# Estimating functions for blind separation when sources have variance-dependencies

Motoaki Kawanabe[1] and Klaus-Robert Müller[1,2]

[1] Fraunhofer FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany
[2] University of Potsdam, August-Bebel-Strasse 89, 14482 Potsdam, Germany
{nabe,klaus}@first.fhg.de

**Abstract.** The blind separation problem where the sources are not independent, but have variance-dependencies is discussed. Hyvärinen and Hurri[1] proposed an algorithm which requires no assumption on distributions of sources and no parametric model of dependencies between components. In this paper, we extend the semiparametric statistical approach of Amari and Cardoso[2] under variance-dependencies and study estimating functions for blind separation of such dependent sources. In particular, we show that many of ICA algorithms are applicable to the variance-dependent model as well. Our theoretical consequences were confirmed by artificial and realistic examples.

## 1 Introduction

Independent component analysis (ICA) is based on the assumption that the observed signals are linear superpositions of mutually independent source signals. Let us denote the $n$ source signals by $\boldsymbol{s}(t) = (s_1(t), \ldots, s_n(t))^\top$ in a vector formula, and the observed signals by $\boldsymbol{x}(t) = (x_1(t), \ldots, x_m(t))^\top$. The mixing process can be expressed as the equation

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

provided that it is not contaminated by any noise, where $A = (a_{ij})$ denotes the mixing matrix. For simplicity, we consider the case where the number of source signals equals that of observed signals ($n = m$).

Among many extensions of the basic ICA models, several researchers have studied the case where the source signals are not independent [3–6]. The dependencies either need to be exactly known beforehand, or they are simultaneously estimated by the algorithms. Recently, a novel idea called double-blind approach was introduced by Hyvärinen and Hurri[1]. In contrast to previous research, their method requires no assumption on the distributions of the sources and no parametric model of dependencies between the components. It is only assumed that the sources are dependent solely through their variances and that the sources have temporal dependencies.

A statistical basis of ICA was established by Amari and Cardoso[2]. They pointed out that the ICA model is an example of semiparametric statistical models[7, 8] and studied estimating functions for it. In particular, they showed that the quasi maximum likelihood (QML) estimation and the natural gradient learning give a correct solution regardless of the true source densities. In this paper, we extend their approach to the blind source separation (BSS) problem considered in [1]. Investigating estimating functions for the model, we show that many of ICA algorithms based on the independence assumption work properly, even if there exist variance-dependencies.

This paper is organized as follows. After explaining our framework in Section 2 and Section 3, estimating functions for the variance-dependent BSS model are studied in Section 4. There, the quasi maximum likelihood estimation is taken as an example, while properties of other ICA algorithms are summarized in Section 5. We carried out numerical experiments with artificial and realistic examples (Section 6). Although only the double-blind algorithm gave correct solutions in the example described in [1], many ICA algorithms also worked for the other datasets.

## 2 Variance-dependent BSS model

Hyvärinen and Hurri[1] introduced the following framework. Let us assume that each source signal $s_i(t)$ is a product of non-negative activity level $v_i(t)$ and underlying i.i.d. signal $z_i(t)$, i.e.

$$s_i(t) = v_i(t) z_i(t). \tag{2}$$

In practice, the activity levels $v_i$'s are often dependent among different signals. In their formulation, each observed signal is expressed as

$$x_i(t) = \sum_{j=1}^{n} a_{ij} v_j(t) z_j(t), \qquad i = 1, \ldots, n, \tag{3}$$

where $v_i(t)$ and $z_i(t)$ satisfy:

(i) $v_i$'s and $z_j$'s are independent,
(ii) each $z_i(t)$ is i.i.d. in time, $z_i$ and $z_j$ are mutually independent,
(iii) $z_i(t)$ have zero mean and unit variance.

No assumption on the distribution of $z_i$ is made except (iii). Regarding the general activity levels $v_i$'s, $v_i(t)$ and $v_j(t)$ are allowed to be statistically dependent, and furthermore, no particular assumption on these dependencies are made (double-blind situation). We refer to this framework as the variance-dependent BSS model in this paper.

They also proposed an algorithm which can separate the sources under the variance-dependent BSS model. Let $\boldsymbol{u}(t)$ be the preprocessed signal of $\boldsymbol{x}(t)$ by spatial whitening. Their method maximizes the objective function

$$J(W) = \sum_{i,j} [\, \widehat{\mathrm{cov}}([\boldsymbol{w}_i^\top \boldsymbol{u}(t)]^2, [\boldsymbol{w}_j^\top \boldsymbol{u}(t - \Delta t)]^2)\, ]^2$$

over an orthogonal matrix $W = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)^\top$, where $\widehat{\mathrm{cov}}$ denotes the sample covariance. It was proved that the objective function $J$ is maximized when $WA$ equals a signed permutation matrix, if $K_{ij} = \mathrm{cov}(s_i^2(t), s_j^2(t - \Delta t)\,)$ is of full rank. This method works quite well, provided that there exist temporal variance-dependencies and the data is not spoiled by outliers.

## 3 Semiparametric statistical models and estimating functions

Amari and Cardoso[2] established a statistical basis of the ICA problem. They pointed out that the standard ICA model

$$p(X|B,\kappa_{\boldsymbol{s}}) = |\det B|^T \prod_{t=1}^{T}\prod_{i=1}^{n} \kappa_{s_i}\{\boldsymbol{b}_i^\top \boldsymbol{x}(t)\} \tag{4}$$

is an example of semiparametric statistical models [7, 8], where $X = (\boldsymbol{x}(1),\dots,\boldsymbol{x}(T)\,)$ is the whole data sequence, $B = (\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)^\top = A^{-1}$ is the demixing matrix to be estimated and $\kappa_{\boldsymbol{s}}(\boldsymbol{s}) = \prod_{i=1}^{n} \kappa_{s_i}(s_i)$ is the density of the sources $\boldsymbol{s}$. As the function $\kappa_{\boldsymbol{s}}$ in (4), semiparametric models contain infinite dimensional or functional nuisance parameters which are difficult to estimate. Moreover, they even disturb inference on parameters of interest.

In the variance-dependent BSS model, the sources $\boldsymbol{s}(t)$ are decomposed of two components, the normalized signals $\boldsymbol{z}(t) = (z_1(t),\dots,z_n(t)\,)^\top$ and the general activity levels $\boldsymbol{v}(t) = (v_1(t),\dots,v_n(t)\,)^\top$. Since the former have mutual independence in the origin of ICA model, the density of the data $X$ is factorized as

$$p(X|V;B,\kappa) = |\det B|^T \prod_{t=1}^{T}\prod_{i=1}^{n} \frac{1}{v_i(t)}\ \kappa_i\left\{\frac{\boldsymbol{b}_i^\top \boldsymbol{x}(t)}{v_i(t)}\right\}, \tag{5}$$

when $V = (\boldsymbol{v}(1),\dots,\boldsymbol{v}(T)\,)$ is fixed. Therefore, the marginal distribution can be expressed as

$$p(X|B,\kappa,\nu) = \int p(X|V;B,\kappa)\nu(V)\mathrm{d}V, \tag{6}$$

where the density $\nu$ of $V$ becomes an extra nuisance function.

Estimating functions are a tool for constructing valid estimators in such semiparametric models. Let us consider a general semiparametric model $p(\boldsymbol{x}|\boldsymbol{\theta},\kappa)$, where $\boldsymbol{\theta}$ is an $r$-dimensional parameter of interest and $\kappa$ is a nuisance parameter. An $r$-dimensional vector valued function $\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})$ is called an estimating function, when it satisfies the following conditions for any $\boldsymbol{\theta}$ and $\kappa$,

$$\mathrm{E}[\,\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\,|\boldsymbol{\theta},\kappa] = \boldsymbol{0}, \tag{7}$$

$$|\det Q| \neq 0, \qquad \text{where } Q = \mathrm{E}\left[\,\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\,\middle|\,\boldsymbol{\theta},\kappa\right], \tag{8}$$

$$\mathrm{E}\left[\,\|\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})\|^2\,\middle|\,\boldsymbol{\theta},\kappa\right] < \infty, \tag{9}$$

where $\mathrm{E}[\cdot|\boldsymbol{\theta},\kappa]$ means the expectation over $\boldsymbol{x}$ with the density $p(\boldsymbol{x}|\boldsymbol{\theta},\kappa)$ and $\|\cdot\|$ denotes Euclidean norm [9]. Suppose that i.i.d. samples $\boldsymbol{x}(1),\dots,\boldsymbol{x}(T)$ are obtained from the model $p(\boldsymbol{x}|\boldsymbol{\theta}^*,\kappa^*)$. If such a function exists, an M-estimator is obtained by solving the estimating equation

$$\sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{x}(t),\hat{\boldsymbol{\theta}}) = \boldsymbol{0}. \tag{10}$$

The estimator $\hat{\boldsymbol{\theta}}$ is consistent regardless of the true nuisance parameter $\kappa^*$, when the sample size $T$ goes to infinity.

## 4 Estimating functions for blind separation

Estimating functions for the ICA model (4) were discussed by Amari and Cardoso[2] and Cardoso[10]. In this case, the parameter of interest is the $n \times n$ matrix $B = A^{-1}$ and hence it is convenient to write estimating functions in $n \times n$ matrix form $F(\boldsymbol{x}, B)$. Amari and Cardoso[2] showed that the quasi maximum likelihood method is a semiparametric algorithm based on estimating functions.

In the variance-dependent BSS model, in contrast to the ICA model studied by Amari and Cardoso[2], the data sequence $X = (\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)\,)$ is not i.i.d. in time, but might have temporal dependencies . Therefore, we have to consider more general functions $\bar{F}(X, B)$ of the whole sequence $X$. General estimating functions $\bar{F}(X, B)$ must satisfy

$$\mathrm{E}[\,\bar{F}(X,B)\,|B,\kappa,\nu] = 0, \tag{11}$$

$$|\det Q| \neq 0, \qquad \text{where } Q = \mathrm{E}\left[\left.\frac{\partial \mathrm{vec}\{\bar{F}(X,B)\}}{\partial \mathrm{vec}(B)}\,\right|\,B,\kappa,\nu\right], \tag{12}$$

$$\mathrm{E}\left[\,\|\bar{F}(X,B)\|_F^2\,\middle|\,B,\kappa,\nu\right] < \infty, \tag{13}$$

for all $(B, \kappa, \nu)$. An $M$-estimator $\hat{B}$ is derived from the estimating equation

$$\bar{F}(X, \hat{B}) = 0. \tag{14}$$

Suppose that the data $X$ is subject to $p(X|B^*, \kappa^*, \nu^*)$ defined by (5) and (6). It is known that the M-estimator $\hat{B}$ is consistent and asymptotically normal.

**Theorem 1.** *If the function $\bar{F}(X, B)$ satisfies the conditions (11)∼(13) and appropriate regularity conditions, the $M$-estimator $\hat{B}$ derived from the equation (14) is asymptotically Gaussian distributed, i.e.* $\mathrm{vec}(\hat{B}) \sim N(\mathrm{vec}(B^*), \mathrm{Av})$*, where*

$$\begin{aligned}
\mathrm{Av} = \mathrm{Av}(B^*, \kappa^*, \nu^*) \;&=\; Q^{-1}\,\Sigma\,(Q^{-1})^\top, \\
\Sigma = \Sigma(B^*, \kappa^*, \nu^*) \;&=\; E\left[\mathrm{vec}\{\bar{F}(X,B^*)\}\,\mathrm{vec}\{\bar{F}(X,B^*)\}^\top \,\middle|\, B^*, \kappa^*, \nu^*\right], \\
Q = Q(B^*, \kappa^*, \nu^*) \;&=\; \mathrm{E}\left[\left.\frac{\partial \mathrm{vec}\{\bar{F}(X,B^*)\}}{\partial \mathrm{vec}(B)}\,\right|\,B^*, \kappa^*, \nu^*\right].
\end{aligned} \tag{15}$$

Now let us describe our main result. We can show that the function

$$\bar{F}(X, B) = \sum_{t=1}^{T} F(\boldsymbol{x}(t), B) \tag{16}$$

constructed from an estimating function $F(\boldsymbol{x}, B)$ for the ICA model becomes a candidate of estimating functions for the variance-dependent BSS model.

**Theorem 2.** *The function $\bar{F}(X, B)$ defined in (16) satisfies the two conditions (11) and (13), provided that $F(\boldsymbol{x}, B)$ is an estimating function for the ICA model (4).*

Because it is difficult to check the other condition (12) in the general form, let us consider the quasi maximum likelihood estimation

$$\bar{F}^{\mathrm{QML}}(X, B) = \sum_{t=1}^{T} \left[ I - \boldsymbol{\varphi}\{\boldsymbol{y}(t)\}\, \boldsymbol{y}^{\top}(t) \right], \tag{17}$$

as an example in the class (16), where $\boldsymbol{\varphi}(\boldsymbol{y}) = (\varphi_1(y_1), \ldots, \varphi_n(y_n))^{\top}$ is a vector of nonlinear functions.

**Theorem 3.** *Suppose that the conditions*

$$\sum_{t=1}^{T} \mathrm{E}\left[ m_i\{v_i(t)\} \right] + T \neq 0, \qquad \forall i, \tag{18}$$

$$\det \begin{pmatrix} \sum_{t=1}^{T} \mathrm{E}[\, k_i\{v_i(t)\} v_j^2(t) \,] & T \\ T & \sum_{t=1}^{T} \mathrm{E}[\, k_j\{v_j(t)\} v_i^2(t) \,] \end{pmatrix} \neq 0, \qquad \forall i \neq j, \tag{19}$$

*hold, where*

$$k_i\{v_i(t)\} = \mathrm{E}\left[\, \dot{\varphi}_i\{v_i(t) z_i(t)\} \mid V; B, \kappa \right], \tag{20}$$

$$m_i\{v_i(t)\} = v_i^2(t)\, \mathrm{E}\left[\, \dot{\varphi}_i\{v_i(t) z_i(t)\}\, z_i^2(t) \,\middle|\, V; B, \kappa \right], \tag{21}$$

*and $\dot{\varphi}_i$ is the derivative of $\varphi_i$. Then, the function $\bar{F}^{\mathrm{QML}}(X, B)$ satisfies the conditions (11)∼(13) and becomes an estimating function. Under appropriate regularity conditions, the quasi maximum likelihood estimator $\hat{B}^{\mathrm{QML}}$ derived from the equation $\bar{F}^{QML}(X, \hat{B}^{\mathrm{QML}}) = 0$ is consistent regardless of the true nuisance functions $(\kappa^*, \nu^*)$.*

## 5 Statistical properties of ICA algorithms

Although we concentrated on estimating functions of the form (16) in the previous section, we can deal with more general functions and investigate other ICA algorithms within the framework of estimating functions or asymptotic estimating functions[10] as well. Here we examined the unbiasedness condition (11) under the variance-dependent BSS model (Results are summarized in Table 1). In fact, this condition holds at least asymptotically in many algorithms. If the other conditions are satisfied, these algorithms give valid solutions regardless of the nuisance densities $(\kappa^*, \nu^*)$. We remark that our extension also enables us to analyze algorithms based on temporal structure such as TDSEP/SOBI[11, 12].

## 6 Numerical experiments

We carried out at first experiments with several artificial datasets. We applied the quasi maximal likelihood methods QML-t and QML-3 (-t and -3 denote $\tanh$ and cubic nonlinearity, resp.), the double-blind algorithm 'DB' [1], JADE, FastICA-t and FastICA-3,

**Table 1.** Unbiasedness condition of other ICA algorithms

| algorithm | unbiasedness | inapplicable cases |
|---|---|---|
| FastICA[13] | yes | Gaussian sources. |
| double-blind[1] | asymptotically | same variance-structures or no temporal variance-dependency |
| JADE[14] | asymptotically | Gaussian sources |
| TDSEP/SOBI[11, 12] | yes | always |
| nonstationary[15] | yes | unclear |

TDSEP/SOBI [11, 12] and the 'sepagaus' algorithm for nonstationary signals[15], For evaluating the results, we used the index defined in Amari et al.[16]

$$\mathrm{AmariIndex}(B, A^*) = \sum_{i=1}^{n} \left\{ \frac{\sum_{j=1}^{n} C_{ij}}{\max_k C_{ik}} - 1 \right\} + \sum_{j=1}^{n} \left\{ \frac{\sum_{i=1}^{n} C_{ij}}{\max_k C_{kj}} - 1 \right\}, \tag{22}$$

where $A^*$ is the true mixing matrix and $C = BA^*$. If $B = PD(A^*)^{-1}$ with a permutation matrix $P$ and a diagonal matrix $D$, then $\mathrm{AmariIndex}(B, A^*) = 0$.

In all artificial datasets, five source signals of various types were generated and the data were observed after mixing with multiplying a random $5 \times 5$ matrix. We prepared eight artificial datasets ar_subG, ar_uni, sin_supG, sin_subG, com_supG, com_subG, exp_supG, and uni_subG. For the activity levels, the abbreviation 'ar' means that the random vector $\boldsymbol{v}(t)$ was the absolute value of a multivariate AR(1) process. The activity levels of 'sin' datasets were sinusoidal functions with different frequencies, while those of 'com' were ones with same frequency. In the case of 'exp' and 'uni' datasets they were linear transformations of i.i.d. Laplace and uniform random vectors, respectively. For the normalized signals, 'uni' and 'supG' denote uniform and Laplace random variables, while 'subG' sequences were signed fourth roots of uniform random variables.

**Table 2.** AmariIndex of the estimators. The values are the medians of 100 replications.

| | QML-t | QML-3 | DB | JADE | FastICA-t | FastICA-3 | TDSEP | sepagaus |
|---|---|---|---|---|---|---|---|---|
| ar_subG | 8.25 | 11.32 | 0.52 | 10.79 | 9.25 | 12.52 | 15.07 | 1.19 |
| ar_uni | 0.30 | 27.77 | 0.70 | 0.66 | 0.38 | 0.73 | 14.92 | 0.85 |
| sin_supG | 0.17 | 29.97 | 0.79 | 0.43 | 0.23 | 0.41 | 15.31 | 0.08 |
| sin_subG | 19.21 | 0.32 | 0.27 | 0.31 | 0.68 | 0.33 | 15.70 | 0.08 |
| com_supG | 0.39 | 28.37 | 6.45 | 0.84 | 0.48 | 0.87 | 16.02 | 1.28 |
| com_subG | 26.53 | 0.14 | 22.05 | 26.49 | 27.04 | 26.65 | 16.23 | 27.08 |
| exp_supG | 0.35 | 28.43 | 7.63 | 1.24 | 0.44 | 1.20 | 16.47 | 1.28 |
| uni_subG | 27.38 | 0.13 | 18.56 | 0.17 | 0.18 | 0.18 | 16.20 | 27.08 |
| sss | 0.03 | 3.82 | 0.02 | 0.02 | 0.19 | 0.09 | 0.01 | 0.01 |
| v12 | 0.01 | 3.73 | 0.21 | 0.19 | 0.17 | 0.08 | 0.14 | 0.01 |

As Hyvärinen and Hurri[1] showed, almost all algorithms except DB did not give a proper solution in ar_subG. However, DB showed poor performance, when (i) the variance-structures are the same or (ii) there is no temporal dependency. As expected, TDSEP did not work for any data, because there are no temporal correlations. QML-t is applicable to supergaussian cases, while QML-3 can be used for subgaussian data. The other algorithm returned acceptable results except in the difficult case com_subG.

Then, we also studied speech signals as more realistic examples. In the first example 'sss'[3], speakers counts from $1$ to $10$ in English and in Spanish, respectively (see the left panels of Figure 1). In the second experiment 'v12'[4], we took two speech signals from Japanese text, and modified the second so that the two sequences have large variance-dependency(see the right panels of Figure 1). The correlation of the variances in each example is substantially positive, i.e. $0.65$ and $0.74$, respectively. The results are shown in Table 2, too. All algorithm except QML-3 gave a proper answer. On these realistic examples, TDSEP also worked, because the statistical model (5) and (6) did not hold perfectly.

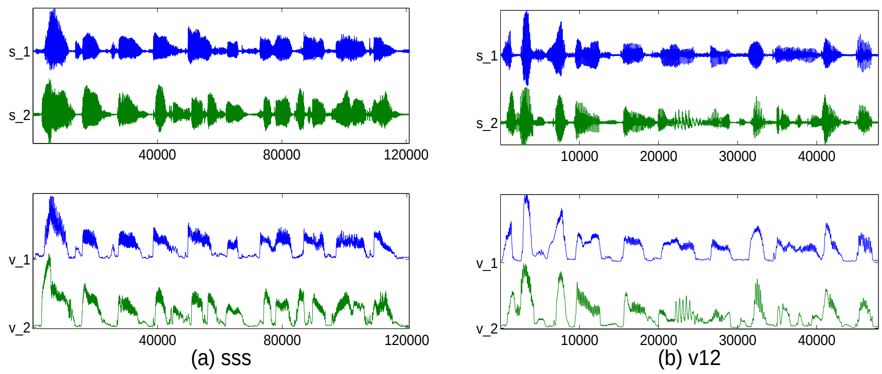


**Fig. 1.** The sources $\boldsymbol{s}(t)$ (upper panels) and the estimators $\boldsymbol{v}(t)$ of their activity levels with an appropriate smoother (lower panels).

## 7 Conclusions

In this paper, we discussed semiparametric estimation for blind separation, when sources have variance-dependencies. Extending the semiparametric statistical approach[2] under variance-dependencies, we investigated estimating functions for the variance-dependent BSS model. In particular, we proved that the quasi maximum likelihood estimator is derived from such an estimating function, and hence consistent regardless of the true nuisance densities. Although we omitted details in this paper, we also analyzed other ICA algorithms within the framework of (asymptotic) estimating functions and showed that many of them can separate sources with coherent variances. The theoretical results

[3] `http://inc2.ucsd.edu/~tewon/`

[4] `http://www.islab.brain.riken.go.jp/~mura/ica/v1.wav` and `v2.wav`

were confirmed by artificial and realistic examples with speech signals. Further research aims necessary to find good applications of the current framework.

**Acknowledgement.** The authors acknowledge A. Ziehe, S. Harmeling, F. Meinecke and N. Murata for valuable discussion and EU PASCAL for partial funding.

## References

1. Hyvärinen, A., Hurri, J.: Blind separation of sources that have spatiotemporal variance dependencies. Signal Processing (2004) to appear.
2. Amari, S., Cardoso, J.F.: Blind source separation—semiparametric statistical approach. IEEE Trans. on Signal Processing **45** (1997) 2692–2700
3. Cardoso, J.F.: Multidimensional independent component analysis. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'98), Seattle, WA (1998)
4. Hyvärinen, A., Hoyer, P.O., Inki, M.: Topographic independent component analysis. Neural Computation **13** (2001)
5. Bach, F.R., Jordan, M.I.: Tree-dependent component analysis. In: Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference (UAI-2002). (2002)
6. Valpola, H., Harva, M., Karhunen, J.: Hierachical models of variance sources. (2003)
7. Bickel, P., Klaassen, C., Ritov, Y., Wellner, J.: Efficient and Adaptive Estimation for Semiparamtric Models. John Hopkins Univ. Press, Baltimore, MD (1993)
8. Amari, S., Kawanabe, M.: Information geometry of estimating functions in semiparametric statistical models. Bernoulli **3** (1997) 29–54
9. Godambe, V., ed.: Estimating Functions. Oxford Univ. Press, New York (1991)
10. Cardoso, J.F.: Estimating equations for source separation. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97). Volume 5., Munich, Germany (1997) 3449–3452
11. Ziehe, A., Müller, K.R.: TDSEP – an efficient algorithm for blind separation using time structure. In Niklasson, L., Bodén, M., Ziemke, T., eds.: Proc. of the 8th Int. Conf. on Artificial Neural Networks (ICANN '98), Berlin, Springer Verlag (1998) 675 – 680
12. A. Belouchrani, K. Abed Meraim, J.F.C., Moulines, E.: A blind source separation technique based on second order statistics. IEEE Trans. on Signal Processing (1996) 1009–1020
13. Hyvärinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Computation **9** (1997) 1483–1492
14. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non gaussian signals. IEE Proceedings-F (1993) 362–370
15. Pham, D.T., Cardoso, J.F.: Blind separation of instantaneous mixtures of non-stationary sources. In: Proc. of ICA2000, Helsinki, Finland (2000) 187–193
16. Amari, S., Cichocki, A., Yang, H.: A new learning algorithm for blind source separation. In: Advances in Neural Information Processing Systems 8. MIT Press (1996) 757–763