The Removal of Blink and Saccade Artifact in EEG recordings by Independent Component Analysis

Jie Dong, Tao Wang, Ai-tao Zhang, Hong-ya Dai
School of Biomedical Engineering
Southern Medical University
Guangzhou, China

***Abstract*—Pervasive electroencephalographic (EEG) artifacts are associated with eye movement. This study shows that Independent Component Analysis (ICA) using Joint Approximate Diagonalization of Eigenmatrice (JADE) algorithm can be applied to removing ocular artifacts in EEG. The nine channels of EEGs were recorded from three young healthy subjects with additional VEOG and HEOG channels in blinking and saccade conditions respectively. Seven EEG recordings were selected for artifacts removal tests with two EOG recordings indicating the possible interferences. Ocular artifacts for both conditions were successfully presented by an independent component, which was then eliminated to reconstruct the artifact-free EEGs. This study demonstrates that JADE algorithm can be an effective tool in correcting EOG interference with multichannel EEG recordings.**

Keywords*—*Electrencepholgram; Ocular artifacts; Independent Component Analysis;

I. INTRODUCTION

Eye movements are a major source of contamination of the electroencephalogram (EEG). This is because eye movements cause a change in the electric fields that surround the eyes, and these distort the electric fields over the scalp. So we require to accurately control the eye movement of subjects, but it's difficult to do it. Because significant eye movement due to the subjects' inability or reluctance to follow the instructions may also be a problem, especially with certain populations (e.g., children with attention deficit/hyperactivity disorder, schizophrenic patients). Consequently, ocular artifact must be removed from the EEG.

Ocular artifacts generated by two (or more) distinct mechanisms can be recorded at the head surface. Blink artifacts are attributed to alterations in conductance arising from contact of the eyelid with the cornea. The majority of this signal propagates through the head and decreases rapidly with distance from the eyes. Eye movements which are caused by retinal dipole movement generate another type of electric signal. As the retino-corneal axis rotates during eye movements, the orientation of this dipole in three-dimensional space also rotates, resulting in changes in electric potential.

Eye blinks and saccades are studied as artifacts of EEG recordings. The fraction of the EOG potential transmitting to the scalp EEG electrodes as artifacts, is calculated for potentials generated during both blinks and eye movement. The 'propagation' of eye-movement voltages to scalp sites is dependent on a number of factors (such as the filter properties of the subject's skull, scalp and neuronal tissues), but as an approximation, the voltage is inversely related to the square of the distance from the eyes [1]. In general, about 0.20 of the vertical eye-movement voltage reaches Fz, with the amount decreasing posteriorly to about 0.05 at occipital sites, and with little effect of laterality.

An alternative approach for ocular artifact reduction is based on blind source separation techniques (BSS). Independent Component Analysis (ICA) [2] has been commonly applied to both EEG source isolation and artifact correction [3] [4] [5] [6]. ICA-based analysis methods are able to separate out EEG signals from an estimate of the overlapping projections of the artifact in all scalp EEG electrodes, assuming that sources are statistically independent. ICA preserves more brain activity than other correction techniques (i.e., PCA), and is able to separate a wide variety of EEG artifacts simultaneously, without needing of reference channels. JADE is a fundamental algorithm based on 4th-order cumulants, and might be more effectively on the low-dimensional data than other ICA algorithms.

The aim of this study is to present an applicable method for removing the blink and saccade artifacts in EEG recordings by ICA with the JADE algorithm. The typical signal evolutions of the major processes are presented for a visual validation. A qualitative measurement to describe the difference between the original signals and artifact-removal ones is also proposed to evaluate the presented method.

II. METHOD

A. Data model

We consider the n-dimensional vector of sensor signals defined as $x(t)=[x_1(t),\cdots,x_n(t)]^T$ generated by an unknown linear model:

$$x(t) = As(t) \tag{1}$$

where $s(t)=[s_1(t),\cdots,s_m(t)]^T$ is the m-dimensional vector whose elements are referenced to as sources. T indicates the matrix transposition. The matrix A is called a mixing matrix. The source signals, $s(t)_j$, j=1,…, m are assumed independent.

The mixing matrix A is a function of the geometry of the sources and the derivations in an EEG recording and the conductivity properties of the brain, cerebrospinal fluid, skull and scalp. Although this matrix is unknown, we assume it to be constant, or slowly changing (to preserve some local constancy).

The most important problem of ICA is to estimate the independent signals from their mixtures, or the equivalent problem of finding the separating matrix B that satisfies:

$$y(t) = Bx(t) \tag{2}$$

Independent component analysis is an optimization problem, there is no unique solution, only a measure of independence in the best sense of the criterion for the approximate solution to $y(t)$ with each component to be independent of each other. So $y(t)$ is the estimated value of $s(t)$, but it is well known that due to lack of prior information, the problem of source separation has two inherent ambiguities: permutation ambiguity and scaling ambiguity.

B. *JADE algorithm*

Joint Approximate Diagonalization of Eigenmatrice (JADE) is a batch processing algorithm based on 4th-order cumulants. Most BSS algorithms can adapt to the case with only sub-Gaussian or super-Gaussian signals. While JADE algorithm is able to separate both sub-Gaussian and super-Gaussian sources, it is possible to extract the signals that are statistically independent, and to identify the spatial map that is associated with the artifactual components.

Before blind source separation, we should preprocess the data. The step consists of whitening the data by *sphering matrix* W in order to transform the mixing matrix A into a unitary matix U. Then we retrieving this unitary matrix by joint diagonalizing a set of eigen matrices of data quadricovariance referred to JADE.

To a n-dimensional random vector z with 4th-order cumulants, a quadricovariance Q is associated. It is defined as the linear matrix-to-matrix mapping: $M \rightarrow N = Q(M)$ where M and N are matrices related by:

$$[N]_{ij} = [Q_z(M)]_{ij} = \sum_{kl} cum(z_i, z_j, z_k, z_l)[M]_{kl}, \quad i, j = 1, 2, \cdots, n \tag{3}$$

where $cum(z_i,z_j,z_k,z_l)$ denotes the 4th-order cumulants of z [7]. It is shown that since the set of the $n \times n$ matrices is a n^2-dimensional linear space, there exist n^2 real numbers λ_r and n^2 orthonormal matrices M_r, verifiying $Q(M_r) = \lambda_r M_r$, $r = 1,\ldots,n^2$. Note that Q is actually a 4th-order tensor and the matrices are the eigenmatrices of Q associated to its eigenvalues λ_r. It is proved [7] that the quadricovariance Q has exactly rank n so that only n out of n^2 eigenvalues are non-zero.

Under the linear data model of (1) and hypothesis of independent sources, the eigenmatrices of the quadricovariance of the whitened data have the following structure,

$$M_r = UD_rU^T, D_r = Diag[\cdots, u_p^T M_r u_p, \cdots] \tag{4}$$

where $r = 1,\ldots,n^2$ and u_p denotes the p-th column of matrix U. Note that the n non-zero eigenvalues associated to the eigenmatrices M_r correspond to the kurtosis of the sources. According to (3), the missing unitary matrix U can be obtained by a joint diagonalization of the n eigenmatrice M_r [7]. Once the unitary matrix U is obtained, the mixing matrix is estimated by $A = UW$ and the unmixing matrix is then given by U^TW^T.

C. *Removing EOG artifact by JADE*

This investigation adopts a matrix notion system such that x is a matrix with row vectors used to represent multichannel EEG signals. We can then obtain a linear transform matrix B, where the output $y(t) = Bx(t)$ represent the time courses of activation of the ICA components. The columns of the inverse matrix, B^{-1}, give the projection strengths of the ICA components onto the scalp sensors. According to these strengths, we can use two-dimensional interpolation to draw the scalp topography of corresponding ICA component. For EEG, each component is a linear weighted sum of the brain activities at all the electrodes (the weights being a row of B). The independent components comprise these activations and the associated scalp maps give the projection weights of the components back into the electrode space. The scalp topographies of the components provide evidence for their biological origins (e.g., eye activity should project mainly to frontal sites). The components that are judged as artifacts will be eliminated by setting these components to zeros, which then resulting in a matrix $y'(t)$ that retains only components mixed by desired EEG signals. The 'corrected' EEG signal can then be derived as $x'(t) = B^{-1}y'(t)$, where B^{-1} indicates the matrix inverse. Figure 1 illustrates the whole processing steps.

The performance of the proposed method is validated quantitatively by evaluating the similarity between the raw and the 'corrected' EEG data. Specifically it is called Relative Root Mean Squared Error (RRMSE):

$$RRMSE = \sqrt{\frac{E\{\|x - x'\|^2\}}{E\{\|x\|^2\}}} = \sqrt{\sum_{p=1}^{P}(x(m,n) - x'(m,n))^2 \Big/ \sum_{p=1}^{P} x(m,n)^2}$$

where $x(t)$ is the raw EEG signals recorded at different electrodes, and $x'(t)$ is the 'corrected' EEG data by subtraction of artifactual independent components. $E\{\cdot\}$ denotes the mathematical expectation value of the bracketed quantity. The P is the total EEG samples.

D. *EEG data recording paradigms*

The nine channels of EEGs were recorded in a lab from three young healthy subjects using the 10-20 international system with additional VEOG and HEOG channels. They were all referred to right mastoid. The electrode impedances were kept below 5kΩ. Analogue filters were set between 0.5-100Hz. Raw EEG data were stored at a sampling rate of 500Hz. The

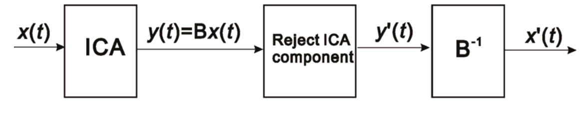

Figure 1. A block digram for the denoising process

data were recorded with Neuroscan SysAmp2 system. During recording, the subjects were instructed to do the following eye-movements: 0-1min closed the eyes; 1-2min blinked normally without excessive effort; 2-3min closed the eyes; 3-4min saccade from one fixed point to another fixed point. In this study, we only choose two EEG conditions, i.e., eye blinks and saccades for artifacts removal tests.

III. RESULT

All the EEG data were segmented into 12 pieces, each contains 5 secs for both conditions. The ICA based artifact removal computation was conducted piece by piece. Figure 2 shows an instance of a piece of raw EEGs (2500 samples) during 1-2min (in blinking task) of a recording block. We can see that EOG peaks caused by eye movements clearly appear on FP1, FP2, and Fz. At the same time, it shows that the HEOG amplitude is about 500μV which is almost 5 times larger than VEOG amplitude, therefore, HEOG is supposed to have more influence on scalp electrodes than VEOG does.

Figure 3 shows the ICA components of Figure 2 obtained by using JAED algorithm. It is obvious that IC2 with some transient short-pulse should be the main component of ocular activity. This hypothesis is verified by the scalp topography of corresponding ICA component activations as show in Figure 4.

Figure 4 shows the scalp topographies of corresponding ICA component activations of EEG data shown in Figure 2. The second topography which corresponds to the projection strength of IC2 to the electrodes shows that the main energy concentrated in FP1, FP2 and other areas near the eyes, and declining rapidly from forehead to the occipital electrodes. It is therefore concluded that IC2 reflects the main contribution of blink EOG interference sources. By the same way, the other topographies shows that IC3 and IC5 are distributed across the whole head, which might reflect random noise without specific sites of sources; and IC1, IC4, IC6 and IC7 are distributed mostly in the front, left, right and back of the head, suggesting useful EEG waves free from ocular artifact.

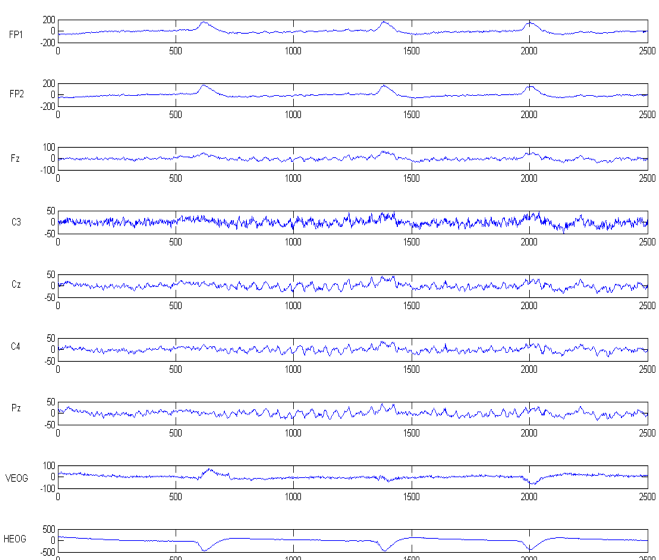

Figure 2. A 5-sec (2500 samples) portion of the 1-2min (blinks) recorded EEG signals

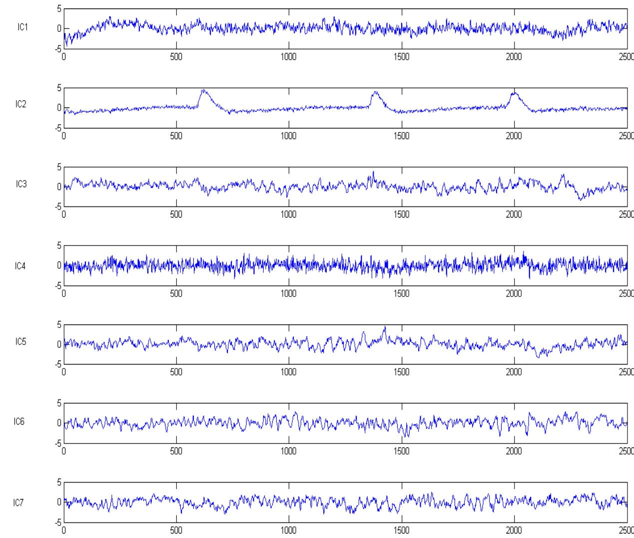

Figure 3. Corresponding ICA component activations of EEG data shown in Figure 2

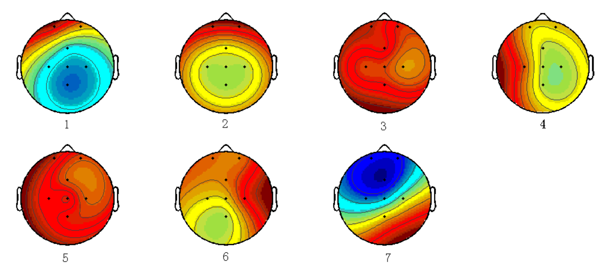

Figure 4. The scalp topography of corresponding ICA component shown in Figure 3

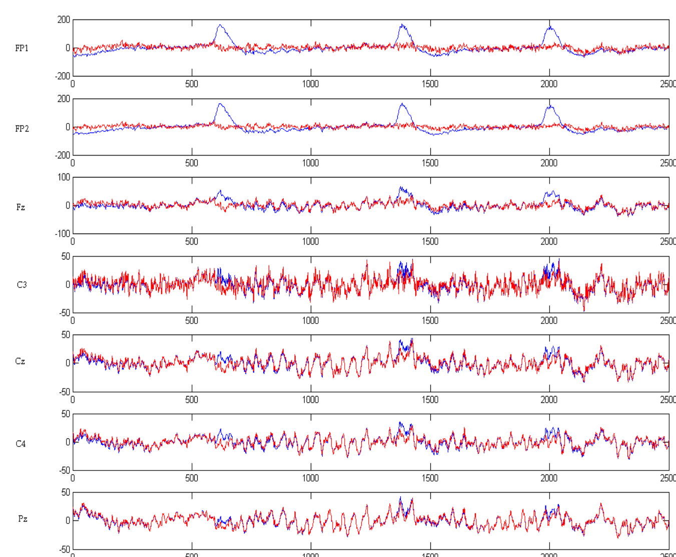

Figure 5. The 'corrected' EEG signals(red lines) obtained by removing selected EOG noise components and the raw EEG signal(blue lines)

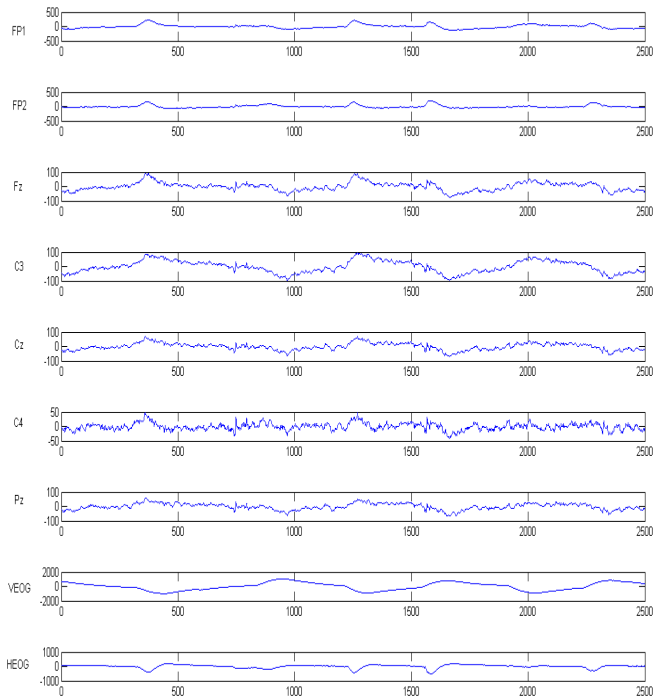

Figure 6. A 5-sec (2500 samples) portion of the 3-4min (saccades) recorded EEG times series

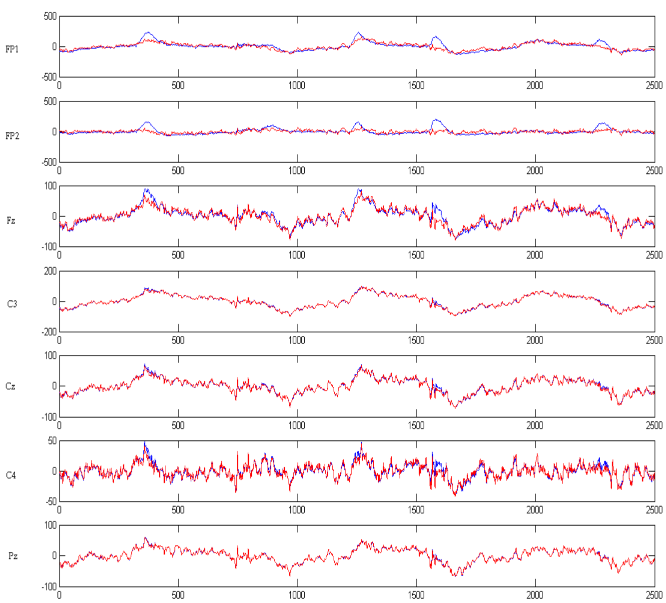

Figure 7. The 'corrected' EEG signals(red lines) obtained by removing selected EOG noise components and the raw EEG signal(blue lines).

The 'corrected' EEGs obtained by removing selected EOG noise components, i.e., IC2 from the raw data are shown in Figure 5. It can be seen that the peaks caused by eye movements on the electrode FP1 and FP2 are mostly removed.

The same operations were carried out for data in which the saccade eye movement is performed. Figure 6 shows a 5-sec portion of the 2-3min (saccades) raw EEGs. We can see that EOG artifacts are manifested in FP1 and FP2 electrodes. Other channels, such as, Fz, C3 and Cz, are also affected more or less by activities from the eye movements. Compared to blinking data, the VEOG amplitude is of -2~2mV, and the HEOG is of -1~1mV, meaning that they both large enough to have great influence on the scalp electrodes.

The ICA method is also able to identify a component that originated in the site of EOGs. When it is ruled out from EEG reconstruction, the 'corrected' EEG thus obtained as shown in Figure 7.

For blinking condition, the mean RRMSEs over all EEG pieces of the electrode FP1 and FP2 are 0.5694±0.2821 and 0.6263±0.269 respectively, while they are less than 0.01 for all the other channels. As to the saccade condition, the RRMSEs of the electrode FP1 and FP2 is 0.4271±0.2058 and 0.5177±0.0884 respectively. Since the larger values of RRMSE indicate more portion of signal contents which are considered here as the EOG artifacts, are removed, it is therefore concluded that the ICA with JADE algorithm is efficient in filtering out EOG artifacts that mainly contaminate frontal channels, and preserving the original details of EEG information.

IV. Discussion

This study shows that ICA using JADE algorithm can be applied to removing ocular artifacts in EEG if the reference EOG signals are available. However, there appear to be a difference in terms of RRMSE between the removal of blink and saccade artifact. The ICA removed more artifacts produced by eye blinking than that by saccade. The reason might lie in the intrinsic properties of EOGs, such as the magnitude of EOG activities, the generation mechanism and the eye movement timing, duration and frequency etc. Since blinking is a relative transient and simple movement for eyes to accomplish, the observed signals also present clear patterns, ICA method is thus performed better in the blinking condition.

We do not have a criterion against which to measure the success of the correction methods. Consequently, Verleger et al. proposed that corrected waveforms should have face validity: they should look reasonable [8]. Although this does not offer an explicit test of a method's effectiveness, we believe that it is the most useful form of validation because it utilizes the experience of the experimenter. In this article, we displayed typical signal evolution of the major processes to present a visual validation; we also used the RRMSE as a qualitative measurement to evaluate the similarity of the raw and the 'corrected' EEG data.

V. Conclusion

This article also presents an applicable method for removing the ocular artifacts from EEG records by JADE. Ocular artifacts for both eye blinks and saccades are successfully extracted in one independent component, which are then ready to be eliminated before the corrected EEG reconstruction. This study demonstrates that JADE algorithm can be an effective tool in correcting EOG interference with multichannel EEG recordings.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (60771035). Correspondence to T. Wang: tatwang@gmail.com

REFERENCES

[1] B. Rockstroh, T. Elbert, N. Birbaumer, W. Lutzenberger, Slow brain potentials and behavior. Baltimore: Urban and Schwarzenberg, pp. 55-57, 1982.

[2] P. Comon, "Independent component analysis, a new concept?" Signal Process, vol. 36, pp. 287–314, 1994.

[3] J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda, "Independent component analysis as a tool to eliminate artefacts in EEG: a quantitative study," J. Clin. Neurophysiol, vol. 20, pp. 249–257, 2003.

[4] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," Psychophysiology, vol. 41, pp. 313–325, 2004.

[5] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. Mckeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," Psychophysiology, vol. 37(2), pp. 163–178, 2000.

[6] Y. Tran, A. Craig, P. Boord, and D. Craig, "Using independent component analysis to remove artifact from electroencephalographic measured during stuttered speech," Med. Biol. Eng. Comput., vol. 42, pp. 627–633, 2000.

[7] J. F. Cardoso and A. Souloumiac, "An efficient technique for blind separation of complex sources," in Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA, 1993.

[8] R. Verleger, T. Gasser, J. Mocks, "Correction of EOG artifacts in event-related potentials of the EEG: aspects of reliability and validity," Psychophysiology , vol. 19, pp. 472-480, 1982.