

SEPARATION OF A CLASS OF CONVOLUTIVE MIXTURES A CONTRAST FUNCTION APPROACH

C. Simon, Ph. Loubaton, C.Vignat

Laboratoire Systèmes de Communication - UMLV
Champs sur Marne - 5, bvd Descartes
77 454 Marne la Vallee Cedex
FRANCE
simonc,loubaton,vignat@univ-mlv.fr

C. Jutten

LIS/INPG
44, rue Viallet
38 031 Grenoble Cedex
FRANCE
chris@tirf.inpg.fr

G. d'Urso

EDF/DER
6 Quai Watier
78 401 Chatou Cedex
FRANCE

ABSTRACT

In this paper, we address the problem of the separation of convolutive mixtures in the case where the non Gaussian source signals are not necessarily filtered versions of i.i.d. sequences. In this context, we show that the contrast functions, used in the linear process source case, still allow to separate the sources by a deflation approach. Some particular properties of higher order cumulants based contrast functions are also given.

1. INTRODUCTION

The field of blind source separation has raised growing interest in the last decade. In this context, M non observable, mutually independent non Gaussian source signals $s(n) = [s_1(n), \dots, s_M(n)]$ are mixed by an *unknown* $N \times M$ ($N \geq M$) linear and time invariant filter $H(z) = \sum_{k \in \mathbf{Z}} H(k)z^{-k}$ to give a N -variate observed signal

$$y(n) = \sum_{k \in \mathbf{Z}} H(k)s(n-k) = [H(z)]s(n) \quad (1)$$

The components of $y(n)$ can be, for example, the signals received on an array of N sensors. The blind source separation problem consists in recovering, from the observation $y(n)$, the contribution of each source on each sensor, i.e. signals $c_{k,l}(n) = [H_{k,l}(z)]s_l(n)$ for $k = 1, \dots, N$ and $l = 1, \dots, M$.

This problem was mainly investigated under the hypothesis that the source signals are non Gaussian, independent and identically distributed (i.i.d.) sequences. In this case, it is possible to recover each source signal $s_l(n)$, up to a delay and a scaling factor by minimizing a contrast like function (see e.g. [6, 9, 2, 7]). In order to retrieve the signals $c_{k,l}(n)$ for each k and l , [10] proposed to use the following scheme : having an estimate $\hat{s}_l(n)$ of source signal $s_l(n)$, signal $c_{k,l}(n) = [H_{k,l}(z)]s_l(n)$ is estimated as the filtered version $[\hat{H}_{k,l}(z)]\hat{s}_l(n)$ of $\hat{s}_l(n)$ for which $E|y_k(n) - [\hat{H}_{k,l}(z)]\hat{s}_l(n)|^2$ is minimum w.r.t. $\hat{H}_{k,l}(z)$.

As remarked in [10], these approaches can be extended to the case where the source signals s_l are non Gaussian *linear processes*, i.e. filtered versions of i.i.d. sequences w_l . In effect, the above procedures allow to retrieve estimates \hat{w}_l of w_l ; each estimate of

$c_{k,l}(n) = [H_{k,l}(z)]s_l(n)$ is then built by seeking the filtered version of \hat{w}_l which is closest to $y_k(n)$ in the mean square sense.

However, the hypothesis that each source signal is a linear process is rather restrictive. We note indeed that most of stationary signals can be represented as the output of a filter driven by a *decorrelated*, but not necessarily i.i.d. sequence. For example, in the context of digital communications, the output of an error correcting encoder driven by an i.i.d. sequence is often a decorrelated but not an i.i.d. sequence. Our purpose is to deal with the source separation problem in the case where the source signals are non Gaussian signals which do not necessarily coincide with filtered versions of i.i.d. sequences. Some previous works have addressed this case by using various techniques (e.g. [11, 8, 3]). The purpose of this study is to show that the approaches based on the optimization of a contrast function do work when the deflation approach (introduced in the linear process case in [6] and developed in [9] where it is called iterative) is used. Our results thus legitimate the use of contrast functions for separation of convolutive mixtures in fairly general cases.

This paper is organized as follows: in section 2, we adapt the classical definition (see [5]) of a contrast function to our purpose and provide some important examples. In section 3, we explain how the maximization of a contrast function allows to solve iteratively the blind source separation problem. In section 4, we study more specifically the contrast function defined as the squared kurtosis; we show that, as in the case of linear process source signals ([6, 9]), this contrast function does not show spurious local maximum. Finally, some simulation examples are given in section 5.

2. CONTRAST FUNCTIONS: DEFINITIONS AND EXAMPLES

Let \mathcal{X} be a set of non Gaussian scalar random variables.

Definition 1 A contrast C on \mathcal{X} is a mapping from \mathcal{X} to \mathbb{R}^+ satisfying the following assumptions :

- (i) For each $x \in \mathcal{X}$, $C(x)$ depends only on the probability distribution of x .
- (ii) Let $(x_i)_{i=1,m}$ (eventually $m = \infty$) be m independent (but not necessarily identically distributed) random variables of

This work has been supported by EDF/DER

\mathcal{X} and let $(a_i)_{i=1,m}$ be m coefficients such that $\sum_{i=1}^m a_i x_i$ belongs to \mathcal{X} . Then,

$$C\left(\sum_{i=1}^m a_i x_i\right) \leq \max_{i=1,m} C(x_i) \quad (2)$$

(iii) Equality holds in (2) if and only if $a_i = a_{i_0} \delta(i - i_0)$ where i_0 denotes one of the indices for which $\max_{i=1,m} C(x_i) = C(x_{i_0})$.

We note that this definition differs from that of Donoho [5] only in that we do not require the sequence $(x_i)_{i=1,m}$ to be identically distributed. Consequently, it is not surprising that most contrast functions used in the one-input / one-output blind deconvolution context still satisfy the requirements of definition 1. For example, it can be easily proved (see [5, 1]) that the opposite of the Shannon entropy, defined by $-S(x) = E[\log p_x(u)] = \int p_x(u) \log p_x(u) du$, is a contrast function on the set \mathcal{X}_1 of all unit variance random variables. Another example, motivated by ref. [7], is the following : let ϕ be a convex increasing function admitting a unique global minimum on \mathbb{R}^+ , denote by $\text{cum}_q(x)$ the order q ($q > 2$) cumulant of x and by \mathcal{X}_{10} the subset of all $x \in \mathcal{X}_1$ for which $\text{cum}_q(x) \neq 0$. Then, $C_\phi(x) = \phi(\text{cum}_q(x))$ is a contrast function on the set \mathcal{X}_{10} .

3. APPLICATION TO THE DEFLATION APPROACH FOR SOURCE SEPARATION

We first recall that the principle of the deflation approach consists in extracting the sources one by one. More precisely, a $1 \times N$ filter $g(z)$ is first computed in such a way that the scalar signal $r(n) = [g(z)]y(n)$ coincides with a filtered version of one of the source signals, say $s_1(n)$. The following step consists in finding a $N \times 1$ filter $t(z)$ for which the variance of the signal $\tilde{y}(n) = y(n) - [t(z)]r(n)$ is minimum. As signal $r(n)$ is supposed to be independent on signals $s_l(n)$ for $l \neq 1$, it is clear that each component $[t_k(z)]r(n)$ of $[t(z)]r(n)$ coincides with the contribution $[H_{k,1}(z)]s_1(n)$ of $s_1(n)$ on the k^{th} sensor. Therefore, $\tilde{y}(n)$ is a convolutive mixture of signals $s_l(n)$ for $l \neq 1$. The same procedure can be applied to signal $\tilde{y}(n)$ to extract the contribution of a second source signal on each sensor. All other sources are extracted by iterating this scheme.

[6, 9, 10] showed that, if the source signals are linear processes, the maximization of a normalized cumulant of $r(n) = [g(z)]y(n)$ allows to extract a particular filtered version of one of the source signals: the source generating i.i.d. sequence. The above iterative approach can thus be used to solve the blind source separation problem in the linear process source signal case.

We now extend these results to the case where the source signals are not filtered versions of i.i.d. sequences. As we seek a filtered version of the sources, we can assume, without any restriction, that each $s_l(n)$ is a unit variance, centered and decorrelated (but not i.i.d.) sequence. If $f(z) = [f_1(z), \dots, f_M(z)] = \sum_{n \in \mathbb{Z}} f(n)z^{-n}$ is a $1 \times M$ filter, let us denote $\|f\| =$

$[\sum_{n \in \mathbb{Z}} f(n)f(n)^*]^{\frac{1}{2}}$ its \mathcal{L}^2 norm. Let C be a contrast function defined on the set \mathcal{X}_1 of all unit variance random variables. Then, we have the following result:

Theorem 1 Assume that, for each, $k = 1, \dots, M$,

$$\sup_{\|f_k\|=1} C([f_k(z)]s_k(n)) < +\infty \quad (3)$$

and that there exists at least one function $f_{k,*}$ in the unit sphere of \mathcal{L}^2 for which¹

$$\sup_{\|f_k\|=1} C([f_k(z)]s_k(n)) = C([f_{k,*}(z)]s_k(n)) \quad (4)$$

Be $g(z)$ any $1 \times N$ filter for which $E([g(z)]y(n))^2 < +\infty$, denote $r(n) = [g(z)]y(n)$ and consider the following function:

$$\Xi : g(z) \mapsto C\left[\frac{r(n)}{\sqrt{E(r^2(n))}}\right] \quad (5)$$

Then, the global maximum of Ξ is finite and is reached by at least one filter, say $g_*(z)$. More importantly, the corresponding scalar signal $r_*(n) = [g_*(z)]y(n)$ is a filtered version of one of the source signals.

Proof. We first remark that $r(n)$ can be written as $r(n) = [f(z)]s(n)$, where $f(z)$ is the $1 \times M$ filter defined by $f(z) = g(z)H(z)$. As the source signals $s_k(n)$ are assumed to be unit variance and decorrelated, the spectral density matrix of the M -variate signal $s(n)$ is reduced to I_M (i.e. the $M \times M$ identity matrix). Hence, $E[r^2(n)] = \|f\|^2$. The normalized random variable $r(n)/\sqrt{E(r^2(n))}$ thus equals $[f(z)]/\|f\|s(n)$. Therefore, maximizing $C[r(n)/\sqrt{E(r^2(n))}]$ with respect to $g(z)$ is equivalent to maximizing $C([f(z)]s(n))$ over the set of $1 \times M$ unit norm filters $f(z)$.

Let $f(z) = (f_1(z), \dots, f_M(z))$, $\|f\| = 1$, be a unit norm $1 \times M$ filter. Then²

$$[f(z)]s(n) = \sum_{k=1}^M \|f_k\| \left[\frac{f_k(z)}{\|f_k\|} \right] s_k(n)$$

Put $r_k(n) = \left[\frac{f_k(z)}{\|f_k\|} \right] s_k(n)$. Then, it is clear that the $(r_k(n))_{1 \leq k \leq M}$ are independent unit variance random variables. Therefore, by item (ii) of Definition 1,

$$C([f(z)]s(n)) \leq \max_{1 \leq k \leq M} C\left(\frac{f_k(z)}{\|f_k\|} s_k(n)\right) \quad (6)$$

Let $\lambda_{k,*} < +\infty$ be defined as $\lambda_{k,*} = \sup_{\|f_k\|=1} C([f_k(z)]s_k(n))$ and denote by $f_{k,*}$ one argument of this maximization problem. Be k_0 such that $\lambda_{k_0,*} = \max_{1 \leq k \leq M} \lambda_{k,*}$. Then, we claim that

$$\sup_{\|f\|=1} C([f(z)]s(n)) = \lambda_{k_0,*} \quad (7)$$

To show this, we first remark that, for each unit norm $1 \times M$ filter $f(z)$,

$$C([f(z)]s(n)) \leq \max_{1 \leq k \leq M} C\left(\frac{f_k(z)}{\|f_k\|} s_k(n)\right) \leq \lambda_{k_0,*}$$

Let $f_*(z)$ be the $1 \times M$ unit norm filter defined by $f_*(z) = (0, \dots, 0, f_{k_0,*}(z), 0, \dots, 0)$. Then, $C([f_*(z)]s(n)) = \lambda_{k_0,*}$. So, (7) holds.

¹As \mathcal{L}^2 is an infinite dimensional set, this hypothesis is not necessarily fulfilled, even for a regular function C , simply because the unit sphere of \mathcal{L}^2 is not compact. However, this assumption seems reasonable.

²the random variable $\left[\frac{f_k(z)}{\|f_k\|} \right] s_k(n)$ should be considered null if $f_k(z) = 0$

In order to complete the proof of the theorem, it remains to establish that if $f'_*(z)$ is another $1 \times M$ unit norm filter that verifies

$$C([f'_*(z)]s(n)) = \lambda_{k_0,*} \quad (8)$$

then all the components of $f'_*(z)$ but one are zero. For this, we remark that

$$C([f'_*(z)]s(n)) \leq \max_{1 \leq k \leq M} C\left(\left[\frac{f'_{k,*}(z)}{\|f'_{k,*}\|}\right]s_k(n)\right) \leq \lambda_{k_0,*}$$

Therefore, identity (8) implies that

$$C([f'_*(z)]s(n)) = \max_{1 \leq k \leq M} C\left(\left[\frac{f'_{k,*}(z)}{\|f'_{k,*}\|}\right]s_k(n)\right)$$

As $[f'_*(z)]s(n) = \sum_{1 \leq k \leq M} \|f'_{k,*}\| \left[\frac{f'_{k,*}(z)}{\|f'_{k,*}\|}\right]s_k(n)$, this holds if and only if all coefficients $(\|f'_{k,*}\|)_{k=1,M}$ but one are zero (see item (iii) of Definition 1), i.e. iff the signal $[f'_*(z)]s(n)$ is a filtered version of a single source signal ■

This shows that, under the assumptions (3) and (4), it is possible to extract filtered versions of the source signals by the deflation approach introduced in [6] and [9]. We feel that these ideas cannot be generalized to non iterative source separation approach based on a contrast function (such as [2]). Due to the lack of space, we do not discuss this point here. Finally, we note that our results can be extended straightforwardly to the complex valued signals case. In particular, the numerical results presented in section 5 are obtained in the complex case.

4. PROPERTIES OF THE CONTRAST FUNCTION

$$CUM_4^2(X)$$

If the source signals are linear processes, the contrast functions defined from the normalized cumulants are attractive because they are free of spurious local maxima [9]. In this section, we show that this property remains valid in our more general context. Without restriction, we work on the set \mathcal{X}_{10} of all unit variance random variables having non zero fourth-order cumulants. Hence, the square of the normalized fourth-order cumulant coincides with function $C_2(x) = (\text{cum}_4(x))^2$. This contrast is relevant in our context if, for each k , $\lambda_{k,*}^{(2)} = \sup_{\|f_k\|=1} C_2([f_k(z)]s_k(n))$ is non zero. This extra assumption is assumed from now on³. We now state the following result:

Theorem 2 *Be $g(z)$ any $1 \times N$ filter for which $E(\|[g(z)]y(n)\|^2) < +\infty$, denote $r(n) = [g(z)]y(n)$ and consider the following function:*

$$\Xi : g(z) \mapsto C_2\left[\frac{r(n)}{\sqrt{E(r^2(n))}}\right] \quad (9)$$

Let $g_(z)$ be a local maximum of (9). Then, the signal $r_*(n) = [g_*(z)]y(n)$ is a filtered version of one of the source signals.*

Proof As in the proof of theorem 1, we set $f(z) = g(z)H(z)$, and remark that the study of the local maxima of (9) is equivalent to the study of the local maxima of the restriction to the unit sphere $\|f\| = 1$ of the function $\Psi(f) = C_2([f(z)]s(n))$. In order to study the maxima of Ψ on the unit sphere, we first give the following lemma which is an immediate generalization of a result of [3].

³our procedure is still relevant if one of the $\lambda_{k,*}^{(2)}$ is zero

Lemma 1 *Let $(r_k)_{1 \leq k \leq M}$ be M independent unit variance random variables among which at least one has a non zero fourth order cumulant. Let us consider the function Φ defined on the unit sphere of \mathbb{R}^M by $\Phi(a) = C_2(\sum_{1 \leq k \leq M} a_k r_k)$ where $a = (a_1, \dots, a_M)$. Then, the local maxima of Φ are the vectors a_* given by $a_{k,*} = \pm \delta(k - k_0)$, where k_0 is any index for which $c_4(r_{k_0}) \neq 0$.*

We now prove Theorem 2. Let $f_*(z)$ be a local maximum of Ψ restricted to the unit sphere. Choose $r_{k,*}(n) = \left[\frac{f_{k,*}(z)}{\|f_{k,*}\|}\right]s_k(n)$ and $a_* = (a_{1,*}, \dots, a_{M,*}) = (\|f_{1,*}\|, \dots, \|f_{M,*}\|)$ (recall that $r_{k,*}(n) = 0$ if $f_{k,*}(z) = 0$). We first note that the fourth order cumulant of the variable $r_{k,*}(n)$ cannot be zero for all k , otherwise, $\Psi(f_*)$, which is a local maximum, would be equal to 0. Let Φ_* be the function defined on the unit sphere of \mathbb{R}^M by

$$\Phi_*(a) = C_2\left(\sum_{1 \leq k \leq M} a_k r_{k,*}(n)\right)$$

As $f_*(z)$ is a local maximum of Ψ , the vector a_* is a local maximum of Φ_* . By lemma 1, $a_{k,*} = \pm \delta(k - k_0)$, i.e. all the components of $f_*(z)$ but one are identically zero. ■

The reader may check that these results still hold if $C_2(x) = (\text{cum}_4(x))^2$ is replaced with $\phi(\text{cum}_4(x))$, where ϕ is a convex increasing function admitting a unique minimum at 0. The results of this section also easily extend to the complex case.

5. IMPLEMENTATION AND SIMULATIONS

In this section, we illustrate our theoretical results on simple examples. We set $M = N = 2$. The first source signal $s_1(n)$ is generated as the centered version of the output of a binary convolutive encoder driven by an i.i.d. sequence. This signal is a decorrelated, but not an i.i.d. sequence. The second source signal $s_2(n)$ is obtained as the centered version of the output of an instantaneous non linear device (an exponential device) driven by a complex valued linear process. The two source signals are mixed by a complex, degree 1, 2×2 non minimum phase filter $H(z)$ to produce the observation $y(n) = [H(z)]s(n)$, where $s(n) = [s_1(n), s_2(n)]^T$.

The source separation procedure is based on the maximization of the contrast

$$\Xi(g) = \left| \frac{\text{cum}_4(r(n))}{(E(r(n)^2))^2} \right|^2$$

where $r(n) = [g(z)]y(n)$. We note that in the particular case $M = N = 2$ considered here, the first step of the deflation procedure allows in principle to separate the two sources : the maximization of $\Xi(g)$ allows to produce a filtered version $r(n)$ of one source signal (say the first source signal $s_1(n)$). Next, the criterion $E(\|y(n) - [t(z)]r(n)\|^2)$ is minimized over the set of all 2×1 filter. The 2-dimensional signal $[t(z)]r(n)$ clearly coincides with the contribution of source 1 on the sensor array, while the error $y(n) - [t(z)]r(n)$ represents the contribution of the second source on the sensor array.

Let us give some details concerning the practical implementation of the procedure. In practice, the minimization of $\Xi(g)$ is performed over the set of all non causal filter $g(z) = \sum_{k=-L}^L g_k z^{-k}$ of fixed degree L , assumed to be large enough. The same trick is used to minimize $E(\|y(n) - [t(z)]r(n)\|^2)$. We denote by \mathbf{g} the vector $\mathbf{g} = (g_{-L}, \dots, g_0, \dots, g_L)^T$ associated with filter $g(z)$. For each filter $g(z)$, the contrast function $\Xi(g)$ is estimated by its

natural empirical estimate $\hat{\Xi}(g)$ obtained by replacing $E|r(n)|^2$ by $\mathbf{g}^T \hat{R} \bar{\mathbf{g}}$ and $cum_4(r(n))$ by

$$1/P \sum_{n=0}^{P-1} |g^T Y(n)|^4 - 2 \left(\mathbf{g}^T \hat{R} \bar{\mathbf{g}} \right)^2 - |\mathbf{g}^T \hat{S} \mathbf{g}|^2$$

Here, we have set $Y(n) = (y(n+L), \dots, y(n-L))^T$, $\hat{R} = 1/P \sum_{n=0}^{P-1} Y(n)Y(n)^*$ and $\hat{S} = 1/P \sum_{n=0}^{P-1} Y(n)Y(n)^T$. Finally, $\bar{\mathbf{g}}$ stands for the complex conjugate of \mathbf{g} , while P represents the sample size.

We note that $\hat{\Xi}(g)$ is scale invariant. Therefore, a certain normalization has to be used. Here, we minimize $\hat{\Xi}(g)$ under the constraint that $\mathbf{g}^T \hat{R} \bar{\mathbf{g}} = 1$. For this, we use a standard iterative gradient procedure followed at each step by a renormalization. The gradient algorithm is initialized at the constant filter $g(z) = (1, 1)$.

We finally present some numerical evaluations. P is equal to 10 000, and the parameter L is fixed to 10. In order to evaluate the performance of the source separation procedure, we first plot in the following figures 300 samples of contributions on the second sensor (Figure1) and of their reconstructions (Figure2).

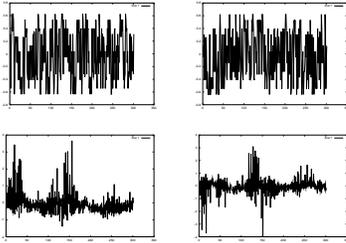


Figure 1: contributions on 2^{nd} sensor (real parts on LHS, imaginary ones on RHS)

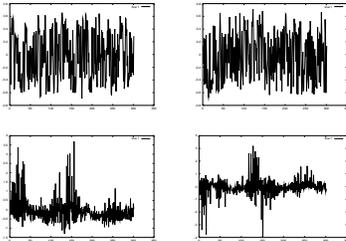


Figure 2: reconstructions on 2^{nd} sensor (real parts on LHS, imaginary ones on RHS)

We also evaluate the so-called *separation rate*: denote by $\hat{c}_{k,l}(n)$ the estimated contribution of source l on sensor k . Then, we define the separation rate of the l^{th} contribution on the k^{th} sensor by:

$$SEPR_{k,l} = \frac{\sum_n |\hat{c}_{k,l}(n) - c_{k,l}(n)|^2}{\sum_n |c_{k,l}(n)|^2}$$

In the case of our present example, we have $SEPR_{2,1} = 0.00355$ and $SPEPR_{2,2} = 0.00252$. In order to measure the “degree of improvement” of our method, we note that, before the source separation procedure, $SEPR_{bef_{2,1}} = 3.3228$ and $SEPR_{bef_{2,2}} = 0.28431$ where $SEPR_{bef_{k,l}} = \frac{\sum_n |y_k(n) - c_{kl}(n)|^2}{\sum_n |c_{kl}(n)|^2}$.

6. CONCLUSION

In this paper, we have addressed the source separation problem in the case where the source signals are not necessarily (non Gaussian) linear processes. We have shown that if a deflation approach is used, it is possible to solve this problem by maximizing a standard contrast function. Moreover, we have proved that as in the case of linear process sources, contrasts based on higher order cumulants have no spurious local maximum. Finally some simulations illustrating our results have been presented.

7. REFERENCES

- [1] P. Comon, “Independent component analysis, a new concept”, *Signal Processing*, n° 36, 1994, pp.287-314.
- [2] P. Comon, “Contrasts for multichannel blind deconvolution”, *IEEE SP Letters*, vol. 3, n° 7, July 1996, pp. 209-211.
- [3] N. Delfosse, Ph. Loubaton, “Adaptive blind separation of independent sources: a deflation approach”, *Signal Processing*, n° 45, 1995, pp. 59-83.
- [4] N. Delfosse, Ph. Loubaton, “Adaptive blind identification of convolutive mixtures”, *Proc. Int. Conf. Asilomar 95*, Asilomar, November 1995, pp. 341-345.
- [5] D. Donoho, “On minimum entropy deconvolution”, *Applied Time Series Analysis II*, pp. 565-609, Academic Press, 1981.
- [6] Ph. Loubaton, Ph. Regalia, “Blind deconvolution of multivariate signals : a deflation approach”, *Proc. ICC-93*, Geneva, June 1993, pp. 1160-1164.
- [7] E. Moreau, J.-C. Pesquet, “Generalized contrasts for multichannel blind deconvolution of linear systems”, *IEEE SP Letters*, vol. 4, n° 6, June 1997, pp. 182-183.
- [8] H.L. Nguyen Thi, C. Jutten, “Blind separation for convolutive mixtures”, *Signal Processing*, vol. 45, 1995, pp. 209-229.
- [9] J.K. Tugnait, “Identification and deconvolution of multichannel linear non Gaussian processes using higher order statistics and inverse filter criteria”, *IEEE Trans. Signal Processing*, vol. 45, March 1997, pp. 658-672.
- [10] J.K. Tugnait, “On blind separation of convolutive mixtures of independent signals”, *Proc. Signal Processing Workshop on Statistical and Array Processing*, Corfu, Greece, June 1996, pp. 312-315.
- [11] D. Yellin, E. Weinstein, “Criteria for multichannel signal separation”, *IEEE Trans. Signal Processing*, vol. 42, August 1994, pp. 2158-2168.