ALTERNATIVE STATISTICAL GAUSSIANITY MEASURE USING THE CUMULATIVE DENSITY FUNCTION

Yolanda Blanco; Santiago Zazo; J. C. Principe

ETS Ingenieros de Telecomunicación - Universidad Politécnica de Madrid. Spain

Computational NeuroEngineering Laboratory- University of Florida. USA.

Phone: 341-3367280; Fax: 341- 3367350; e-mail: yolanda@gaps.ssr.upm.es

ABSTRACT

This paper discusses a novel method called 'ICA with OS' ('Independent Component Analysis with Order Statistics') to solve the vital problem of Blind Source Separation. The key of the procedure is a new alternative Gaussianity measure estimated by Order Statistics of the cdf (cumulative density function) instead of the common pdf (probability density function) moments. The maximization of this measure performs the extraction of one source in each of the successive stages of a deflation procedure based on Givens rotations.

I. INTRODUCTION

Blind Source Separation (BSS) is a problem whose solution is vital in numerous applications in communications such as in teleconferencing or in a mobile communications environment where interferences and echoes have to be cancelled.

The goal of BSS is to extract *N* unknown sources from a set of *N* of their mixtures; only the independence condition between the source signals is assumed, and the final objective is to force the maximization of an independence measure at the output signals to extract the original sources. In this sense many methods share an analysis based on the joint pdf and its related cumulants to obtain independence properties [1,2,3]. Other methods use a non-linearity (based on certain conditions over the pdf) in order to minimize the joint entropy (Infomax) [4].

Other approaches like the 'kurtosis method' [5,7] are interested in separating only one source from the mixture in every stage of a deflation procedure [6]. This kind of methods decreases the Gaussianity of the analyzed single output to provide the separation according to the Central Limit Theorem [7].

In our approach we propose a novel statistical Gaussianity measure whose maximization drives the source extraction. This measure is based on the cdf instead of the pdf (like in [5,7]) and it is developed through the concept of order statistics –OS-; consequently a novel cost function (CF) using certain OS has been deduced (see [8,9,10]).

Some of the advantages of 'ICA with OS' method are: higher quality than cumulant based methods (JADE[13]) with small sample sizes; it improves the performance for mixtures of subGaussian sources; the implementation, performed very efficiently with search strategies, is based on sample ordering.

The multistage deflation procedure will be reviewed briefly in the following section. The third section will concern with the fundamentals of 'ICA with OS' driving to the new CF. Afterwards we resume an efficient iterative algorithm to reach the extreme of the CF even for hybrid mixtures. Finally some comparative results with other methods, an several examples of sources separation conclude the paper.

II. MODEL AND DEFLATION PROCEDURE

The processing model in a generic *NxN* scenario is shown in figure *1*, where linear instantaneous mixtures of *N* independent unknown sources (with zero mean and unit variance) are collected by the sensors. A first step performs spatial decorrelation in order to decrease the number of parameters to be found.

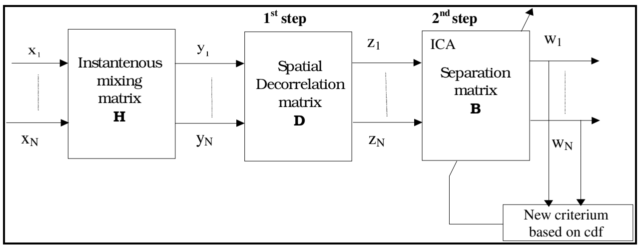


Figure1: Model of the separation procedure in two steps

The global resultant mixture of the spatial decorrelation preprocessing **V=DH** (see *figure(1)*) is an unknown orthogonal matrix . Therefore a second step is needed to perform BSS in order to update the separation matrix **B** until the separation is carried out when $\mathbf{B}=\mathbf{V}^H$

Our method is based in the following fact: any orthonormal real matrix $\mathbf{V}(NxN)$ can be factored in the product of $p=N(N-1)/2$ elementary rotation matrixes $\mathbf{R}_{ij}(\alpha_{ij})$, where α_{ij} are the unknown rotation angles :

$$\mathbf{V} = \prod_{\substack{(i,j)\\ i=1,\dots,N-1\\ j>i}} \mathbf{R}_{ij}(\alpha_{ij}) \qquad (1)$$

where $\mathbf{R}_{ij}(\alpha_{ij})$ express the rotation of an angle α_{ij} from the i axe towards the j axe, the (k,l) element of this matrix is:

$$\mathbf{R}_{ij}^{kl}(\alpha_{ij}) = \begin{cases} \cos(\alpha_{ij}) \quad if \quad k=l=i \quad or \quad k=l=j \\ sin(\alpha_{ij}) \quad if \quad k=i \quad and \quad l=j \\ -sin(\alpha_{ij}) \quad if \quad k=j \quad or \quad l=i \\ 1 \quad if \quad k=l\neq i \quad or \quad k=l\neq j \\ 0 \quad rest \end{cases} \qquad (2)$$

Equation 3 shows the factorization of the matrix **B** at different stages of the deflation procedure

$$\mathbf{B}=\mathbf{B}_{N-1}\,\mathbf{B}_{N-2}\,......\mathbf{B}_2\,\mathbf{B}_1 \qquad (3)$$

At each t $(t=\{1,....,N-1\})$ stage the matrix to be update is:

$$\mathbf{B}_t = (\prod_{j>t} \mathbf{R}_{tj}(\beta_{tj}))^H \qquad (4)$$

Let us call $\boldsymbol{\beta}_t=(\{\beta_{tj}\})_{j>t}$ the vector of parameters (or rotation angles) involved in each t stage (*Eq.4*) .

According to *Eq. (3)* the problem has been decomposed in $N-1$ stages : In each t^{th} stage of the deflation procedure the t^{th} output channel $\boldsymbol{w}_t$ is processed in order to obtain the specific set of elementary rotation matrices in *Eq.(4)* with the corresponding extraction of one source in that t^{th} channel (for more detail see [9,6])

Consequently a Cost Function J_t $(\boldsymbol{\beta}_t)$ is needed in each t stage which presents a global maximum (or minimum) for the correspondent set of parameters solution; when the extreme is reached one of the original sources would be estimated , let us call $\boldsymbol{s}_t$ this source estimation.

Figure (2) shows an scheme of the generic deflation procedure in terms of CF's.

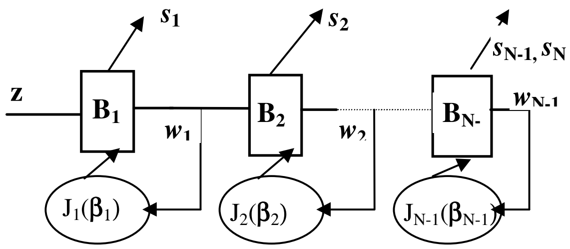


Figure2: Scheme of the generic deflation procedure

In the following section we will deduce a novel CF from a Gaussianity measure.

III. STATISTICAL MEASURE OF GAUSSIANITY BASED ON THE CDF

A first approximation to our method was proposed in [8]. Let us focus the fundamentals from a more appropriate perspective in this section.

III.I Previous ideas

1.<u>Concept of Order Statistics</u>

Let $(w_{(1)}<w_{(2)}< ...< w_{(n)})$ be an ordered list obtained from the discrete signal $\boldsymbol{w}$: $w[1]....w[n]$. The k-sample $w_{(k)}$ is labeled as the O.S. of order k. The main feature we want to extract from the O.S. is that it is in fact an asymptotic non parametric estimator of the cdf :

$$w_{(k)} \,/\, \hat{F}_w(w_{(k)}) = k/n \qquad (5)$$

Around 500 ordered samples are enough to estimate very robustly and easily any cdf, while with the same number of samples the accuracy is much poorer with the pdf estimation by means of the Gram-Charlier expansion (using cumulants [1]); this fact was the original motivation to exploit the properties of the cdf instead of the pdf to obtain the separation.

<u>2.- Central Limit Theorem (CLT)</u>

The CLT states that the linear combination of independent random variables always increases the Gaussianity of the constrained distribution. Blind Source Separation must simply decrease the Gaussianity of the output analysis channel to extract one of the original non-Gaussian sources. *Figure (3)* illustrates the CLT implication for an arbitrary orthogonal mixture of two Laplacian sources. Dashed lines represent the theoretical distributions (*pdf's top, cdf's bottom)* of one output channel for an arbitrary matrix **B**. The dotted line represents the equivalent Gaussian distribution with the same variance. The distribution of the Laplacian source is represented as the solid line.

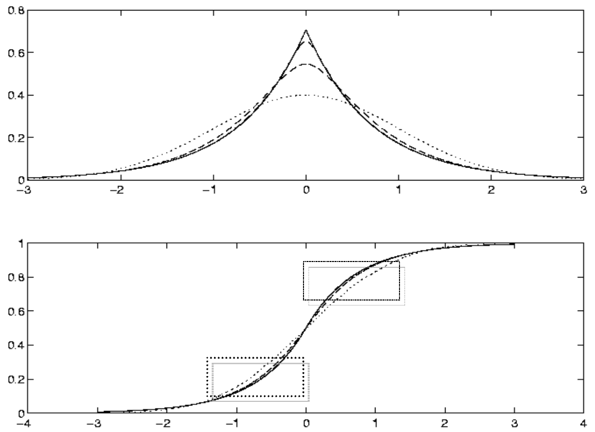


Figure 3, Top: theoretical pdf's in one output channel for a mixture of two laplacian sources. Bottom: Theoretical cdf's for the same signals

III.II Gaussianity Measure based on the cdf

From the previous section, we conclude that Gaussianity is an appropriate distance between the analyzed signal and the equivalent Gaussian distribution; several methods [5,7] estimate Gaussianity based on cumulants. Our contribution is to utilize order statistics instead.

As it can be observed in *figure 3-bottom* ,there are certain regions in the cdf's representation (marked with squares) that show discriminability between the signals involved. We are able to exploit this property because OS provides a practical way of estimating the cdf.

To illustrate the method let us take an example of a *NxN* instantaneous mixture where at least one of the sources has superGaussian distribution (a Laplacian distribution in this case). However any kind of mixture offers similar behavior.

Let us call $\boldsymbol{w}$ the processed t^{th} output channel in each t^{th} stage of the deflation procedure (*section II*). *Figure (4)* represents the estimation of the involved cdf's through the OS (see *Eq 5*) for the interval corresponding to the upper right part in *figure 3-bottom:* The dashed line is the estimated cdf ($F_w(\boldsymbol{w})$) of the signal for an arbitrary value of $\mathbf{B}_t(\boldsymbol{\beta}_t)$; The dotted line is the estimation of the equivalent Gaussian cdf of the same power ($F_G(\boldsymbol{w})$) . Also the cdf of the unknown original Laplacian source ($F_x(\mathrm{w})$) is represented as the solid line.

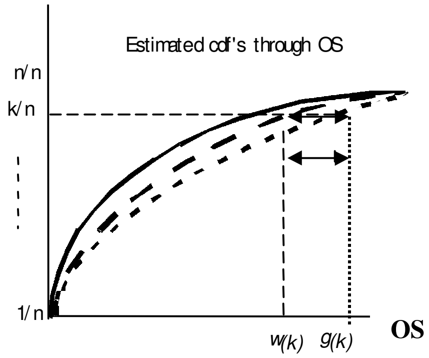


Figure 4: Gaussianity measure through the *k*-OS

Let us observe that the cdf's of the $\boldsymbol{w}$ channel are always above the cdf of the Gaussian distribution in this interval and the distance between them is maximum when the separation condition holds, that is when $\boldsymbol{w} = \boldsymbol{x}$ ($\boldsymbol{x}$ is one original source). In this sense the distance marked in the *figure 4* with the double arrow can be taken like a Gaussianity measure G_m:

$$Gm = \left| w_{(k)}(\beta_t) - g_{(k)} \right| \quad for\ certain\ k-OS \qquad (6)$$

In *Eq.(6)* $w_{(k)}$ is the *k*-OS of $\boldsymbol{w}$, and $g_{(k)}$ is the *k*-OS of the Gaussian distribution. For appropriate values of *k*, *Gm* increases when the Gaussianity of $\boldsymbol{w}$ decreases; therefore the source would be extracted at the maximum distance *Gm.*
Gm is maximum when $w_{(k)}$ reaches its minimum value, however, if the source is subGaussian the polarity changes and $w_{(k)}$ would be maximum.
Therefore the following cost function could be deduced from *Gm*:

$$\underset{\beta_t}{max/min} \left| w_{(k)}(\beta_t) \right| \quad \begin{matrix} max: subGaussian \\ min: superGaussian \end{matrix} \qquad (7)$$

For symmetric distributions we can define a cost function:

$$J_t(\beta_t) = \left| w_{(k)}(\beta_t) - w_{(l)}(\beta_t) \right| \quad for\,(k,l)\ \ with\ \ k+l=1 \qquad (8)$$

Let us remark that J_t presents a minimum when the extracted source is superGaussian and a maximum when the source is subGaussian for *(k,l)* in intervals around *(75%,25%).*

Experimentally we verified that the values of *(k,l)*= *(75%,25%)* can be chosen as the optimum range of the OS if the involved distributions are unknown. This conclusion has been reached after a detailed study over the variance of the estimator in *Eq(7)* for several values of *k.*

III.III Improving the Gaussianity measure

As we have just seen, the previous cost function *(Eq.8)* is based on the estimation of the cdf of the r.v. $\boldsymbol{w} : F_w(\boldsymbol{w})$
The improved measure is based on the estimation of the new normalized r.v. $\boldsymbol{y}$:

$$F_y(\boldsymbol{y}) \ \ with \ \ \boldsymbol{y} = \frac{\boldsymbol{w}}{max(\boldsymbol{w})} \tag{9}$$

The relationship between both distributions is a scale transformation:

$$F_y(\boldsymbol{w}) = F_w(\boldsymbol{w} \ \ max(\boldsymbol{w})) \tag{10}$$

$F_y(\boldsymbol{y})$ presents the same properties in relation with its equivalent Gaussian distribution $F_G(\boldsymbol{y})$ (see *figure(4)*). The advantage is the increased curvature of $F_y(\boldsymbol{y})$ because the new scale limits the curves between -1 and +1 at the horizontal axis, so that the discrimination measure in *figure(4)* is more noticeable.
Therefore the cost function in *Eq(8)* can be modified in terms of $\boldsymbol{y}$:

$$J_t(\boldsymbol{\beta}_t) = \left| y_{(k)}(\boldsymbol{\beta}_t) - y_{(l)}(\boldsymbol{\beta}_t) \right| \ \ for\ (k,l) = (75\%, 25\%) \tag{11}$$

Let us recall that *max(w)* is the OS of higher order $k=n$ (n is the number of processed samples); the improved cost function has been deduced substituting $\boldsymbol{y}$ (*Eq(9)*) in *Eq(11):*

$$J_t(\boldsymbol{\beta}_t) = \left| \frac{w_{(k)}(\boldsymbol{\beta}_t) - w_{(l)}(\boldsymbol{\beta}_t)}{w_{(n)}(\boldsymbol{\beta}_t)} \right| \ for\ (k,l) = (75\%, 25\%) \tag{12}$$

We have compared the *Amari's performance index* [11] from both cost functions (*Eq. 12, Eq. 8*) and the results are better for the normalized case.

IV. AN ALGORITHM TO FIND THE EXTREME OF THE COST FUNCTION

IV.I Iterative Algorithm

According to the previous section, separation requires the optimization of the cost function presented in *Equation (12)* at each t stage. The procedure to reach the extreme would be the same for each t stage, only the dimension M ($M=N-t$) of the set of involved parameters ($\boldsymbol{\beta}_t=(\{\beta_{tj}\})_{j>t}$) changes.

The proposed iterative algorithm is based on an efficient searching strategy over a set S of discrete values of the vector $\boldsymbol{\beta}_t$. Let us suppose that S has enough number of elements and accuracy in order to calculate the discrete values of J_t for each S element and then to estimate their extremes. This procedure is not practical because S too high dimension when the number of parameters M increases.

Therefore let us propose an iterative procedure where a set S_i, with only a few elements, changes with each iteration i. Discrete values for the cost function $(J_t(\boldsymbol{\beta}_t))_i$ at the i iteration are calculated for each $\boldsymbol{\beta}_t$ element of S_i in the following way: we obtain the output channel $\boldsymbol{w}$ for each matrix $\mathbf{B}_t(\boldsymbol{\beta}_t)$ (*Eq(4)*) ; then the *25%,75%* and n- OS are computed simply by reordering the samples and extracting the samples at those positions to calculate the value of J_t for this $\boldsymbol{\beta}_t$ through *Eq(12)* .
In the *ith* iteration the extreme of $(J_t(\boldsymbol{\beta}_t))_i$ provides an increment for the vector solution according to:

$$\begin{gathered} \boldsymbol{\beta}_t[i+1] = \Delta\boldsymbol{\beta}_t[i] + \boldsymbol{\beta}_t[i] \\ where \quad \Delta\boldsymbol{\beta}_t[i] = \underset{\boldsymbol{\beta}_t}{\overset{max}{min}} \ (J_t(\boldsymbol{\beta}_t))_i \end{gathered} \tag{13}$$

Nevertheless the 'iterative algorithm' still requires a relatively high computational cost when the number of sources increases, therefore we have proposed a descent (ascent) gradient technique to reach the minimum (maximum) of the cost function [10]. The gradient expression in the TITO scenario has been deduced in [10]. The generalization to the MIMO scenario is under analysis, and it is out of the scope of the present paper.

IV.II Hybrid Mixtures

There are many cases where the type of distributions of the involved sources are unknown or different kind of distributions are involved. In these situations an analysis of the form of the *Kurtosis Function* [5] has been implemented in order to find out the source characteristic (sub-super Gaussian) and therefore the choice of the extreme characteristic of J_t. The following point has been borrowed from [5]:

"The *kurtosis function* as a function of the set of M parameters in each stage, has non-negative maximum for superGaussian components and non-positive minimum for subGaussian components ".

Therefore a pre-processing before each stage has been implemented in order to estimate the shape of the *kurtosis function* $K(\boldsymbol{\beta}_t)$ where several *kurtosis values* are estimated for a small set of discrete values of the vector $\boldsymbol{\beta}_t$ according to equation (14)*:*

$$K(\beta_t) = \frac{E\{\boldsymbol{w}^4(\beta_t)\}}{(E\{\boldsymbol{w}^2(\beta_t)\})^2} \qquad (14)$$

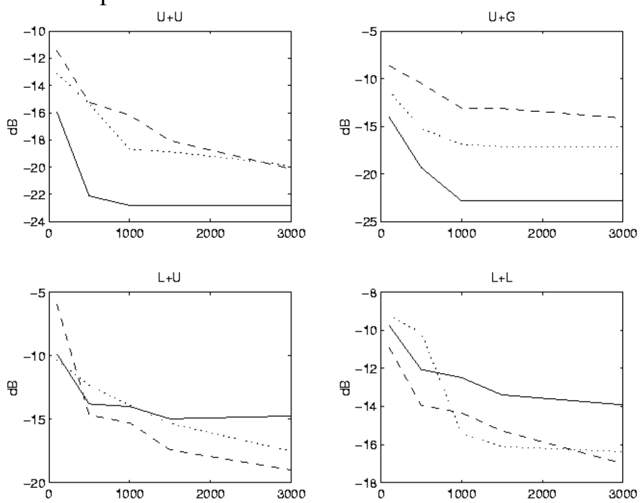
In spite of the fact that the estimation of the *kurtosis function Eq(14)* has not enough accuracy, the evolution of this function follows the real *kurtosis function*, and this is enough to know the kind of extreme. Recall that if the estimation of the *kurtosis function* presents a positive maximum our cost function presents a minimum for the solution and vice versa.

V. RESULTS

In order to compare the performance of the 'ICA with OS' methods with other representative methods (extended INFOMAX[4] and JADE[13]) several kind of sources have been mixed through the orthogonal matrix $V=R_{xy}(45^o)$. We have considered this mixture because it implies the maximum and equivalent rotation for the involved two sources. 'ICA with OS' use the improved cost function from *Eq(12)* and it has been optimizaed by means of the full algorithm in *section IV.*

In *figure(6)* we present the comparative results: *Amari's performance index in dB* [11] (averaged for ten realizations) front to number of processed samples.

The following conclusions can be deduced from *figure(6)*: -
1.- 'ICA with OS' (based on order statistics) works better than 'JADE' (based on cumulants) for small window size (100,500) samples.
2.- When the involved sources are subGaussians, 'ICA with OS' improves the quality of the separation other traditional methods, while the other methods show a better behavior for the superGaussian cases.



Figure(6): Comparative performance index. Solid line: 'ICA with OS', Dashed line: 'INFOMAX'*, Dotted line: 'JADE'*. (L:Laplacian, U:Uniform, G: Gaussian)

Figure(7) corroborates that 'ICA with OS' presents a better behavior for the subGassians cases; three image signals have been mixed through an arbitrary **H**, and have been separated successfully by our method, while Infomax quality separation is not good enough. (100 pixels have been processed by both procedures).

Figure(7): 'ICA with OS' and 'INFOMAX' for an arbitrary mixture of three images. Top: Original images; Second row: mixtures, Third row: Separated images though ICA with OS (-11 dB); Four row: Estimated images though INFOMAX (-2 dB).

Bellow, an example of a mixture of two *nPAM* discrete sources and a Laplacian noise source is presented in order to show the capability of the method in a communication environment and for hybrid mixtures. In this case an excellent *performance index* of -17.5 dB has been obtained with 500 processed samples.

*Simulations with "ext_ica.m" from *www/cnl.salk.edu/-tewon* and "jade.m" from *sig.enst.fr:80/-cardoso/guidesepson.html*

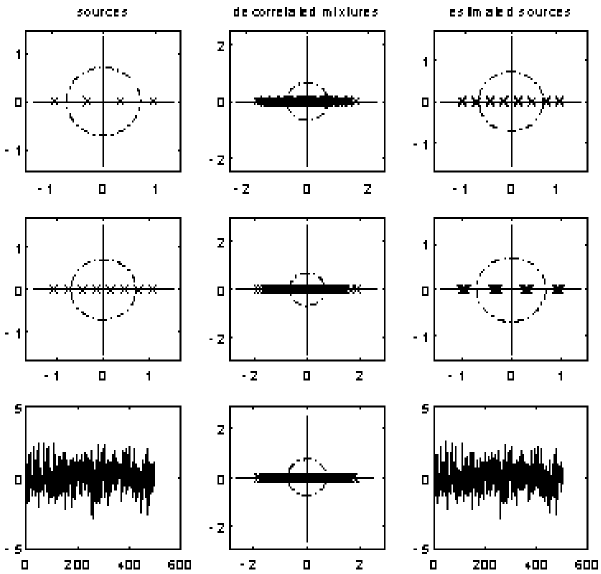


Figure (8): Mixture of 4-PAM with 8-PAM and laplacian noise

VI CONCLUSIONS

The novel method 'ICA with OS' has been presented in this paper under the perspective of an alternative measure of Gaussianity based on the cdf instead of the pdf.

We have shown that our method separates any kind of mixtures even hybrid ones with very few samples, presenting the best quality in the subGaussian scenarios. Besides the implementation is simply based on sample ordering applied in a deflation procedure. Another advantage of this method is that it is very efficient against outliers because the contaminated samples will be placed at the upper OS and therefore are filtered by our procedure. Our present line of research is the generalization of the method to convolutive mixtures. The separation in the frequency domain [12] is the most attractive alternative: we have already extended the procedure to the complex instantaneous mixtures in order to apply the method to each of the frequency bins obtained through the Short Time Fourier Transform (STFT) of the observable signals. Anyway this topic is still under study and the principal inconvenient is the permutation problem.

VII REFERENCES

[1].-Pierre Common . Independent component analysis, A new concept. *Signal Processing,* 36(1994) 287-314.

[2].-C.Jutten and J. Herault. Blind separation of sources, part I: An adapative procedure based on neuromimetic architecture . *Signal processing*, 24, vol 1. page 1-10

[3].- D. T. Pham, P. Gharat, C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. *Proc. EUSIPCO* 1992, 771-774

[4].-A.J. Bell and T.J Sejnowski . An information maximization approach to Blind Separation and Blind Deconvolution. *Neural Computation,* 7,6, 1129-1159

[5].-S.Y. Kung, C.Mejuto. Extraction of Independent Components from Hybrid mixture: knicnet learning algorithm and applications.*ICASSP'98* . vol II, 1209,1211.

[6].- N. Delfosse, P. Loubaton. Adaptive separation of independent sources: A deflation approach.
Proc. ICASSP, 94. page IV-41-IV-44

[7]H.C. Wu. "Blind source separation using information measures in time and frequency domain". *Phd. Thesis. CNEL . University of Florida. 1999.*

[8].-S. Zazo, Y Blanco, JM Páez-Borrallo. Order Statistics: A new approach to the problem of blind source separation. *Workshop ICA'99. Aussois. France.*page 413-418

[9] Y Blanco, S. Zazo, JM Páez-Borrallo. A Multistage Deflation Algorithm for Blind Source Separation based on ICA with OS. *Fifth Workshop on Emergent Technologies on Communications.* Bayona (Spain) 1999

[10] Y Blanco, S. Zazo, JM Páez-Borrallo. Adaptive processing of blind source separation through 'ICA with OS'. Accepted by *ICASSP'00*. June 2000.

[11].-.S. Amari . A. Cichocki . H. H. Yang. A new learning algorithm for blind signal separation. *Advances in neural `information processing systems*, 8, 757-763

[12]Ch. Servier. Feasibility of source separation in frequency domain . In proceedings of *ICASSP'98*. pp 2085-2088

[13]JF Cardoso. A.Soulimac. Blind beamforming for non gaussian signals. *IEE proceedings-F*, vol 140, no6, pp 362-370, dec.93