Evaluating the efficacy of an automated procedure for EEG artifact removal

Yvonne Tran, Ranjit A. Thuraisingham, Ashley Craig, and Hung Nguyen, *Senior Member, IEEE*

Abstract—**Electroencephalography (EEG) signals are often contaminated with artifacts arising from many sources such as those with ocular and muscular origins. Artifact removal techniques often rely on the experience of the EEG technician to detect these artifact components for removal. This paper presents the results comparing an automated procedure (AT) against visually (VT) choosing artifactual components for removal, using second order blind identification (SOBI) and canonical correlation analyses. The results show that the resulting EEG signal after artifact removal for the AT and VT were comparable in both variance amongst electrodes and spectral energy. The AT technique is objective, faster and easier to use, and shown here to be comparable to the standard technique of visually detecting artifact components.**

I. Introduction

Electroencephalography (EEG) involves the recording of electrical activity from scalp electrodes produced from underlying firing of neurons in the brain. It contains valuable information on the spontaneous electrical activity from the brain reflecting motor, sensory and cognitive function. However, EEG signals are often contaminated by many different types of artifact and disturbance caused by eye blinks, eye movement, muscle activity, line noise and heart signals, making analyses of the underlying processes difficult [1]. For example, the localization of the neuronal source using source localization techniques becomes more difficult in the presence of signals from non-neuronal sources. The artifact overlap with the electrical signals from the brain, thereby confounding the analysis of EEG signals. Many artifact removal techniques have been explored and shown to be effective, for example, using blind source separation (BSS) in methods such as independent component analysis (ICA) [1]. However, these methods rely on the user having a high level of expertise in identifying EEG signals. To remove artifactual components the person needs to be able to identify the different types and sources of artifact components against EEG components. This paper presents the results of comparing an automated artifact removal method [2] with the current standard method of subjectively choosing artifact components using visual detection. The automatic procedure is based on the autocorrelation of the signal. BSS techniques using Second Order Blind Identification (SOBI), and has been used to identify ocular artifacts for removal [2,3,4,5] and Canonical Correlation Analysis (CCA) has been used for muscle artifact removal [2,6,7]. SOBI is particularly suited to separating ocular sources and CCA for muscle artifact, from the total EEG signal. Automatic artifact removal was based on set parameters using a fractal [8], spectra [9], spatial and Renyi entropy [10] method. To test for reliability in the automatic method components chosen, artifact is compared with components chosen by an experienced EEG technician [11].

Manuscript received April 6, 2009. This work was supported in part by the Australian Research Council (ARC) Linkage Granting Scheme under Grant LP0560590.

Y. Tran is with the Key University Research Centre in Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney (phone: 612-9808-0525; fax: 612-9809-9037; e-mail: Yvonne.Tran@uts.edu.au).

R. A. Thuraisingham, is consultant for the Rehabilitation Studies Unit, University of Sydney, (e-mail: ranjit@optusnet.com.au).

A. Craig is with the Rehabilitation Studies Unit, University of Sydney, (e-mail: a.craig@med.usyd.edu.au and nwij@mail.usyd.edu.au)

H. T. Nguyen is with the Faculty of Engineering and Information Technology, University of Technology, Sydney (email: Hung.Nguyen@uts.edu.au)

II. Methods

A. Second Order Blind Identification (SOBI)

In these BSS methods, the signals at the sensors

$$\boldsymbol{X}(t) = [x_1(t), x_2(t), \ldots\ldots. x_K(t)]^T,\ t = 1,\ldots N$$

with N =number of samples, K= number of sensors , is considered to be a linear mixture of unknown K source signals

$$S(t) = [s_1(t), s_2(t), \ldots\ldots s_K(t)]^T,$$

This can be written as

$$X(t) = AS(t) \qquad \text{-----(1)}$$

where A is the unknown mixing matrix. The goal is to estimate the mixing matrix A and recover the original source signals $S(t)$. This is obtained by generating the de-mixing matrix W that approximates A^{-1}. The unknown source signals $S(t)$ is approximated as

$$Z(t) = WX(t) \qquad \text{------(2)}$$

In SOBI, the source signals are assumed to be temporarily uncorrelated to each other but have non–zero time delayed autocorrelations. Under these assumptions the SOBI computes the mixing matrix as the matrix that jointly diagonalizes a set of p correlation matrices $R(\tau_i) = E[X(t)X(t-\tau_i)^T]$, where i=1,…., p and E [ ] is the expectation operator. In our work we used p=$N/3$

In ICA based methods, the estimation of complex statistical measures requires independence, however, in SOBI only second order statistics are required, making SOBI relatively easy to evaluate and robust for modeling errors [8].

B. Canonical Correlation Analysis (CCA)

In canonical correlation analysis, the sources are assumed to be mutually uncorrelated and maximally correlated with a predefined function. The predefined function is defined here to be:

$$Y(t) = X(t+1)$$

This function is a temporarily delayed version of the original data matrix. The method used in CCA for the separating the mixed signals *X(t)* is to find the linear combination of the measured EEG signals in the *K* sensors that has the maximum correlation with the linear combination of *Y(t).* This is solved using the Matlab routine 'cannoncorr', which estimates the demixing matrix *W* and the sources *Z(t).*

Both CCA and SOBI were carried out using moving data windows. This copes with the non-stationarity property of the EEG signals. The optimum window temporal length should be such that it covers enough data samples to estimate reliably the mixing matrix and learn about the artifact. Although there is no easy and objective way of selecting the optimum window length, the presence of relatively stable spatial patterns for ocular artifact prompts the use of a longer window length for SOBI [8]. On the other hand, for the removal of the shorter duration EMG bursts, a shorter window length is used in CCA [12]. For SOBI, we used the full data size of 0.058 **K***K s* (60s) samples as the window length while a size 0.0097**K***Ks* (10s) was chosen for CCA.

C. Automatic Artifact Removal Procedure

Once the sources are separated then several methods were adopted to identify artifact. To identify ocular artifact, fractal [8], spectra [9] with spatial characteristics, and Renyi entropy [10] methods were applied. Of these, the spectral method with spatial characteristics was found to be the most reliable and was consequently used here. The information on the spatial characteristics of the sources was obtained from the mixing matrix *A.* Eye blinks produce high amplitude, low frequency activity (1-3 Hz) which was located mainly in the frontal electrodes. Morphologically, eye blinks appear as large amplitude in downward or upward deflections mostly in the frontal electrodes, such as FP1, FP2 sites. Lateral eye movements often appear as low frequency movements in electrodes F7, F8.

In addition to the eye artifact, sources located mainly in one electrode were also detected and eliminated at the SOBI stage. The basis for this decision is that neural activities are unlikely to appear exclusively in a single scalp sensor. The skull and scalp act as a low pass spatial filter and neural activity spreads across channels via volume conduction. If such a spread does not occur, the most likely reason is that the activity has been generated outside of the skull, and thus can be considered artifact. Morphologically this appears as single or multiple sharp waveforms. At other times when the impedance change is not so abrupt it may mimic low voltage arrhythmic delta waves.

Once the ocular artifact and additional artifact located mainly in one electrode are identified in SOBI, the columns of the mixing matrix *A,* which represent the activations of the artifactual sources, are set equal to zero and the cleaned data reconstructed.

$$X_{clean}(t) = A_{clean}z(t) \qquad \text{-----(3)}$$

The cleaned data from SOBI is then used as input to CCA. It was observed that the muscle artifact is well separated from components related to the brain activity and that they are found in the lowest components of CCA [6]. Muscle artifact is further identified by using the average and relative power in the typical EEG and EMG bands [12]. Identification of these muscle artifacts was easy since these artifacts were well separated in the CCA analysis. As before, once these artifacts were identified, the corresponding columns in the mixing matrix was set to zero and cleaned data was obtained using equation (3).

D. Comparison between Automatic (AT) and Visual Technique(VT)

Automatic artifact removal (AT) was performed using data obtained from a 32 channel EEG recording of 40 second duration. The 32-channels were recorded following the International 10-20 Montage system and the channels numbers (1-32) are in the following order (FP1, AP3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, PZ, PO3, O1, OZ, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, FP2, FZ, CZ). The data used was sampled at 256 Hz. The multi-channel EEG data collected and preprocessing was conducted using EEGlab [1]. In EEGlab, the following preprocessing was carried out: removal of non–EEG channels; linear trend in the data set removed using short IIR and highpass filter with a cut off frequency of 1 Hz; notch filtered in the frequency bands 45 to 55 Hz and 95 to 110 Hz to remove line noise and its harmonics; and data was then average referenced.

Once the components were generated using the SOBI method, the component data was sent to an EEG technician, with experience using ICA for removal of artifact, to identify the artifact components in the files using a subjective visual technique (VT). Comparisons from n=9 files from nine separate participants were analyzed. Comparisons were mainly performed using variance amongst electrodes as an index of artifact removal. A further n=30 files were analyzed using the automatic method to demonstrate its effectiveness and this data is presented in this paper.

One of the problems when artifact is removed from contaminated EEG is to determine how well the artifact algorithm has performed. One possible approach to check this in multi-channel EEG data is to check the variance amongst the multi-channels. If EEG channels contained neuronal activity only, one would expect that due to volume conduction, neuronal activity will be dispersed throughout the different EEG channels. This would ensure that the time series recorded at the various channels will be similar and hence the variance amongst the electrodes will be small. On the other hand if artifact external to brain activity is recorded, this similarity will be lost. The time series recorded at the various electrodes in this case will be different and the variance amongst the electrodes will be large. This is the principle idea behind using variance to monitor artifact removal in this paper.

In this procedure comparisons were made with the value in matrix X, of size (N x M) where N represents the electrodes and M the number electric potential values recorded at each electrode, with another matrix Y of similar size. The values contained in the matrix X are the values before artifact removal and the values contained in Y are after artifact removal. However the values in X and Y can also represent the different stages of artifact removal procedure. The time intervals at which the potential values are recorded in X and Y, was determined by the sampling frequency. For each electrode i, the variance *v(i)* is calculated as follows:

$$v_x(i) = \mathrm{var}(X(i, j = 1:M)) \qquad i=1,\ldots N \quad \text{----(1)}$$

$$v_y(i) = \mathrm{var}(Y(i, j = 1:M)) \qquad i=1,\ldots N \quad \text{----(2)}$$

III. Results

Table 1 shows the results comparing the components chosen using SOBI in the automatic technique (AT) compared with visually choosing (VT) the artifactual components by an EEG signal expert. The majority of the components chosen as artifact were common using both techniques. As SOBI was applied for the comparison, the VT would choose some components as artifact, which were line noise or muscular in nature, that was not detected using AT. CCA was used by AT to detect muscle artifact. For example in Subject 1, components 13, 16, 19, 20, 22 were muscle artifact chosen using VT, which were not selected by the AT method. However, AT was able to detect components 23 and 24 as components arising from one electrode source, which the VT was unable to find.

TABLE I
COMPARISON OF AUTOMATIC (AT) AND VISUAL (VT) TECHNIQUE

Subject No.	Components using AT	Components using VT
1	1, 5, 18,21, 23, 24, 26 30, 31, 32	1,5,13, 16, 18, 19, 20,21, 22, 26, 30, 31, 32
2	1, 3, 9, 10, 15, 18, 22, 23, 25, 26, 29, 30, 31, 32	1, 3, 8, 9, 10, 15, 22,23, 25, 26, 30, 31,32
3	1, 8, 12, 14, 19, 22, 24, 25, 26, 27, 30, 31, 32	1, 11, 14, 18, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32
4	1,2,3, 6, 10, 11, 13 19, 21, 22,26, 27,28, 29,30,31,32	1, 16, 17, 18, 21, 22, 26,27, 28, 29, 30, 31, 32
5	1, 2, 8, 13, 21 22, 30, 31, 32	1, 2, 6, 20, 21,22, 25, 30,32
6	1,10,14,17,20,21,23,26,27 ,30, 31,32	1, 14, 16, 17,18, 23, 26, 31,32
7	1, 4, 13, 21, 23, 25, 26, 30, 31, 32	1, 10, 13, 14, 15, 19, 20, 21, 23, 25, 26, 30, 31, 32
8	1, 7, 11, 14,15,18, 20, 24, 26, 27, 28,29,30,31, 32	1, 5, 10, 15, 16, 17, 21, 24, 26, 27, 28, 29, 30, 31, 32
9	1, 6, 8,12,15,18, 21, 22, 23, 24, 25, 26, 30, 31, 32	1, 20, 22, 23, 24, 26, 30, 31, 32

Fig 1. shows the overall artifact removel in both the AT (Red) and VT (Green) compared with the preprocessed raw EEG data (Blue) using the variance amonst the electrodes technique. The AT and VT plots are similar to each other with only very small differences in the variance, compared to the raw EEG. The data for the figure was from Subject No. 1 who was representative of the other eight subjects.

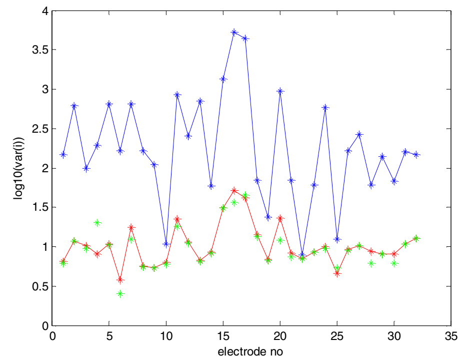

Fig. 1. Comparisons in variance in all 32 channels between Raw EEG (Blue), with AT (Red) and VT (Green) in representative Subject 1

Fig 2. Shows the mean electrode variance in the 32 channels in all nine subjects. Once again AT (Red) and VT (Green) are comparable as they overlap, showing that similar amounts of artifact were removed using both techniques compared to the raw EEG (Blue).

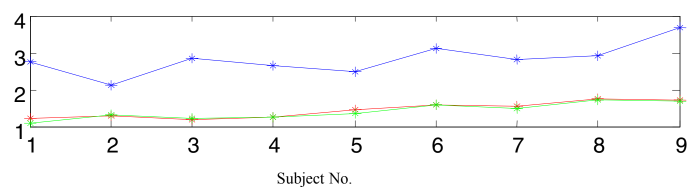

Fig. 2. Comparison in variance over 32 channels in N=9 subjects. The y-axis shows the $\log_{10}$ value of the mean variance. The raw EEG (Blue) had substantially larger variance values compared to the EEG signal after AT (Red) and VT (Green).

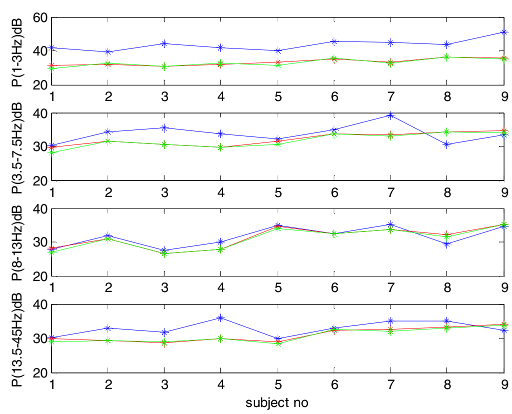

Fig. 3. Shows the spectral energy in delta (Top), theta (2nd Top), alpha (2nd bottom) and beta (bottom) bands over 32 channels in N=9 subjects. Blue line= raw EEG; Red line= AT; Green line= VT.

Spectral analyses was also computed to evaluate the artifact removal from both AT and VT. Fig 3. shows that AT and VT were comparable. There was a reduction in spectral amplitude mostly in slow frequencies such as delta (1-3Hz) and theta (3.5-7.5 Hz) demonstrating a reduction in ocular activity. Alpha activity (8-13Hz) remained mostly unchanged from the raw EEG. There was also a reduction in beta (13-45 Hz) activity from the raw EEG, this shows reduction in muscular artifact. The AT program was used to evaluate a further 30 subjects. Results were similar to that seen in the n=9 data above in both variance of electrodes and spectral analyses. Fig. 4 & 5. Shows the variance and spectral results in the total N=39 participants, respectively.

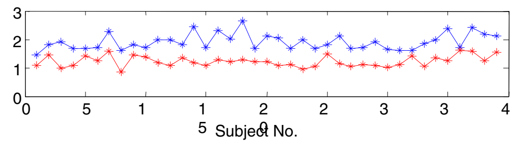

Fig. 4. Comparison in variance over 32 channels in N=39 subjects. The y-axis shows the $\log_{10}$ value of the mean variance. The raw EEG (Blue) had substantially larger variance values compared to the EEG signal after AT (Red).

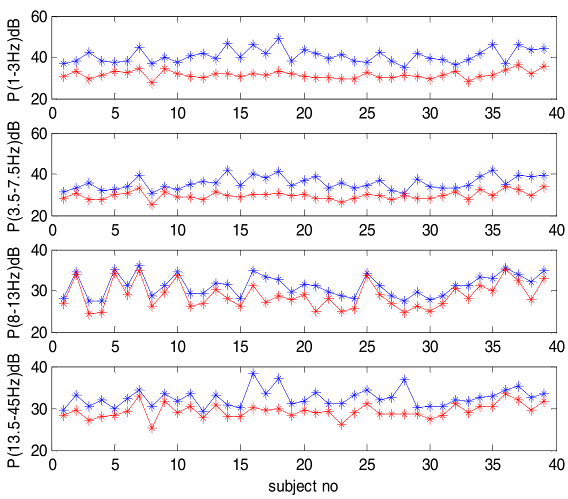

Fig. 3. Shows the spectral energy in delta (Top), theta (2nd Top), alpha (2nd bottom) and beta (bottom) bands over 32 channels in N=39 subjects. Blue line= raw EEG; Red line= AT.

IV. Discussion

This paper presented an AT used to remove various artifact from the EEG signal. Individually, techniques such as ICA, SOBI and CCA have been shown to be effective in removing EEG artifact [1,2,5,6], however, they often require a high level of expertise in EEG signals by the user. By automating the artifact removal procedure, the technique applied has been shown to become faster, easier to use, and more objective and valid in the manner in which artifact components are chosen. To evaluate the efficacy of the AT, we compared the artifact components chosen by the AT against a VT whereby components were chosen by an experienced EEG technician. The results show that components chosen by both techniques overlapped well. As only SOBI was used to choose components the overlap was seen in mostly ocular artifact and artifact arising from one electrode site. To test the two techniques further, a method measuring the variance amongst the electrodes and spectral analyses were used. The results show both AT and VT had decreased variance compared to the raw preprocessed EEG. AT and VT variances were similar to each other as the plots overlap. Spectral analyses showed reduction from the raw EEG in low frequencies such as delta and theta activity indicating possible removal of ocular artifacts and reduction in beta activity indicating possible removal of muscular artifact.

Acknowledgment

The contribution of German Gomez-Herrero of Tampere University of Technology, Institute of Signal Processing, Tampere, Finland has been pivotal in the study reported here. The SOBI and CCA code used here was written by him. His advice and insight in assisting us with the problems faced is much appreciated.

References

[1] A. Delorme and S. Makeig "EEGLAB: An open source toolbox for analysis of single trial EEG dynamics including independent component analysis," *J. Neurosci. Meth.,* vol 134, pp 9-21, 2004.

[2] G. Gomez_Herrero, Transform methods for electroencephalography (EEG): Automatic Artifact Removal (AAR) Toolbox for Matlab (Version 1.3), http://www.cs.tut.fi/~gomezher/projects/eeg/aar.htm

[3] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Sig. Proc.* vol 45, pp434-444 (1997).

[4] A.C. Tang, M.T. Sutherland and C.J. McKinney, "Validation of SOBI components from high density EEG," *NeuroImage*, Vol 25, pp539-553, 2004.

[5] C. A. Joyce, I.F. Gorodnitsky and M. Kutas, "Automatic removal of eye movements and blink artifacts from EEG data using blind component separation," *Psychophysiol.* vol 41, pp 313-325, 2004.

[6] W. De 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. Bio-Med. Eng.*, vol53, pp2583-2587, 2006.

[7] W. De Clercq, A. Vergult, B. Vanrumste, J. Van Hees, A. Palmini, W. VanPaesschen and S. Van Huffel, " A new muscle artifact removal technique to improve the interpretation of the ictal scalp electroencephalogram," *Proc. IEEE-EMBS*, Shangai, China pp1136-1139, 2005.

[8] G. Gomez_Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel and W. VanPaesschen, "Automatic removal of ocular artifacts in the EEG without a reference EOG channel", *Proc. NORSIG*, Reykjavik, Iceland pp 130-133, 2006.

[9] A. Delorme, T. Sejnowski and S. Makeig, "Enhanced detection of artifacts in EEG data using higher order statistics and independent component analysis", *NeuroImage*, vol 34, pp1443-1449, 2007.

[10] A. Greco, N.Mammone, F.C. Morabito and M. Versaci, "Kurtosis, Renyi's entropy and independent component scalp maps for the automatic artifact rejection from EEG data", *Int. J. of Biomed. Sciences*, vol 1, pp1306-1312, 2005.

[11] Y. Tran, A. Craig, P. Boord, and D. Craig, "Using independent component analysis to remove artifact from EEG measured during stuttered speech," *Med. & Biol., Eng. & Comp.*, vol 42 pp627-633, 2004.

[12] G. Gomez_Herrero, Transform methods for electroencephalography (EEG): Sample EEG recording from epileptic patients, http://www.cs.tut.fi/~gomezher/projects/eeg/aar.htm