesting in the literature. Many of the approaches minimize artifacts, by taking ocular and muscle artifacts separately into consideration. There are many methods for correcting either ocular or muscle artifacts tailoring for a particular application. In this paper, the canonical correlation analysis (CCA) is used as blind source separation (BSS) technique for minimization of both ocular and muscle artifacts simultaneously. For ocular artifact reduction, we have used wavelet-based enhancement of CCA components. The performance of the method is tested on real EEG signals using power spectrum as a quantitative measure.

II. METHODS

The BSS approaches are increasingly being used in biomedical signal processing involving analysis of multivariate time series data such as EEG [17]. In this approach, the observed multichannel signals are assumed to reflect a linear combination of several sources which are associated with underlying physiological processes, artifacts and noise. The BSS approach aims to recover a set of unobserved source signals using only a set of observed mixtures of sources. Typically, the observations are measured at the output of an array of sensors such as EEG electrodes, where each sensor receives a different combination of the source signals.

The standard formulation of the linear instantaneous BSS problem consists of a source mixture model, in which M observed signals $X(t) = [x_1(t), x_2(t), ..., x_M(t)]^T$, $t = 1, 2, ..., N$, where T denotes transposition, are a linear combination of K unobserved source signals $S(t) = [s_1(t), s_2(t), ..., s_K(t)]^T$. This can be expressed as $X(t) = AS(t)$, where A is an $M \times K$ matrix whose columns represent the source sensor projections. The number of sources, their waveform and sensor projection A are unknown. The objective of BSS is to determine K, $S(t)$ and A from $X(t)$ using only minimal assumptions.

The number of sources is often assumed to be the number of sensors, that is $K = M$. To determine the source waveform estimates $\hat{S}(t)$ which is approximately equal to $S(t)$, requires estimate of a mixing matrix $\hat{A}$ which is approximately equal to A, which is used to form an approximate $P \times M$ demixing matrix $\hat{W}$ to obtain source waveform estimates $\hat{S}(t) = \hat{W}X(t)$. Finding a matrix estimate $\hat{A}$ generally involves an optimization problem with a goal of minimizing mutual dependence of the source waveform estimates with reference to either signal time structure, or higher order statistics or information theoretic measures.

A. BSS using CCA

There are many ways to solve the BSS problem depending on the definition of contrast function. The ICA method tries to make the signals as non-Gaussian as possible. However, in PCA and most of the ICA algorithms, the temporal correlations are not taken into consideration for solving contrast functions. The samples in the time may be rearranged arbitrarily and the method will give the same solution. This may seem as strength, but the fact is that almost all the temporal information in the signal is thrown away. In contrast to ICA, the CCA based BSS method utilizes the autocorrelation in the source signals as contrast function [14].

In real-world situation, most of the signals have certain autocorrelation, and the multichannel EEG signals also have temporal or spatial structure causing an autocorrelation in the signal. Considering only statistical distribution of the sample values as in most of ICA cases, ignoring the temporal or spatial relations within the source signals, one may discard the relevant temporal information. But the CCA can solve the BSS problem with much less computational effort by utilizing the temporal autocorrelation structure of the source signals, in particular for the EEG data.

The CCA is used to measure the linear relationship between two multidimensional variables, by finding two bases one for each variable, and the bases are optimal with respect to correlations. Thus, the CCA finds two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized [13], [15]. Since BSS is an ill posed problem, some statistical constraints on the sources have to be added to solve it. The CCA solves the problem by forcing the sources to be mutually uncorrelated and maximally correlated with a predefined function. The proposed BSS using CCA approach requires the sources to be uncorrelated and to be maximally correlated with a given function, in this case which is time shifted version of the original signal.

Consider the observed data matrix $X(t)$ and its temporally delayed version $Y(t) = X(t-1)$. The CCA finds two sets of basis functions, one for X and another for Y, such that the correlations between the projections of the variables on to these basis vectors are mutually maximized. The total covariance matrix is given by

$$C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = E\left[\begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}^T\right],$$

where C_{xx} and C_{yy} are the within-set covariance matrices of X and Y respectively, and C_{xy} is the between-sets covariance matrix ($C_{xy} = C_{yx}^T$), and E denotes expectation operation. The canonical correlation between X and Y can be found by solving the eigen-value equations, $C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}\hat{w}_x = \rho^2\hat{w}_x$, $C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}\hat{w}_y = \rho^2\hat{w}_y$, where, the eigen values ρ^2 are the squared canonical correlations and the eigen vectors $\hat{w}_x$ and $\hat{w}_y$ are the normalized canonical correlation basis vectors. Since the solutions are related, only one of the eigen value equations needs to be solved to get the demixing matrix W. The estimates of the sources are given by $\hat{S}(t) = \hat{W}_x^T X(t)$. The CCA gives the source signals that are uncorrelated with each other, maximally autocorrelated and ordered by decreasing autocorrelation.

The sum of uncorrelated signals will have an autocorrelation function which is less than or equal to the maximum of the autocorrelation functions of the individual signals. Since CCA maximizes the correlation, it will not choose a mixture of the source signals, since that would give less correlation than if one of the source signals in the data is chosen. The CCA approach is used as a BSS technique to separate ocular and muscle artifacts in multichannel EEG recordings. It is to be noted that simultaneous reduction of ocular and muscle artifacts in EEG using CCA has not been yet studied.

B. Muscle Artifact Correction by Component Elimination

The artifact contaminated EEG data is applied to CCA algorithm to get component signals. Due to broad frequency spectrum of muscle noise in EEG recordings, muscle artifacts tend to have properties of temporally white noise, thus having low autocorrelation. As a consequence, the muscle artifact source components are always those with the lowest autocorrelation. We manually identify the components contributing to the muscle artifact by visualizing correlation structure. Then the EEG is reconstructed by projecting the selected non-artifact components back into the scalp. We refer this technique to as component elimination method.

The muscle noise rectified EEG is given by $X_m(t) = \hat{A}S_m(t)$, where $\hat{A} = \hat{W}^{-1}$ is the estimate of mixing matrix and $S_m(t)$ is the components after making those components zeros which contribute to muscle artifact. However, the muscle artifact corrected EEG data $X_m(t)$ may have ocular artifact which is to be rectified. Hence the components $S_m(t)$ are subjected to the next step where the ocular artifact sources are considered for further processing.

C. Ocular Artifact Correction by Component Filtering

Most EEG artifact removal techniques either consider only ocular artifacts or only muscle artifacts, while, to our knowledge, not much effort has been made on the simultaneous removal of ocular and muscle artifacts. However, we use CCA to minimize both ocular and muscle artifacts simultaneously.

It is possible that the high frequency cerebral activity may leak into the components marked as artifactual, and that may be lost if the component is rejected during reconstruction [18]. To overcome the leakage of cerebral components to artifactual components, we introduce filter to the ocular artifactual components to retain the high frequency cerebral activity in them. For this purpose, we have used wavelet based filtering to recover the neural activity in the ocular components, thereby enhancing the CCA. We refer this method to as wavelet filtered component inclusion method. We also quantify the spectral distortions of EEG for CCA and wavelet enhanced CCA.

The ocular artifact components are manually identified after CCA decomposition and subjected to wavelet filter to retain the high frequency neural part. Hence no ocular artifact component is thrown away in the reconstruction process. We refer this technique to as wavelet filtered component inclusion method (CCA+WF). The wavelet decomposition separates the artifact component into low frequency ocular artifact part and high frequency cerebral part. The ocular artifact is considered as independent (uncorrelated) with EEG data, which can be well separated in the wavelet domain using standard threshold based rule [16].

In the proposed method, the source components having ocular artifact are subjected to undecimated wavelet transform using Daudechies 4 filter. The filtering is based on the threshold rule, where the wavelet coefficients greater than the prescribed threshold are made zero. The threshold is calculated as $TH = \sigma\sqrt{2\log(L)}$, where $\sigma^2 = median(|D_c|)/0.6745$ and L is the length of epoch considered, and D_c is the set of wavelet coefficients in different scales [16]. Thus, all the components having ocular artifacts are subjected to wavelet enhancement. The clean EEG

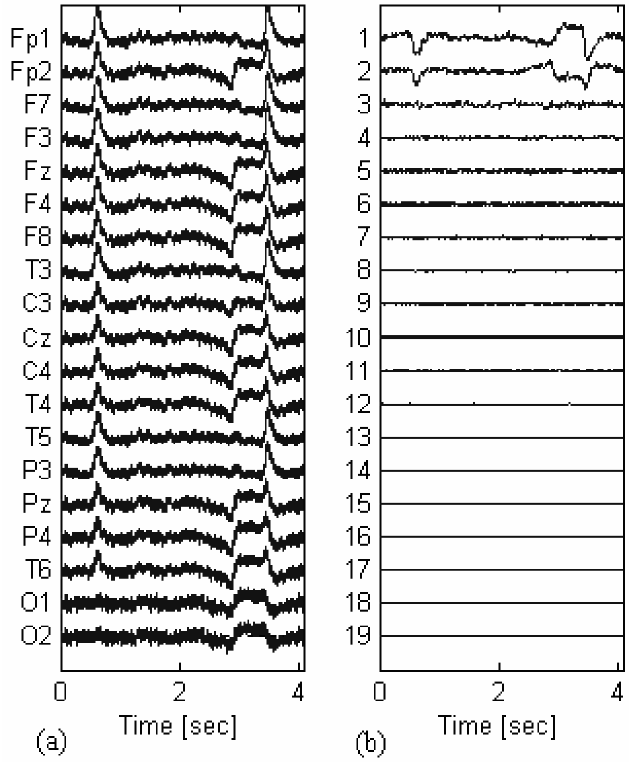

Fig. 1. (a) An epoch of real EEG data contaminated with eye blinks and movement artifacts, and (b) CCA components of EEG data.

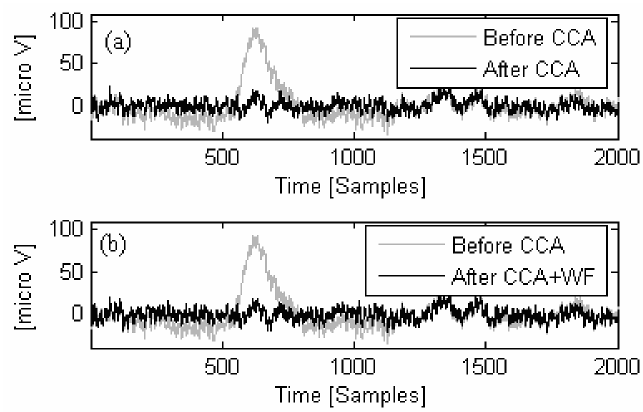

Fig. 2. A part of Fp1 channel of reconstructed EEG data after (a) component elimination method, (b) wavelet filtered component inclusion (CCA+WF) method.

data is obtained as $X_{clean}(t) = \hat{A}S_{mo}(t)$, where $S_{mo}(t)$ is the ocular artifact corrected components from $S_m(t)$. Thus, both muscle and ocular artifact are simultaneously processed in the components after decomposition using CCA.

III. Results and Discussion

To demonstrate the effectiveness of the CCA+WF artifact correction method, the method is applied on different sets of real EEG data. In the first case, real EEG data has been

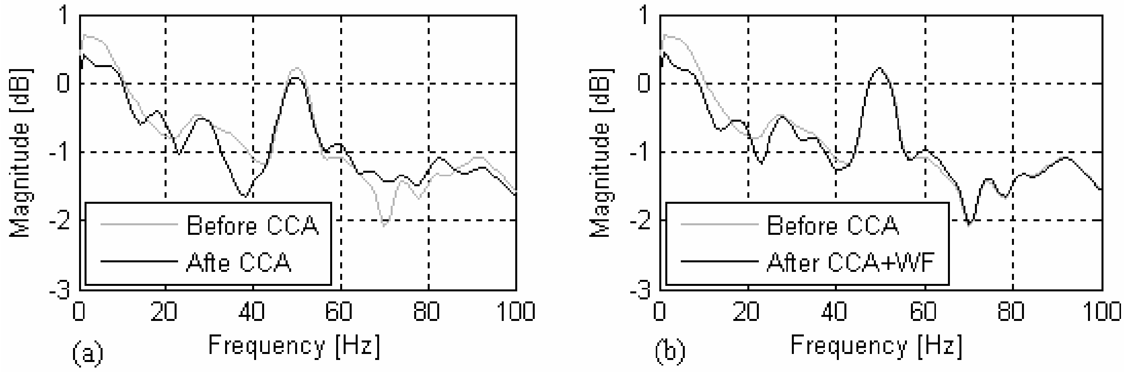

Fig. 3. Power spectrum of Fp1 channel of reconstructed EEG data using (a) component elimination method, (b) wavelet filtered component inclusion (CCA+WF) method.

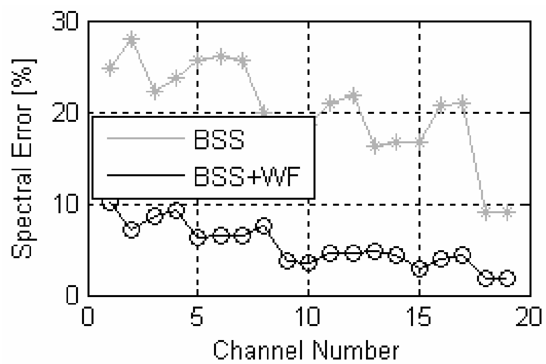

Fig. 4. Percentage spectral error (in frequency range 13-70 Hz) of different channels for 0-2 sec interval of EEG data. The legend BSS denotes CCA.

acquired using Neuroscan EEG system with Fz as reference. The data is recorded according to the 10-20 standard of electrode placement, digitized with a sampling rate of 250 Hz and a resolution of 12 bits. A band pass filter of 0.5 to 70 Hz has been used to limit the frequency band. We have chosen only 19 channels of EEG recordings without taking electrooculogram (EOG) into consideration for the analysis.

Fig. 1(a) shows an epoch of 19 channel real EEG data contaminated with eye blinks and head movement artifacts. The eye blink artifact is getting reduced as moving towards the occipital region. The components obtained by applying CCA algorithm on this data are also shown in Fig. 1(b). The components accounting for blink and movement artifact are present at the top components (in this case first and second). In most of the ICA based ocular artifact correction methods, the ocular artifact source signals are made zero and then reconstructed the clean EEG data (component elimination method). The blink and movement artifacts are very much localized and cerebral information is present out side this interval, which will be lost if entire components are eliminated. A portion of the reconstructed EEG data of Fp1 channel using the two approaches is shown in Fig. 2.

The power spectrum of the Fp1 channel before and after ocular artifact correction is shown in Fig. 3. The Welch-based method is used for estimating power spectrum of the EEG signals. From the plot, it is clear that the component elimination method has removed high frequency information in the range of 30 to 40 Hz and 65 to 75 Hz. This spectral

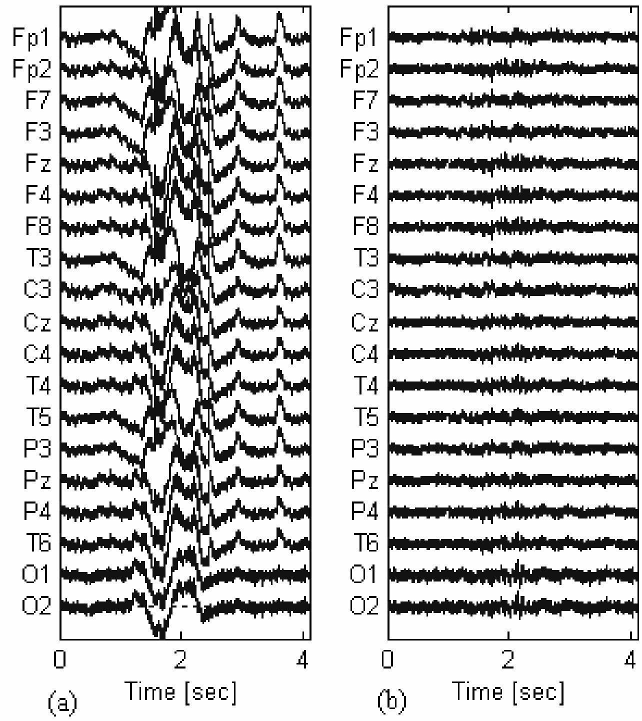

Fig. 5. (a) An epoch of real multichannel EEG contaminated with ocular and movement artifacts, (b) artifact minimized EEG using wavelet enhanced CCA (CCA+WF) method.

domain distortion may introduce errors in interpreting quantitative EEG analysis. However, the wavelet filtered component inclusion method (CCA+WF) has retained most of the high frequency information leading to less spectral distortion. To quantify the comparison between the two methods, percentage spectral error is plotted in the Fig. 4 for all the channels. For calculating mean spectral error, we have used 13-70 Hz frequency range, because in this range ocular artifact is having very less frequency overlap with clean EEG. The error is calculated between original raw EEG data and reconstructed EEG data. The wavelet enhanced CCA method has less percentage spectral error, and as the distance form the eye is increased towards occipital region, the spectral error is reduced. Fig. 5 shows another EEG data segment contaminated with severe movement artifact and artifact

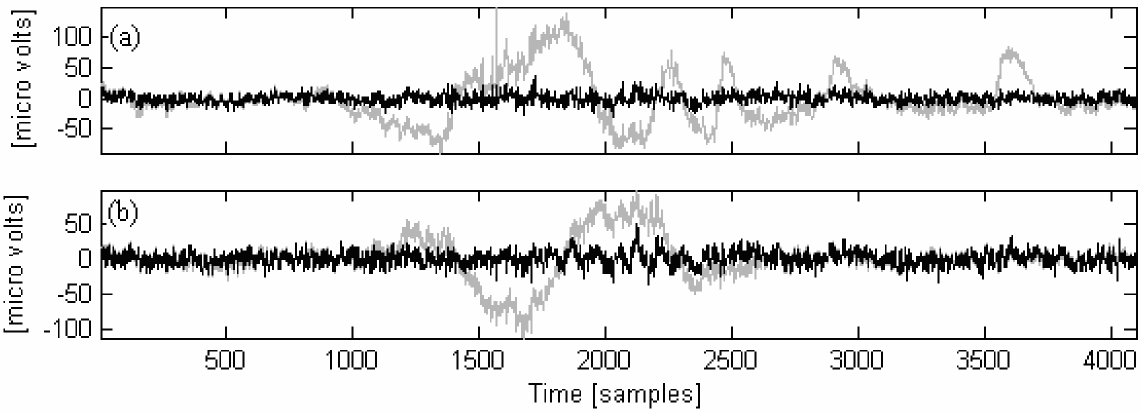

Fig. 6. Enlarged version of the signals of Fig. 5, (a) Fp1 channel, and (b) O2 channel, before (gray) and after (black) artifact minimization.

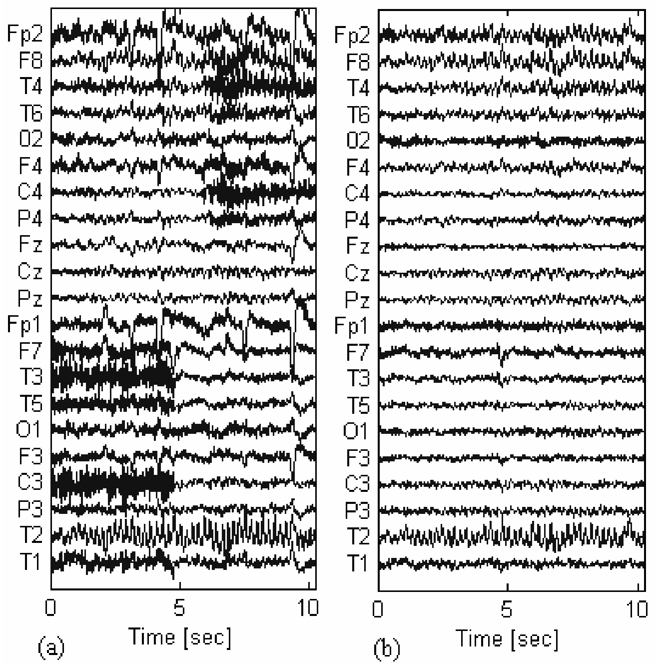

Fig. 7. (a) An epoch of 10 sec ictal EEG with eye blink and muscle artifacts, (b) artifact minimized EEG using wavelet enhanced CCA (CCA+WF) method.

corrected EEG data using wavelet enhanced CCA method. Enlarged version of the plot is shown in Fig. 6 for Fp1 and O2 channels. It is observed that the wavelet enhanced CCA method has effectively corrected the ocular artifacts without altering much of the high frequency cerebral activity of EEG.

In another case, the wavelet enhanced CCA method is tested on ictal EEG data of 21 channels. Sampling frequency of the data is 250 Hz. In the data, as shown in Fig. 7(a), predominant epileptic seizure activity can be observed on channels Fp2, F8, T4 and T2. The Fp1 and Fp2 channels are contaminated with eye blink artifacts. Muscle artifacts can be observed in the time interval of 0-3.9 sec on channels F7, T3, T5, C3, T1 and in 5-10 sec on channels F8, T4, F4, C4 and P4. In Fig. 7(b) results obtained by applying the wavelet enhanced CCA method to the ictal EEG data is shown. The reconstructed EEG data after exclusion of muscle artifact components and inclusion of

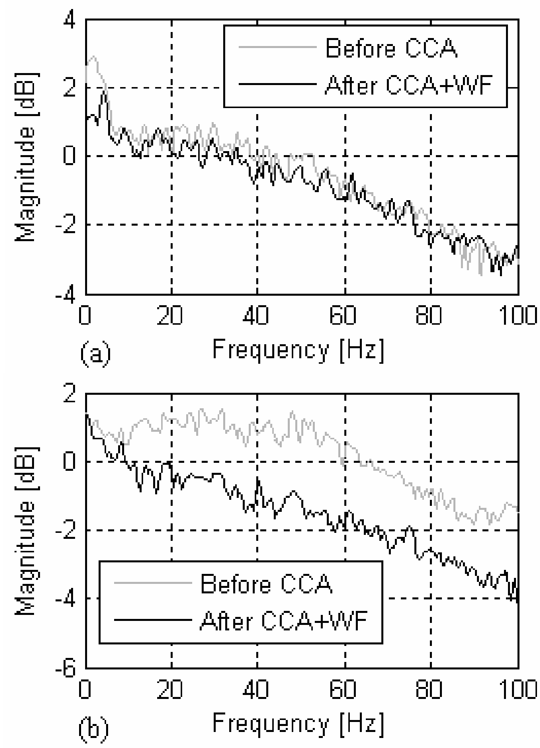

Fig. 8. Power spectrum of (a) Fp2 channel (contaminated with eye blinks) in 0-5 sec interval, (b) T4 channel (contaminated with muscle artifact) in 5-10 sec interval.

wavelet filtered ocular artifact components is also shown. It is observed from the Fig. 7(b) that the ictal activity is not distorted by the application of the wavelet enhanced CCA method. Fig. 8 shows power spectrum of the result for two different channels Fp2 while T4 having different types of artifacts. The Fp2 channel has ocular artifact and the T4 channel has muscle artifact. The wavelet enhanced CCA method retained the high frequency spectral power, eliminating ocular artifact. In the other case, it has effectively removed the high frequency muscle noise contribution. The enlarged plots of EEG signals before undergoing wavelet enhanced CCA algorithm and after artifact correction are shown in Fig. 9. From the plot, it is clear that the eye blink and muscle noise are eliminated without altering the ictal activity.

In all the cases discussed above, we have manually selected the artifact components after CCA decomposition. Further research is needed to investigate the selection of artifact components automatically in a larger subject group. The components are ordered according to their autocorrelation in CCA. The ocular artifact components having maximum variance will be at the top and the muscle artifact components having less autocorrelation will be sorted at the bottom. This makes the method easy to automatically identify the artifact components. Moreover, the CCA method is computationally much faster than iterative ICA algorithms [15]. Hence this method can be made applicable for real time applications in clinical settings.

IV. Conclusion

We have presented a method for simultaneous minimization of ocular and muscle artifacts using CCA as a BBS technique. In this method, the CCA is used as a BSS technique for decomposition of multichannel EEG into components. The muscle artifact components are identified and discarded, and the ocular artifact components are subjected to wavelet filtering in order to retain high frequency cerebral part of EEG. Finally, the clean EEG is reconstructed to have both ocular and muscle artifacts minimized simultaneously. The performance of the method is tested on real EEG recordings using power spectrum as a quantitative measure. From the study, it is observed that the method has given good performance in rectifying both ocular and muscle artifacts, minimally affecting cerebral activity in EEG. The method does not require any reference signal for artifact minimization. This method can be used as preprocessing step to facilitate further signal processing of scalp EEG recordings, which could again improve the performance of many nonlinear measures in the EEG data analysis.

References

[1] Croft RJ, Barry RJ. Removal of ocular artifact from the EEG: a review. Clin Neurophysiology 2000; 30: 5-19.
[2] Narasimhan SV, Dutt DN. Application of LMS adaptive filtering for muscle artifact (noise) cancellation from EEG signals. Computers Elect Engg 1996; 22(1): 13-20.
[3] Croft RJ, Barry RJ. EOG correction: a new perspective. Electroenc and Clin Neurophys 1998; 107: 387-394.
[4] Schlogl A, Keinrath C, Zimmermann D, Scherer R, Leeb R, Pfurtscheller G. A fully automated correction method of EOG artifacts in EEG recordings. Clin Neurophysiol 2007; 118: 98-104.
[5] Wallstorm GL, Kass RE, Miller A, Cohn JF, Fox NA, Automatic correction of ocular artifacts in the EEG: a comparison of regression based and component based methods. Int J of Psychophys 2004; 53: 105-119.
[6] Bell AJ, Sejnowski TJ. An information maximization approach to blind separation and blind deconvolution. Neural Comput 1995; 7: 1129-1159.
[7] Jung T-P, Humphries C, Lee T-W, Makeig S, McKeown MJ, Iragui V, Sejnowski TJ. Extended ICA removes artifacts from electroencephalographic recordings. Advances in Neural Information Processing 1998; 10: 894-900.
[8] Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V, Sejnowski TJ. Removing electroencephalographic artifacts by blind source separation. Psychophysiol 2000; 37: 163-178.
[9] Sadasivan PK, Dutt DN. SVD based technique for noise reduction in electroencephalographic signals. Signal Proc 1996; 55: 179-189.
[10] Sadasivan PK, Dutt DN. Development of Newton-type adaptive algorithms for minimization of EOG artifacts from noisy EEG signals. Signal Proc 1997; 62: 173-186.
[11] Ille N, Berg P, Scherg M. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. J Clin Neurophys 2002; 19: 113-124.
[12] James CJ, Gibson J. Temporally constrained ICA: an approach to artifact rejection in electromagnetic brain signal analysis. IEEE Trans Biomed Eng 2003; 50(9): 1108-1116.
[13] Clercq WD, Vergult A, Vanrumste B, Paesschen WV, Huffel SV. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. IEEE Trans Biomed Eng 2006; 53: 2583-2587.
[14] Friman O, Cedefamn J, Lundberg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magnetic Resonance in Medicine 2001; 45: 323-330.
[15] Friman O, Borga M, Lundberg P, Knutsson H. Exploratory fMRI analysis by autocorrelation maximization. NeuroImage 2002; 16(2): 454-464.
[16] Donoho DL. Denoising by soft-thresholding. IEEE Trans Inform Theory 1995; 41(3): 613-627.
[17] Hyvorinen A, Oja E. Independent component analysis: algorithm and applications. Neural Netw 2000; 13: 411-430.
[18] Castellanos NP, Makarov VA. Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis. Journal of Neurosc Methods 2006; 158(2): 300-312.

B. S. Raghavendra was born in Bobbi-Thirthahalli, India. He received B.E. degree from Bangalore University in 2001, and M. Tech. degree in Electronics and Communication Engineering from National Institute of Technology Karnataka, Surathkal, India, in 2004. Currently he is pursuing his Ph.D. in the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. His research interests include biomedical signal processing, nonlinear dynamics, and computational neuroscience.

D. Narayana Dutt obtained his Ph.D degree in Electrical Communication Engineering from the Indian Institute of Science, Bangalore, India. He has been a member of faculty in the same Institute for the past several decades. He worked earlier in the areas of acoustics and speech signal processing. He has been working for the past several decades in the area of applications of digital signal processing to the analysis of biomedical signals. He has been collaborating with researchers from NIMHANS, Bangalore, in applying signal processing techniques to various problems in the area of neuroscience. He worked in the Biomedical Engineering Research Centre at Nanyang Technological University in Singapore, as a visiting faculty and worked in the application of digital signal processing to cardiovascular signals. He has published a large number of papers in this area in leading international journals.