33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# An Automatic Ocular Artifacts Removal Method Based On Wavelet-Enhanced Canonical Correlation Analysis

Chunyu Zhao, Tianshuang Qiu

*Abstract*—**In this paper, a new method for automatic ocular artifacts (OA) removal in EEG recordings is proposed based on wavelet-enhanced canonical correlation analysis (wCCA). Compared to three popular ocular artifacts removal methods, wCCA owns two advantages. First, there is no need to identify the artifact components by subjective visual inspection, because the first canonical components found by CCA for each dataset, also the most common component between the left and right hemisphere, are definitely related to artifacts. Second, quantitative evaluation of the corrected EEG signals demonstrates that wCCA removed the most ocular artifacts with minimal cerebral information loss.**

## I. Introduction

Electroencephalograph (EEG) recordings are frequently contaminated by ocular artifacts (eye movements and eye blinks), muscle noise, heartbeat and line noise, due to the limitation of EEG signal recording technique. These artifacts, especially the ocular artifacts (OA), often complicate the interpretation of the EEG recordings. Electrooculography (EOG) artifacts can generally be of orders of magnitude greater than the brain-generated electrical potentials. EOG has a spectral overlap with the underlying EEG and cannot be removed using conventional filtering. Besides, OA is most prominent over the anterior head regions [1].

Blind source separation (BSS) is a widely explored approach for correcting the ocular artifacts, which finds the effect of the artifact signals onto each electrode, and then subtracts the artifacts based on the weights from those electrodes. There are mainly two problems for BSS based methods. First, ocular sources extracted by independent component analysis (ICA) always contain some cerebral activities especially in anterior electrodes, while it did not happen with second order blind identification (SOBI) [2]. To overcome this problem, several studies have combined ICA and wavelet de-noising to remove ocular artifacts from EEG signals [3-4]. Besides, comparative studies have shown that SOBI performs better than ICA for removing the EOG artifact [5]. The second problem is the subjectivity of identifying the ocular artifact component by time-consuming visual inspection. To overcome that problem, some studies have attempted to find some rules using statistical properties such as kurtosis or entropy [6-7]. However, the performance of the combination methods mentioned above to remove the ocular artifacts is not satisfying and still needs improvement.

In this paper, a new method for ocular artifact removal in EEG recordings is presented: wavelet enhanced canonical correlation analysis (wCCA). First, the CCA is applied to the mixed signal in a new way according to the differences of the spatial distribution between the EEG signals and the EOG signals. There is no need to identify the artifact component by subjective visual inspection, because the first canonical components found by CCA for each dataset, also the most common component between the left and right hemisphere, are definitely related to artifacts. Then wavelet thresholding is employed to recover the cerebral activities leaked into this artifact component. The performance of the proposed method is tested on semi-simulated data, and compared to three popular OA removal methods (CCA, SOBI and wavelet-ICA) in terms of correlation coefficient and signal-to-artifact ratio (SAR). Furthermore, the method is illustrated on a real spontaneous EEG recording contaminated by obvious ocular artifacts for visual inspection.

## II. Methods

### A. DWT Method

The multi-resolution DWT (discrete wavelet transform) and IDWT (inversse discrete wavelet transform) can be implemented by cascades of several two-channel analysis and synthesis filterbanks, respectively. Through the lowpass filter an approximation signal is extracted, whereas by the highpass filter a detail signal is taken out.

### B. Common CCA Method

In most BSS-CCA cases, let $\mathbf{X}(t)$ be the observed data matrix with $K$ mixtures and $N$ samples, and let $\mathbf{Y}(t)$ be a temporally delayed version of the original data matrix $\mathbf{Y}(t) = \mathbf{X}(t-1)$. When the mean of each row from the data matrices $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ is removed, consider the linear combinations of the components in $\mathbf{X}$ and $\mathbf{Y}$, called the canonical variates

Manuscript received March 22, 2011. This work is supported partly by the National Science Foundation of China under the Grant 60940023 and the Plan of Liaoning Science and Technology Department of China under the Grant No.2010216008.

Chunyu Zhao was with the Department of Biomedical Engineering, Dalian University of Technology, Dalian, 116024, China. She is now with the School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, 19104, USA(e-mail: zhaocy.dut@gmail.com).

Tianshuang Qiu is with the Department of Biomedical Engineering, Dalian University of Technology, Dalian, 116024, China (corresponding author, e-mail: qiutsh@dlut.edu.cn).

$$\mathbf{u} = w_{x_1}\mathbf{x}_1 + ... + w_{x_k}\mathbf{x}_k = \mathbf{w}_\mathbf{x}^T\mathbf{X}$$
$$\mathbf{v} = w_{y_1}\mathbf{y}_1 + ... + w_{y_k}\mathbf{y}_k = \mathbf{w}_\mathbf{y}^T\mathbf{Y} \quad (1)$$

CCA finds the weight vectors $\mathbf{w}_\mathbf{x} = [w_{x_1}, ..., w_{x_k}]^T$ and $\mathbf{w}_\mathbf{y} = [w_{y_1}, ..., w_{y_k}]^T$ that maximize the correlation $\rho$ between the variates $\mathbf{u}$ and $\mathbf{v}$. The first variates constructed by CCA, $\mathbf{u}_1(t) = \mathbf{w}_\mathbf{x}^T\mathbf{X}(t)$ and $\mathbf{v}_1(t) = \mathbf{w}_\mathbf{y}^T\mathbf{Y}(t)$ are maximally correlated with each other, which can also be seen as the most common components existing in $\mathbf{X}(t)$ and $\mathbf{Y}(t)$. By putting the obtained components equal to the sources, BSS-CCA finds sources which are uncorrelated with each other, maximally autocorrelated and ordered by decreasing autocorrelation index $\rho_i$ [8]. When BSS-CCA is applied to the EEG, then the sources, or components, contributing to the EEG and the EOG are derived respectively, and the cleaned EEG data can be got by setting the component representing the ocular artifact source equal to zero. This is exactly the idea of the CCA-based ocular artifact removal method. However, it is inevitable that there are some commonly-existing cerebral activities leaked into the artifact component, and cancelling it will cause relevant cerebral information loss. Besides, without reasonable explanation of the physiology characteristic of the first canonical components found by CCA, it still needs visual inspection for artifact component identification.

### *C. wCCA Method*

Typically, EEG observations are obtained from the output of a multitude of scalp electrodes, where each sensor receives a different combination of the EOG sources and EEG sources. It is known that the EEG sources from different sites can be quite different. The sensorimotor rhythms originating from very localized areas in the cortex is just an example: the left and right hemisphere have quite different kind of waves (event related synchromization (ERS)/desynchronization (ERD)). On the other hand, the EOG amplitude is attenuated approximately with the square of the distance, with similar kind of waves across the cortex. Inspired by the differences of spatial distribution, a new version of BSS-CCA application is tried: let $\mathbf{X}(t)$ be the EEG recordings of the left hemisphere, and $\mathbf{Y}(t)$ be the EEG recordings of the right hemisphere. Besides, the vertical EOG signal is added to the end of both $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, in order to enhance the ocular artifacts' percentage in the first found canonical components. As we have described above, the first canonical components constructed by the CCA between the two sets of variables, are maximally correlated with each other, which is exactly the most common component between $\mathbf{X}(t)$ and $\mathbf{Y}(t)$. Obviously, combined with the physiology characteristic of the EEG and EOG signals, the ocular artifacts must be captured in the first canonical components, as well as some commonly-existing cerebral signals. Through this kind of BSS-CCA, we skillfully avoid the selection of artifact component manually, which poses quite a problem in most components based methods. There are inevitably some brain activities in this component for each dataset, and canceling it causes cerebral information loss. Thus the wavelet decomposition procedure is introduced as a thresholding: all wavelet coefficients above a certain threshold are set to zero, and then the resulting structure is used for the inverse wavelet transformation. The signal for each decomposition level is compared with their corresponding thresholds to find the artifacts. The wavelet function used in the current study is db4, which has been frequently used in many EEG studies. Here the simplest fixed form threshold [3-5] is used:

$$K_i = \sqrt{2\log(N_i)}\sigma_i \quad (2)$$

where $N_i$ is the number of samples in the level $i$, $D_i(t)$ is the detail signal of level $i$, $\sigma_i = \text{median}(D_i(t))/0.6745$, and $\text{median}(A)$ means the median value of $A$. For the highest level of analyzing signal, $D_i(t)$ is replaced by the approximation signal of that level ( $A_i(t)$ ) in the computation for $\sigma_i$. The main steps of wCCA-based ocular artifacts removal method are described as follows:

1) The canonical components for each dataset are obtained through CCA decomposition to the raw EEG signals.

2) The first found canonical components extracted by CCA, which are definitely the ones related to ocular artifacts for each set respectively, are transformed through DWT.

3) The wavelet coefficients for each level are compared with their corresponding thresholds, and all the wavelet coefficients above a certain threshold are set to zero.

4) The corrected artifact components are obtained through the wavelet reconstruction, using the nonselected details and the cleaned details.

5) The wCCA-corrected EEG signals are obtained through CCA reconstruction.

## III. RESULTS

### *A. Simulation Data*

The spontaneous EEG signals are recorded through a 32-channel electrocap, sampled at 250Hz. Vertical and horizontal EOG (VEOG and HEOG) signals are recorded through 4 electrodes around the eyes. After a careful inspection of no obvious artifacts, a 4s consecutive epoch obtained from 8 EEG channels according to the international 10/20 system (FP1, FP2, F3, F4, C3, C4, O1 and O2) are assigned as EEG sources. Another 4s epoch from the same subject with only obvious ocular artifacts is assigned as EOG sources for simulation. The ocular contamination is simulated by means of the addition of VEOG and HEOG sources (weighted by their corresponding propagation factors) to the EEG sources [9], composing the semi-simulated data. TABLE I shows the weights used in this paper. We took five equally different weight pairs accordingly to simulate different signal.

Table I

VEOG and HEOG Propagation Factors used to Simulate Data

| Channel | VEOG | HEOG |
|---|---|---|
| Fp1 | 0.969 ± 0.075 | 0.029 ± 0.046 |
| Fp2 | 0.983 ± 0.083 | -0.063 ± 0.041 |
| F3 | 0.495 ± 0.101 | 0.085 ± 0.027 |
| F4 | 0.463 ± 0.107 | -0.013 ± 0.026 |
| C3 | 0.223 ± 0.069 | 0.073 ± 0.025 |
| C4 | 0.202 ± 0.081 | -0.073 ± 0.023 |
| O1 | 0.027 ± 0.029 | 0.009 ± 0.023 |
| O2 | 0.027 ± 0.029 | -0.007 ± 0.021 |

*B. Validation of Correction*

Among so many criteria to test the validation of OA removal methods [9-10], two goals are summarized: (1) testing the degree of the removal of the ocular artifacts; (2) testing the quality of recovering of the brain signals. Combining the two goals, the proposed method can be evaluated effectively and objectively. The semi-simulated data are used so that we can estimate the recovering quality in time domain. In this paper, similarities of waveforms between the EEG sources and the corrected EEG data are evaluated by calculating the correlation coefficients between them. Besides, in order to quantify the degree of the removal of the ocular artifacts, the differences between the corrected EEG signals and the mixed EEG signals are originally measured through signal-to-artifact ratio (SAR) defined as follows:

$$\mathrm{SAR} = 10 \cdot \log \frac{\mathrm{mean}(\mathrm{EEG}_{\mathrm{corrected}})^2}{\mathrm{mean}(\mathrm{EEG}_{\mathrm{mixed}} - \mathrm{EEG}_{\mathrm{corrected}})^2} \qquad (3)$$

The proposed method is compared with three existing OA removal methods: CCA, ICA and wavelet-enhanced ICA [5].

Fig.1 shows the mixed EEG signals, the EEG sources and the corrected EEG signals after applying different OA removal techniques. Visual comparison of the proposed correction methods on two anterior channels Fp1 and F3 can be observed. We can see that, both wCCA and SOBI extract ocular artifacts with less loss of cerebral information than the other two methods.

Table II shows the averaged correlation values between the sources and the corrected EEG signals, corresponding to three brain areas: anterior, central, and posterior. We caculated the correlation coessicients of five different signals.The highest correlations as a whole are obtained when the wCCA algorithm is applied, despite the neglectably inferior performance in posterior areas. Because it is known that ocular artifacts decrease rapidly with the distance from the eyes, and the most serve interference occurs in the frontal areas. So the task of removing ocular artifacts without altering the cerebral activities is more challenging and important for frontal channels than posterior ones, so the proposed method performs much better than the other methods as a whole.

Table III shows the differences between the corrected EEG signals and the mixed EEG signals in terms of SAR (take the third weight pairs as an example). The highest SAR improvements are obtained after applying the wCCA algorithm, which means that wCCA removes far more artifacts than the other methods. And combining the results of the correlation coefficients criterion, the conclusion that wCCA-based method removes the most ocular artifacts from the mixed EEG data, and at the same time preserves the most neural signals can be made.

Table II

Correlation Coefficients for OA Removal Methods

| Area | EOG removal methods | | | | |
|---|---|---|---|---|---|
| | non-corrected | CCA | SOBI | wICA | wCCA |
| Anterior | 0.411 ± 0.016 | 0.847 ± 0.022 | 0.861 ± 0.012 | 0.685 ± 0.011 | **0.959 ± 0.004** |
| Central | 0.615 ± 0.074 | 0.906 ± 0.077 | 0.922 ± 0.055 | 0.841 ± 0.068 | **0.975 ± 0.006** |
| Posterior | 0.986 ± 0.013 | 0.985 ± 0.011 | 0.986 ± 0.007 | **0.990 ± 0.002** | 0.986 |
| All EEG Channels | 0.606 ± 0.013 | 0.896 ± 0.032 | 0.907 ± 0.021 | 0.800 ± 0.016 | **0.970** |

Table III

Signal-to-artifact Ratio for OA Removal Methods

| Area | EOG removal methods | | | | |
|---|---|---|---|---|---|
| | non-corrected | CCA | SOBI | wICA | wCCA |
| Anterior | -8.1475 | 5.2879 | 4.0281 | 2.1637 | **10.9728** |
| Central | -4.2194 | 4.0991 | 8.2817 | 4.5175 | **12.7002** |
| Posterior | 16.7085 | 16.6946 | 19.2960 | 17.3671 | **23.6340** |
| All | -0.9515 | 7.8424 | 8.9085 | 6.5530 | **14.5699** |

*C. Real Data*

Fig. 2 shows an example of 3.5s-epoch EEG signals before and after applying the wCCA-based ocular artifact removal method. The effects of ocular removal method on different EEG leads are shown in Fig.2 (b). The leads chosen for Fig. 2 are representative due to their distance to the eyes, corresponding to the left hemisphere. By visual inspection, this example demonstrates that wCCA algorithm removes efficiently the ocular artifacts from spontaneous real EEG data and preserving the neural signals.

## IV. Conclusion

In this paper, a new method for automatic OA removal in EEG recordings based on wavelet-enhanced CCA (wCCA) is presented. Making full use of the differences between the spatial distribution of EEG and EOG signals, wCCA applies the CCA to decompose the signals into canonical components and then applies wavelet de-noising for artifact removal on the artifact components for each dataset. The performance of wCCA method is evaluated on semi-simulated data, and outperforms three popular OA removal methods: CCA, ICA and wICA, removing the most ocular signal, with minimal cerebral information loss. Besides, the performance of the proposed method is illustrated on a real spontaneous EEG recording with obvious ocular artifacts. The ocular artifacts are successfully removed, with little EEG signal alterations.

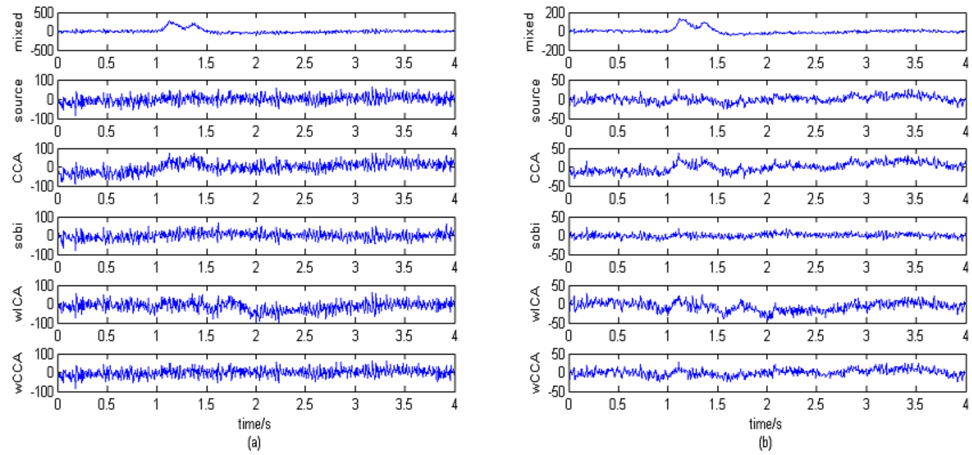


Fig.1 Visual comparison of the results corresponding to the mixed, the source and the corrected EEG signals by different ocular artifacts removal methods. Only Two channels are plotted as examples: (a) Fp1 (b) F3 because these frontal channels are contaminated most severely.

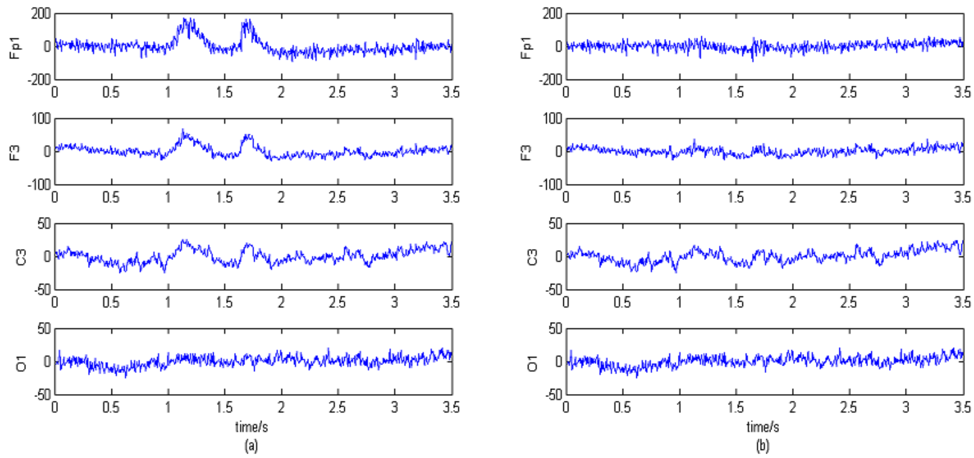


Fig. 2 An example of 3.5s-epoch EEG signals before and after applying the automatic wCCA-based ocular artifacts removal method (a) the raw EEG signals containing obvious ocular artifacts; (b) the corrected EEG signals.

## V. ACKNOWLEDGMENT

The authors would like to thank the reviewers for their invaluable comments and suggestions.

## REFERENCES

[1] M. Fatourechi, A. Bashashati, R. K. Ward and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey", Clinical Neurophysiology, vol. 118, no. 3, pp. 480-494, 2007.

[2] M. Congedo, C. Gouy-Pailler and C. Jutten, "On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics", Clinical Neurophysiology, vol. 119, no. 12, pp. 2677-2686, 2008.

[3] V.A. Makarov and N.P. Castellanos, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis", Neuroscience Methods, vol. 158, no. 2, pp. 300-312, 2006.

[4] G. Inuso, F. La Foresta, N. Mammone, and F.C. Morabito, "Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings", in Proc. Int. Conf., pp. 1524-1529, 2007.

[5] N. G. Siew-Cheok and P. Raveendran, "Enhanced u rhythm extraction using blind source separation and wavelet transform", IEEE Trans Biomed. Eng., vol. 56, no. 8, pp. 2024-2034, 2009.

[6] G. Barbati, C. Porcaro, F. Zappasodi, et al, "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals", Clinical Neurophysiology, vol. 115, no. 5, pp. 1220-1232, 2004.

[7] 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, no.4, pp. 1443-1449, 2007.

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

[9] S. Romero, M.A. Mananas and M.J. Barbanoj, "A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case", Computers in Biology and Medicine, vol. 38, no. 3, pp. 348-360, 2008.

[10] B.Noureddin, P. D. Lawrence and G. E. Birch, "Time-frequency analysis of eye blinks and saccades in EOG for EEG artifact removal", in Proc. 3rd. Int. Conf. IEEE. EMBS., pp. 564-567, 2007.