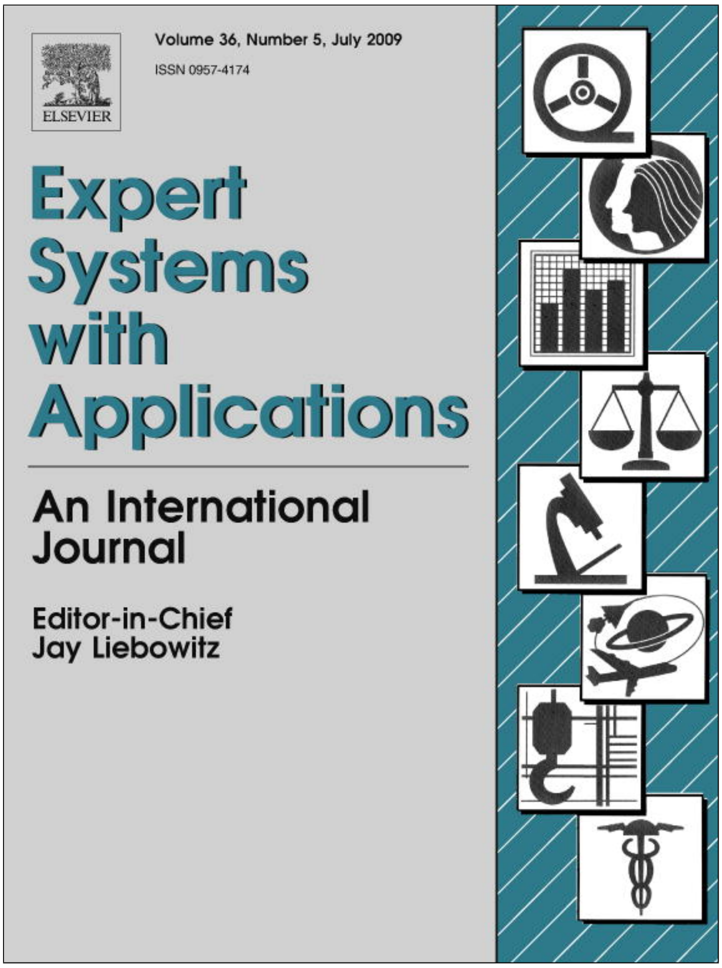
Volume 36, Number 5, July 2009
ISSN 0957-4174
ELSEVIER
Expert Systems with Applications
An International Journal
Editor-in-Chief
Jay Liebowitz

Expert Systems with Applications 36 (2009) 8855–8861

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Blind source separation with dynamic source number using adaptive neural algorithm

Tsung-Ying Sun [a,*], Chan-Cheng Liu [a], Shang-Jeng Tsai [a], Sheng-Ta Hsieh [b]

[a] Department of Electrical Engineering, National Dong Hwa University, 97401 Hualien, Taiwan, ROC
[b] Department of Electrical Engineering, Oriental Institute of Technology, 220 Taipei County, Taiwan, ROC

ARTICLE INFO

Keywords:
Adaptive learning rate
Adaptive neural algorithm
Blind source separation
Dynamic number of sources

ABSTRACT

A difficult blind source separation (BSS) issue dealing with an unknown and dynamic number of sources is tackled in this study. In the past, the majority of BSS algorithms familiarize themselves with situations where the numbers of sources are given, because the settings for the dimensions of the algorithm are dependent on this information. However, such an assumption could not be held in many advanced applications. Thus, this paper proposes the adaptive neural algorithm (ANA) which designs and associates several auto-adjust mechanisms to challenge these advanced BSS problems. The first implementation is the on-line estimator of source numbers improved from the cross-validation technique. The second is the adaptive structure neural network that combines feed-forward architecture and the self-organized criterion. The last is the learning rate adjustment in order to enhance efficiency of learning. The validity and performance of the proposed algorithm are demonstrated by computer simulations, and are compared to algorithms with state of the art. From the simulation results, these have been confirmed that the proposed ANA performed better separation than others in static BSS cases and is feasible for dynamic BSS cases.

1. Introduction

In general, the goal of blind source separation (BSS) is to distinguish source signals from mixed signals received by sensors. The adjective "blind" emphasizes the fact that, the source signals are not observed, and, no information is available about the mixing situation. But, the assumption is often held physically that the source signals are mutually independent. Recently, BSS in signal processing has received considerable attention from researchers, due to its numerous promising applications in the areas of biomedical signal processing, digital communications and speech signal, sonar, image processing, and monitoring (Cichocki & Unbehauen, 1996; Schiessl et al., 2000; Tonazzini, Bedini, & Salerno, 2006; Wei, Woo, & Dlay, 2007; Yilmaz & Rickard, 2004; Zhang & Kassam, 2004).

Since the pioneering work of Herault and Juten (1986), a variety of algorithms have been proposed for different BSS subjects. In general, the existing algorithms can be divided into several major categories according to criterion: independent component analysis (Cichocki, Karhunen, Kasprzak, & Vigário, 1999; Cichocki & Unbehauen, 1996; Herault & Juten, 1986; Liu, Sun, Lin, & Chou, 2006; Schiessl et al., 2000; Tonazzini et al., 2006; Wei et al., 2007; Ye, Zhu, & Zhang, 2004; Yilmaz & Rickard, 2004; Zhang & Kassam, 2004), density model-based algorithms (Amari, Chen, & Cichocki, 1997; Comon, 1994; Lee, Girolami, & Sejnowski, 1999; Vigliano & Uncini, 2003), algebraic algorithms (Belouchrani, Abed-Meraim, Cardoso, & Moulines, 1997; Even & Moisan, 2005; Li & Wang, 2002; Lou & Zhang, 2003), information-theoretic algorithms (Pajunen, 1998; Pham, 2002; Pham & Vrins, 2005) and time-frequency (or sparseness) based algorithms (Araki, Sawada, Mukai, & Makino, 2007; Lee, Lewicki, Girolami, & Sejnowski, 1999; Liu, Sun, Li, Hsieh, & Tsai, 2007; Lv & Zhang, 2006; Yilmaz & Rickard, 2004).

A common assumption, the numbers of sources are known a priori, has been applied by most of BSS algorithms in the past. However, this assumption could not be ensured in some advanced applications. The mechanics identifying the number of sources was developed upon time-frequency based algorithms, such as the one discussed in Liu et al. (2007). However, the method could not perform an instant identification because of its inherent shorts from huge computation. Neural network (NN) based algorithms are often utilized in operations of on-line separation, because of its ability to produce output signals almost instantly with its simple operations of gradient iteration. For taking into account the nonstationary environment, an adaptive algorithm is necessary when dealing with BSS problems with dynamic source numbers.

In the aspect of NN based algorithm, in order to have an exploitable dimension of neural network, the numbers of sources n is assumed a priori. Typically it should be equal to the number of

* Corresponding author. Tel.: +886 0 3863 4078; fax: +886 0 3863 4060.
E-mail address: sunty@mail.ndhu.edu.tw (T.-Y. Sun).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.11.011

sensors and outputs m, termed the determined state. In this state, the natural gradient descent is derived and the property of stability has been theoretically demonstrated by Cichocki & Amari (2003). Under-determined neural algorithms, those that have fewer sensors than sources ($m < n$), have been proposed in Li and Wang (2002), Cao and Liu (1996). But when there is insufficient prior information related the mixing matrix or source, the source signals cannot be successfully recovered. The over-determined neural algorithm has more sensors than sources, i.e. $m > n$. Learning divergence arises from the high correlations of similar or redundant signals from the extra sensors. To overcome the problem, several neural network architectures, together with the associated learning algorithms, were proposed in Cichocki et al. (1999). However, when a pre-whitening layer is applied during estimating the source number to reduce the dimension of the data vectors, separation results may be poor for ill-conditioned mixing matrices or weak source signals. Similar disadvantages exist if possible data compression takes place in the separation layer instead of the pre-whitening layer. A modified natural gradient algorithm using the adaptive scheme of self-normalized matrix has been suggested recently by Ye et al. (2004). Unfortunately, its separating performance is often incommensurate since the adjusting speed of its self-normalized matrix cannot always be applied to every case.

Supervised learning algorithms suffer from a large degree of system complexity because any arbitrary system has to be learned. A large amount of information is required as input to fine tune the system in order to avoid under or overestimation. This kind of prior knowledge is not always available; this restricts the successful application of fixed structure systems. In order to find the optimal system, a NN with self-organized structure was proposed (Tenorio & Manoel, 1990).

The concept behind the proposed method is a system capable of instantly identifying the number of sources, and adjusts the size of NN accordingly to ensure optimal performance during separation. In 2006 we proposed a self-organized neural algorithm to solve the problem (Liu et al., 2006). According to its properties, the gradient descent value will increase to infinity while $m > n$, the rule of self-organization will allow the dimension of network to grow gradually at a fixed period, until infinity occurs, and then trim the dimensions one by one until the gradient value steadies again. However, there is no best value when defining a growth period, for example, the algorithm will take a long time to become over-determined when a large value is given; otherwise, it will spend too much time trimming the redundant dimensions when a small value is given. Though the ability of this algorithm to handle a static BSS problem is verified, it could not solve a problem with dynamic source numbers.

In this paper, the focus is on advanced BSS problems which involve unknown and dynamic number of sources. Considering the researchers' previous experiences in BSS study, an adaptive neural algorithm (ANA) is further proposed. This algorithm is based on the feed-forward NN proposed by Cichocki & Unbehauen (1996), and utilizes an improved cross-validation technique to instantly estimate the number of sources. The size of the NN will be updated accordingly in respect to the estimation. In order to allow greater efficiency when learning in the dynamic NN structure, an adaptive learning rate for the NN by reference to gradient information was proposed. The validity and efficiency of the proposed ANA is demonstrated by several BSS simulations. Ye's algorithm (Ye et al., 2004) is implemented to perform comparisons on the separation performance. Note that the method proposed in Cichocki et al. (1999) is not able to tackle dynamic BSS problem because its source number estimation – which requires calculation of a large number of mixed signals – cannot be performed instantaneously. This method is not implemented in this paper because it is not appropriate for the function in these simulations.

The remainder of this paper is organized as follows: Section 2 presents the BSS problem formulation, the estimator of source numbers, and the cross-validation technique. Section 3 presents the on-line estimator, the self-organized criterion, and the adjustment of learning rate. The separation performances of the ANA and an existing algorithm are contrasted in several computer simulations, shown in Section 4. Section 5 contains a brief conclusion.

2. Problem formulation

2.1. The description of general BSS

In a situation where the sources are unobservable, $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is the zero-mean vector and is mutually (spatially) statistically independent (or as independent as possible), where n denotes the number of sources and $t = 1, \ldots, N$ is the instant time of sampling. The available sensor vector $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$, where m is the number of sensors, is given by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

where $\mathbf{A} \in \mathbf{R}^{m\times n}$ is a non-singular and unobservable matrix has non-zero determinant. The assumption $m = n$ is held for convenience when explaining. The goal of BSS is to recover the waveforms of each source from individual output $\mathbf{y}(t)$ such that

$$\mathbf{y}(t) = \mathbf{DPs}(t) = \tilde{\mathbf{P}}\mathbf{s}(t) \tag{2}$$

where $\mathbf{D} \in \mathbf{R}^{n\times n}$ is a diagonal scaling matrix with non-zero entries and $\mathbf{P} \in \mathbf{R}^{n\times n}$ is any permutation matrix and $\tilde{\mathbf{P}}$ is the generalized permutation matrix with exactly one non-zero element in each row and each column (Cichocki et al., 1999). Typically, the mapping elements $\mathbf{W}(t)$ of a NN, called separating matrix is regarded as an inverse of $\mathbf{A}$. Then, the separated signals are obtained through the following:

$$\mathbf{y}(t) = \mathbf{W}(t)\mathbf{x}(t) \tag{3}$$

Generally, weights of NN are updated by the gradient decent, and the objective function is a nonlinear correlation criterion which is derived from independent component analysis (ICA) Cao & Liu, 1996. Since source signals are zero-mean and mutually independent, the generalized covariance matrix is represented as

$$\mathbf{R_{fg}} = E\{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\} - E\{\mathbf{f}(\mathbf{y})\}E\{\mathbf{g}^T(\mathbf{y})\} \tag{4}$$

where $f(y_i)$ and $g(y_i)$ are different odd nonlinear activation functions. The diagonal elements of $\mathbf{R_{fg}}$ are non-zero and all other terms are zero. The terms $E\{f(y_i)\}E\{g^T(y_i)\}$ will be zero if the probability density functions (pdf) of each source are even. The choice of activation functions, $f(y_i)$ and $g(y_i)$, is based on the statistically distribution characteristic of source signals, refer to literature (Cao & Liu, 1996) for further detail.

The first updating equation is proposed by Herault & Juten (1986), as the following:

$$w_{ij}(t) = w_{ij}(t-1) + \mu \cdot \Delta w_{ij}(t) \tag{5}$$

and

$$\Delta w_{ij}(t) = f(y_i(t))g(y_j(t)) \tag{6}$$

where $\{w_{ij}\}$ are elements of the separating matrix $\mathbf{W}$, and μ denotes a learning rate, a lower value is usually given. According to (6), Amari utilized feed-forward NN and proposed the robust updating equation for ill-condition cases, as the following:

$$\Delta w_{ij}(t) = [\lambda_{ij} - f(y_i(t))g(y_j(t))]w_{ij}(t) \tag{7}$$

where $\boldsymbol{\lambda} = \{\lambda_{ij}\}$ denotes the self-normalized matrix (typically $\boldsymbol{\lambda} = \mathbf{I}$).

2.2. Estimation of the number of sources

The cross-validation techniques have been primarily applied in multivariate statistics (Krzanowski, 1987). The basic idea of a cross-validation method is that it divides the data into several subgroups. One group is used to determine some features of the data, and the other groups are used to verify the features. Using this technique, a criterion for estimating the number of sources based on the error of estimated noise variance was proposed by Cao, Murate, Amari, Cichocki, & Takeda (2003).

Let $\mathbf{C}$ denote the covariance matrix given by $\mathbf{C} = \mathbf{xx}^T$. The estimation can be obtained as

$$\Psi = diag(\mathbf{C} - \hat{\mathbf{A}}\hat{\mathbf{A}}^T) \tag{8}$$

where $\hat{\mathbf{A}} = \mathbf{U}_{\hat{n}}\boldsymbol{\Lambda}_{\hat{n}}^{1/2}$, and $\boldsymbol{\Lambda}_{\hat{n}}$ is a diagonal matrix whose elements are the $\hat{n}$ largest eigenvalues of $\mathbf{C}$. The columns of $\mathbf{U}_{\hat{n}}$ are the corresponding eigenvectors. The data matrix $\mathbf{x} \in \mathbf{R}^{m \times N}$ is then divided into several disjoint groups, such as $\mathbf{x}_i \in \mathbf{R}^{m \times N/K}$, where the group number $i = 1,\ldots,K$. There will be a larger error between the noise variance $diag(\Psi_i)$ and its estimated $diag(\Psi_m)$, where $i \neq m$, making it obvious when the estimation of the source number $\hat{n}$ has not been matched to its true value. Based on this property, the criterion for each conjectured source number $\hat{n}$ was defined as

$$Error(\hat{n}) = \frac{1}{K}\sum_{i=1}^{K} tr[diag(\Psi_i^{(\hat{n})}) - diag(\Psi_m^{(\hat{n})})]^2 \tag{9}$$

It is necessary to compute all the possible estimates of the source number, the boundary of $1 \leqslant \hat{n} \leqslant (1/2)(2m + 1 - \sqrt{8m+1})$ is suggested in Cao et al. (2003).

There are other available methods for approximating the numbers of sources, such as principal component analysis (PCA) or singular value decomposition (SVD). According to our experiences, the error ratio for the estimations and the cost of computation are slightly less evident with cross-validation than PCA or SVD. Consequently, it is the reason why this study bases its system on cross-validation to develop an instant estimator.

3. The adaptive neural BSS algorithm

The relationship between the unobservable outside situation and our algorithm is illustrated as Fig. 1. The source number estimator informs the adaptive structure NN what size of structure is best. The learning rate adjustment evaluates a proper value according to gradient, and returns it to the updating function of the learning criterion. In this study, ambient noises from the environment are considered, but the subtle noises from transmission are ignored.

3.1. Improvement of cross-validation

According to the cross-validation mentioned in the last section, the unknown number of sources can be calculated in batch. To satisfy the requirements of on-line learning, this paper used the recursive method to calculate the mean and the covariance matrix. In the beginning, a sampling window length wd is defined. While the number of sampling mixtures is less than wd (i.e. $t \leqslant wd$), the covariance matrix is recursively calculated as following (Lou & Zhang, 2003):

$$\mathbf{M}(t) = \frac{t-1}{t}\mathbf{M}(t-1) + \frac{1}{t}\mathbf{x}(t) \tag{10}$$

$$\Delta\mathbf{x}(t) = \mathbf{M}(t) - \mathbf{M}(t-1) \tag{11}$$

$$\mathbf{C}(t) = \frac{t-1}{t}[\mathbf{C}(t-1) + \Delta\mathbf{x}(t)\Delta\mathbf{x}(t)^T] + \frac{1}{t}[\Delta\mathbf{x}(t) - \mathbf{M}(t)][\Delta\mathbf{x}(t) - \mathbf{M}(t)]^T \tag{12}$$

where $\mathbf{M}$ is the mean of $\mathbf{x}$. While $t > wd$, past wd mixtures, $\hat{\mathbf{x}} = \{\mathbf{x}(t-(wd+1)),\ldots,\mathbf{x}(t)\}$, used to generate the mean and the covariance matrix, therefore, early mixtures will not affect present estimation when the number of sources has changed.

At each iteration, the instant noise variance $\Psi^{(\hat{n})}$ is obtained from (8), where $\hat{n} = 1,\ldots,m$. Then, at the tth instant the estimated number of source can be obtained by

$$\tilde{n}(t) = \min\{\arg\min_{\hat{n}} J(\hat{n})\} \tag{13}$$

where

$$J(\hat{n}) = \sum[\Psi^{(\hat{n})} - \Psi^{(m)}]^2 \tag{14}$$

where $\Psi^{(m)}$ is the last of instant noise variances.

3.2. Self-organized criterion

In the field of neural network research, how to define a proper structure size is an important issue. NN's with a small structure size cannot perform complex mappings. A large sized NN has not only a high running cost but also unstable performance. The size of the structure is usually defined by trial and error in general applications; however, a self-organized structure is proposed for the dynamic problems such as those discussed in Tenorio and Manoel (1990), Morris and Garvin (1996), Park, Huh, Kim, Seo, and Park (2005). The method proposed in this study utilizes a fully connected feed-forward NN which includes an input layer and an output layer. The two layers contain the same number of nodes, $\tilde{n}$. Then, a pair of nodes from both layers will be generated or removed at the time $\tilde{n}$ changes.

The self-organized procedure at each iteration is to evaluate the values of $\tilde{n}(t)$ at the beginning. When $\tilde{n}(t) > \tilde{n}(t-1)$, which implies the number of source signals is increasing, then the size of the separating matrix will be enlarged by

$$w_{ij}(t) = \begin{cases} w_{ij}(t-1), & \text{as } i \leqslant \tilde{n}(t) \text{ and } j \leqslant \tilde{n}(t), \\ r, & \text{otherwise.} \end{cases} \tag{15}$$

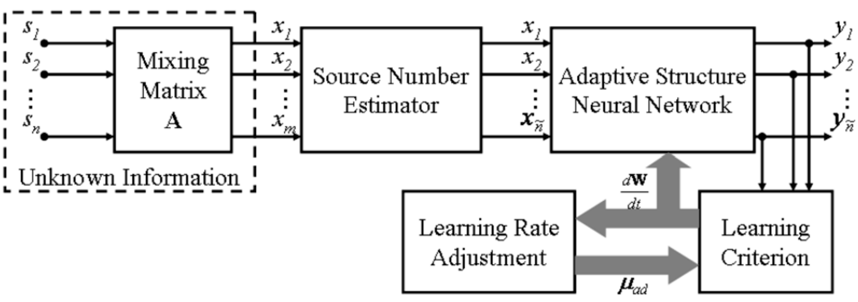

Fig. 1. The framework of the proposed adaptive neural algorithm.

where r is a random value between [−1,1], newly produced elements are added to the end of the matrix $\mathbf{W}(t-1)$. While $\tilde{n}(t) < \tilde{n}(t-1)$, which implies the number of source signals is decreasing, then the size of the separating matrix will be trimmed by

$$\mathbf{W}(t) \leftarrow \{w_{ij}(t-1) | i \leqslant \tilde{n}(t) \text{ and } j \leqslant \tilde{n}(t)\} \tag{16}$$

The redundant elements of the separating matrix are removed from the end of the matrix. If $\tilde{n}(t) = \tilde{n}(t-1)$, then the size of separating matrix can no longer be tuned, i.e. $\mathbf{W}(t) = \mathbf{W}(t-1)$.

3.3. Learning rate adjustment

Since NN's structure size is tuned according to the varying number of sources, the weights will be retrained repeatedly. The speed and stability of NN learning are affected by the size of learning rate. Consequently, an adjustment to evaluate an adaptive learning ratio between 0 and 1 has been proposed, Since the BSS is unsupervised, there is no definitive guide. We therefore evaluate the ratio with gradient information, MD_{gd} and MA_{gd}. MD_{gd} evaluates the average of variation between two neighbour gradients, as

$$MD_{gd} = \frac{1}{\tilde{n}^2(t)} \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} |\Delta w_{ij}(t) - \Delta w_{ij}(t-1)| \tag{17}$$

Then, MA_{gd} indicates the instant mean of MD_{gd}, that is calculated by

$$MA_{gd}(t) = \frac{t-1}{t} MA_{gd}(t-1) + \frac{1}{t} MD_{gd} \tag{18}$$

The adaptive ratio μ_{ad} is further generated as

$$\mu_{ad}(t) = \frac{1}{MA_{gd}(t)+1} \in (0,1] \tag{19}$$

Finally, the separating criterion (7) can be rewritten as the following:

$$\Delta w_{ij}(t) = \mu_{ad}(t) \lfloor \lambda_{ij} - f(y_i(t)) g(y_j(t)) \rfloor w_{ij}(t) \tag{20}$$

Generally, a reduced learning rate could make a suitable training state for NN. According to this idea, our design further bases on the practical learning state of NN to derive a rectification of learning rate. Consequently, it is helpful to enhance efficiency for the unsupervised BSS problem.

4. Simulations and results

The simulations include two kinds of BSS cases with static and dynamic number of sources. These cases incorporate six stationary source signals which are mutually independent and zero-mean:

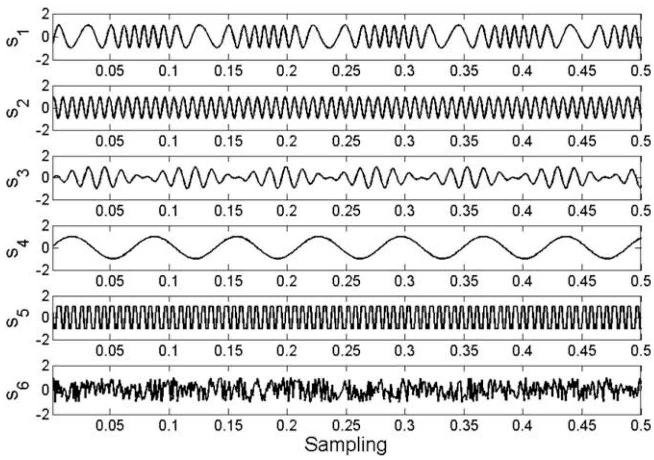

Fig. 2. The waveform of source signals.

$$\begin{cases} s_1(t) = \sin(500t + 5\cos(60t)) \\ s_2(t) = \sin(800t) \\ s_3(t) = \sin(450t)\sin(40t) \\ s_4(t) = \sin(90t) \\ s_5(t) = \text{sign}(\cos(2\pi 155t)) \\ s_6(t) = \text{Noise with uniformly distributed in } [-1,1] \end{cases} \tag{21}$$

where s_1: phase-modulated signal, s_2: high-frequency sinusoidal signal, s_3: amplitude-modulated signal, s_4: low-frequency sinusoidal signal, s_5: sign signal, s_6: noise signal (Cichocki and Unbehauen, 1996; Ye et al., 2004). The sample rates are 1 kHz. The waveforms are displayed in Fig. 2. The mixing matrix $\mathbf{A}$ is a full column rank, and its elements are randomly selected from [0,1].

The cross-talking error proposed by Amari et al. (1997) is utilized to evaluate the separation performance

$$PI = \frac{1}{m} \sum_{p=1}^{m} \left(\sum_{q=1}^{n} \frac{|c_{pq}|}{\max_k |c_{pk}|} - 1 \right) + \frac{1}{n} \sum_{q=1}^{n} \left(\sum_{p=1}^{m} \frac{|c_{pq}|}{\max_k |c_{kq}|} - 1 \right) \tag{22}$$

where $\{c_{pq}\} = \mathbf{WA}$ is the combination of the separating and mixing matrix. A small PI value implies a competent separation.

We implement Ye's algorithm (Ye et al., 2004) in each simulation to compare with the ANA using three different learning rates ($\mu_1 = 0.001$, $\mu_2 = 0.005$ and $\mu_{ad} \times \mu_2$). The involved parameters of Ye's algorithm are given by referring to Ye et al. (2004). In ANA, there are eight sensors (i.e. $m = 8$), the initial size of structure is 1 (i.e. $\tilde{n}(0) = 1$), and the length of the cross-validation slide window is $wd = 30$. Since all of the sources in (21) are sub-Gaussian signals, we take on the advice of Cao and Liu (1996), and pick activation functions for (20):

$$\begin{cases} f(y_i(t)) = \text{sign}(y_i(t)) \\ g(y_i(t)) = \tanh(10 y_i(t)) \end{cases} \tag{23}$$

All simulation results will be presented as average values which are generated from 50 independent runs of some algorithm for each simulation.

4.1. BSS with static number of sources

This simulation includes two cases whose number of sources is constant. The first is a $n = 3$ case which has the first three source signals $s_i(t)$ of (21), where $i = 1,\ldots,3$. Similarly, the second is a $n = 5$ case, and the index of source signals is $i = 1,\ldots,5$.

Presented in Figs. 3 and 6 are the results of both cases after 10,000 iterations, and the curves of estimation of source signals by the ANA with adaptive learning rates. The waveforms of sepa-

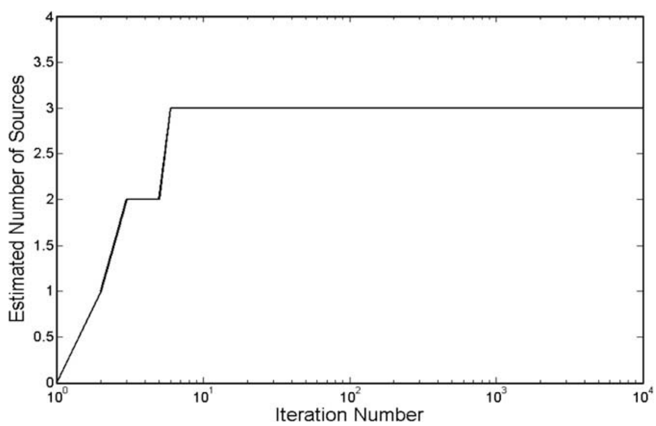

Fig. 3. The curve of estimated number of sources by the ANA with $\mu_{ad} \times \mu_2$ in the $n = 3$ case.

rated signals are presented in Figs. 4 and 7 which display the last 500 separated samples. Figs. 5 and 8 present four curves of the average *PI* which are obtained by Ye's algorithm and the ANA with different learning rates. The average and variance of final *PI* values are listed in Table 1 for comparing.

4.2. BSS with dynamic number of sources

This is a dynamic BSS simulation which uses time-variant numbers of sources in the following equations:

(1) $n = 4, 3, 6$ *Case:* the numbers of sources decrease, then will increase:

$$n = \begin{cases} 4, \text{as } IN \leqslant 6000 \\ 3, \text{as } 6000 < IN \leqslant 12,000 \\ 6, \text{as } 12,000 < IN \leqslant 18,000 \end{cases} \tag{24}$$

(2) $n = 3, 6, 2$ *Case:* the numbers of sources increase, then will decrease:

$$n = \begin{cases} 3, \text{as } IN \leqslant 6000 \\ 6, \text{as } 6000 < IN \leqslant 12,000 \\ 2, \text{as } 12,000 < IN \leqslant 18,000 \end{cases} \tag{25}$$

where *IN* denotes the iteration number, and *n* source signals is the first *n* of (21). Both cases have three states of *n* which switch every 6000 iterations.

After the two simulation cases finish, the number of sources obtained by the ANA with $\mu_{ad} \times \mu_2$ are presented in Figs. 9 and 12, the waveforms of the last 200 separated signals of each state are presented in Figs. 10 and 13. Several windows have no results because those outputs do not exist. The four *PI* curves obtained from different methods are compared in Figs. 11 and 14. The average and variance of the final *PI* values are further presented in Table 1. It should be noted that the values of Ye's algorithm in the table are all blank because it is always diverging when the number of sources is increasing.

5. Discussion

From the above simulation results, it can be confirmed at first that the on-line source number estimator is functional and has a short identifying time (under 20 iterations which is smaller than the value of *wd*). The value of *wd* is selected by experience. If *wd* is not big enough, the estimation of cross-validation cannot correctly evaluate the source numbers.

About the convergence of the *PI* average, Ye's algorithm always converges at a high value because the self-normalized matrix can-

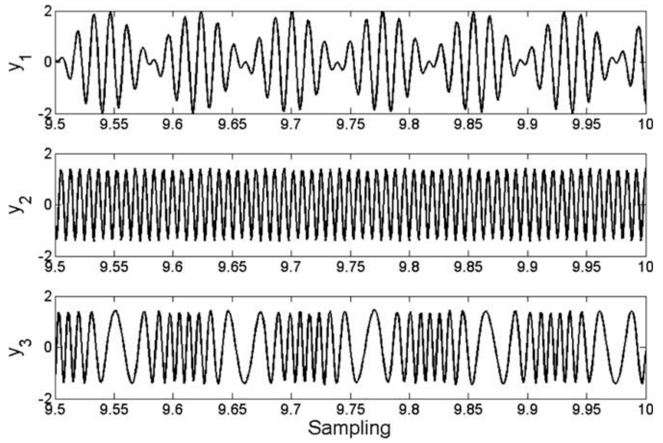

Fig. 4. The waveforms of the separated signals by the ANA with $\mu_{ad} \times \mu_2$ in the $n = 3$ case.

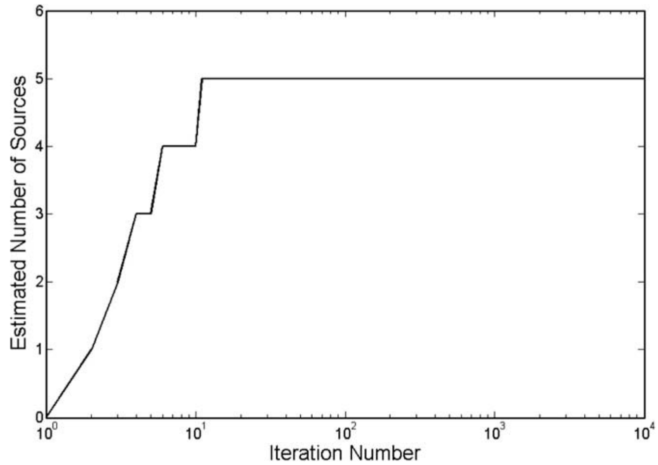

Fig. 6. The curve of estimated number of sources by the ANA with $\mu_{ad} \times \mu_2$ in the $n = 5$ case.

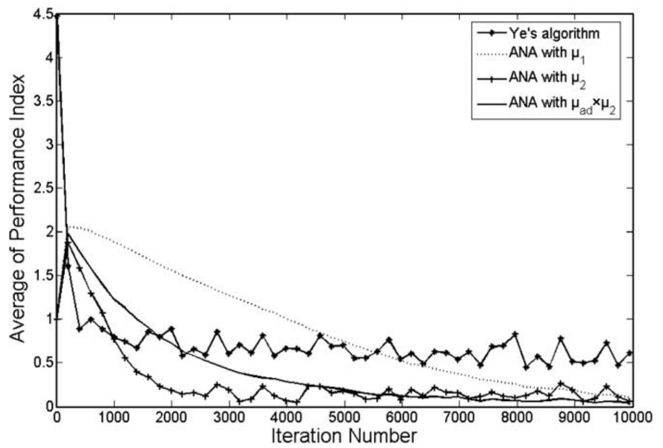

Fig. 5. The average performance indexes of Ye's algorithm and the ANA with different learning rates (50 independent runs of the $n = 3$ case).

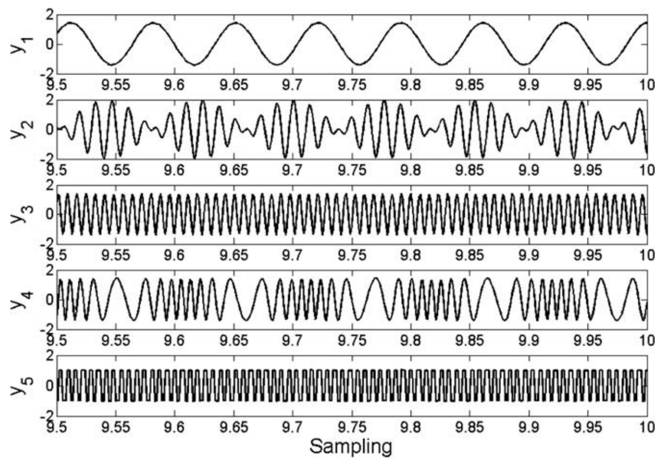

Fig. 7. The waveforms of the separated signals by the ANA with $\mu_{ad} \times \mu_2$ in the $n = 5$ case.

not find the perfect value during its evolution. For ANA, the *PI* curve decreases monotonously when a small learning rate μ_1 is used. Inversely, the curve fluctuates during decrease and cannot steady when a big learning rate μ_2 is used. However, when the adaptive learning rate $\mu_{ad} \times \mu_2$ is used, the *PI* speedily and stably converges at a lower value than others.

From Table 1, it can be observed that both algorithms have similar performance in variance, but the ANA has a far superior value on average.

Table 1
Comparison of final performance index.

Case	*PI*	ANA with $\mu_{ad} \times \mu_2$	Ye's algorithm
$n = 3$	Average	0.0476	0.6322
	Variance	0.0014	0.0116
$n = 5$	Average	0.0863	0.8639
	Variance	0.0116	0.0119
$n = 4, 3, 6$	Average	0.1794	None
	Variance	0.0202	None
$n = 3, 6, 2$	Average	0.0175	None
	Variance	9.0621e−006	None

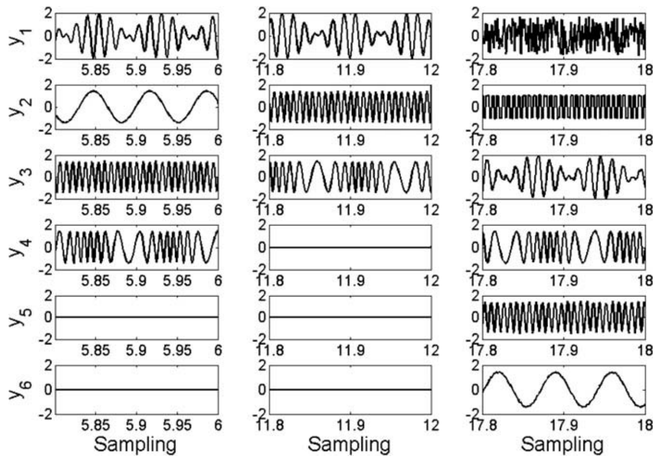

Fig. 10. The waveforms of the separated signals by the ANA with $\mu_{ad} \times \mu_2$ in the $n = 4, 3, 6$ case (the windows of columns from left to right correspond to all outputs of neural network in the iteration intervals [5801–6000], [11,801–12,000] and [17,801–18,000]).

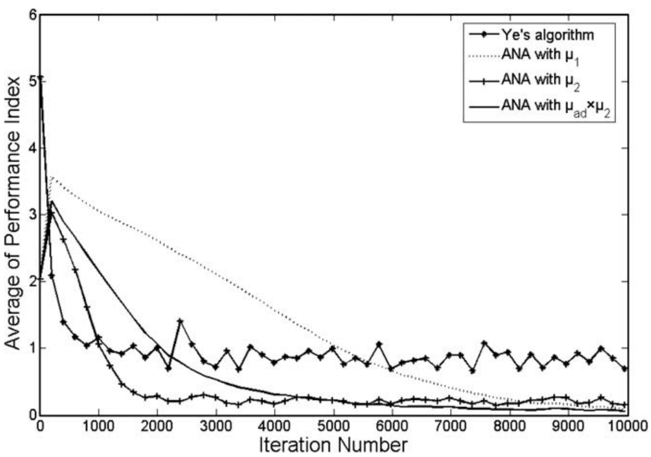

Fig. 8. The average performance indexes of Ye's algorithm and the ANA with different learning rates (50 independent runs of the $n = 5$ case).

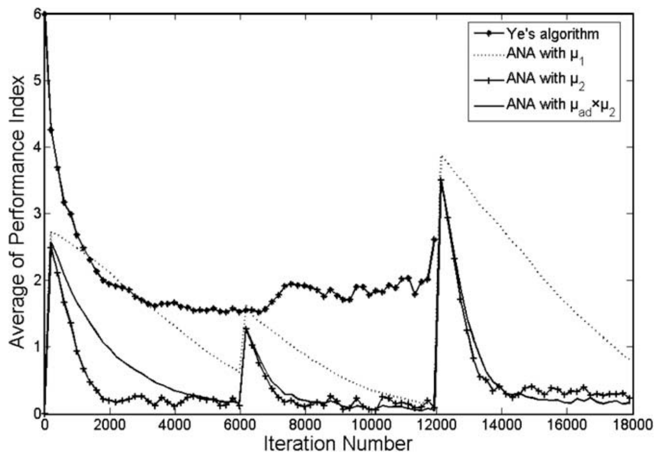

Fig. 11. The average performance indexes of Ye's algorithm and the ANA with different learning rates (over 50 independent runs of the $n = 4, 3, 6$ case).

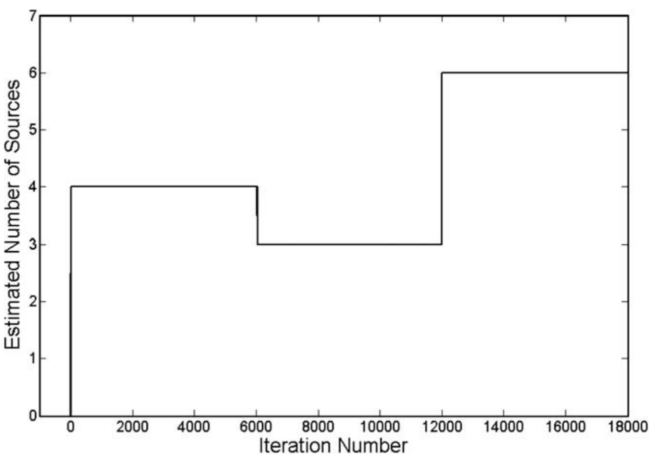

Fig. 9. The curve of estimated number of sources by the ANA in the $n = 4, 3, 6$ case.

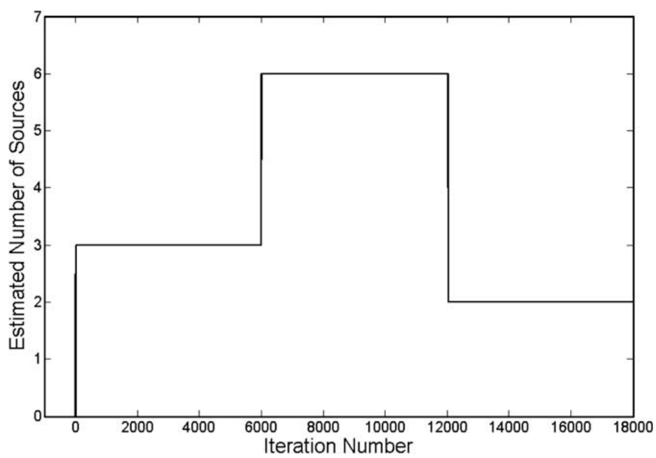

Fig. 12. The curve of estimated number of sources by the ANA in the $n = 3, 6, 2$ case.

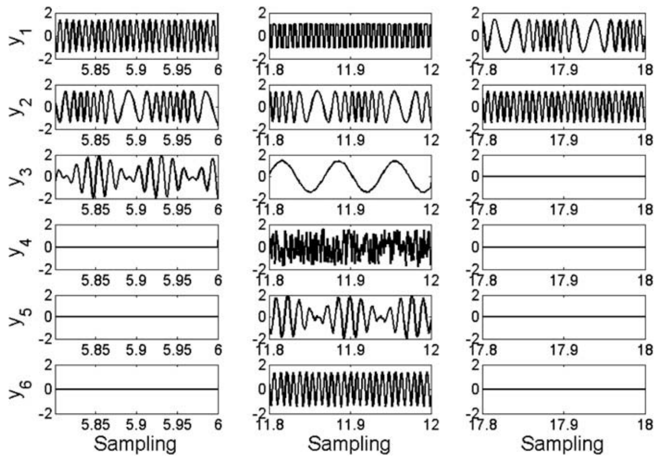

Fig. 13. The waveforms of the separated signals by the ANA with $\mu_{ad} \times \mu_2$ in the $n = 3, 6, 2$ case (the windows of columns from left to right correspond to all outputs of neural network in the iteration intervals [5801–6000], [11,801–12,000] and [17,801–18,000]).

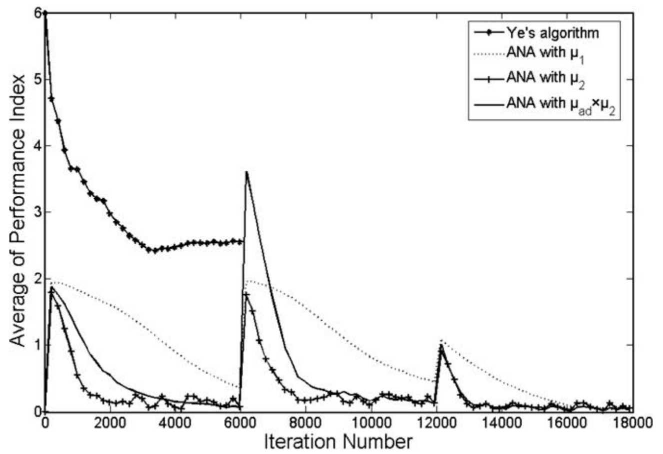

Fig. 14. The average performance indexes of Ye's algorithm and the ANA with different learning rates (over 50 independent runs of the $n = 3, 6, 2$ case).

6. Conclusion

An adaptive neural algorithm to solve BSS problems with unknown and dynamic source numbers is proposed in this paper. Estimation of the source numbers are based on a system improved from cross-validation, and allows fast evaluation and discovery of the actual value. Next, the self-organized criterion guides the dimension of NN into the determined state. The learning rate adjustment further enhances the speed and stability of NN convergence.

According to the simulation results, the ANA is able to conquer the supplied BSS problems with a separation performance superior to that of Ye's algorithm, what's more, the proposed adaptive learning rate benefits from fast and stable training of NN when using fixed learning rates.

References

Amari, S., Chen, T. P., & Cichocki, A. (1997). Stability analysis of learning algorithms for blind source separation. *Neural Networks, 10*(8), 1345–1351.

Araki, S., Sawada, H., Mukai, R., & Makino, S. (2007). Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors. *Signal Processing, 87*, 1833–1847.

Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Acoustics, Speech and Signal Processing, 45*(2), 434–444.

Cao, X. R., & Liu, R. W. (1996). General approach to blind source separation. *IEEE Transactions on Acoustics, Speech and Signal Processing, 44*(3), 562–571.

Cao, J., Murate, N., Amari, S., Cichocki, A., & Takeda, T. (2003). A robust approach to independent component analysis of signals with high-level noise measurements. *IEEE Transactions on Neural Networks, 14*(3), 631–645.

Cichocki, A., & Amari, S. (2003). *Adaptive blind signal and image processing*. New York: Wiley.

Cichocki, A., Karhunen, J., Kasprzak, W., & Vigário, R. (1999). Neural networks for blind separation with unknown number of sources. *Neurocomputing, 24*(1–3), 55–93.

Cichocki, A., & Unbehauen, R. (1996). Robust neural networks with on-line learning for blind identification and blind separation of sources. *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications, 43*, 894–906.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing, 36*(3), 287–314.

Even, J., & Moisan, E. (2005). Blind source separation using order statistics. *Signal Processing, 85*(9), 1744–1758.

Herault, J., & Juten, C. (1986). Space or time adaptive signal processing by neural network models. In J. S. Denker (Ed.), *AIP Conf. Snowbird, UT. Neural Networks for Computing* (pp. 206–211). New York: Amer. Inst. Phys.

Krzanowski, W. J. (1987). Cross-validation in principal component analysis. *Biometrics, 43*, 575–584.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Computation, 11*(2), 409–433.

Lee, T. W., Lewicki, M. S., Girolami, M., & Sejnowski, T. J. (1999). Blind source separation of more sources than mixtures using overcomplete representations. *Signal Processing Letters, 6*(4), 87–90.

Li, Y., & Wang, J. (2002). Sequential blind extraction of instantaneously mixed sources. *IEEE Transactions on Acoustics, Speech and Signal Processing, 50*(5), 997–1006.

Liu, C. C., Sun, T. Y., Li, K. Y., Hsieh, S. T., Tsai, S. J., (2007). Blind sparse source separation using cluster particle swarm optimization technique. In *IASTED international conference on artificial intelligence and applications* (pp. 289–294).

Liu, C. C., Sun, T. Y., Lin, C. L., & Chou, C. P. (2006). A self-organized neural network for blind separation process with unobservable sources. *International Symposium on Intelligent Signal Processing and Communication Systems*, 177–180.

Lou, S. T., & Zhang, X. D. (2003). Fuzzy-based learning rate determination for blind source separation. *IEEE Transactions on Fuzzy Systems, 11*(3), 375–383.

Lv, Q., & Zhang, X. D. (2006). A unified method for blind separation of sparse sources with unknown source number. *IEEE Signal Processing Letters, 13*(1), 49–51.

Morris, R. D., & Garvin, A. D. M. (1996). Fast probabilistic self-structuring of generalized single-layer networks. *IEEE Transactions on Neural Networks, 7*(4), 881–888.

Pajunen, P. (1998). Blind source separation using algorithmic information theory. *Neurocomputing, 22*(1–3), 35–48.

Park, J. H., Huh, S. H., Kim, S. H., Seo, S. J., & Park, G. T. (2005). Direct adaptive controller for nonaffine nonlinear systems using self-structuring neural networks. *IEEE Transactions on Neural Networks, 16*(2), 414–422.

Pham, D. T. (2002). Mutual information approach to blind separation of stationary sources. *IEEE Transactions on Information Theory, 48*(7), 1935–1946.

Pham, D. T., & Vrins, F. (2005). Local minima of information-theoretic criteria in blind source separation. *IEEE Signal Processing Letters, 12*(11), 788–791.

Schiessl, I., Stetter, M., Mayhew, J. E. W., McLoughlin, N., Lund, J. S., & Obermayer, K. (2000). Blind signal separation from optical imaging recordings with extended spatial decorrelation. *IEEE Transactions on Biomedical Engineering, 47*(5), 573–577.

Tenorio, da M., Manoel, F. (1990). The self organizing neural network algorithm: adapting structure for optimum supervised learning. In *The twenty-third annual hawaii international conference on system sciences* (vol. 1. pp. 187–195).

Tonazzini, A., Bedini, L., & Salerno, E. (2006). A Markov model for blind image separation by a mean-field EM algorithm. *IEEE Transactions on Image Processing, 15*(2), 473–482.

Vigliano, D., & Uncini, A. (2003). Flexible ICA solution for nonlinear blind source separation problem. *Electronics Letters, 39*(22), 1616–1617.

Wei, C., Woo, W. L., & Dlay, S. S. (2007). Nonlinear underdetermined blind signal separation using Bayesian neural network approach. *Digital Signal Processing*(17), 50–68.

Ye, J. M., Zhu, X. L., & Zhang, X. D. (2004). Adaptive blind separation with an unknown number of sources. *Neural Computation, 16*, 1647–1660.

Yilmaz, O., & Rickard, S. (2004). Blind separation of speech mixtures via time–frequency masking. *IEEE Transactions on Acoustics, Speech and Signal Processing, 52*(7), 1830–1847.

Zhang, Y. C., & Kassam, S. A. (2004). Robust rank-EASI algorithm for blind source separation. *IEE Proceedings Communications, 151*(1), 15–19.