SEPARATION OF CONVOLUTIVE MIXTURES OF TEMPORALLY-WHITE SIGNALS: A NOVEL FREQUENCY-DOMAIN APPROACH

Adriana Dapena, Mónica F. Bugallo, Luis Castedo

Departamento de Electrónica e Sistemas
Universidade da Coruña
Campus de Elviña s/n, 15.071. A Coruña. SPAIN
Tel: ++34-981-167000, Fax: ++34-981-167160, E-mail: adriana@des.fi.udc.es

ABSTRACT

In this paper we propose a novel approach for separating convolutive mixtures in the frequency domain. This approach involves the solution of several instantaneous mixing problems and the elimination of the indeterminacies which appear because the sources may be extracted in a different order or with different amplitudes in some frequency bins. In order to separate each instantaneous mixture, we will extend the criterion proposed in [4]. We also show that both the permutation and the amplitude indeterminacies can be removed using second-order statistics when the sources are temporally-white.

Keywords:- Blind source separation, convolutive mixtures, blind deconvolution, frequency-domain approach.

1. INTRODUCTION

The separation of convolutive mixtures of statistically independent signals (sources) is a fundamental problem in signal processing that arises in a large number of applications. The problem is termed *blind* when the sources are recovered without resorting to an a priori knowledge of the sources or the mixing system [5, 8]. This problem can be solved in the frequency domain by interpreting a convolutive mixture as several instantaneous mixtures which may be separated using many existing algorithms (see [6] and references therein). However, since each problem is solved independently, the sources may be recovered in a different order (permutation indeterminacy) and with different amplitude (amplitude indeterminacy) in some frequency bins. Removing both indeterminacies is crucial because the sources in the time domain are obtained from the outputs in all the frequencies. During the last years, several criteria have been proposed to remove the permutation problem [2, 7, 8, 9] but few solutions to the amplitude indeterminacy have been presented [8].

In this paper we extend the algorithm proposed in [4] to the frequency domain. We prove that the attractors of the algorithm correspond to points where the perfect separation of the sources is achieved. In addition, the permutation indeterminacy is solved by clustering the outputs according to their cross-correlations. Then, the amplitude indeterminacy is corrected taking into account the values predicted by the stability analysis.

This paper is structured as follows. Section 2 presents the signal model. Section 3 introduces a compact representation of the observations in the frequency domain which will be used in the paper. In Section 4 we extend the separating algorithm proposed in [4]. In Section 5 we present a solution to the permutation and amplitude indeterminacies which exploits the temporally-white property of the sources. Section 6 presents several computer simulations which show the behavior of the proposed system. Finally, Section 7 is devoted to the conclusions.

2. SIGNAL MODEL

We will consider the following signal model. Let $\mathbf{s}(n) = [s_1(n), ..., s_N(n)]^T$ be the vector of N sources whose exact probability density functions are unknown. We assume that the sources are stationary, complex-valued, zero-mean, temporally-white, non-Gaussian distributed, statistically independent and have the same kurtosis sign. The sources propagate along an open medium and arrive at an array of M sensors. The output of the sensors, denoted by $\mathbf{x}(n) = [x_1(n), ..., x_M(n)]^T$, provides a convolutive combination of the N sources. The observations are typically related to the sources through the following expression

$$\mathbf{x}(n) = \sum_{k=-\infty}^{\infty} \mathbf{A}(k)\mathbf{s}(n-k) \qquad (1)$$

where $\mathbf{A}(k)$ is an unknown $M \times N$ matrix representing the mixing system. Without any loss of generality we can sup-

This work has been supported by FEDER funds (1FD97-0082) and Xunta de Galicia (PGIDT00PXI10504PR).

pose that all the sources have a normalized power equal to one.

It is interesting to note that in the frequency domain, the convolutive mixture (1) takes the form

$$\mathbf{x}[\omega] = \mathbf{A}[\omega]\mathbf{s}[\omega] \quad (2)$$

where $\mathbf{x}[\omega]$, $\mathbf{s}[\omega]$ and the matrix $\mathbf{A}[\omega]$ represent the observations, the sources and the mixing coefficients in the frequency domain, respectively. From (2), we can interpreted a convolutive mixture as several instantaneous mixtures. Therefore, we can recover the sources at each frequency by using a MIMO (Multi-Input Multi-Output) system with output

$$\mathbf{y}[\omega] = \mathbf{W}^H[\omega]\mathbf{x}[\omega] \quad (3)$$

where $\mathbf{W}[\omega]$ is the $M \times N$ coefficients matrix in the frequency ω. This matrix can be adapted using many existing algorithms (see [6] and references therein).

Combining both (2) and (3) together, we can express the outputs as follows

$$\mathbf{y}[\omega] = \mathbf{G}[\omega]\mathbf{x}[\omega] \quad (4)$$

where $\mathbf{G}[\omega] = \mathbf{W}^H[\omega]\mathbf{A}[\omega]$ is the overall mixing/separating matrix. Sources are optimally recovered when each output extracts a single and different source. This means that the optimum matrix $\mathbf{G}[\omega]$ has the form

$$\mathbf{G}[\omega] = \Delta[\omega]\mathbf{P}[\omega] \quad (5)$$

Note that if at each frequency the separating system is adapted independently, the sources can be recovered in a different order (permutation indeterminacy) and with different amplitudes (amplitude indeterminacy).

3. SHORT-TIME FOURIER TRANSFORM OF THE OBSERVATIONS

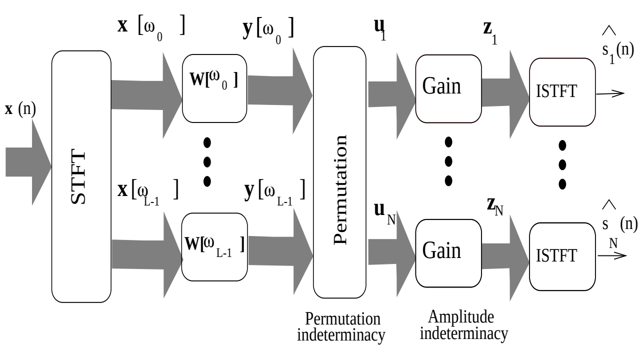

Fig. 1. Separating system

To recover the sources from convolutive mixtures we propose to use the system shown in Figure 1. The first stage consists of applying the Short-Time Discrete Fourier Transform (STFT) to moving windows of observations. We split each particular observation, $x_j(t)$, in R non-overlapped segments of K points, i.e., $\mathbf{x}_j(t_r) = [x_j(t_r), x_j(t_r+1), ..., x_j(t_r+K-1)]^T$ where $t_r = rK,\ r = 0\cdots, R-1$ denotes the window position. Subsequently, we compute the L-points DFT ($L \geq K$) of each window and obtain

$$x_j[\omega_k, t_r] = \sum_{m=0}^{K-1} x_j(t_r+m)e^{j\omega_k m} \quad k = 0, ..., L-1 \quad (6)$$

where $\omega_k = 2\pi k/L$ denotes the frequency bin. In a compact form, we can write the observation in the frequency domain as follows

$$x_j[\omega_k, t_r] = \mathbf{f}_k^T\, \mathbf{x}_j^e(t_r) \quad k = 0, ..., L-1 \quad (7)$$

where $\mathbf{f}_k = [1, ..., e^{j2\pi k(K-1)/L}, ..., e^{j2\pi k(L-1)/L}]^T$ is the Fourier transform vector and $\mathbf{x}_j^e(t_r) = [x_j(t_r), ..., x_j(t_r + K-1), 0, ..., 0]^T$ is an extended version of $\mathbf{x}_j(t_r)$ obtained by appending $L-K$ zeros.

In practice, we assume that the mixing system can be modeled as a FIR (Finite Impulse Response) filter of order F and, therefore, the mixing matrices $\mathbf{A}[\omega_k]$ are formed by coefficients

$$a_{ij}[\omega_k] = \mathbf{f}_k^T \mathbf{a}_i^e \quad (8)$$

where $\mathbf{f}_k = [1, ..., e^{j2\pi k(F-1)m/L}, ..., e^{j2\pi k(L-1)/L}]^T$ is the vector corresponding to the L-points DFT, $\mathbf{a}_i^e = [a_i(0), ..., a_i(F-1), 0, ..., 0]^T$ is an extended version of the i-th column of the mixing matrix $\mathbf{A}$. We also assume that the sources are transmitted in blocks of F samples and, therefore, the sources in the frequency domain are given by

$$s_j[\omega_k, t_r] = \mathbf{f}_k^T \mathbf{s}_j^e(t_r) \quad (9)$$

where $\mathbf{f}_k = [1, ..., e^{j2\pi k(F-1)m/L}, ..., e^{j2\pi k(L-1)/L}]^T$ and $\mathbf{s}_j^e(t_r) = [s_j(t_r), s_j(t_r+1), ..., s_j(t_r+F-1), 0, ..., 0]^T$. According to the properties of the DFT, we can express the relationship between observations and sources (2) as follows

$$\mathbf{x}[\omega_k, t_r] = \mathbf{A}[\omega_k]\mathbf{s}[\omega_k, t_r] \quad (10)$$

In order to simplify the notation, we will use $\mathbf{x}[k]$ instead of $\mathbf{x}[\omega_k, t_r]$, $\mathbf{A}[k]$ instead of $\mathbf{A}[\omega_k]$ and $\mathbf{s}[k]$ instead of $\mathbf{s}[\omega_k, t_r]$.

It is important to remark that the transformation of a convolutive mixture in several instantaneous mixtures by means of the STFT is strictly true only if we append zeros to the sources. This occurs in several applications such as in digital communications where transmissions occur in bursts and there exists a guard time between frames of data.

4. SEPARATING CRITERION

As explained in Section 3, we can interpret the convolutive mixture (1) as L instantaneous mixtures in the frequency domain given by (10). In order to solve each individual problem, the observations at each frequency bin are processed by a MIMO linear system with coefficients $\mathbf{W}[k],\ k=0,...,L-1$ and outputs $y_i[k],\ i=1,...,N,\ k=0,...,L-1$ (see Figure 1). At each frequency bin, we propose to select the matrix $\mathbf{W}[k]$ by maximizing the following cost function [4]

$$J(\mathbf{W}[k]) = \sum_{i=1}^{N} J_i(\mathbf{W}[k]) - \gamma \sum_{j=1}^{N}\sum_{\substack{i=1\\ i\neq j}}^{N} J_{ij}(\mathbf{W}[k]) \quad (11)$$

where the parameter γ is a real and positive constant,

$$\begin{aligned} J_i(\mathbf{W}[k]) &= |K_{y_i[k]y_i[k]}| - f(E[|y_i[k]|^2]) \\ &= |E[|y_i[k]|^4] - 2E^2[|y_i[k]|^2] - |E[y_i[k]^2]|^2| \\ &- \quad f(E[|y_i[k]|^2]) \end{aligned} \quad (12)$$

and

$$\begin{aligned} J_{ij}(\mathbf{W}[k]) &= |K_{y_i[k]y_j[k]}| \\ &= |E[|y_i[k]|^2|y_j[k]|^2] - E[|y_i[k]|^2]E[|y_j[k]|^2] \\ &- \quad |E[y_i[k]y_j^*[k]]|^2 - |E[y_i[k]y_j[k]]|^2| \end{aligned} \quad (13)$$

The term $K_{y_i[k]y_j[k]} = E[|y_i[k]|^2|y_j[k]|^2] - E[|y_i[k]|^2]E[|y_j[k]|^2] - |E[y_i[k]y_j^*[k]]|^2$ in (12) and (13) denotes the fourth-order cross-cumulant between $y_i[k]$ and $y_j[k]$. The main difference of the cost function (11) with respect to the criterion proposed in [4] is that now the function $f(x)$ in expression (12) is a derivable real-valued function chosen so that each of the following functions

$$p_i(x) = \frac{|K_{s_i[k]s_i[k]}|}{E[|s_i[k]|^2]}\frac{x^2}{F} - f(x), \quad i=1,...,N \quad (14)$$

has a single maximum for $x>0$. Expression (14) depends on the kurtosis and the second order moments of the sources in the frequency domain. However, since we are assuming that the sources are temporally-white, stationary and have power equal to one, we can write

$$\begin{aligned} K_{s_i[k]s_i[k]} = FK_{s_i(n)s_i(n)} = F\rho_i \\ E[|s_i[k]|^2] = F\ E[|s_i(n)|^2] = F \end{aligned} \quad (15)$$

where $\rho_i = K_{s_i(k)s_i(k)}/E[|s_i(k)|^2]$ is the normalized kurtosis of source $s_i(k)$. Substituting (15) in (14), we obtain that the function $f(x)$ must be chosen so that each of the following functions

$$p_i(x) = |\rho_i|\frac{x^2}{F} - f(x), \quad i=1,...,N \quad (16)$$

has a single maximum for $x>0$. For instance, consider the second order polynomial

$$f(x) = \alpha x^2 - 2\beta x \quad (17)$$

It is straightforward to check that each function $p_i(x) = (|\rho_i|/F - \alpha)x^2 + 2\beta x$ has a single maximum in

$$x_i = \frac{F\beta}{F\alpha - |\rho_i|}, \quad i=1,...,N \quad (18)$$

when $\alpha > \frac{|\rho_i|}{F},\ i=1,...,N$ and $\beta>0$. It is apparent that for the second-order polynomial (17) we only need to select a sufficiently large value of α.

4.1. Stability Analysis

In this subsection, we demonstrate that the only maxima of the proposed cost function correspond to the perfect separation of the sources. In addition, a stochastic algorithm

$$\mathbf{W}_t[k] = \mathbf{W}_{t-1}[k] + \mu\ \hat{\nabla}_{\mathbf{W}[k]} J(\mathbf{W}_{t-1}[k], \mathbf{x}_t[k]) \quad (19)$$

contains the same attractors and attraction domains as the deterministic equation provided that μ is sufficiently small, $\mathbf{x}(n)$ is stationary and $\nabla_{\mathbf{W}[k]} J(\mathbf{W}[k])$ is a regular and reasonably well-behaved function of its arguments [1]. Here, μ is the step-size parameter and $\hat{\nabla}_{\mathbf{W}[k]} J(\mathbf{W}_{t-1}[k], \mathbf{x}_t[k])$ is an estimation of the gradient of the cost function (11) depending on both the matrix coefficient at the instant $t-1$, $\mathbf{W}_{t-1}[k]$, and the observations at the instant t, $\mathbf{x}_t[k]$. As a consequence, the analysis guarantees the stability of a gradient algorithm that maximizes the cost function (11). Due to lack of space, in this paper we do not present the algorithm but it can be obtained like in [4].

First, we will write the cost function in terms of the global matrix $\mathbf{G}[k]$. Recall that the outputs of the MIMO system are linear combinations of the sources

$$y_i[k] = \sum_{l=1}^{N} g_{il}[k]s_l[k], \quad i=1,...,N \quad (20)$$

Considering (15) and (20), the statistical moments involved in $J(\mathbf{W}[k])$ can be expressed as

$$\begin{aligned} E[|y_i[k]|^2] &= \sum_{l=1}^{N}|g_{il}[k]|^2 E[|s_l[k]|^2] = F\sum_{l=1}^{N}|g_{il}[k]|^2 \\ K_{y_i[k]y_j[k]} &= \sum_{l=1}^{N}|g_{il}[k]|^2|g_{jl}[k]|^2 K_{s_l[k]s_l[k]} \\ &= F\sum_{l=1}^{N}|g_{il}[k]|^2|g_{jl}[k]|^2|\rho_l| \\ K_{y_i[k]y_i[k]} &= \sum_{l=1}^{N}|g_{il}[k]|^4 K_{s_l[k]s_l[k]} = F\sum_{l=1}^{N}|g_{il}[k]|^4|\rho_l| \end{aligned} \quad (21)$$

Substituting (21) in (11), the cost function $J(\mathbf{W}[k])$ becomes

$$\psi(\mathbf{G}[k]) = \sum_{i=1}^{N}\left(\sum_{l=1}^{N} F|\rho_l||g_{il}[k]|^4 - f(F||\mathbf{g}_i[k]||^2)\right) - \gamma F \sum_{\substack{i,j=1\\ i\neq j}}^{N}\sum_{l=1}^{N}|\rho_l||g_{il}[k]|^2|g_{jl}[k]|^2 \quad (22)$$

where $||\mathbf{g}_i[k]||^2 = \sum_{l=1}^{N}|g_{il}[k]|^2$. Comparing this cost function and the criterion presented in [4], we observe that the only difference is that in (22) the constant F appears. Thus, it is straightforward to extend the results in [4] to the new cost function.

Using the complex derivative operator, we obtain that the first derivatives of the cost function (22) are given by

$$\begin{aligned}\frac{\partial\psi}{\partial g_{im}[k]} &= 2F|\rho_m|g_{im}^*[k]|g_{im}[k]|^2\\ &- Fg_{im}^*[k]f'(F||\mathbf{g}_i[k]||^2)\\ &- 2F\gamma|\rho_m|g_{im}^*[k]\sum_{\substack{j=1\\ i\neq j}}^{N}|g_{jm}[k]|^2\\ & i,m=1,...,N \end{aligned} \quad (23)$$

Recall that at the separating point, the global matrix $\mathbf{G}[k]$ takes the form $\mathbf{G}[k] = \Delta[k]\mathbf{P}[k]$ and the first derivatives (23) can be written as follows

$$\begin{aligned}\frac{\partial\psi}{\partial\delta_{ii}[k]} &= F|\rho_i|\delta_{ii}^*[k]|\delta_{ii}[k]|^2 - F\delta_{ii}^*[k]f'(F|\delta_{ii}[k]|^2)\\ &= F\delta_{ii}^*[k]p_i'(F|\delta_{ii}[k]|^2) = 0,\ i=1,...,N\end{aligned}$$

where $p_i'(x)$ is the first derivative of $p_i(x)$. Since each function $p_i(x)$ contains a single maximum for $x > 0$, we conclude that the expression (24) vanishes when $x_i = F|\delta_{ii}[k]|^2$ is the maximum of $p_i(x)$. Using the same reasoning as in [4], it is easy to conclude that this point is a maximum of the cost function (22). In addition, from the analysis in [4], we conclude that the condition $\gamma > 1$ is sufficient to ensure that the other stationary points of the cost function (22) are minima or saddle points.

In summary, the analysis above guarantees that the only maxima of the proposed cost function correspond to the point where the sources are separated. The analysis also predicts that the magnitude of the entries of the matrix $\Delta[k]$ are $|\delta_{ii}[k]|^2 = x_i/F$ where x_i is the maximum of $p_i(x)$. For instance, from (18) we obtain that the cost function (22) has a maximum in

$$|\delta_{ii}[k]|^2 = \frac{x_i}{F} = \frac{\beta}{F\alpha - |\rho_i|}\ i=1,...,N \quad (24)$$

5. TIME-DOMAIN RECOVERING

The last stage in the separating system shown in Figure 1 consists of recovering the sources in the time domain. Both the permutation and the amplitude indeterminacy must be solved before applying the Inverse STFT (ISTFT) to the outputs.

5.1. Permutation indeterminacy

Assuming that the output $y_i[m]$ extracts the source $s_p[m]$ and $y_l[k]$ extracts the source $s_t[k]$, the cross-correlation between both outputs is given by

$$E[y_i[m]y_l^*[k]] = g_{ip}[m]\ g_{lt}^*[k]\ E[s_p[m]s_t^*[k]] \quad (25)$$

Using the equation (9) we obtain

$$E[y_i[m]y_l^*[k]] = g_{ip}[m]g_{lt}^*[k]\mathbf{f}_m^T E[\mathbf{s}_p^e(t_r)(\mathbf{s}_t^e(t_r))^H]\mathbf{f}_k^*$$

where $\mathbf{f}_k = [1,...,e^{j2\pi k(F-1)/L},...,e^{j2\pi k(L-1)/L}]^T$, $\mathbf{s}_j^e(t_r) = [s_j(t_r),...,s_j(t_r+F-1),0,...,0]^T$. Since, the sources are temporally white and stationary, we obtain

$$\begin{aligned}E[y_i[m]y_l^*[k]] &= g_{ip}[m]g_{lt}^*[k]E[s_p(n)s_t^*(n)]\hat{\mathbf{f}}_m^T\hat{\mathbf{f}}_k^*\\ &= g_{ip}[m]g_{lt}^*[k]E[s_p(n)s_t^*(n)]\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_m\end{aligned} \quad (26)$$

where $\hat{\mathbf{f}}_k = [1,...,e^{j2\pi km/L},...,e^{j2\pi k(F-1)/L}]^T$. Supposing $\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_m \neq 0$, we deduce that expression (26) will be non-zero when $y_i[m]$ and $y_l[k]$ extract the same source, i.e., $y_i[m] = g_{ip}[m]s_p[m]$ and $y_l[k] = g_{lp}[k]s_p[k]$.

From the above explanation, we can devise a method to avoid the permutation indeterminacy. We compute the cross-correlation between each output at the first frequency bin and the outputs in the other frequency bins, $E[y_i[0]y_l^*[k]],\ i,l = 1,...,N,\ k = 1,...,L-1$. For each output at the first frequency bin, we select the outputs in the other frequency bins with the maximum cross-correlation. Then, we cluster the outputs corresponding to the same source

$$\begin{aligned}\mathcal{U}_i &= \{u_i[0] = y_i[0], u_i[1] = \max_{y_l[1]}|E[y_i[0]y_l^*[1]]|,...,\\ & u_i[L-1] = \max_{y_l[L-1]}|E[y_i[0]y_l^*[L-1]]|\}\end{aligned} \quad (27)$$

This criterion can be used only when $\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_0 \neq 0,\ k = 1,...,L-1$. Since

$$\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_0 = \sum_{p=0}^{F-1}\left(e^{-j\frac{2\pi k}{L}}\right)^p = \frac{1-e^{-j\frac{2\pi kF}{L}}}{1-e^{-j\frac{2\pi k}{L}}} \quad (28)$$

this occurs when kF/L is not an integer number for $k = 1,...,L-1$.

Frequency bin	$\|E[y_1[0]y_1^*[k]]\|$	$\|E[y_1[0]y_2^*[k]]\|$
$k=1$	1.8278	0.1899
$k=2$	1.2352	0.1765
$k=3$	0.5439	0.1245
$k=4$	0.5240	0.1180
$k=5$	1.1886	0.1901
$k=6$	0.1508	1.8746

$|E[y_2[0]y_j^*[k]]|$

Frequency bin	$\|E[y_2[0]y_1^*[k]]\|$	$\|E[y_2[0]y_2^*[k]]\|$
$k=1$	0.3726	1.5053
$k=2$	0.2514	1.0344
$k=3$	0.0252	0.5457
$k=4$	0.0416	0.4156
$k=5$	0.2450	1.0475
$k=6$	1.7227	0.1263

Table 1. Cross-correlations $|E[y_i[0]y_j^*[k]]|$

$\delta_{11}[0]\delta_{11}^*[k]$

Frequency bin	Estimated value	True value
$k=1$	1.0143 + 0.0091i	0.9852 - 0.0178i
$k=2$	0.8229 + 0.5513i	0.8360 + 0.5895i
$k=3$	-0.3905 + 1.1581i	-0.3318 + 0.9851i
$k=4$	-0.4441 - 1.0904i	-0.3934 - 0.8873i
$k=5$	0.6445 - 0.7023i	0.6555 - 0.7496i
$k=6$	1.0114 - 0.2436i	0.9996 - 0.2057i

$\delta_{22}[0]\delta_{22}^*[k]$

Frequency bin	Estimated value	True value
$k=1$	0.7570 + 0.3533i	0.8282 + 0.3747i
$k=2$	0.3621 + 0.7464i	0.3254 + 0.8634i
$k=3$	-0.8355 + 0.8977i	-0.5356 + 0.7908i
$k=4$	-0.5343 - 0.7658i	-0.4528 - 0.8218i
$k=5$	0.6321 - 0.5532i	0.6332 - 0.6187i
$k=6$	0.9499 - 0.1080i	0.9250 - 0.0947i

Table 2. True and estimated value of $\delta_{ii}[0]\delta_{ii}^*[k]$

5.2. Amplitude indeterminacy

In order to solve the amplitude indeterminacy, we will force that all the outputs $u_i[k]$ into a set $\mathcal{U}_i$ have the same amplitude than $u_i[0]$. Let us consider $u_i[k] = \delta_{ii}[k]s_i[k]$, $i = 1,..,N$. We start with $z_i[0] = u_i[0] = \delta_{ii}[0]s_i[0]$ and, then, we compute the other outputs $z_i[k]$, $k = 1,...,L-1$ as follows

$$z_i[k] = \frac{\delta_{ii}[0]\delta_{ii}^*[k]}{|\delta_{ii}[k]|^2}u_i[k], \quad k = 1,...,L-1 \tag{29}$$

Substituting (29) in $u_i[k] = \delta_{ii}[k]s_i[k]$, it is easy to show that $z_i[k]$ has the same amplitude as $z_i[0]$

$$z_i[k] = \frac{\delta_{ii}[0]\delta_{ii}^*[k]\delta_{ii}[k]}{|\delta_{ii}[k]|^2}s_i[k] = \delta_{ii}[0]s_i[k] \tag{30}$$

Note that the term $|\delta_{ii}[k]|^2$ in (29) can be found using (24). Therefore, in order to evaluate (29), we only need to determine the term $\delta_{ii}[0]\delta_{ii}^*[k]$. Since the sources have unit power, we obtain

$$\begin{aligned} E[u_i[0]u_i^*[k]] &= \delta_{ii}[0]\delta_{ii}^*[k]\hat{\mathbf{f}}_0^T E[\mathbf{s}_i^e(t_r)(\mathbf{s}_i^e(t_r))^H]\hat{\mathbf{f}}_k^* \\ &= \delta_{ii}[0]\delta_{ii}^*[k]\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_0 \end{aligned} \tag{31}$$

As a consequence,

$$\delta_{ii}[0]\delta_{ii}^*[k] = \frac{E[u_i[0]u_i^*[k]]}{\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_0} \tag{32}$$

Finally, substituting (32) in (29), we deduce that the output $z_i[k]$ can be obtained as follows

$$z_i[k] = \frac{E[u_i[0]u_i^*[k]]}{\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_0|\delta_{ii}[k]|^2}u_i[k], \quad k = 1,...,L-1 \tag{33}$$

where $\hat{\mathbf{f}}_k = [1,...,e^{j2\pi k(F-1)/L}]^T$, $\hat{\mathbf{f}}_0 = [1,1,....,1]^T$, $|\delta_{ii}[k]|^2$ is the theoretical amplitude predicted by the stability analysis and $E[u_i[0]u_i^*[k]]$ is the cross-correlation between two outputs into the set $\mathcal{U}_i$.

5.3. Inverse Short-Time Fourier Transform

Finally, the ISTFT is applied to the outputs $z_i[k,t_r]$ for recovering the sources

$$\hat{s}_i(t_r+m) = \frac{1}{L}\sum_{k=0}^{L-1} z_j[k,t_r]e^{-j\frac{2\pi km}{L}}, \quad m = 0,...,F-1$$

where $t_r = rF$ with $r = 1,...,R$ is the window position.

6. SIMULATION RESULTS

We have generated 10,000 samples of a 4-QAM and a 16-QAM which have been partitioned in blocks of $F = 5$ samples. We have appended $P-1 = 2$ zeros to each block. The mixing system has been obtained by truncating to $F = 5$ the following transfer matrix:

$$\mathbf{H}(z) = \begin{bmatrix} \frac{0.6+z^{-1}}{1+0.6z^{-1}} & -0.8\frac{0.6+z^{-1}}{1+0.6z^{-1}} \\ 0.8\frac{0.5+z^{-1}}{1+0.5z^{-1}} & \frac{0.5+z^{-1}}{1+0.5z^{-1}} \end{bmatrix} \tag{34}$$

To recover the sources from the observations we have used $L = F + P - 1 = 7$ frequency bins. The coefficients of the separating systems have been computed using the proposed algorithm with the non-linear function (17), $\alpha = 1$, $\beta = 4$, $\gamma = 1.5$ and step-size parameter $\mu = 0.01$.

Note that the permutation and the amplitude indeterminacies can be removed using the proposed methods because kF/L, $k = 1,...,6$ is not an integer number. Table 1 shows the absolute value of the cross-correlations

$E[y_i[0]y_j^*[m]], i, j = 1, 2,\ m = 1, ..., 6$, which have been computed using $2,000$ samples of the outputs. It is interesting to note that the cross-correlation obtained in the frequency bins $k = 3$ and $k = 4$ are small in comparison with the other frequency bins. The reason of this disparity is that the value of $\mathbf{f}_k^H \mathbf{f}_0 = 0.4450\ k = 3, 4$ is small. From this table, we have clustered the outputs in two sets

$$\mathcal{U}_1 = \{u_1[0] = y_1[0], u_1[1] = y_1[1], u_1[2] = y_1[2], u_1[3] = y_1[3], u_1[4] = y_1[4], u_1[5] = y_1[5], u_1[6] = y_2[6]\}$$

and

$$\mathcal{U}_2 = \{u_2[0] = y_2[0], u_2[1] = y_2[1], u_2[2] = y_2[2], u_2[3] = y_2[3], u_2[4] = y_2[4], u_1[5] = y_2[5], u_1[6] = y_1[6]\}$$

In the next step we have solved the amplitude indeterminacy by evaluating the expression (32) using the theoretical amplitudes obtained from (24), $|\delta_{11}[k]|^2 = 1$ and $|\delta_{22}[k]|^2 = 0.9259$, and the estimation of $\delta_{ii}[0]\delta_{ii}^*[k]$ shown in Table 2. In order to validate our method, Table 2 also shows the true values obtained from the final separating matrix and the mixing matrix (34). We can see the similarity between the estimated and the true values. Finally, Figure 2 shows the recovered sources in the time domain (part (a)), the sources obtained when only the permutation problem is solved (part (b)) and the sources obtained without solving the amplitude nor the permutation problem (part (c)). It is apparent that both indeterminacies must be removed to recover the sources.

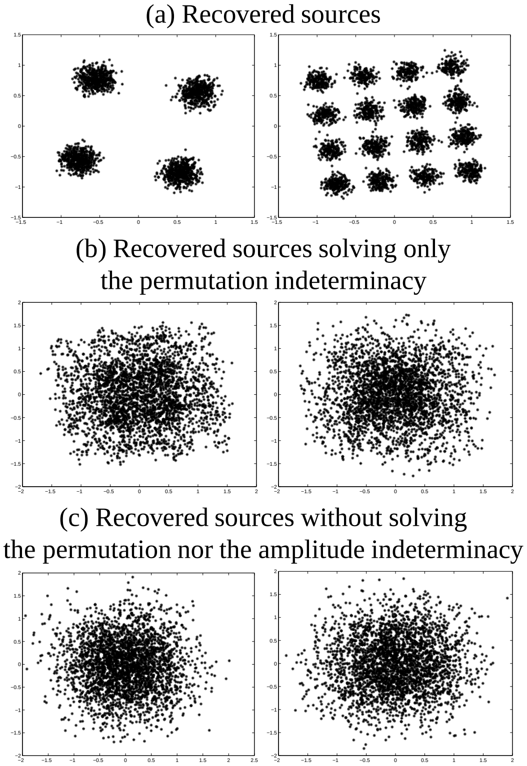

Fig. 2. Recovered sources

7. CONCLUSIONS

We have proposed a new approach for separating temporally-white signals in the frequency domain. The sources at each frequency bin are recovered using a gradient algorithm which maximizes a cost function obtained by extending the criterion in [4]. We have demonstrated that the attractors of the gradient algorithm correspond to the desired solution. The permutation problem has been solved by clustering the outputs according to their cross-correlation. Finally, the amplitude indeterminacy has been corrected taking into account the values predicted by the stability analysis.

8. REFERENCES

[1] A. Benveniste, M. Métivier and P. Priouret, Adaptive Algorithms and Stochastic Approximations, Applications Mathematics, Springer-Verlag, 1990.

[2] V. Capdevielle and C. Serviere and J.L. Lacoume, “Blind Separation of Wide-band Sources in the Frequency Domain”, in *Proceedings of ICASSP 1995*, pp. 2080-2083, 1995

[3] J. F. Cardoso, “Blind Signal Separation: Statistical Principles”, *Proceedings of IEEE*, vol. 86, no. 10, pp. 2009-2025, Oct. 1998

[4] A. Dapena and L. Castedo, “Stochastic Gradient Adaptive Algorithms for Blind Source Separation”, *Signal Processing*, No. 75, pp. 11–27, 1999

[5] G. Gelle, M. Colas and C. Serviere, “BSS for Fault Detection and Machine Monitoring Time or Frequency Domain Approach?”, in *Proceedings of ICA’2000*, pp. 555-560, Helsinki, June 2000

[6] Lee T-W, *Independent Component Analysis*, Kluwer Academic Publishers, 1998.

[7] C. Mejuto, A. Dapena and L. Castedo, “Frequency-domain Infomax for Blind Separation of Convolutive Mixtures”, in *Proceedings of ICA’2000*, pp. 315-320, Helsinki, June 2000.

[8] N. Murata, S. Ikeda, A. Ziehe, “An Approach to Blind Source Separation Based on Temporal Structure of Speech Signals” , to appear in *Neurocomputing*.

[9] P. Smaragdis, “Blind Separation of Convolved Mixtures in the Frequency Domain”, in *Proceeding of IWANN’98*, Spain, February 1998