Neural Comput & Applic (1997)6:127–141

Neural Computing & Applications

Blind Source Separation with Convolutive Noise Cancellation

W. Kasprzak, A. Cichocki and S. Amari

Frontier Research Program RIKEN, Brain Information Processing Group, Saitama, Japan

On-line adaptive learning algorithms for cancellation of additive, convolutive noise from linear mixtures of sources with a simultaneous blind source separation are developed. Associated neural network architectures are proposed. A simple convolutive noise model is assumed, i.e. the unknown additive noise in each channel is a (FIR) filtering version of environmental noise, where some convolutive reference noise is measurable. Two approaches are considered: in the first, the noise is cancelled from the linear mixture of source signals as pre-processing, after that the source signals are separated; in the second, both source separation and additive noise cancellation are performed simultaneously. Both steps consist of adaptive learning processes. By computer simulation experiments, it was found that the first approach is applicable for a large amount of noise, whereas in the second approach, a considerable increase of the convergence speed of the separation process can be achieved. Performance and validity of the proposed approaches are demonstrated by extensive computer simulations.

Keywords: Adaptive noise cancellation; Blind separation; FIR filters; ICA; Learning algorithms; Neural networks

1. Introduction

Blind Source Separation (BSS) has recently become an active research area both in statistical signal processing and unsupervised neural learning [1–8]. The goal of BSS is to extract statistically independent but otherwise unknown source signals from their linear mixtures without knowing the mixing coefficients. BSS techniques have many potential applications in, for example, data communications, speech processing and medical signal processing.

Most approaches to BSS assume that sensor signals are noiseless or noise is considered as one of the primary sources [4–7]. In the past we have developed efficient and robust learning algorithms for blind separation of 1-D and 2-D signals (images) [1,8]. We have demonstrated by computer simulations the high performance and efficiency of the proposed algorithms. These algorithms can extract all source signals even if some of them are extremely weak or the mixing matrix is very ill conditioned, assuming that additive noise $n_i(t)$ at each sensor is equal to zero or is negligibly small. In fact, we have assumed that the noise signal is one of the unknown primary source signals which could be separated from other sources. Of course, more than one source could be noise, but one of them can be Gaussian noise if it is necessary to extract all source signals including noises.

A more realistic and practical model of BSS considers that different unknown noise signals, possibly representing coloured noise, are added to each sensor signal [9–11]. Such a situation appears in most real-life (real-world) problems. Now the following problem arises: how can signals be efficiently separated if additive noise can no longer be neglected? Alternatively, the problem can be stated as: how to cancel or suppress the additive noise. Thus, our problem is how to modify existing algorithms to be still valid and efficient with additive noise. The main objective of this paper is to investigate this problem, and to propose some constructive solutions.

Correspondence and offprint requests to: Professor A. Cichocki, Frontier Research Program RIKEN, Brain Information Processing Group, Hirosawa 2-1, Wako-shi, Saitama 351-01, Japan (e-mail: cia@zoo.riken.go.jp)

In this paper we propose an adaptive approach to simultaneous source separation and cancellation of additive, convolutive noise from many-source signals. In the basic *demixing* model, we simultaneously separate signals and subtract additive noise by employing an adaptive FIR filter in each channel. In an alternative model, we first attempt to reduce or cancel noise, and then perform the blind separation of sources. Each step is performed by adaptive learning algorithms, based on the mutual stochastic independent principle and generalized energy minimization of the output signals.

2. Basic Model of Noise and Mixture

2.1. The Mixing Model

In this paper we consider the extended BSS mixing model that includes additive noise [9–11] (Fig. 1(a))

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) \quad (1)$$

where $\boldsymbol{s}(t) = [s_1(t),\ldots,s_m(t)]^{\mathrm{T}}$ is a vector of *unknown*,

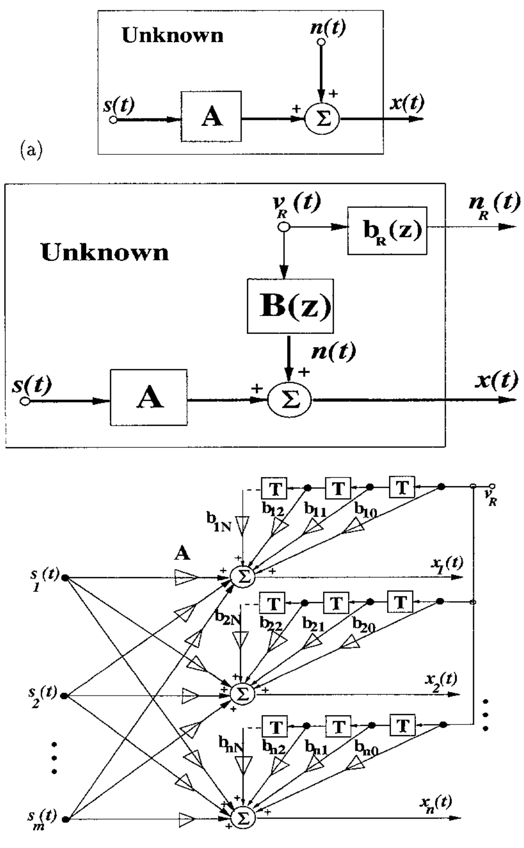

Fig. 1. The model of source mixing with additive noise. (a) General model, (b) a mixing model with convolutive noise, (unknown environment noise and its secondary (convolutive) reference noise), (c) more detailed model of the additive convolutive noise.

independent primary sources, $\boldsymbol{A}$ is a $m \times n$ unknown mixing matrix, $\boldsymbol{x}(t) = [x_1(t),\ldots,x_n(t)]^{\mathrm{T}}$ is the *observed* (*measured*) vector of sensor signals, and $\boldsymbol{n}(t) = [n_1(t),n_2(t),\ldots,n_n(t)]^{\mathrm{T}}$ is an additive noise vector. We assume that the noise signals $n_i(t)$ are decorrelated to source signals $s_i(t)$, $\forall i$.

In general, the problem is quite difficult because we have $n+m$ unknown signals (where m is the number of sources and n is the number of sensors). Hence, the problem is highly under-determined, and without any *a priori* information about the mixture model and/or noise it is very difficult or even impossible to solve it [9,10].

2.2. The Noise Model

However, in many practical situations we can measure or model the environmental noise. Such noise we will denote further as *reference noise* $\nu_R(t)$. For example, in an acoustic *cocktail party* problem we could measure such noise during a short salience period (when no persons speak), or we could measure and record such noise by an extra isolated microphone. In a similar way, we could measure noise in biomedical applications like EEG or ECG by extra electrodes, appropriately positioned.

This reference or environmental noise $\nu_R(t)$ influences each sensor, but it could be added to a mixture of signals with different strength. Moreover, noise could reach each sensor with some delay due to the finite time propagation of signals. For this reason in this work we model the additive and convolutive noise by the Finite Impulse Response or *Moving Average* (FIR or MA) model [9,12] (Fig. 1(b)), i.e.

$$n_i(t) = \sum_{j=0}^{N} b_{ij}(z)\nu_{Rj}(t) = \sum_{j=1}^{P}\sum_{k=0}^{N} b_{ijk}\nu_{Rj}(t-kT) \quad (2)$$

where $z^{-1} = e^{-sT}$ is unit delay and ν_{Rj} are reference noises.

Such a model is generally accepted as a realistic (real-world) model in both areas of signal and image processing [9,12]. In this model, we assume that the known reference noises $\boldsymbol{\nu}_R = [\nu_{R1},\nu_{R2},\ldots,\nu_{RP}]^{\mathrm{T}}$ are added to each sensor (mixture of sources) with different unit delays T and various but unknown coefficients $b_{ijk}(t)$. In other words, we assume that additive noise is a convolution and superposition of some reference noises ν_{Rj} (see Fig. 1(c))

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{B}(z)\boldsymbol{\nu}_R(t) \quad (3)$$

where $\boldsymbol{B}(z) = [b_{ij}(z)]_{n\times P}$ with

$$b_{ij}(z) = b_{ij0} + b_{ij1}z^{-1} + \ldots + b_{ijN}z^{-N} \quad (4)$$

and $\boldsymbol{\nu}_R = [\nu_{R1},\nu_{R2},\ldots,\nu_{RP}]^{\mathrm{T}}$.

For simplicity, we assume that only one single reference noise ν_R ($b_{ij}(z) = b_j(z)$) is available (see Fig. 1). However, we could easily extend our approach for the case of an arbitrary number of reference noises.

2.3. Measured Reference Noise

As already mentioned, to propose a realistic solution for source separation in a noisy environment, we should be able to measure a reference noise ν_R directly, or some of its convolutive form n_R. Here we assume that a convolutive reference noise $n_R(t)$ can be measured independently from the sensors measuring the noisy source mixtures (see Fig. 1(b)). Noise $n_R(t)$ is also modelled as the convolution of the unknown environment noise $\nu_R(t)$, i.e.

$$n_R(t) = \sum_{j=0}^{N_R} b_{Rj}\nu_R(t - jT) = \boldsymbol{b}_R(z)\nu_R(t) \tag{5}$$

Thus, a general mixing model (Fig. 1(b)) contains the following unknown elements: the matrix $\boldsymbol{A}$, matrix $\boldsymbol{B}(z)$ and the vector $\boldsymbol{b}_R(z)$. It is assumed that the number of sources and the number of time delay units N (i.e. maximum order of FIR filters) are completely unknown (however, the number of sensors must be larger or equal to the number of sources).

3. De-mixing and De-noising Neural Network Models

In this section we describe the basic model for simultaneous source separation and noise cancellation, as well as a simplified model. Also, their multi-layer versions are proposed.

3.1. Basic Model

In the basic approach two learning steps are simultaneously performed: the signals are separated from their linear mixture and the additive noise is estimated and subtracted (Fig. 2(a,b)). Thus, the output signals are derived as

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{Wx}(t) - \boldsymbol{H}(z)n_R \\ &= \boldsymbol{WAs}(t) + \boldsymbol{WB}(z)\nu_R - \boldsymbol{H}(z)\boldsymbol{b}_R(z)\nu_R \end{aligned} \tag{6}$$

where $\boldsymbol{H}(z) = [h_1(z),\ldots,h_n(z)]^{\mathrm{T}}$ with

$$h_i(z) = h_{i0} + h_{i1}z^{-1} + h_{i2}z^{-2} + \ldots + h_{iM}z^{-M} \tag{7}$$

In such a simplified case, the de-mixing model can be described by a set of equations (see Fig. 2(b) and Fig. 3)

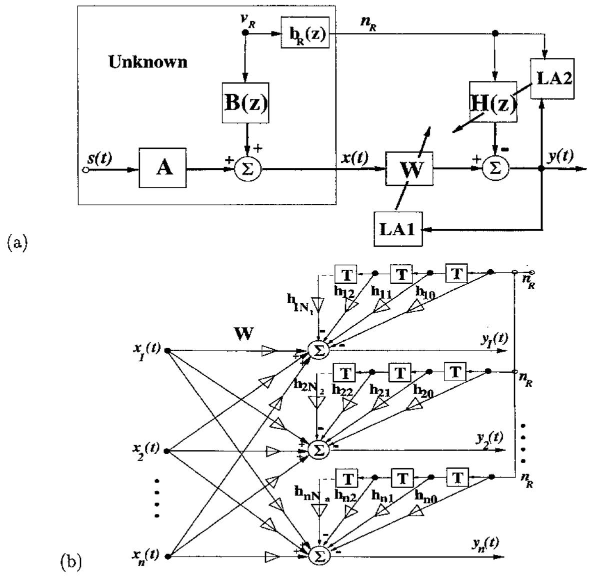

Fig. 2. The basic de-mixing model. (a) Applying two learning algorithms *LA*1, *LA*2 simultaneously, (b) detailed structure of the de-mixing and de-noising models.

$$y_i(t) = \tilde{y}_i(t) - n_i(t) = \tilde{y}_i(t) - \sum_{j=0}^{M} h_{ij}n_R(t - jT) \tag{8}$$

where $\tilde{y}_i(t) = \Sigma_{j=1}^{n} w_{ij}(t)x_j(t)$, ($i = 1,2,\ldots,m$).

For simplicity, let us assume that signals from $\boldsymbol{y}(t)$ are properly scaled and ordered in accordance with $\boldsymbol{s}(t)$. Then $\boldsymbol{y}(t) \simeq \boldsymbol{s}(t)$ if

$$\boldsymbol{H}(z)\boldsymbol{b}_R(z) = \boldsymbol{WB}(z) \tag{9}$$

$$\text{i.e.} \sum_{k=1}^{N} h_i(z)b_R(z) = \sum_{j=1}^{m} w_{ij}b_j(z), \forall i$$

$$\text{and } \boldsymbol{WA} = \boldsymbol{I} \tag{10}$$

The number of time delay units M in the de-noising model should be at least equal to the sum

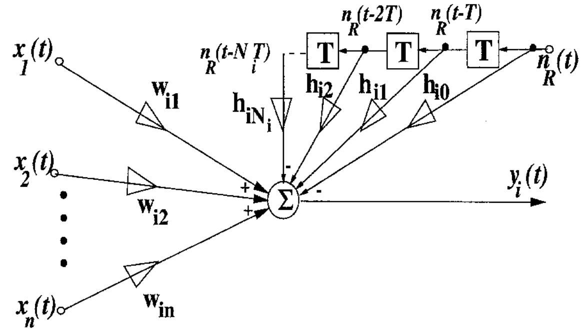

Fig. 3. Simple model of single neuron for simultaneous blind extraction and noise cancellation.

of corresponding numbers N, N_R in the mixing model (i.e. $M \geq (N+N_R)$), but in practice, especially if $N_R > 1$, it should be much larger.

3.2. Alternative Simplified Model

In a simplified model the two learning steps are performed in sequence. Here, as pre-processing we first attempt to cancel the noise contained in the mixture, and then to separate the sources (Fig. 4). This is the simplest way to deal with noise cancellation. Its drawback is that in real life problems we never expect to be able perfectly to cancel noise.

Thus, the output signals are derived from

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{W}[\mathbf{x}(t) - \mathbf{H}(z)\mathbf{n}_R] \\ &= \mathbf{W}\mathbf{A}\mathbf{s}(t) + \mathbf{W}[\mathbf{B}(z) - \mathbf{H}(z)\mathbf{b}_R(z)]\nu_R \end{aligned} \tag{11}$$

so it is obvious that $\mathbf{y}(t) \simeq \mathbf{s}(t)$ if (again, problems of signal scaling and permutation are ignored)

$$\mathbf{W}\mathbf{A} = \mathbf{I} \quad \text{and} \quad \mathbf{H}(z)\mathbf{b}_R(z) = \mathbf{B}(z) \tag{12}$$

3.3. Improved Multi-layer Models

To improve the learning performance multi-layer neural neworks could be used for both basic and simplified models (Fig. 5). These models perform separation and noise elimination by using multi-layer networks with pre- or post-processing noise cancellation steps.

At first, employing many layers is justified if we want to apply a local learning rule (discussed in the next section) for the separation of mixtures in which some signals are very weak or a mixing matrix $\mathbf{A}$ is ill-conditioned. Secondly, a multi-layer model might be a proper solution if the initialisation and decay factor of learning rates $\eta(t)$ and $\tilde{\eta}$ have not been chosen optimally. To optimally choose them, the noise level contained in the mixtures should be known in advance, otherwise specific methods for the learning rate adaptation itself may be considered [3].

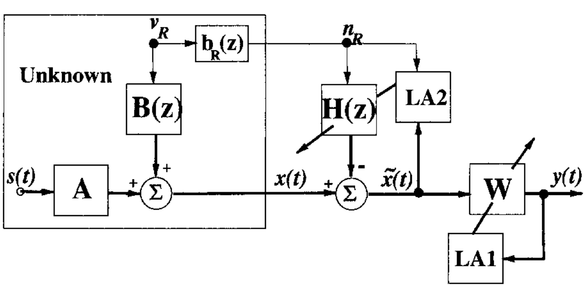

Fig. 4. Alternative (simplified) model for blind separation with noise cancellation as pre-processing.

4. Neural Network Learning Algorithms

In this section we describe the adaptive learning rules that perform simultaneous noise cancellation and blind source separation, according to a specified demixing model.

4.1. Separation Learning Rules

The dependency among output signals is measured by the Kullback–Leibler divergence between the joint and the product of the marginal distribution of the outputs [1,2]

$$D(\mathbf{W}) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\pi_{i=1}^{n} p_i(y_i)} \, d\mathbf{y} \tag{13}$$

where $p_i(y_i)$ is the marginal probability density function (pdf).

This Kullback–Leibler divergence $D(\mathbf{W})$ is related to the MI of $\mathbf{y}$ (mutual information of the outputs) given by [1]

$$\begin{aligned} D(\mathbf{W}) &= -H(\mathbf{y}) + \sum_{i=1}^{n} H(y_i) \\ &= -\log|\det(\mathbf{W})| + \sum_{i=1}^{n} \log[p_i(y_i)] \end{aligned} \tag{14}$$

where $H(\mathbf{y}) = -\int p(\mathbf{y})\log[p(\mathbf{y})]d\mathbf{y}$ and $H(y_i) = -\int p_i(y_i)\log[p_i(y_i)]dy_i$ is the marginal entropy. By applying the standard stochastic gradient descent algorithm

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \eta(t) \frac{\partial D}{\partial \mathbf{W}} \tag{15}$$

we obtain the following learning algorithm [6]

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \eta(t) \, [\mathbf{I} - \mathbf{f}[\mathbf{y}(t)] \, \mathbf{y}^{\mathrm{T}}(t)]\mathbf{W}^{-\mathrm{T}}(t) \tag{16}$$

where $\eta(t)$ is the learning rate and $\mathbf{f}(\mathbf{y}(t))$ is a vector of nonlinear activation functions. Typically, $f_i(y_i) = |y_i|^p \mathrm{sign}(y_i)$ ($p > 1$, typically $p = 2{,}3$) for sub-Gaussian source signals, and ($0 \leqslant p < 1$) for sub-Gaussian signals.

To avoid matrix inversion and to improve performance, we can apply a natural gradient approach introduced by Amari [1]

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \eta(t) \frac{\partial D}{\partial \mathbf{W}} \mathbf{W}^{\mathrm{T}}\mathbf{W} \tag{17}$$

which leads [2] to the robust form of basic learning algorithm

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \eta(t)\,[\boldsymbol{I} - f[\boldsymbol{y}(t)]\,\boldsymbol{y}^{\mathrm{T}}(t)]\boldsymbol{W}(t) \tag{18}$$

Using a modified form of stochastic gradient descent technique

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) - \eta(t)\,\frac{\partial D}{\partial \boldsymbol{W}}\boldsymbol{W}^{\mathrm{T}} \tag{19}$$

we obtain a simple local learning rule

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \eta(t)\,\{\boldsymbol{I} - f[\boldsymbol{y}(t)]\boldsymbol{y}^{\mathrm{T}}(t)\} \tag{20}$$

or

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)\,\{\delta_{ij} - f_i[y_i(t)]y_j^{\mathrm{T}}(t)\} \tag{21}$$

4.2. Noise Cancellation: Generalised Adaptive Delta Rule

To develop an adaptive learning algorithm for updating on-line coefficients $h_{ij}(t)$, we can apply the concept of the minimisation of the generalised output energy of output signals $\tilde{\boldsymbol{x}}(t) = [\tilde{\boldsymbol{x}}_1(t), \tilde{\boldsymbol{x}}_2(t), \ldots, \tilde{\boldsymbol{x}}_n(t)]^{\mathrm{T}}$. In other words, we can formulate the following cost function (generalised energy):

$$J(\boldsymbol{h}) = \sum_{i=1}^{n} \rho_i(\tilde{x}_i) \tag{22}$$

where $\rho_i(\tilde{x}_i)$ is a suitably chosen loss function, typically [13]

$$\rho_i(\tilde{x}_i) = \frac{1}{\beta}\log\cosh(\beta\tilde{x}_i) \quad \text{or} \quad \rho_i(\tilde{x}_i) = \frac{1}{p}|x_i)|^p \tag{23}$$

and

$$\tilde{x}_i(t) = x_i(t) - \sum_{j=1}^{n} h_{ij} n_R(t - jT), \quad \forall i \tag{24}$$

Minimisation of this cost function according to stochastic gradient descent leads to a learning algorithm (see Fig. 3)

$$\begin{aligned} h_{ij}(t+1) &= h_{ij}(t) - \tilde{\eta}(t)\frac{\partial J(\boldsymbol{h})}{\partial h_{ij}} \\ &= h_{ij}(t) + \tilde{\eta}(t) f_R[\tilde{x}_i(t)] n_R(t - jT) \end{aligned} \tag{25}$$

where $f_R(\tilde{x}_i(t))$ is a suitably chosen nonlinear activation (error) function defined as

$$f_R(\tilde{x}_i(t)) = \frac{\partial \rho_i(\tilde{x}_i)}{\partial \tilde{x}_i} \tag{26}$$

Typically, $f_R(\tilde{x}_i) = \tanh(\beta\tilde{x}_i)$ or $f_R(\tilde{x}_i) = \tilde{x}_i^3$. The optimal choice of activation function depends on the distribution of noise and source signals. For Gaussian noise it is optimal to set $f_R(\tilde{x}_i) = \tilde{x}_i$, for super-Gaussian signals one can use $f_R(\tilde{x}_i) = \tanh(\beta\tilde{x}_i)$, and for sub-Gaussian, $f_R(\tilde{x}_i) = |\tilde{x}_i|^p \mathrm{sign}(\tilde{x}_i), p = 2,3,4$.

4.3. Learning Rules for Multi-layer Model

The above-developed on-line adaptive learning rules can easily be generalised or extended for the more sophisticated models shown in Fig. 4(a,b). It should be noted that the synaptic weights h_{ij} can be updated according to the general rule (Fig. 5)

$$h_{ij}^{(l)}(t+1) = h_{ij}^{(l)}(t) + \tilde{\eta}(t) f_R[u_i^{(l)}(t)] n_R(t - jT) \tag{27}$$

where $u_i^{(l)}(t) = y_i^{(l)}(t)$ for the model of Fig. 4(a), and $u_i^{(l)}(\mathrm{t}) = \tilde{\mathrm{x}}_{(l)}(\mathrm{t})$ for the model shown in Fig. 4(b) $(l = 1,2,\ldots,K)$.

The synaptic weights of the separation net are updated during the leanring process as

$$\begin{aligned} \boldsymbol{W}^{(l)}(t+1) &= \boldsymbol{W}^{(l)}(t) + \\ &\eta(t)\,\{\boldsymbol{I} - f[\boldsymbol{y}^{(l)}(t)]\ [\boldsymbol{y}^{(l)}(t)][\boldsymbol{y}^{(l)}(t)]^{\mathrm{T}}\} \end{aligned} \tag{28}$$

or in scalar form as

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) + \eta(t)\,[\delta_{ij} - f_i(y_i^{(l)}(t))y_j^{(l)}(t)]] \tag{29}$$

5. Simulation Results

We have tested the proposed approach for various demixing models and for both image and sound sources. Convolutive noises were modelled using 10th order FIR filters. As we applied large additive noise, at first the performance of the proposed learning algorithms has been verified by perception of output signals, either by observing the output images or sound signal waveforms or by listening to them. To estimate the separation and noise cancellation quality more quantitatively, we assume that we know the original sources. Then the quality of results obtained is provided in two ways:

1. Individually for each source i, by estimating the best $SNR(Y_k, S_i)$ (signal to noise ratio) or $PSNR$ (peak signal to noise ratio) between some reconstructed source Y_k and the corresponding original source S_i

$$SNR = -10\log_{10}(MSE) \tag{30}$$

$$PSNR = 10\log_{10}\left(\frac{A^2}{MSE}\right) \tag{31}$$

where MSE is the mean square error of separated source S_i

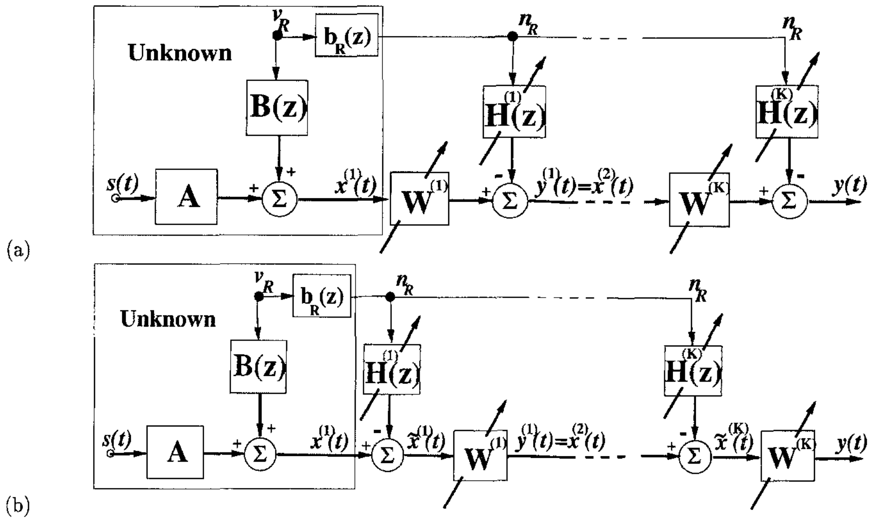

Fig. 5. Alternative improved multi-layer neural network models for blind separation with noise cancellation.

$$MSE = \frac{1}{T}\sum_{k=1}^{T}(y_{jk} - s_{jk})^2 \tag{32}$$

$A = s_{\max} - s_{\min}$ is the amplitude interval of source signal, and T is the number of samples.

2. For the whole separated signal set, by calculation of a normalised error index EI, which is defined as

$$EI = \sum_{j=1}^{m}\left(\sum_{i=1}^{n}|p_{ij}|^2 - 1\right) \tag{33}$$

The p_{ij}-s are entries of a normalised matrix $\boldsymbol{P}(t) = \boldsymbol{W}(t)\boldsymbol{A} \in R^{n\times m}$. Every non-zero row of matrix $\boldsymbol{P}$ is normalised by its highest value, i.e. $\max_i|p_{ik}| = 1, \forall i$.

In the overdetermined case, i.e. if $m > n$, the index EI is evaluated for an n-elementary subset of output signals, containing all reconstructed sources.

5.1. Direct Measurement of Reference Noise

In the first experiment, four (unknown) natural images (or unknown sound sources) were mixed by an ill-conditioned mixing matrix $m = 4$, $n = 4$ (Figs 6,8)

$$\boldsymbol{A}_1 = \begin{bmatrix} 1.0 & 0.7 & 0.4 & 0.5 \\ 0.9 & 0.9 & 0.45 & 0.6 \\ 0.8 & 0.85 & 0.5 & 0.55 \\ 0.7 & 0.9 & 0.6 & 0.45 \end{bmatrix}$$

with condition number cond $(\boldsymbol{A}_1) = 106.68$.

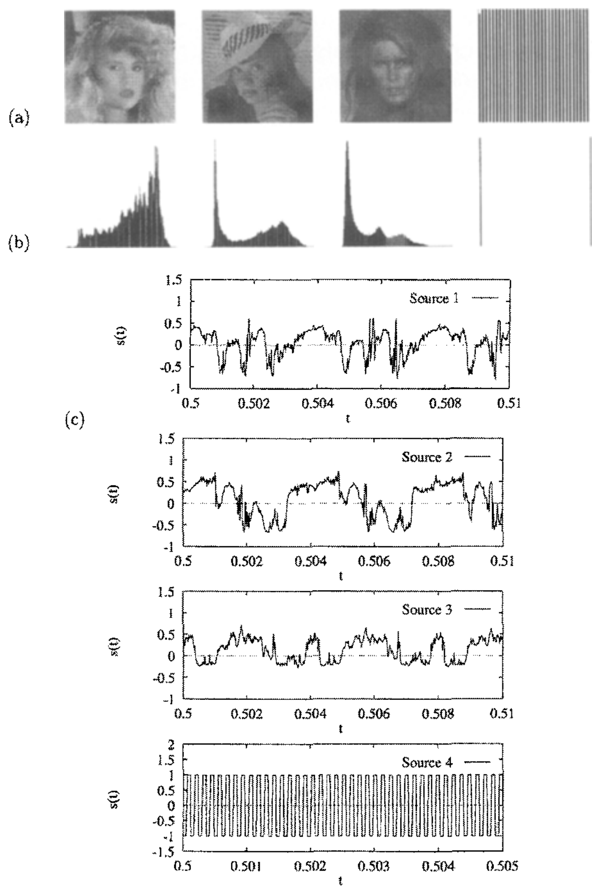

Fig. 6. Four original images (assumed to be completely unknown). (a) The images, (b) their histograms, (c) exemplary 1-D source signals after image scanning.

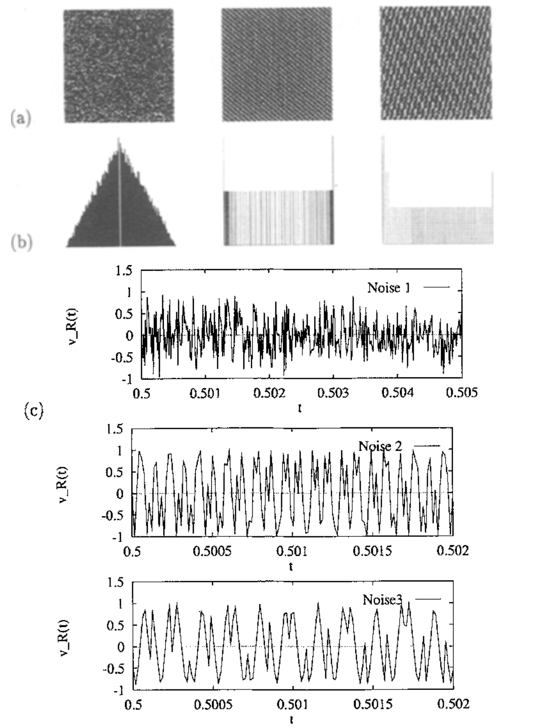

Fig. 7. Three exemplary environment noises used in our computer experiments for generation of additive convolutive nosises. (a) The noise images, (b) their histograms, (c) 1-D noise signals after image scanning.

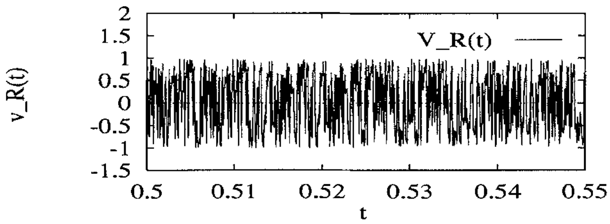

Fig. 9. An environment noise signal used in our sound separation tests for generation of additive convolutive noises.

Table 1. Statistical characteristics of normalised images and different reference noises used in experiments: m_1 – mean value, μ_2 – variance, μ_3 skewness, κ_4 – normalised kurtosis

Image	m_1	μ_2	μ_3	κ_4
Face 1	0.00	0.1086	–0.0283	–0.315
Face 2	0.00	0.2270	–0.0156	–1.473
Face 3	0.00	0.0619	0.0134	–0.449
Stripes	0.00	1.000	0.00000	–2.000
Noise signals: ν_R				
Noise 1	0.00	0.1688	–0.0004	–0.600
Noise 2	0.00	0.5040	0.0000	–1.502
Noise 3	0.00	0.3503	0.0213	–1.291

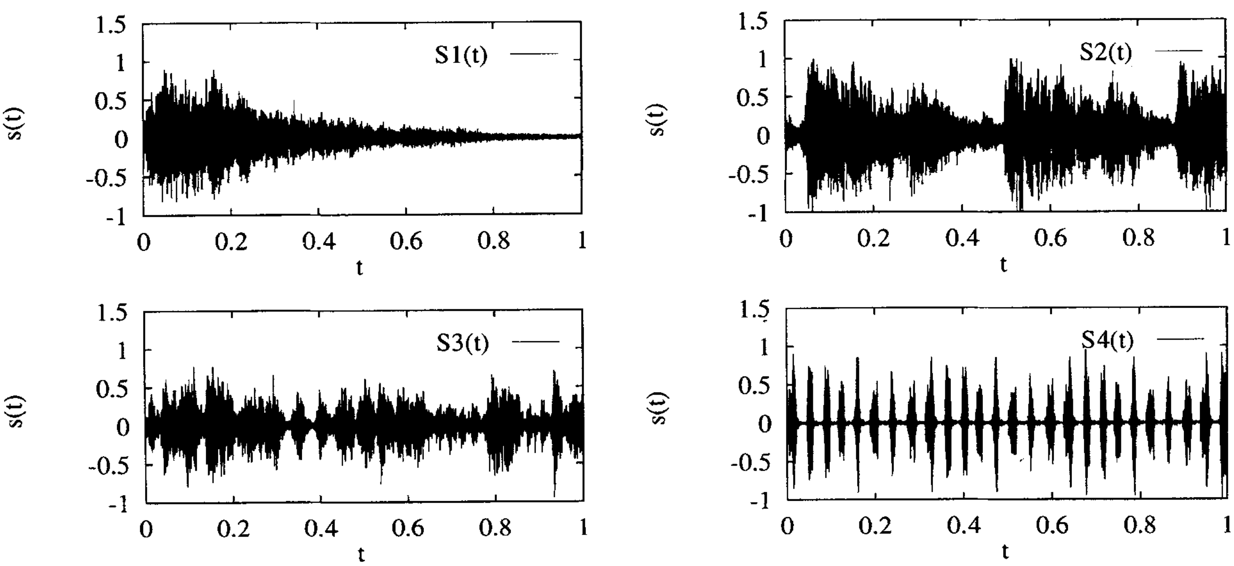

Fig. 8. Four sound sources used in our computer tests (gong, music, laughter, bird chirp).

In the first experiment we assume the environment noise is directly available as the reference noise, i.e. $n_R = \nu_R$ and $\boldsymbol{b}_R = 1$. (Figs 7,9).

Convolutive noises were modelled using four 10th order FIR filters (N=10), where all of them are introducing large additive noise. Their coefficients in vector form are randomly chosen as

$$\boldsymbol{b}_1 = [1.2,1.4,1.1,1.0,0.9,0.8,0.7,0.55,0.40,0.3]$$
$$\boldsymbol{b}_2 = [1.3,1.2,1.5,1.1,0.95,0.84,0.77,0.65,0.54,0.4]$$
$$\boldsymbol{b}_3 = [1.1,1.2,1.3,1.15,0.99,0.74,0.87,0.85,0.94,1.0]$$
$$\boldsymbol{b}_4 = [1.2,1.0,0.8,0.55,0.28,0.14,0.37,0.48,0.64,0.9]$$

In the de-mixing model, the number M of delay units was chosen to be equal to 25. Both the basic

Table 2. Source image and noise image correlations (in %), i.e. $E\{s_i s_j\} \times 100$

Image sources	Image sources			Noise images $\nu_R(t)$		
	Face 2	Face 3	Stripes	Noise 1	Noise 2	Noise 3
Face 1	22.60	0.275	0.709	0.316	0.165	0.195
Face 2	–	14.05	0.226	0.304	0.040	0.914
Face 3	–	–	0.875	0.323	0.111	0.158
Stripes 3	–	–	–	0.149	0.0002	0.025

Table 3. Statistical characteristics of sound sources and environmental noise used in experiments: m_1 – mean value, μ_2 – variance, μ_3 skewness, κ_4 – normalised kurtosis

Sound	m_1	μ_2	μ_3	κ_4
Gong (S1)	0.00	0.0211	0.0000	5.514
Music (S2)	0.00	0.0590	0.0012	1.198
Laughter (S3)	0.00	0.0277	–0.0004	1.952
Chirp	0.00	0.0397	0.0002	4.314
Noise	0.00	0.4915	0.0071	−1.661

Table 4. Sound correlations (in %), i.e. $E\{s_j s_j\} \times 100$

Sound sources	Sound sources			Noise
	Music	Laughter	Chirp	
Gong Music	0.499	1.330	0.247	0.343
Laughter	–	0.234	0.037	0.362
Chirp	–	–	0.193	0.007
	–	–	–	0.016

and simplified demixing models of Figs 2 and 3 gave very good noise cancellation and source separation results. Already after one epoch of signal data, the weights in $\boldsymbol{h}$ achieved an equilibrium point, e.g. final estimation of the deconvolution matrix $\boldsymbol{H}(z)$ for model in Fig. 2 was

$\boldsymbol{h}_1$ = [1.2022,1.4011,1.0967,0.9982,0.8962,
0.7949,0.6999,0.5470,0.3998,0.2989,
–0.0118,–0.0052,–0.0037,–0.0028,
0.0017,–0.0014,–0.0009,0.0000,
–0.0125,–0.0070,–0.0084,–0.0067,–0.0019,
–0.0016,–0.0022]

$\boldsymbol{h}_2$ = [1.3016,1.2035,1.4990,1.1013,0.9489,
0.8345,0.7707,0.6486,0.5409,0.4009,
–0.0112,–0.0036,–0.0020,–0.0033,0–0021,
–0.0006,–0.0010,0.0016,
–0.0122,–0.0051,–0.0068,–0.0069,–0.0016,
–0.0019,–0.0034]

$\boldsymbol{h}_3$ = [1.1017,1.2028,1.2985,1.1510,
0.9889,0.7351,0.8707,0.8487,0.9413,
1.0011,–0.0104,–0.0035,–0.0022,–0.0028,0.0024,
–0.0000,–0.0005,0.0014,
–0.0112,–0.0048,–0.0059,–0.0061,
–0.0013,–0.0018,–0.0032]

$\boldsymbol{h}_4$ = [1.2018,1.0025,0.7986,
0.5514,0.2799,0.1356,0.3705,0.4789,0.6419,
0.9018,–0.0096,–0.0032,–0.0020,–0.0024,
0.0027,0.0003,–0.004,0.0013,
–0.0103,–0.0043,–0.0049,–0.0058,
–0.0014,–0.0021,–0.0035]

From the above data, it is evident that the estimation error for each matrix element is below 1%. The exemplary results for this simple reference noise case are illustrated in Figs 10 and 11.

The quality factors SNR, $PSNR$ and the error index EI of separated sources are given in Tables 5 and 7.

5.2. Convolutive Reference Noise

In the next experiments we assume a general reference noise measurement model in which $\boldsymbol{b}_R$ is some non-zero vector, and $n_R(t)\quad \boldsymbol{b}_R(z)\nu_R(t)$. We apply a 10-order FIR filter $\boldsymbol{b}_R$ = [1.5,1.3,1.2,0.92,0.75, 0.70,0.75,0.86,0.94,1.0].

It is assumed that these coefficients are completely unknown. In the de-mixing model the number M of delay units was selected as 100. If compared to the previous case of a simplified reference noise measurement, the learning process is slower and the noise cannot be fully cancelled. Although in the pre-processing model of noise cancellation we were able to suppress the additive noise in the mixture of source signals, in the subsequent source separation stage this noise is again amplified to some visible form in one of the outputs. Other $m-1$ outputs correspond to $(m-1)$ estimated sources. However, the quality of such separated $(m-1)$ sources strongly depends upon the condition number of the mixing matrix. In the case of the matrix $\boldsymbol{A}_1$ only three sources have been extracted with rather poor quality. For another mixing matrix $\boldsymbol{A}_2$, due to its lower condition number a successful separation of three sources (from a total of four) with high quality was possible. Illustrative results are given in Figs 12(a,b) and 13(a,b), where $\boldsymbol{A}_2$ is

$$\boldsymbol{A}_2 = \begin{bmatrix} 1.0 & 0.7 & 0.6 & 0.3 \\ 0.5 & 0.9 & 0.7 & 0.4 \\ 0.2 & 0.4 & 0.4 & 0.8 \\ 0.5 & 0.8 & 0.6 & 0.9 \end{bmatrix}$$

Its condition number is cond(A_2) = 38.78.

To successfully extract all sources under additive noise, at least one auxiliary sensor may be useful. We have investigated the influence of the number of sensors on the perfomance of the proposed learning algorithms. For example, the de-mixing process provided good results, even if the environment noise was not directly available, when applied to mixtures obtained from following mixing matrices A_3,A_4 (and with additive noise introduced at each sensor)

$$A_3 = \begin{bmatrix} 1.0 & 0.7 & 0.6 & 0.3 \\ 0.5 & 0.9 & 0.7 & 0.4 \\ 0.2 & 0.4 & 0.4 & 0.8 \\ 0.5 & 0.8 & 0.6 & 0.9 \\ 0.5 & 0.6 & 0.8 & 0.8 \end{bmatrix}, \ \mathrm{cont}(A_3) = 13.475$$

$$A_4 = \begin{bmatrix} 1.0 & 0.7 & 0.3 \\ 0.5 & 0.9 & 0.4 \\ 0.2 & 0.4 & 0.8 \\ 0.5 & 0.8 & 0.9 \\ 0.5 & 0.6 & 0.8 \end{bmatrix}, \ \mathrm{cond}(A_4) = 7.575$$

For illustration of the results obtained, see Figs 12(c-d) and 13(c–d). Performance results for convolutive reference noise and various mixing matrices by A_1–A_4 are summarized in Tables 6 and 8.

5.3. Discussion of Results

It should be emphasised that performance depends on learning rates $\eta(t)$, $\tilde{\eta}(t)$ – on a proper setting of their initial values and decay factors – and on the chosen nonlinear activation functions f, f_R. The sample results show that for a given set of parameter values, both models of Figs 2 and 3 performed well, and gave similar results.

The basic demixing model ensures a faster learning, i.e. the learning rate for the separation matrix can be decreased much faster (by a factor of 10 or even more) than in the simple demixing model.

The pre-processing based noise cancellation scheme is able to eliminate or almost completely cancel the additive noise, i.e. if the original environment noise is measurable, the noise is fully cancelled, but if only a convolutive image of such noise is measurable, the additive noise is nearly eliminated

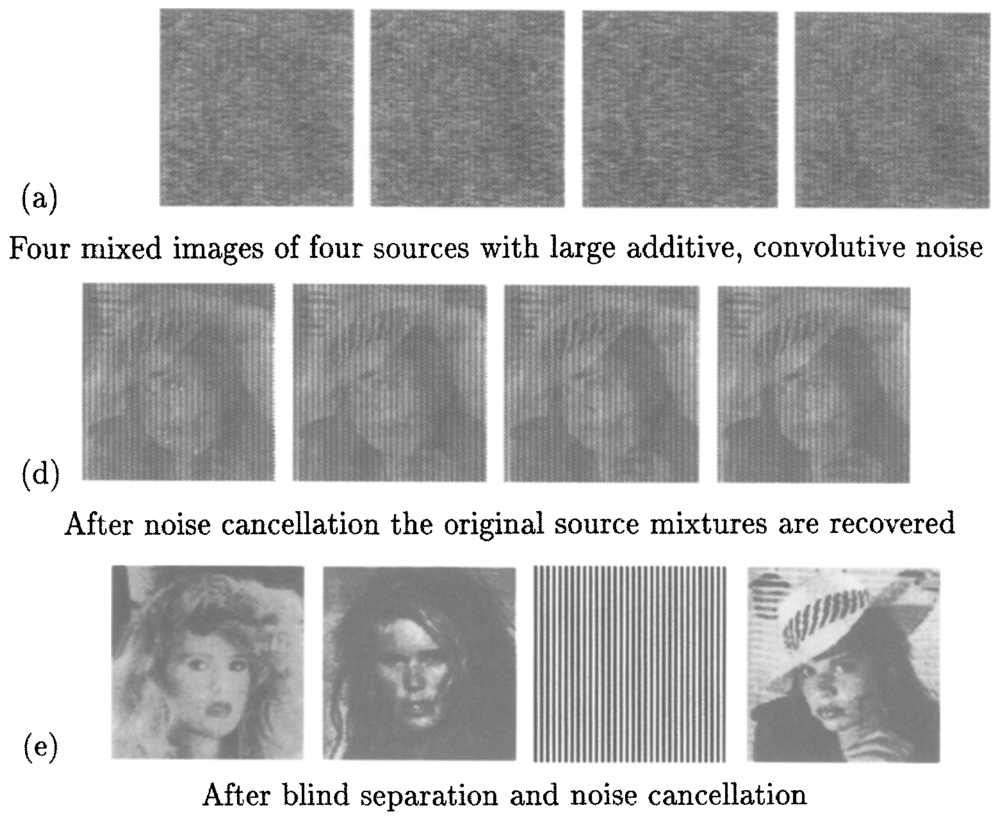

Fig. 10. Example of blind source separation and noise cancellation for image mixtures with additive, convolutive noise, if original environmental noise is directly available using model of Fig. 2.

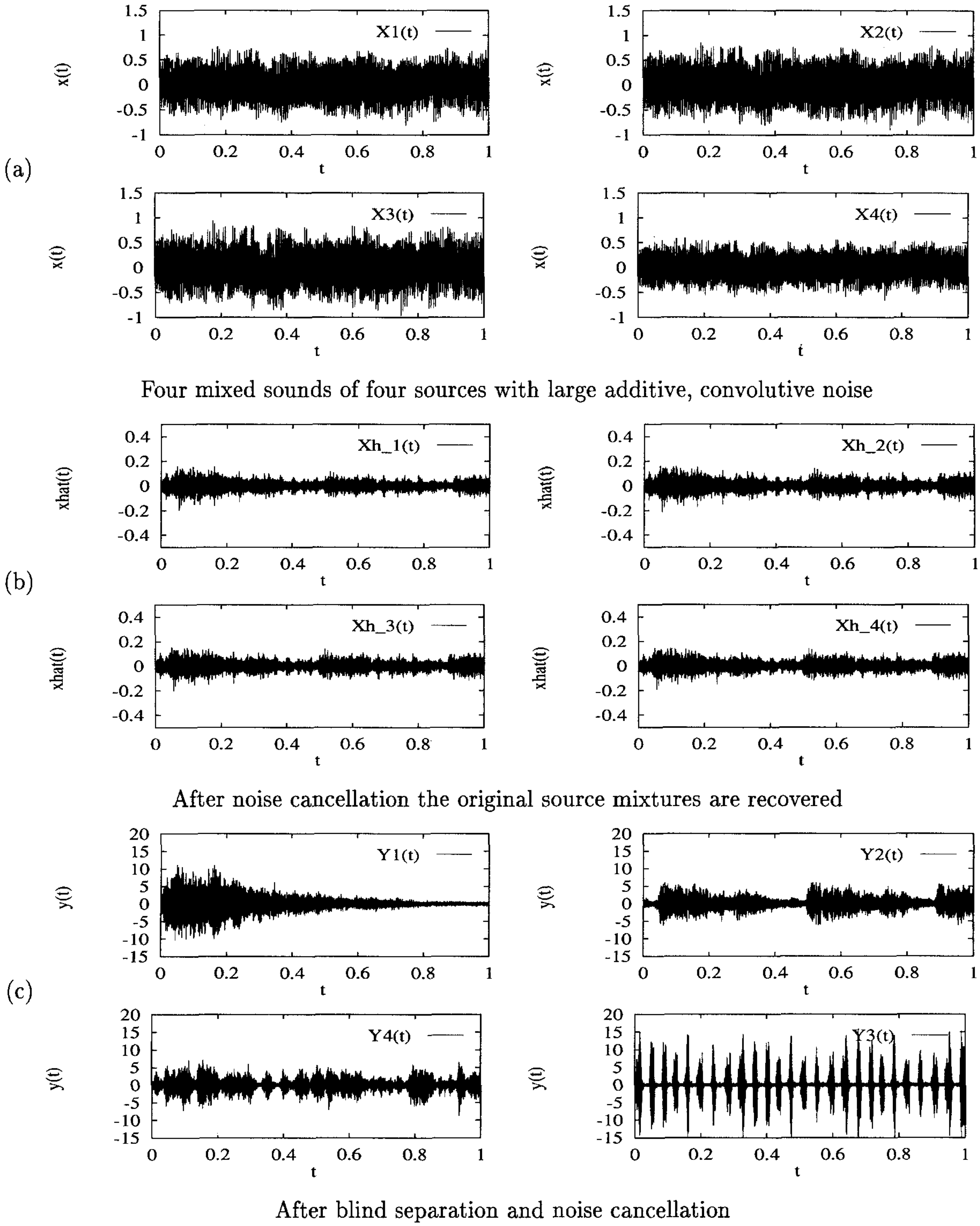

Fig. 11. Example of blind source separation and noise cancellation of a sound mixture with additive, convolutive noise, if original environmental noise is directly available.

Table 5. Quality factors of image separation with noise cancellation in the direct reference noise case

Noise	*EI*	SNR, PSNR [dB]			
		Face 1	*Face* 2	*Face* 3	*Stripes*
		Separation of a noise-free mixture			
none	0.063	8.42, 22.23	28.13, 39.88	10.74, 25.04	42.83, 48.85
		Simple model			
large 1	0.063	8.50, 22.31	26.95, 38.70	10.35, 24.66	32.93, 38.95
		Simultaneous model			
large 1	0.105	8.52, 22.33	17.96, 29.72	6.63, 20.94	33.33, 39.36

Table 6. Quality factors of image separation with noise cancellation in the direct reference noise case

Noise	*EI*	SNR, PSNR [dB]			
		Face 1	*Face* 2	*Face* 3	*Stripes*
		For difficult mixture mixed by $\boldsymbol{A}_1$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.975	6.51, 20.32	6.63, 18.38	–0.79, 10.96	4.13, 10.15
		For mixture mixed by $\boldsymbol{A}_2$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.321	8.90, 22.71	10.78, 22.53	–0.39, 13.91	13.81, 19.83
		For over-determined mixture mixed by $\boldsymbol{A}_3$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.155	5.64, 19.45	20.25, 32.01	9.50, 23.81	15.18, 21.20
		For over-determined mixture mixed by $\boldsymbol{A}_4$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.023	10.57, 24.38	21.39, 33.15	–,–	19.77, 25.79

in this stage. In the subsequent blind separation stage, in the first case the sources are very well separated, whereas in the second case the noise is amplified and appears in one of the outputs, so one source signal is usually lost. If the mixing matrix is very ill-conditioned, the noise can also appear on other outputs, corrupting the $(m-1)$ separated sources. The same behaviour has been observed in the case of simultaneous separation and noise cancellation (see Fig. 2). In practice, for this basic approach the appropriate decay of the learning rates may be more difficult to establish, i.e. without careful study of this problem the separation results will usually be slightly worse than the results of the simplified approach.

If only the convolutive image of the environment noise is available, an overdetermined sensor case is required. Usually, one or two additional sensors are sufficient for successful extraction of all sources.

Summarising the separation ability of our approach, we conclude that:

Table 7. Quality factors of sound separation with noise cancellation in the direct reference noise case

Noise	*EI*	SNR, PSNR [dB]			
		Gong 1	*Music* 2	*Laughter* 3	*Chirp* 4
		Separation of a noise-free mixture			
none	0.00081	38.27, 60.79	28.85, 47.13	36.34, 57.45	48.18, 67.99
		Simple model			
large 1	0.0060	19.90, 42.43	27.11, 45,39	25.31, 46.42	25.22, 45.03
		Simultaneous model			
large 1	0.0095	18.22, 36.20	17.10, 25.65	16.24, 32.42	24.54, 42.96

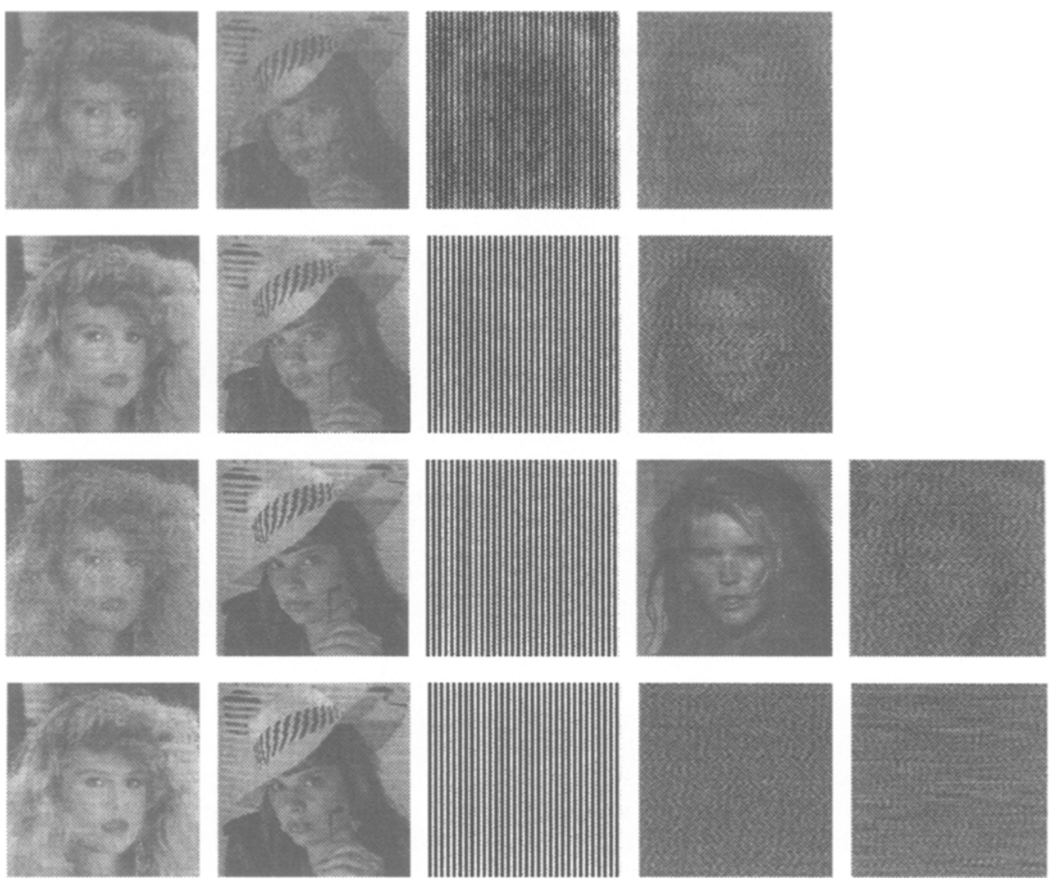
(a)
(b)
(c)
(d)

Fig. 12. Results of image separation with noise cancellation if convolutive reference noise $\boldsymbol{n}_R$ only is available. (a) For ill-conditioned mixing matrix $\boldsymbol{A}_1$, (b) for mixing matrix $\boldsymbol{A}_2$, (c) for mixing matrix $\boldsymbol{A}_3$ with one auxiliary sensor, (d) for mixing matrix $\boldsymbol{A}_4$ with two auxiliary sensors.

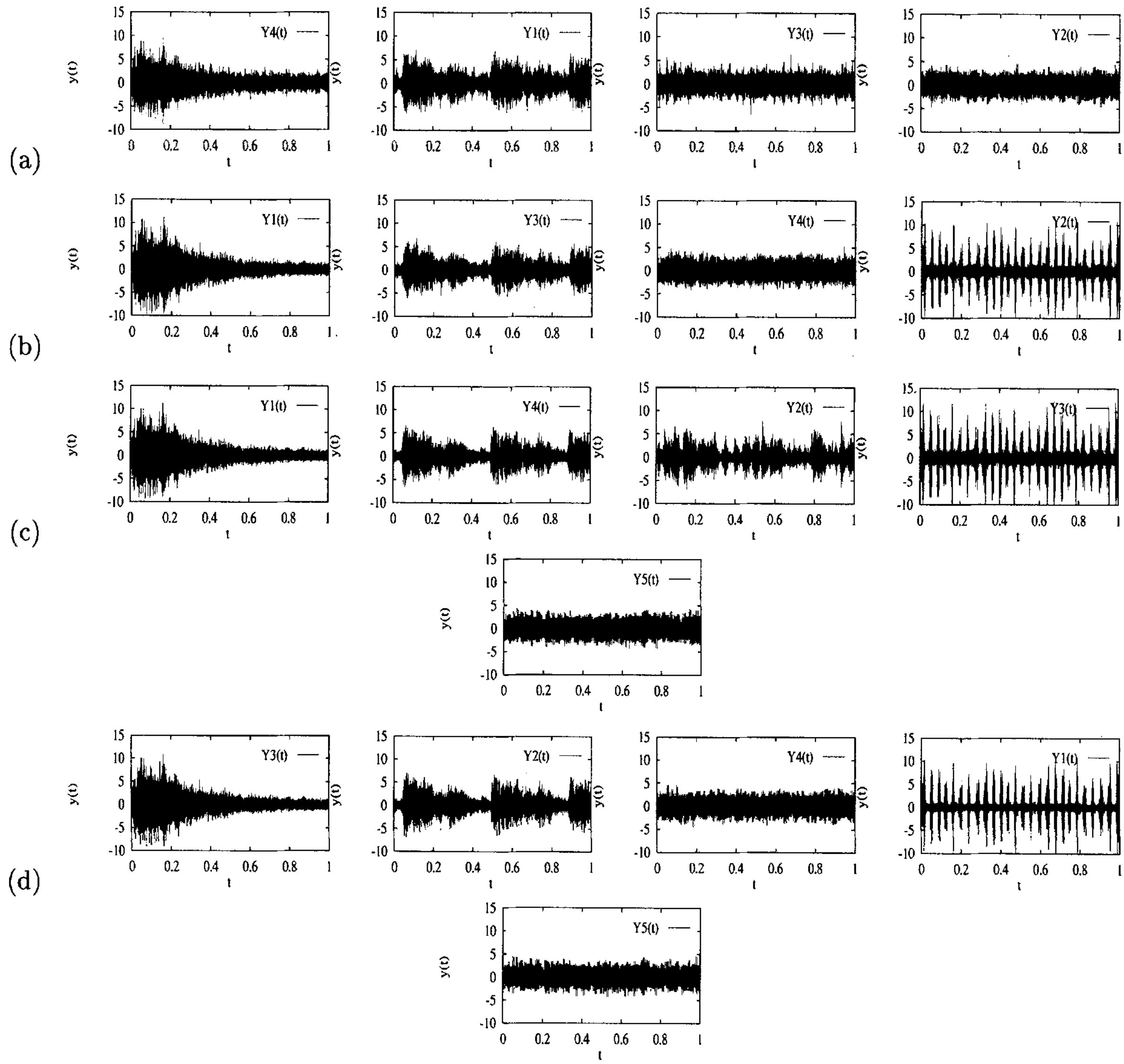

Fig. 13. Results of sound separation with noise cancellation if convolutive reference noise n_R only is available. (a) For ill-conditioned mixing matrix A_1, (b) for mixing matrix A_2, (c) for mixing matrix A_3 with one auxiliary sensor, (d) for mixing matrix A_4 with two auxiliary sensors.

1. Without additive noise our approach is able to separate signals with very ill-conditioned mixing matrices, with condition numbers even larger than 10^6.
2. With direct measurable environment noise, the mixing matrices can have moderate condition numbers of the order 200 or less.
3. With a convolutive image of environment noise, the mixing matrices may have quite low condition numbers, less than 50.

Another promising method is to reconstruct the environmental noise from its convolutive observation using the known techniques for single channel blind equalisation [14].

6. Conclusions

An adaptive approach for the restortion of unknown source signals from their signal mixtures distorted by additive noise was developed. The approach is valid under the assumption that the unknown coloured (or Gaussian) noise can be modelled as a convolutive noise mixture of known reference noises.

The approach was tested on image sources and sound sources, but it is generally applicable to various classes of non-Gaussian signals, and also to speech signals and biomedical signals. Computer experiments are very promising.

Table 8. Quality factors of sound separation with noise cancellation in the case of convolutive reference

Noise	*EI*	SNR, PSNR [dB]			
		Gong 1	*Music* 2	*Laughter* 3	*Chirp* 4
		For difficult mixture mixed by $\boldsymbol{A}_1$			
$\boldsymbol{B}, \boldsymbol{h}_R$	1.4029	4.77, 28.29	6.08, 24.37	–4.59, 16.52	–0.61, 19.20
		For mixture mixed by $\boldsymbol{A}_2$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.7118	9.86, 32.38	7.27, 25.56	–3.76, 17.35	10.84, 30.65
		For over-determined mixture mixed by $\boldsymbol{A}_3$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.1062	10.17, 32.69	14.22, 32.51	10.97, 32.08	11.07, 30.88
		For over-determined mixture mixed by $\boldsymbol{A}_4$			
$\boldsymbol{B}, \boldsymbol{h}_R$	0.0809	10.22, 32.75	14.49, 32.58	–,–	14.63, 34.44

The proposed noise model could be extended to IIR adaptive filters, gamma filters and other more sophisticated nonlinear models of noise, like NARMAX. However, the open problem is to optimally choose the nonlinear model of noise. We hope to solve these problems in the future by using nonlinear neural filters. An even more challenging task is how to proceed if no reference noise or no knowledge about noise statistics is available.

References

1. Amari S-I, Cichocki A, Yang HH. A new learning algorithm for blind signal separation. In: Neural Information Processing Symposium (NIPS'95), vol 8, MIT Press, Cambridge, MA, 1996; 757–763
2. Amari S-I, Cichocki A, Yang HH. Recurrent neural networks for blind separation of sources. In: Proc NOLTA'95, Tokyo, Japan, 1995, 37–42. NTA Research Society of IEICE
3. Cichocki A, Amari S, Adachi M, Kasprzaak W. Self-adaptive neural networks for blind separation of sources. In: 1996 IEEE Int Symposium on Circuits and Systems, (ISCAS'96), vol 2, 1996; 157–160. IEEE
4. Oja E, Karhunen J. Signal separation by nonlinear hebbian learning. Computational Intell 1995: 83–96
5. Cardoso JF, Belouchrani A, and Laheld B. A new composite criterion for adaptive and iterative blind source separation. In: Proc ICASSP-94, vol 4, 1994: 273–276
6. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 1995; 7: 1129–1159
7. Jutten CL. Herault J. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. Signal Process 1991; 24(1): 1–10
8. Cichocki A, Kasprzak W, Amari S. Multi-layer neural networks with a local adaptive learning rule for blind separation of source signals. In: Int Symposium on Nonlinear Theory and its Applications (NOLTA'95), Tokyo, Japan, 1995; 61–65. NTA Research Society of IEICE
9. Van Gerven S. Adaptive Noise Cancellation And Signal Separation with Applications to Speech Enhancement. Dissertation, Katholieke Universiteit Leuven, Departement Elektrotechnik, Departement Elektrotechnik, Leuven, Belgium, 1996
10. Van Gerven S, Van Compernolle D, Nguyen-Thi HL, Jutten C. Blind separation of sources: A comparative study of a 2-nd and a 4-th order solution. In: Signal Processing VII, Proc EUSIPCO-94, Lausanne, Switzerland, 1994: 1153–1156. EUSIP Association
11. Cichocki A, Kasprzak W, Amari S. Adaptive approach to blind source separation with cancellation of additive and convolutional noise. In: Third Int Conf Signal Processing (ICSP'96), vol I, China, October 1996: 412–415. IEEE Press/PHEI Beijing
12. Windrow B, Stearns SD. Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1985
13. Cichocki A, Unbehauen R. Neural Networks for Optimization and Signal Processing John Wiley & Sons, New York, 1994
14. Haykin S (ed.). Blind Deconvolution, Prentice-Hall, Englewood Cliffs, NJ, 1994

Nomenclature

Symbol	*Meaning*
κ_4	the normalised kurtosis of a signal
$\eta(t)$, $\tilde{\eta}(t)$	learning rates
m	number of sources
n	number of sensors
N,M	order of the FIR filters
$\boldsymbol{s}(t)$	m-dimensional vector of (unknown) source signals
$\boldsymbol{x}(t)$	n-dimensional vector of mixed signals (sensors)
$\boldsymbol{y}(t)$	n-dimensional vector of separated output signals (estimated sources)
$\nu_R(t)$	(unknown) primary environment noise signal
$n_R(t)$	secondary reference noise signal
$\boldsymbol{n}(t)$	n-dimensional vector of additive noise signals
$f(\cdot)$,$g(\cdot)$	activation functions in separation rule
$f_R(\cdot)$	activation function in noise cancellation rule
$\boldsymbol{A} = [a_{ij}]_{m\times n}$	(unknown) mixing matrix
$\boldsymbol{B} = [b_{ij}]_{n\times N}$	additive noise generation matrix
$\boldsymbol{H}(t) = [h_{ij}]_{n\times M}$	noise cancellation matrix
$\boldsymbol{W}(t) = [w_{ij}]_{n\times n}$	global de-mixing matrix