Removal of eye blinking artifact from the electro-encephalogram, incorporating a new constrained blind source separation algorithm

L. Shoker S. Sanei W. Wang J. A. Chambers

Centre for Digital Signal Processing, Cardiff School of Engineering, Cardiff University, Cardiff, UK

Abstract—*A robust constrained blind source separation (CBSS) algorithm has been developed as an effective means to remove ocular artifacts (OAs) from electro-encephalograms (EEGs). Currently, clinicians reject a data segment if the patient blinked or spoke during the observation interval. The rejected data segment could contain important information masked by the artifact. In the CBSS technique, a reference signal was exploited as a constraint. The constrained problem was then converted to an unconstrained problem by means of non-linear penalty functions weighted by the penalty terms. This led to the modification of the overall cost function, which was then minimised with the natural gradient algorithm. The effectiveness of the algorithm was also examined for the removal of other interfering signals such as electrocardiograms. The CBSS algorithm was tested with ten sets of data containing OAs. The proposed algorithm yielded, on average, a 19% performance improvement over Parra's BSS algorithm for removing OAs.*

Keywords—*Ocular artifact, Blind source separation, Electro-encephalogram*

Med. Biol. Eng. Comput., 2005, **43**, 290–295

1 Introduction

OCULAR ARTIFACTS (OAs), also called electro-oculograms (EOGs), are the main source of interference within electro-encephalogram (EEG) signals. These pose a significant problem to clinicians and neurologists, because of the large number of data that can be lost owing to their presence. Such OAs refer to the potential difference that is generated when the eye moves in its socket or when a blink occurs. OAs propagate to other recording electrodes and superimpose themselves on the existing EEG. They can be measured by placing electrodes around the eyes. Horizontal eye movement can be measured by placing electrodes on either side of the eyes, whereas vertical movement and blinks can be measured by electrodes placed above and below the eyes.

The interfering eye blinks generate a signal that is in the order of ten times larger than cortical signals. Eye blinks can last between 200 and 400 ms. The eyeball can be considered as a dipole rotating in a socket. This means that, as the eye rotates, the cornea remains at 0.4–1 mV positive with respect to the retina. Rotations of the eyeball in saccadic eye movements cause large, external field variations that can contaminate EEG readings (OVERTON and SHAGASS, 1969). Owing to the magnitude of the blinking artifacts and the high resistance of the scalp, OAs can contaminate the majority of electrodes, even those in the occipital area. Some experiments have been attempted to estimate the propagation factors in the past (GRATTON, 1988).

It is possible to ask patients to fix their eyes on a point, which will reduce the number of eye movements, but involuntary movements, such as eye blinks, are just as troublesome. Asking the patients to suppress eye blinks will distract them from the clinician's instructions and proves to be impossible, for example, when examining children. Closing the eyes results in increased involuntary eye movements. Eye blinks may be in response to a cognitive task, and therefore simply rejecting the data segment will result in the loss of important information.

The main reason why EOGs cannot be simply removed using conventional filtering techniques is because of the spectral overlap between EOG and the underlying EEG. Numerous methods have been employed for removal of OAs that exploit the use of regression analysis, which is incorporated into popular EEG monitoring software, such as Neuroscan. Part of the EOG is subtracted from the EEG such that *Corrected EEG* $=$ *Raw EEG* $- \gamma$*EOG*, where the EOG is measured at the mastoids, which removes the need for horizontal and vertical EOG measurements (ELBERT *et al.*, 1985).

The parameter γ has been determined in numerous ways, such as being the ratio between EEG and EOG. In SCHLOGL and PFURTSHELLER (1999), γ was determined by the covariance between EOG and EEG. However, owing to volume conduction, OAs contain some EEG information that will inevitably be subtracted using the techniques mentioned.

Adaptive filters have been implemented for the removal of EOG artifacts (HE *et al.*, 2004). The vertical and horizontal

Correspondence should be addressed to Mr Leor Shoker; email: shokerl@cf.ac.uk

Paper received 1 April 2004 and in final form 20 December 2004

MBEC online number: 20053996

EOGs were measured and used as reference inputs to the adaptive algorithm. In another approach, CELKA *et al.* (2001) proposed a method that does not require a reference input for removing the EOG artifact. Their adaptive algorithm estimated the EOG by predictive filtering techniques. The mentioned adaptive filtering techniques show promising results; however, they operate on one EEG channel at a time, which can be computationally expensive, especially when a large number of electrodes are used. Moreover, the adaptive filtering techniques do not consider all the information within the EEG signals, and hence, their use in artifact rejection is not efficient.

Another method for removing blinking artifacts from EEG was proposed by BEWRG and SCHERG (1994) using principal component analysis (PCA). It finds the orthogonal directions of greatest variance in the EEG signals. PCA is based on explicit spectral matrix factorisation of the EEG signals, and therefore the application of PCA is generally superior to the traditional, aforementioned regression technique. The main drawback of PCA lies in the fact that neurobiological signals are not believed to be orthogonal, and, hence, OAs will not always be effectively removed (BELL and SEJNOWSKI, 1995).

One area that has sparked interest in the biomedical field is the use of independent component analysis (ICA) in blind source separation (BSS). ICA is a method of estimating the sources given that only the mixtures are available. This is achieved by making as few assumptions as possible about the original sources. One common assumption in most ICA algorithms is that the sources are statistically independent. This is a stronger claim than uncorrelatedness, because it assumes that the joint probability density of the sources can be factorised into the product of marginal densities (HYVARINEN *et al.*, 2001).

With this assumption in mind, many algorithms are designed so that the estimated sources meet, albeit approximately in practice, this criterion. One such algorithm is based on the information maximisation (Infomax) theorem (JUNG *et al.*, 2001). It uses a neural network to segregate individual components by maximising the joint entropy at the output which in turn minimises the mutual information between components.

They later designed a system where EEG signals were segregated, and then the eye blinking effect was removed. The separated signals were then recombined to reconstruct the artifact-free EEG. A similar system used second-order blind identification (SOBI) techniques (JOYCE *et al.*, 2003) and thereby relied on exploiting the temporal structure in the signals.

EEGs are said to be instantaneous mixtures, as potentials are due to emissions of electromagnetic dipoles, and the bandwidth of the signal (and accordingly the required sampling frequency) is very low (bandwidth <50 Hz). This in turn means that the signals measured at the electrodes are received with a negligible delay (linearly mixed), and, hence, an instantaneous type of ICA is used for separation of EEG signals. Although the number of signal sources within the brain is yet unknown, an initial assumption is that the number of sources N is less than the number of electrodes M, i.e. an over-determined system has been considered.

The challenge is to separate the signals into their independent constituent sources while automatically removing the artifact and retaining any diagnostic information about the brain disorder. In this paper, a pre-determined reference is incorporated into the minimisation algorithm, hence yielding an automated artifact rejection system. The significance of the algorithm is also due to its performance in the case of an undetermined number of sources.

2 Joint diagonalisation of correlation matrices

ICA relies on the fundamental assumption that the source signals within $\boldsymbol{s}(t) = [s_1(t),\ s_2(t),\ \ldots,\ s_N(t)]^T$ are statistically independent and zero mean, where t denotes the discrete time sample, and $(\cdot)^T$ is the vector transpose. This means that the joint distribution of the source signals can be factorised into the product of their marginal densities, i.e. $p(\boldsymbol{s}) = \prod_i p_i(s_i)$. The mixtures can be modelled by

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{v}(t) \qquad (1)$$

where $\boldsymbol{A}$ is the $M \times N$ full column rank mixing matrix, N is the number of sources, M is the number of mixtures, and $M > N$; $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$ contains the linear mixtures observed at the electrodes; and $\boldsymbol{v}(t) = [v_1(t), v_2(t), \ldots, v_M(t)]^T$ is the additive zero mean sensor noise. We assume that the sensor noise is temporally uncorrelated, i.e. $E\{\boldsymbol{v}(t)\boldsymbol{v}^T(t-k)\} = \boldsymbol{0}\ \forall\ k \neq 0$, and uncorrelated with the sensor data $E\{\boldsymbol{v}(t)(\boldsymbol{A}\boldsymbol{s}(t))^T\} = \boldsymbol{0}$. The output of the ICA system (i.e. the estimated original sources) is given by

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t) \qquad (2)$$

where $\boldsymbol{y}(t) = [y_1(t), y_2(t), \ldots, y_N(t)]^T$ is the vector of the estimated sources, and $\boldsymbol{W}$ is the $N \times M$ separation matrix. The separation matrix can be found by finding the minimum of a cost function $J(\boldsymbol{W})$, which provides a measure of independence of the estimated sources. Therefore the goal of the diagonalisation algorithm is to find a $\boldsymbol{W}$ that will make a set output covariance matrix $\boldsymbol{R}_Y(k)$ diagonal, $k \in \{1, 2, \ldots, K\}$, where K is the maximum time lag. Hence, minimising $J(\boldsymbol{W})$ will ensure that the estimated sources are as independent as possible. The covariance matrix $\boldsymbol{R}_Y(k)$ to be diagonalised is given by

$$\boldsymbol{R}_Y(k) = \boldsymbol{W}[\boldsymbol{R}_X(k) - \boldsymbol{R}_V(k)]\boldsymbol{W}^T \qquad (3)$$

where, in practice, $\boldsymbol{R}_X(k)$ is the estimate of the time-lagged covariance matrix of the signal mixtures, and $\boldsymbol{R}_V(k)$ is the estimate of the covariance matrix of the sensor noise. As we assume that the noise is spatially uncorrelated, $\boldsymbol{R}_V(k)$ will be a diagonal matrix for $k = 0$ and $\boldsymbol{R}_V(k) = \boldsymbol{0}$ for $k \neq 0$ (PARRA and SPENCE, 2000).

$$\boldsymbol{R}_X(k) - \boldsymbol{R}_V = \boldsymbol{A}\boldsymbol{R}_S(k)\boldsymbol{A}^T - \boldsymbol{R}_V \qquad (4)$$

where $\boldsymbol{R}_S(k)$ is a diagonal covariance matrix of the independent source signals. Following PARRA and SPENCE, 2000, the least squares (LS) estimate of $\boldsymbol{W}$ is

$$\boldsymbol{W}_{opt} = \arg\min_{\boldsymbol{W}} \sum_{t=1}^{T_B} \|E(t)\|_F^2 \qquad (5)$$

where $\|\cdot\|_F^2$ is the squared Frobenius norm, $E(t)$ is the error to be minimised between the covariance of the source signals $\boldsymbol{R}_S(k)$ and the estimated sources $\boldsymbol{R}_Y(k)$, and T_B is the data block length. Therefore a suitable cost function is defined based upon minimising the off-diagonal elements for multiple lagged covariance matrices, as

$$\begin{aligned}\boldsymbol{W}_{opt} &= \arg\min_{\boldsymbol{W}} \sum_{k=1}^{K} J_M(\boldsymbol{W}, k) \\ &= \arg\min_{\boldsymbol{W}} \sum_{k=1}^{K} \|\boldsymbol{R}_Y(k) - diag(\boldsymbol{R}_Y(k))\|_F^2 \qquad (6)\end{aligned}$$

where $diag(\cdot)$ is an operator that zeros the off-diagonal elements of a matrix.

3 Constrained learning

Minimising the cost function in (6) alone is not enough to remove the EOG from the underlying EEG, as there is no constraint to minimise the effect of the EOG. This is very important in places when there is an undetermined number of sources such that the output independent components (ICs) may not represent the actual sources. In this case, minimisation of the cost function should be subject to a constraint. To impose the constraint, a second cost function term is introduced as $J_G = F(E\{\boldsymbol{g}(t)\boldsymbol{y}(t)^T\})$, where the non-linear function $F(\cdot)$ is chosen based on the probability density function (PDF) of the data (LU and RAJAPAKSE, 2001). The non-linear function is chosen to have

$$F(g) = C(g) \approx \int_{-\infty}^{g} P_g(\xi)d\xi \tag{7}$$

where $P_g(\xi)$ is the PDF of the artifact, and $C(g)$ is the cumulative density function (CDF) of the artifact. We choose a function that is as close as possible to the CDF of the artifact so that its derivative will approximate its PDF, as shown in Fig. 1. Then a new function $J_T(\boldsymbol{W})$ is defined, so that

$$\begin{aligned} \boldsymbol{W}_{opt} &= \arg\min_{\boldsymbol{W}} \sum_{k=1}^{K} J_T(\boldsymbol{W}) \\ &= \arg\min_{\boldsymbol{W}} \sum_{k=1}^{K} (J_M(\boldsymbol{W},\ k) + \Lambda J_G(\boldsymbol{W})) \end{aligned} \tag{8}$$

where $\Lambda = \{\Lambda_{ii}\}$ $(i = 1, \ldots, N)$ is the weighted factor that is governed by the correlation between the EOG and EEG signals ($\boldsymbol{R}_{GY}$), defined by

$$\Lambda \simeq P diag(\boldsymbol{R}_{GY}) \tag{9}$$

where $P \in \mathbb{R}^+$ is an adjustable constant. Therefore the cost function to be minimised is the sum of $J_T(\boldsymbol{W})$. We use the natural gradient algorithm (NGA) (HAYKIN, 2002) to find the $\boldsymbol{W}$ that minimises $J_T(\boldsymbol{W})$. The general NGA update equation is

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \Delta\boldsymbol{W}(t) \tag{10}$$

where $\Delta\boldsymbol{W}(t)$ is the incremental update of $\boldsymbol{W}(t)$ given by CICHOCKI and AMARI (2002)

$$\Delta\boldsymbol{W}(t) = -\mu(t)\frac{\partial J_T(\boldsymbol{W})}{\partial \boldsymbol{W}}\boldsymbol{W}\boldsymbol{W}^T \tag{11}$$

The adaptive learning rate $\mu(t)$, as used in WANG *et al.* (2003), is dependent on the spread of the data and the gradient of the total cost function. It is given by

$$\mu(t) = \mu_0\left(\frac{1}{\sum_{k=1}^{K}\|\boldsymbol{R}_X(k)\|_F^2} + \frac{2}{\zeta + \|\Delta J_T(\boldsymbol{W})\|^2}\right) \tag{12}$$

where μ_0 is a positive constant typically < 1, and ζ is a regularisation parameter that prevents the learning rate from being too large when the gradient becomes small. The typical value of the parameter ζ is 0.05, and $\Delta J_T = J_T(\boldsymbol{W}(t-1)) - J_T(\boldsymbol{W}(t))$. Finding the gradient of (8) yields

$$\begin{aligned} \frac{\partial J(\boldsymbol{W})}{\partial \boldsymbol{W}} &= 4\sum_{k=1}^{K}[\boldsymbol{R}_Y(k) - diag(\boldsymbol{R}_Y(k))]\boldsymbol{W}[\boldsymbol{R}_X(k) - \boldsymbol{R}_V] \\ &\quad + \frac{\partial}{\partial \boldsymbol{W}}(\Lambda F(\boldsymbol{R}_{GY})) \end{aligned} \tag{13}$$

When the noise of the system is unknown, its covariance can be estimated in the same fashion:

$$\hat{\boldsymbol{R}}_V(t+1) = \eta\hat{\boldsymbol{R}}_V(t) + (1-\eta)\Delta\hat{\boldsymbol{R}}_V(t) \tag{14}$$

where $\Delta\hat{\boldsymbol{R}}_V(t) = \hat{\boldsymbol{R}}_X(k) - \boldsymbol{W}^{-1}\hat{\boldsymbol{R}}_Y(k)(\boldsymbol{W}^T)^{-1}\hat{\boldsymbol{R}}_X(k)$ and $\hat{\boldsymbol{R}}_Y(k)$ are, respectively, moving window estimates of the observation and output covariance matrices, and $\eta \in (0, 1)$. The adaptation stops when the error falls below an acceptable level, i.e. when $\|\boldsymbol{W}(t-1) - \boldsymbol{W}(t)\|^2 \simeq 0$. In the following Section, we examine the method on a set of simulated signals, a set of EEG contaminated by eye blinking artifact and a set of EEG contaminated by electrocardiogram (ECG).

4 Experiments

In this Section, we apply the CBSS algorithm to simulated signals and real EEG data and analyse the results. The performance of the proposed algorithm is evaluated in terms of the convergence speed and the ability to remove the artifact from each of the components.

4.1 *Simulated source signals*

In the first experiment, we provided a synthetic set of signals affected by a simulated artifact. The synthetic signals, were two speech signals, of 5000 samples long and sampled at 12 kHz. The artifact was generated using a sampled sinc function, as shown in Fig. 2. Here, we assumed that the artifact signal could be easily extracted from the mixtures. The source signals and artifact were artificially mixed using an $M \times N$ matrix. $\boldsymbol{W}$ was initialised to $\boldsymbol{I}$, and the other parameters were set as follows: $P = 0.01$, $\mu_0 = 0.1$ and $\mu_V = 0.1$. As the original sources were available, the mean square error (MSE)

$$\varepsilon^2 = E\|\boldsymbol{y} - \boldsymbol{s}\|^2 = \frac{1}{N}\sum_{i=1}^{N} E\{|y_i(t) - s_i(t)|^2\} \tag{15}$$

was used to evaluate the resemblance between the estimated and the original sources.

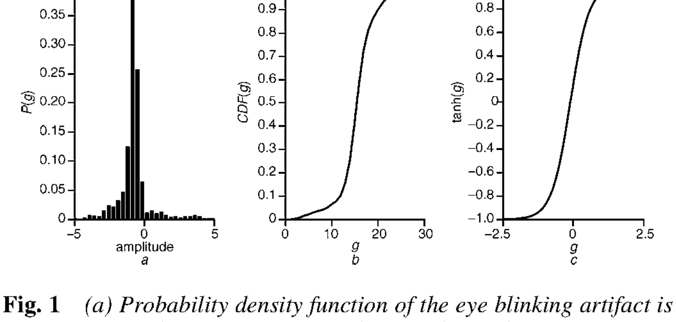

Fig. 1 *(a) Probability density function of the eye blinking artifact is seen to be super-Gaussian. (b) Cumulative density function of the eye blinking artifact closely matches (c) tanh (g) function*

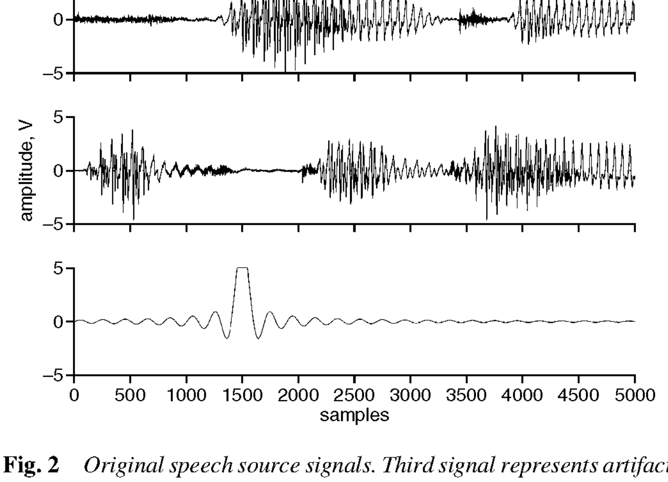

Fig. 2 *Original speech source signals. Third signal represents artifact*

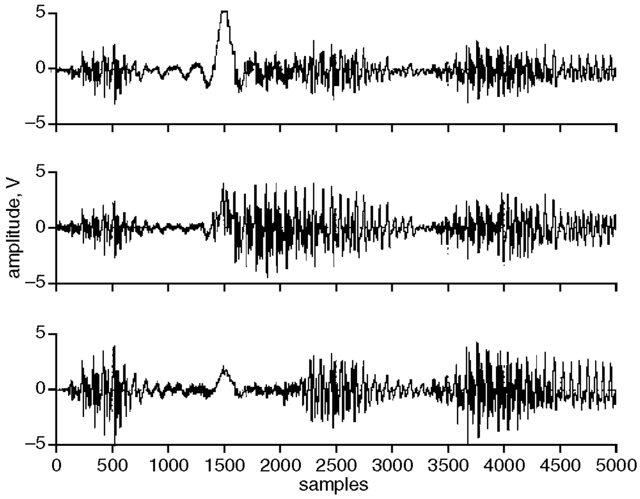

Fig. 3 *Artificially mixed signals*

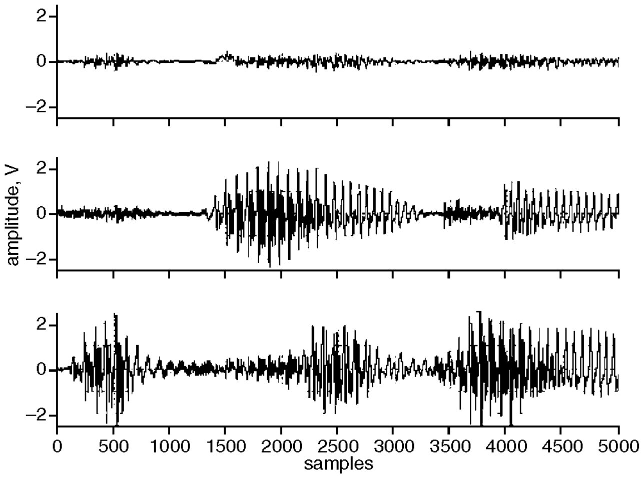

Fig. 4 *Estimated sources with artifact minimized*

The performance of the algorithm was measured by finding the waveform similarity, in decibels, defined by $\varepsilon_{dB}^2 = 10\log_{10}(1-\varepsilon^2)$. We assumed that the signals were zero mean and unit variance. The mixed signals and the estimated sources are shown, respectively, in Figs 3 and 4. By inspection of the estimated sources, it is possible to see that the artifact has been removed from the signals of interest.

The algorithm was tested using ten data sets of synthetic signals mixed with the same mixing matrix. We compared the waveform similarity for Parra's algorithm ($\Lambda = 0$) and the proposed algorithm ($\Lambda \simeq Pdiag(\boldsymbol{R}_{GY})$) for each data set. The waveform similarities for the proposed algorithm and Parra's algorithm were $\varepsilon_{dB}^2 = -0.27$ dB (SD = 0.02 dB) and -0.53 dB (SD = 0.01 dB), respectively. The performance of the algorithm was further examined by comparing the cross-correlation between the estimated sources and the artifact. Table 1 shows the performance improvement over Parra's algorithm. The goal of the algorithm was to minimise the effect of the artifact by minimising the cross-correlation between the estimated sources and the artifact. The artifact component may not be completely eliminated, as the number of iterations in 11 is finite, which means that $\boldsymbol{W}$ approaches $\boldsymbol{W}_{opt}$ as $t \to \infty$. The convergence performance, shown in Fig. 5, was comparable with that of PARRA and SPENCE (2000).

4.2 *Removing the effect of eye blinking from real EEG data*

The CBSS algorithm was further examined using real EEG data. The signals were obtained from the biomedical laboratory in King's College, London, using an EEG amplifier*, and are available from the author. EEG was collected from 16 electrodes placed on the scalp at locations defined by the conventional 10–20 electrode system. The earlobe was used as a common reference for all the channels. The ocular artifact reference signal was obtained from electrodes placed above and below the left or right eye. The data were sampled at 200 Hz and were digitally low-pass filtered with a cutoff frequency of 40 Hz.

We presented ten data sets of 10 s in length containing eye blinking artifacts. We used ten datasets because we found that this was the minimum number to provide reliable results while keeping the experimental work involved within a realistic boundary. Each of the data sets was standardised to have unit variance and zero mean. A threshold was applied to the artifact so that any details concerning other brain signals presented in the EOG would not contribute to the penalty term Λ. The artifact signal then became

$$\begin{cases} g(t) & \text{if } g(t) \geq \tau \forall t \\ 0 & \text{if } g(t) < \tau \forall t \end{cases} \tag{16}$$

The parameter τ was empirically found to be 0.2 for normalised signals. The performance was evaluated by finding the cross-correlation between the artifact and each of the mixtures and comparing it with the cross-correlation between the artifact and the estimated sources. Datasets of EEG sensor data and the artifact reference are shown in Figs 6 and 7, respectively. The resulting separated sources are shown in Fig. 8.

Table 1 Performance of CBSS algorithm is based on measurement of cross-correlation: average cross-correlation between artifact and mixtures is compared with cross-correlation between artifact and estimated sources. In this experiment, artifact is sampled sinc signal. Results are also compared with Parra's algorithm

Average correlation between synthetic artifact and estimated sources by		
Mixtures	Parra	CBSS
0.82 (SD = 0.3)	0.19 (SD = 0.01)	0.09 (SD = 0.01)

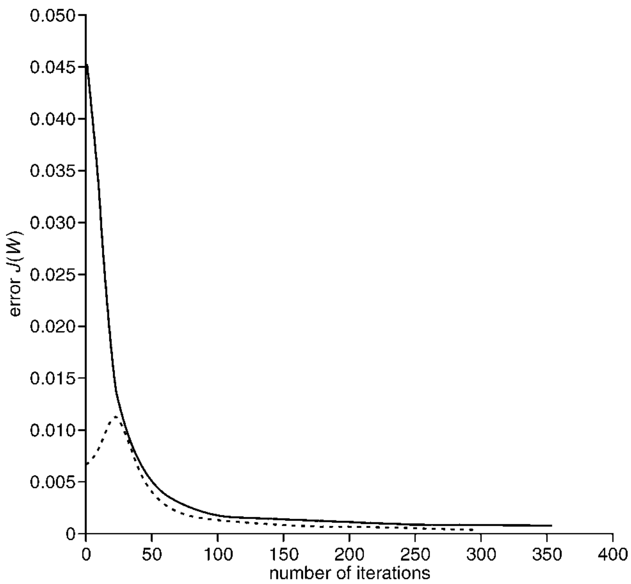

Fig. 5 *Convergence performance of (—) proposed algorithm with that of (···) Parra's algorithm*

*Cadwell Easy II

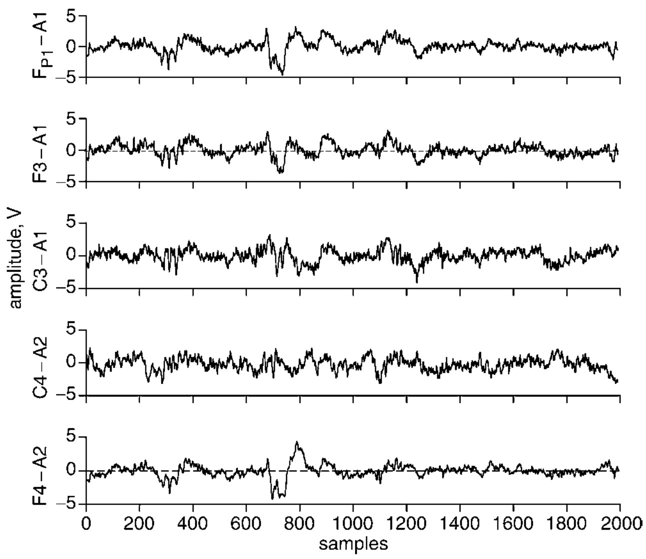

Fig. 6 *Selection of five EEG channels from a 16 channel EEG recording. EEGs on these channels are corrupted by ocular artifact between samples 600 and 900*

From Table 2, it is possible to see that, through application of the constrained algorithm, the cross-correlation between the estimated sources and the artifact has been considerably reduced. The penalty term Λ is adjusted in proportion to the cross-correlation between the artifact and the estimated sources, i.e. $E\{\boldsymbol{g}(t)\boldsymbol{y}(t)_i^T\}$. Therefore the higher the cross-correlation between the estimated source and artifact, the harsher the penalty on that component.

4.3 *Removing the effect of ECG from real EEG data*

The system was also tested on EEG signals contaminated with ECG, and the performance was reported. The ECG was

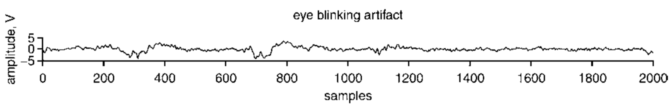

Fig. 7 *Vertical EOG signal measured from right eye*

Table 2 Performance of CBSS algorithm is based on measurement of cross-correlation between EEG and EOG artifact

Average correlation between artifact and		
mixtures	sources estimated by Parra	sources estimated by CBSS
0.75 (SD = 0.02)	0.23 (SD = 0.02)	0.16 (SD = 0.01)

Table 3 Performance of system based on cross-correlation between EEG and ECG

Average correlation between the artifact and		
mixtures	sources estimated by Parra	sources estimated by CBSS
0.76 (SD = 0.23)	0.21 (SD = 0.01)	0.17 (SD = 0.02)

measured using Eindhoven's triangle for the electrode configuration (WAGNER and MARRIOTT, 2001). The ECG data were acquired by the amplifier and sampled at 200 Hz. In this experiment, appropriate values for $\mu_0 = 0.01$, $\mu_V = 0.1$ and $P = 0.01$ were found empirically. The performance of the system in terms of the cross-correlation between the artifact and the estimated output is illustrated in Table 3. An 8 s segment of contaminated EEG is shown in Fig. 9. The EEG after removal of the artifact is shown in Fig. 10, and the reference signal is shown in Fig. 11.

Based on our trials for 20 sets of EEGs, we found that the average correlation for our proposed CBSS algorithm was 0.16, with standard deviation 0.01. As we do not know the distribution of the estimator but we do know the variance (0.01^2), we appeal to Chebychev's inequality $Prob\{|0.16 - R| < \epsilon\} \geq 1 - 0.01^2/\epsilon^2$, where R is the true value of the correlation. We can be 90% sure that we are within ϵ of the true value R.

From Table 3, it is possible to see that the CBSS algorithm has successfully separated the mixtures, and its decorrelation performance in the undetermined case of EEG is, on average, better than that of Parra's algorithm (PARRA and SPENCE

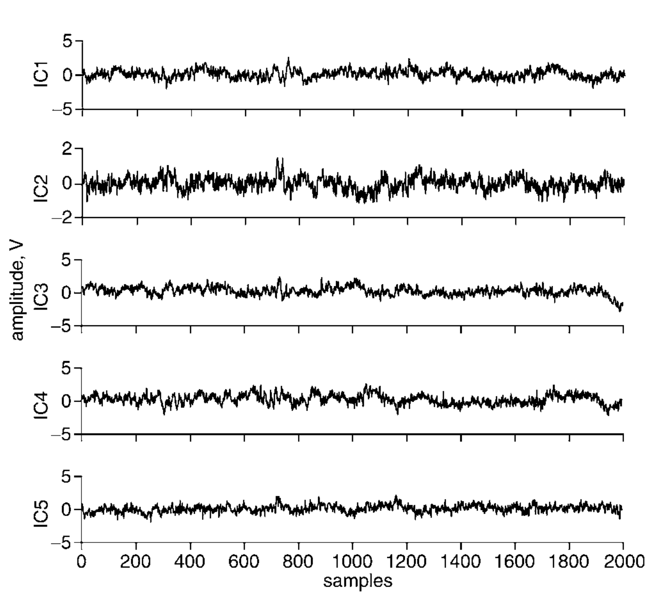

Fig. 8 *Selection of five independent components (ICs) derived from EEG corrupted by ocular artifact. The ICs represent EEG with EOG artifact removed*

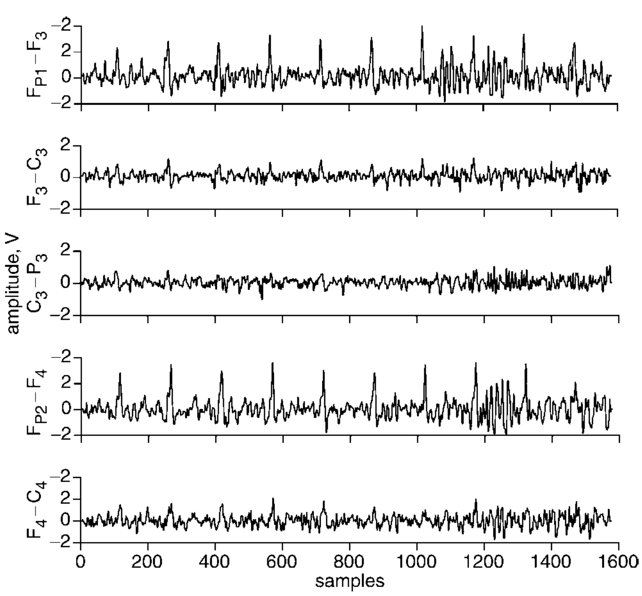

Fig. 9 *Selection of five channels from EEG recording. There is obvious ECG artifact present in upper and lower two channels*

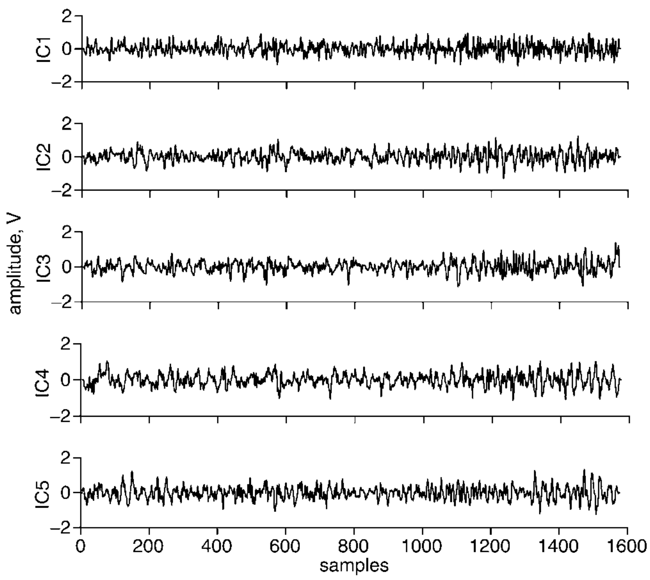

Fig. 10 *Selection of five independent components after CBSS algorithm has removed ECG*

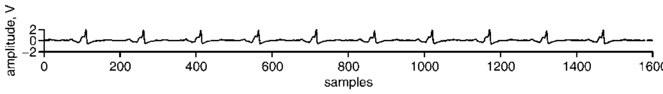

Fig. 11 *Measured ECG reference signal*

2000). The extent to which the artifact has been removed can also be verified by visual inspection of the output (Fig. 10).

5 Conclusions

A constrained BSS system for removing the eye blinking artifact has been developed by the introduction of non-linear penalty functions. The penalty terms incorporate the constraints into the main objective function, thereby converting the constrained problem into an unconstrained problem. The effect of the undesired (interfering) signal is highly reduced, and the desired components are extracted. The quality of the separated signals has been improved. The convergence performance is comparable with that of PARRA and SPENCE (2000).

The result of the algorithm can be extended to the removal of other interferences, such as electrocardiograms (ECGs) and electroglottograms (EGGs) from EEGs. As for the case of on-line EEG processing, the permutation ambiguities must be resolved using algorithms such as in SMARAGDIS *et al.* (1997).

References

BELL, A. J. and SEJNOWSKI, T. J. (1995): 'An information-maximization approach to blind separation and blind deconvolution', *Neural Comput.*, **7**, pp. 1129–1159

BEWRG, P., and SCHERG, M. (1994): 'Dipole modeling of eye activity and its application to the removal of eye artifacts from EEG and MEG', *Clin. Physiol. Physiol. Meas.*, **12**, pp. 49–54

CELKA, P., BOASHASH, B., and COLDITZ, P. (2001): 'Preprocessing and time-frequency analysis of newborn EEG seizures', *IEEE Trans. Neural Netw.*, **20**, pp. 30–39

CICHOCKI, A., and AMARI, S. (2002): 'Adaptive blind signal and image processing: learning algorithms and applications' (John Wiley & Sons Ltd, 2002)

ELBERT, T., LUTZENBERGER, W., ROCKSTROH, B., and BIRBAUMER, N. (1985): 'Removal of occular artifacts from the EEG—a biophysical approach to the EOG', *Electroencephalogr. Clin. Neurophysiol.*, **60**, pp. 455–463

GRATTON, G. (1998): 'Dealing with artifacts: The EOG contamination of the event-related brain potential', *Behav. Res. Methods, Instrum. Comput.*, **30**, pp. 40–41

HAYKIN, S. (2002): 'Adaptive filter theory, 4th edn' (Prentice-Hall, 2002)

HE, P., WILSON, G., and RUSSELL, C. (2004): 'Removal of ocular artifacts from electro-encephalogram by adaptive filtering', *Med. Biol. Eng. Comput.*, **42**, pp. 407–412

HYVARINEN, A., OJA, E., and KARHUNEN, J. (2001): 'Independent component analysis (adaptive & learning systems for signal processing, communications & control S.)' (Thomson Learning, Inc, 2001)

JOYCE, C.A., GORODNITSKY, I., and KAUTAS, M. (2004): 'Automatic removal of eye movement and blink artifacts from EEG data using blind component separation', *Psychophysiology*, **41**, pp. 313–325

JUNG, T., MAKEIG, S., MCKEOWN, M., BELL, A., LEE, T., and SEJNOWSKI, T. J. (2001): 'Imaging brain dynamics using independent component analysis', *IEEE Procses.*, **89**, pp. 1107–1122

LU, W., and RAJAPAKSE, J. (2001): 'ICA with reference'. Proc. 3rd Int. Conf. Independent Component Analysis & Blind Signal Separation, ICA2001, pp. 120–125

OVERTON, D. A., and SHAGASS, C. (1969): 'Distribution of eye movement and eyeblink potentials over the scalp', *Electroenceph. Clin. Neurophysiol.*, **27**, p. 546

PARRA, L., and SPENCE, C. (2000): 'Convolutive blind separation of non-stationary sources', *IEEE Trans. Speech Audio Process.*, **8**, pp. 320–327

SCHLOGL, A., and PFURTSHELLER, G. (1999): 'EOG and ECG minimization based on regression analysis'. Technical Report, Institute for Biomedical Engineering

SMARAGDIS, P. (1997): 'Information theoretic approaches to source separation'. Masters thesis, MIT Media Lab

WANG, W., SANEI, S., and CHAMBERS, J. A. (2005): 'Penalty function based joint diagonalization approach for convolutive blind separation of nonstationary sources', *IEEE Trans. Signal Process*, (to be published).

WAGNER, G. S., and MARRIOTT, H. J. L. (2001): 'Marriott's practical electrocardiography', 10th edn (Lippincott Williams & Wilkins, 2001)

Authors' biographies

LEOR SHOKER is a PhD researcher within the Centre of Digital Signal Processing (CDSP), Cardiff School of Engineering, Cardiff University. His current research focus is brain computer interfacing (BCI) through the exploitation of blind source separation applied to EEG signals. He also has interest in classification using support vector machines and their general application to biomedical signal analysis.

Dr SAEID SANEI is a Senior Lecturer within the CDSP and gained a PhD in Biomedical Signal and Image Processing in 1991 from Imperial College of Science, Technology and Medicine, (ICST), London.

Dr WENWU WANG is a postdoctoral researcher within the CDSP and gained his PhD in Control Systems in 2002 from Harbin Engineering University, China.

PROFESSOR JONATHON CHAMBERS is the Director of the CDSP and holds a Cardiff Professorial Research Fellowship. He gained his PhD in Adaptive Signal Processing in 1990 from ICST.