CASCADED APPROACH FOR MICROSLEEP DATA EXTRACTION

Wai Yie Leong and Danilo P. Mandic

Department of Electronics and Electrical Engineering
Imperial College London, SW7 2AZ, UK
waiyie@ieee.org, d.mandic@imperial.ac.uk

ABSTRACT

The noisy component extraction (NoiCE) algorithm is proposed to blind-extract noisy signals. This is achieved based on a combination of blind extraction structure and a cascaded nonlinear adaptive estimation. Although we use the concept of sequential blind extraction of sources and independent component analysis (ICA), we do not assume that sources are statistically independent. In fact, we show that the proposed cascaded nonlinear filter can be used to extract a signal (a single signal each time) from their noisy mixtures. Computer simulations confirm the validity and performance of the proposed algorithm in noisy microsleep events.

Index Terms— Blind source separation, blind source extraction, adaptive cascaded nonlinear estimation, noisy mixtures

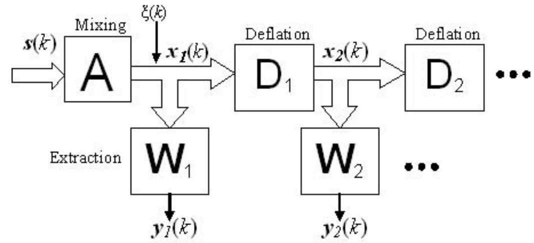

Fig. 1. A general structure of the blind source extraction (BSE).

1. INTRODUCTION

Blind Signal Extraction (BSE) [9, 12, 7] is a technique which aims at extracting *source signals sequentially* from their mixtures. This is achieved without the knowledge of the mixing process and the sources themselves. The importance of using BSE over blind source separation (BSS) becomes clearer in large scale problem (for example, 122 sensors in magnetoencephalographic (MEG) experiments). However, the main concern about BSE is: the extracted signal (from the mixtures) would consist of noise. In noisy backgrounds, standard BSS may not be feasible or could take prohibitively long time. Thus, it may be more convenient to extract only one or a subset of signals of interest (with some desired characteristics) rather than separate simultaneously all the sources (as in blind source separation [8, 1]). Some work on BSE has already been carried out, but most of the approaches focus on noise-free problems [9, 12, 7].

In the BSE model, there are n sources $\mathbf{s}(k) = [s_1(k), s_2(k), \ldots, s_n(k)]^T$, which are mixed via an unknown mixing system $\mathbf{A}$, with added noise; by m sensors we acquire the received mixed signals $\mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_m(k)]^T$, given by

$$\mathbf{x}(k) = \mathbf{A} \cdot \mathbf{s}(k) + \xi(k), \tag{1}$$

with $[\mathbf{A}]_{i,j} = a_{i,j},\ i = 1, \ldots, m,\ j = 1, \ldots, n,\ \xi(k)$ is the noise vector[1]. For convenience, we normally assume that the sources are zero-mean and the elements of $\xi(k)$ are white Gaussian and independent of the source signals.

[1] $\xi(k)$ is derived so that its covariance matrix $\boldsymbol{E}\{\xi(k)\xi^T(k')\} = \sigma(k-k')\boldsymbol{Q}(k)$, $\boldsymbol{Q}(k)$ is an $m \times m$ symmetric and positive definite matrix; $\sigma(k) \in \{0, 1\}$.

We thank Prof. Martin Golz and David Sommer for many helpful discussions. The microsleep source data is provided by University of Applied Sciences, Schmalkalden, Germany.

For independent sources, Liu et al. [10] proposed to remove the effect of noise by specially rearranging the cost function; this was achieved based on an estimate of the variance of the noise. Such a cost function had the same generic form as that for the noise-free case, but the method required some prior knowledge of the noise variance. This way, as the kurtosis of a Gaussian random variable is zero, the kurtosis of an extracted signal, $kt(y_1(k))$ will be the same as in the case with zero noise.

It is therefore clear that in order to make BSE applicable to realistic situations, that is the case of noisy mixtures, there is a need for further investigation into both the effects of noise and the unknown observation noise disturbance after deflation.

To that end, we propose an improvement to the existing BSE algorithms, which helps alleviate problems associated with: a) noisy extraction, b) noise caused by deflation; and provides solutions for BSE of instantaneous noisy mixtures, termed Noisy Component Extraction (NoiCE). Based on a rigorous analysis of the normalised mean square estimation error (MSPE) for a linear estimator based BSE method for noisy mixtures [10], we propose a novel higher-order statistical method based on cascaded nonlinear estimation. Unlike the existing methods for BSE of noisy mixtures, this approach does not require prior knowledge of the noise variance.

The paper is organised in the following manner. In Section 2, we introduce to motivation for the blind extraction structure and the proposed NoiCE learning. In Section 3, we present a description of the proposed algorithm used based on a one-step cascaded nonlinear estimator; the algorithmic design of the network is also summarised. In Section 4, an experimental study of the proposed learning applied to the estimation of the noisy signals is presented. The paper concludes with some final remarks in Section 5.

2. NOISY COMPONENT EXTRACTION (NOICE)

2.1. Extraction Procedure

A general structure of the BSE process for extracting one single source at a time is shown in Fig. 1; there are two principal stages - extraction and deflation [7]. The original mixtures first undergo the extraction stage to have one source recovered; after deflation, the effects of the extracted source are removed from the mixtures. These new "deflated" mixtures then undergo the next extraction process to recover the second source. This process may be repeated to extract the original source signals one by one, until the last source of interest is recovered. To prevent the newly extracted source signal from being extracted again in the next processing unit, we employ cascaded nonlinear estimator of the mixtures that uses information about this signal. We then derive a learning rule that deflates the extracted signal from the mixtures. The approach proposed here applies to both blind extraction and deflation in noisy environments.

2.2. NoiCE with a Nonlinear Estimator in Noisy Environments

Following the practice from radar and laser research, one convenient way to deal with the noise would be to employ a source of nonlinearity within the system. We aim to achieve this by using a nonlinear estimator [4] within the NoiCE structure as shown in Fig. 2, where the single extracted signal $y_1(k) = \mathbf{w}_1^T \cdot \mathbf{x}_1(k)$ is passed through a nonlinear estimator. A standard extraction process with extracting vector $\mathbf{w}_1(k)$ is used in the first step to extract one signal (denoted by $y_1(k)$) from the mixture. To initially extract one of the sources from the noisy environment, the input-output relation of the network are given by:

$$y_1(k) = \mathbf{w}_1^T \mathbf{x}_1(k) = \sum_{j=1}^{n} w_{1j} x_{1j}(k) = \mathbf{g}_1^T \mathbf{s}_1(k) + \mathbf{w}_1^T \xi_1(k) \qquad (2)$$

by $\mathbf{x}_i(k)$, we will denote the deflated mixtures, after the ith deflation, where $\mathbf{x}_1(k) = \mathbf{x}(k)$.

In the next step, a nonlinear filter with nonlinearity Φ typically a nonlinear function[2] $\Phi(\cdot)$, is used to assist the extraction, The use of this nonlinearity is particularly important to support the extraction process by eliminating the effects of the remaining noise [4, 1]. The output of this block can be expressed as

$$y_1(k) = \Phi[y_1(k)] \qquad (3)$$

where $\mathbf{g}_1 = \mathbf{w}_1^T \cdot \mathbf{A}$. In (3), the output $y_1(k)$ is an estimate of the extracted signal $y_1(k)$, e.g., $y_1 = y_1 + sgn(y_1)y_1^2$ or $y_1 = tanh(\gamma y_1)$. Notice that in modern implementations the hard-limiter function (4) is usually replaced by a smooth nonlinear function such as the sigmoid function (5), where the positive scalar γ is used to modify the shape (slope) of $\Phi(\cdot)$. This way, the hard limiter

$$y_1 = \begin{cases} y_1, & if \;\; y_1 \geq 0; \\ 0, & otherwise. \end{cases} \qquad (4)$$

is approximated by a hyperbolic tangent function, with a varying slope γ, given by

$$\Phi[y_1] = sgn(y_1) \approx tanh(\gamma y_1) = \frac{e^{\gamma y_1} - e^{-\gamma y_1}}{e^{\gamma y_1} + e^{-\gamma y_1}}. \qquad (5)$$

[2] A sigmoid vector function is a vector valued function Φ: $R^n \rightarrow (0,1)^n$ with $\Phi(y_1, y_2, \ldots, y_n) = (\Phi(y_1), \Phi(y_2), \ldots, \Phi(y_n))$ where Φ is a univariate sigmoid function.

Clearly, from (4), by changing γ, the nonlinearity can be varied between a linear device and a hard limiter. The effects of $\gamma \rightarrow 0$ can be studied by scaling y_1 by a constant.

$$\lim_{\gamma^2 \rightarrow \infty} [y_1] = y_1 \;\; and \;\; \lim_{\gamma^2 \rightarrow 0} [y_1] = \gamma \sqrt{\pi/2} sign(y_1). \qquad (6)$$

3. CASCADED NONLINEAR ESTIMATION FOR BLIND EXTRACTION

Previous results [7] had shown that blind extraction would suffer from the effects of noise, including both the noise from the environment and noise after deflation. Therefore, in the development of the adaptive BSE algorithm; unlike the existing approaches based on the noise removal directly from the cost function, we need to consider both the effects of noise and the artifacts of extraction and deflation in a most realistic. To overcome this issue, cascaded nonlinear estimators are proposed in this work. Following the theoretical justification from [3, 11], we set out to investigate whether the nonlinear estimation within the BSE structure from Fig. 2 offers advantages in BSE. The commonly used linear methods which obey the superposition principle suffer from serious degradation upon the arrival of samples corrupted with high-amplitude noise. Nonlinear methods, on the other hand, promise to better exploit the statistical characteristics of the noise.

3.1. The Derivation of the Learning Algorithm

Let us therefore focus on the cascaded nonlinear structure from Figure 2. The update of the weight vector $\mathbf{w}_i$ associated with standard BSE (shown in equation (2)) can be performed based on the kurtosis minimization as decreased in [10].

Following the approach from [3, 11], and using (5), together with its derivative $\frac{\partial}{\partial y_i} tanh(y_i) = sech^2(y_i) = \frac{2}{e^{y_i} + e^{-y_i}}$, we can now define a cascaded signed observation $\tilde{y_i}^a$ as

$$\tilde{y}_i^a = sgn[sgn(\tilde{\mathbf{w}}_i)\tilde{y}_i - a] \qquad (7)$$

where a is a real valued scalar ($-\infty < a < \infty$) and the vector $\tilde{\mathbf{w}}_i$=$\mathbf{w}_i$. Note that for $a \leq \tilde{y}_i$ the output samples have values equal to unity, and for $a > \tilde{y}_i$, these values are equal to -1. Thus, from (7)

$$\tilde{y}_i^a = \begin{cases} 1, & for \; a \leq \tilde{y}_i \\ -1, & for \; a > \tilde{y}_i \end{cases} \qquad (8)$$

The "running weights" [3] estimate of the desired output signal, $\tilde{y_1}^a$, is achieved via the following operation

$$\begin{aligned} D &= |\tilde{\mathbf{w}}_i| \diamond sgn(\tilde{\mathbf{w}}_i)\tilde{y}_i|_{i=1}^n, \\ &= |\tilde{\mathbf{w}}_i| \diamond \frac{1}{2} \int sgn[sgn(\tilde{\mathbf{w}}_i)\tilde{y}_i - a] da|_{i=1}^n \end{aligned} \qquad (9)$$

where "$\diamond$" is the replication operator.

Using this as the principal contributor to the update, and since $\mathcal{S}_{(k)}$ is the output of the weighted median at time $k(S_{(k)} = y^a(k))$, the simplified algorithm leading to the following recursion referred to as the "fact weight adaptive algorithm" [3] is given by:

$$w_{1i}(k+1) = w_{1i}(k) + \mu sgn(w_{1i}(k)) \times sgn[sgn(w_{1i}(k)) y_{1i}(k) - \mathcal{S}(k)], \qquad (10)$$

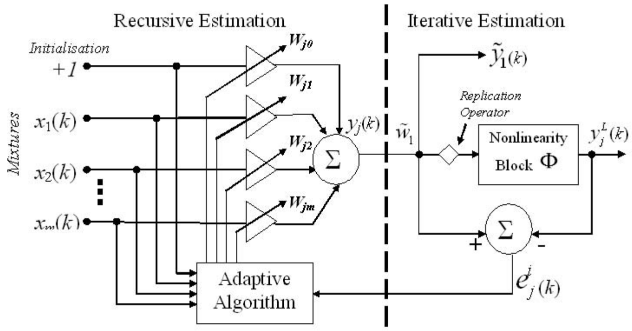

Fig. 2. Structure for noisy component extraction (NoiCE).

for i = 1, 2, ..., n. This algorithm will be applied in the deflation procedure.

After the deflation process, the remaining mixtures become

$$\hat{\mathbf{x}}_1(k) = \mathbf{x}_1(k) - \sum_{i=1}^{n} w_{1i} y_{1i}(k) \, , \tag{11}$$

This completes the derivation of the proposed BSE algorithm for extracting noisy signals (NoiCE).

4. SIMULATIONS

4.1. Experiment I

To verify the performance of the proposed NoiCE learning, a 3×3 mixing matrix **A** was randomly generated and is given by

$$\mathbf{A} = \begin{bmatrix} 0.8521 & 0.4110 & -0.9223 \\ 0.3442 & -0.9947 & -0.4895 \\ -0.1601 & -0.5357 & -0.1420 \end{bmatrix} . \tag{12}$$

Fig. 3(a) shows the three source signals, denoted by s_1 with binary distribution, s_2 sine waveform and s_3 Gaussian distribution, used in simulations. These signals have positive kurtosis ($\beta = 1$). The proposed cascaded nonlinear estimator is adopted in this experiment. Monte Carlo simulations with 500 iterations of independent trials were performed. This way, the initial estimation errors of the three signals were respectively $\{10.0735, 4.0554, 10.0133\}$, the variance of the noise in (1) was set to $\sigma^2 = 0.1$. By applying the proposed NoiCE algorithm, we expect the signals with the smallest estimation error to be first extracted, which is the binary distribution (s_1) and the sine waveform (s_3). The stepsize μ_0 = 0.0007. After the extraction process, the estimation errors for the proposed NoiCE algorithm become $\{-0.0522, 0.0104, -0.0040\}$.

The waveforms of the sequentially extracted signals by the proposed nonlinear estimator method is given in Fig.3(b). Following our analysis (which based on the smallest estimation error), the proposed NoiCE learning first extracted s_1 with binary distribution, followed by s_2 a sine waveform and then s_3 with Gaussian distribution. These three extracted signals matched closely the original source signals.

To further illustrate the qualitative performance of the proposed approach, scatter plots of the original sources and the recovered output signals are displayed in Fig.3(c). These scatter plots illustrate the

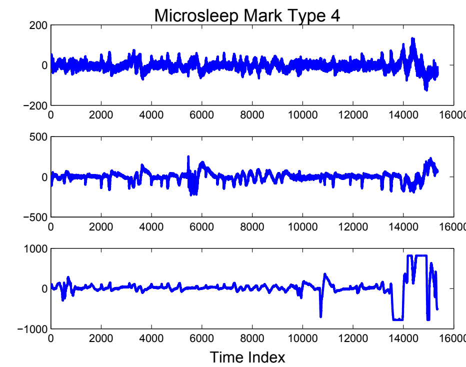

Fig. 4. Mark type 4 microsleep data: (top) FP1 EEG signal, (middle) FP2 EEG signal and (bottom) EOG signal.

Extracted Signals			
Kurtosis	Signal 1	Signal 2	Signal 3
Original Signal	1.0232	1.500	2.9319
Noisy Mixture	2.6887	2.2643	2.1578
MPSE Linear estimator [10]	2.4183	1.2329	1.9569
Proposed NoiCE	1.0844	1.5919	2.6858

Table 1. Kurtosis of the original sources and the kurtosis of the extracted signals using the proposed NoiCE and the normalised MPSE linear estimator method [10].

”qualitative” performance and show the degree of independence between the outputs, where each point on the diagram corresponds to one data vector. If, instead of the proposed nonlinear estimator, the standard normalised MSPE [10] approach was used, it was unable to give satisfactory extraction performance, as shown in Fig.3(c). In addition, it can be seen that the proposed blind extraction algorithm provides, in general, better kurtosis matching of the source and output signals (Table I). Conforming with the above results, the extracted output signals using the proposed method have a closer match to the original sources, as compared to the normalised MSPE [10].

4.2. Experiment II

In this experiment we considered EEG data recorded during shirt lapses of awareness (microsleep) [2, 6, 5]. The task was to extract the Electro-Oculogram artifact from the useful EEG data. Figure 4 shows a 3s portion of the recorded EEG time series collected from the electrodes: (top) FP1, (middle) FP2, and (bottom) EOG; all referred to the left mastoid at mark type 4 (high degree microsleep). From Fig.4, these records show a large amount of correlation, hence bandpass filtering could not have been used to separate them. The proposed blind extraction, on the other hand, separates and extracts the EOG artifact from all the EEG components. From Fig. 5, artifact and nonlinearity of the high microsleep signal was isolated as compared with the ”clean” mark type 0 (tired but awake) data. The ”corrected” data showed a low level noise and best matched the mark type 0 (tired but awake) data.

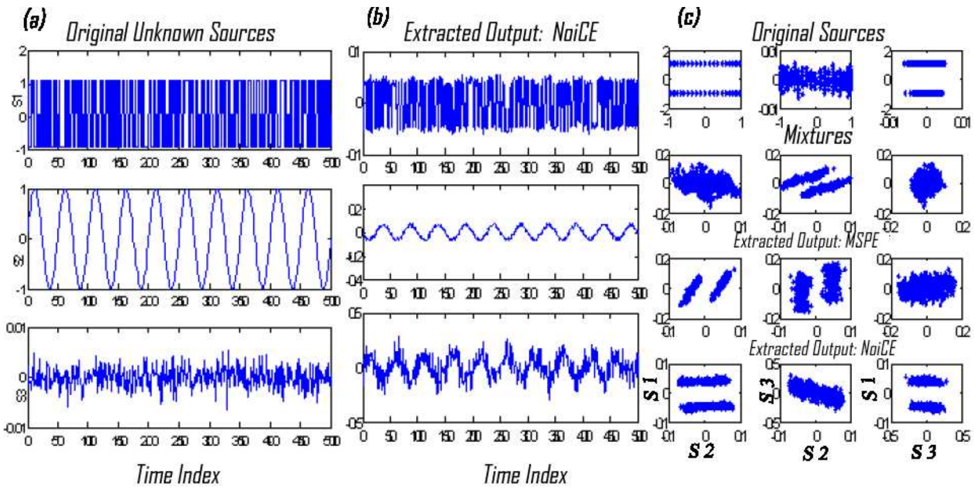

Fig. 3. Source signals used in simulations: (a) The original source signals, s_1 with binary distribution, s_2 sine waveform and s_3 with Gaussian distribution; (b) The extracted output signals based on NoiCE, s_1 with binary distribution, s_2 a sine waveform and s_3 with Gaussian distribution. (c) Scatter plots comparing the independence level of the extracted signals.

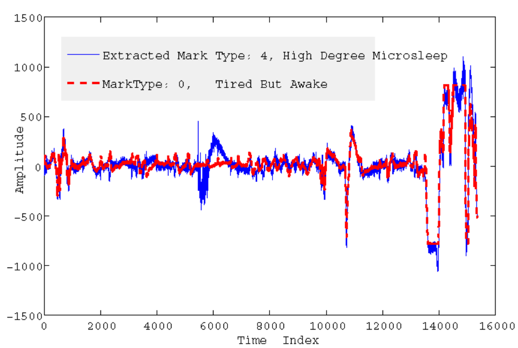

Fig. 5. A comparison between —- Extracted signal (Mark type = 4); - - - Tired but awake signal (Mark type = 0) after BSE.

5. CONCLUSIONS

We have addressed a special class of blind source separation (BSS) algorithms, namely noisy component extraction (NoiCE), by which we can recover a single source or a subset of sources each time, instead of recovering all of the sources simultaneously. We have discussed the neural network model and its associated adaptive learning rule, and have developed a BSE algorithm for noisy mixtures. We have studied the BSE problem in noisy environments and proposed a new NoiCE algorithm based on minimisation of cascaded nonlinear estimation error. Unlike the existing algorithms for noisy BSE, which remove the effects of noisy directly from the cost function, this approach does not require the knowledge of noise variance, or any preprocessing. Simulations have shown that the proposed algorithm can perform satisfactory extraction of the corresponding sources from noisy mixtures.

6. REFERENCES

[1] A.Cichocki and S.I.Amari. *Adaptive Blind Signal and Image Processing*. Locally Adaptive Algorithms for ICA and Their Implementations. John Wiley & Sons, Ltd, England, 2002.

[2] B.Priest, C.Brichard, G.Aubert, G.Liistro, and D.O.Rodenstein. Microsleep during a simplified maintenance of wakefulness test. *American Journal of Respiratory and Critical Care Medicine*, 163:1619–1625, 2001.

[3] G.R.Arce. *Nonlinear Signal Processing: A statistical Approach*. John Wiley & Sons, Inc, New Jersey, 1st edition, 2005.

[4] I.Pitas and A.N.Venetsanopoulos. *Nonlinear Digital Filters Principles and Applications*. Kluwer Academic Publishers, Boston, 1990.

[5] M.Golz and D.Sommer. Classification of eyetracking signals with vector-based neural networks. In *Proc Europ Symp Intell Technol Hybrid Syst Implem Smart Adaptive Syst (eunite 2002)*, pages 130–134, Albufeira, Portugal, 2002.

[6] M.Golz, D.Sommer, and D.Mandic. Microsleep detection in electrophysiological signals. In *Proc 1st Int Workshop Biosignal Processing and Classi-fication (BPC 2005)*, pages 102–109, Portugal, 2005.

[7] N.Delfosse and P.Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 49:59–83, 1995.

[8] P.Comon. Independent Component Analysis, a new concept? *Special Issue on Higher-Order Statistics, Signal Processing*, 36(3):287–314, April 1994.

[9] S.A.Cruces-Alvarez, A.Cichoki, and S.I.Amari. On a new blind signal extraction algorithm: Different criteria and stability analysis. *IEEE Signal Processing Letters*, 9:233–236, August 2002.

[10] W.Liu, D.P.Mandic, and A.Cichocki. Blind source extraction of instantaneous noisy mixtures using a linear predictor. In *The 2006 IEEE International Symposium on Circuits and Systems (ISCAS2006)*, pages 4199–4202, Kos, Greece, May 2006.

[11] Y.Lin, R.K.Yang, G.Moncef, and N.Yrjo. Weighted median filters: A tutorial. *IEEE Transactions on Circuits and Systems II- Analog and Digital Signal Processing*, 43:157–192, March 1996.

[12] Y.Q.Li and J.Wang. Sequential blind extraction of instananeously mixed sources. *IEEE Transctions on Signal Processing*, 50:997–1006, May 2002.