4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

AN ADAPTIVE NONLINEAR FUNCTION CONTROLLED BY ESTIMATED OUTPUT PDF FOR BLIND SOURCE SEPARATION

Kenji Nakayama *Akihiro Hirano* *Takayuki Sakai*

Dept. of Information Systems Eng., Faculty of Eng., Kanazawa Univ.
2-40-20, Kodatsuno, Kanazawa, 920-8667, Japan
nakayama@t.kanazawa-u.ac.jp

ABSTRACT

In blind source separation, convergence and separation performance are highly dependent on a relation between a probability density function (pdf) of the output signals y and nonlinear functions $f(y)$ used in updating coefficients of a separation block. This relation was analyzed based on kurtosis κ_4 of the output signals. The nonlinear functions, $\tanh(y)$ and y^3 have been suggested for super-Gaussian ($\kappa_4 \geq 0$) and sub-Gaussian ($\kappa_4 < 0$) distributions, respectively. Furthermore, an adaptive nonlinear function, which can be continuously controlled, was proposed. The nonlinear function is formed as a linear combination of y^3 and $\tanh(y)$. Their linear weights are controlled by the estimated κ_4. Although the latter can improve separation performance, its performance is still limited especially in difficult separation problems. In this paper, a new method is proposed. Nonlinear functions are directly controlled by the estimated pdf $p(y)$ of the separation block outputs y. $p(y)$ is expressed by a mixture Gaussian model, whose parameters are iteratively estimated sample by sample. $f(y)$ and $p(y)$ are related by the stability condition $f(y) = -(dp(y)/dy)/p(y)$. Blind source separation using 2∼ 5 channel music signals are simulated. The proposed method is superior to the above conventional methods. Three Gaussian functions are enough to express the output pdf.

1. INTRODUCTION

Recently, many kinds of information are transmitted and processed. At the same time, high quality is required. For this reason, signal processing including noise cancelation, echo cancelation, equalization of transmission lines, restoration of signals have been becoming very important technology. In some cases, we do not have enough information about signal sources and interference. Furthermore, their mixing process and transmission process are not well known in advance. Under these situations, blind source separation using statistical property of the signal sources has become important.

Jutten et all proposed a blind source separation algorithm based on statistical independence and symmetrical distribution of the signal sources [1]-[8]. Two kinds of stabilization methods have been proposed for Jutten's method [10],[17]. Furthermore, convolutive mixture models have been discussed [9],[20].

Convergence and separation performances are highly dependent on relation between a probability density function (pdf) of the output signals y and nonlinear functions $f(y)$ and $g(y)$, which are used in updating coefficients in a separation block. Optimum nonlinearity has been discussed based on kurtosis κ_4 of the output signals. Nonlinear functions $f(y) = a\tanh(y)$ and $f(y) = by^3$ have been suggested for super-Gaussian ($\kappa_4 \geq 0$) and sub-Gaussian ($\kappa_4 < 0$) distributions, respectively. Another function is fixed to $g(y) = y$ [14],[15],[16].

In music signals, the kurtosis κ_4 dynamically changes and takes both positive and negative values. In this case, the nonlinear functions must be controlled in an online manner. From the above discussion, one method is to switch two kinds of nonlinear functions based the polarity of κ_4. However, since the kurtosis of music signals dynamically changes, switching method is not stable. In order to solve this problem, an adaptive nonlinear function has been proposed, which is continuously controlled by the kurtosis [19]. The nonlinear function is formed as a linear combination of y^3 and $\tanh(y)$. Their linear weights are controlled by the estimated κ_4. Furthermore, a learned parametric mixture method has been proposed [18]. Although these methods can improve separation performance, it is still insufficient for difficult separation problems, in which the number of channels is large, and power of the interference is close to that of the main signal.

In this paper, a new adaptive nonlinear function is proposed. The nonlinear function is derived from the stability condition $f(y) = -(dp(y)/dy)/p(y)$. $p(y)$ is a pdf of the output signal y. $p(y)$ is estimated though an iterative method. Blind source separation using 2∼ 5 channel music signals are simulated, and usefulness of the proposed methods will be evaluated.

2. NETWORK AND LEARNING ALGORITHM

2.1. Network Structure

In this paper, the feedforward network shown in Fig.1 is taken into account.

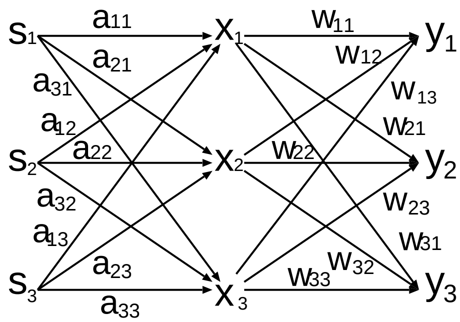

Fig. 1. Feedforward blind source separation.

The signal sources $s_i(n), i = 1, 2, ..., N$ are linearly combined using unknown weights a_{ji}, and are sensed at N points, resulting in $x_j(n)$. A general form is

$$x_j(n) = \sum_{i=1}^{N} a_{ji} s_i(n) \qquad (1)$$

The output of the separation block $y_k(n)$ is given by

$$y_k(n) = \sum_{j=1}^{N} w_{kj} x_j(n) \qquad (2)$$

This relation is expressed using vectors and matrices as follows:

$$\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{s}(n) \qquad (3)$$
$$\boldsymbol{y}(n) = \boldsymbol{W}\boldsymbol{x}(n) \qquad (4)$$
$$\boldsymbol{y}(n) = \boldsymbol{W}\boldsymbol{A}\boldsymbol{s}(n) = \boldsymbol{P}\boldsymbol{s}(n) \qquad (5)$$

$\boldsymbol{A}$ is an unknown mixing matrix and $\boldsymbol{W}$ is a weight matrix of the separation block. $\boldsymbol{P}$ expresses separation performance. If $\boldsymbol{P}$ has a single nonzero element in each raw and each column, and the number of the nonzero elements is equal to N, then the signal sources are completely separated at the outputs $\boldsymbol{y}(n)$.

2.2. Learning Algorithm

Learning algorithms given by [11],[12],[13] are employed in this paper.

$$\begin{aligned}\boldsymbol{W}(n+1) &= \boldsymbol{W}(n) + \eta(n)[\boldsymbol{\Lambda}(n) \\ &- f(\boldsymbol{y}(n))g(\boldsymbol{y}^T(n))]\boldsymbol{W}(n)\end{aligned}$$

$\boldsymbol{\Lambda}(n)$ is any positive definite scaling diagonal matrix. $f(\boldsymbol{y}(n))$ and $g(\boldsymbol{y}^T(n))$ are nonlinear functions, which will be optimized in an online manner based on distribution of the output $\boldsymbol{y}(n)$.

2.3. Relation between Nonlinear Functions and Kurtosis

Relations between nonlinear functions and 4th-order statistics ”kurtosis” have been discussed. Some kinds of nonlinear functions are selected based on kurtosis κ_4 as follows: [16]

$$\text{Kurtosis :} \quad \kappa_4 = \frac{E[(y-\bar{y})^4]}{E^2[(y-\bar{y})^2]} - 3 \qquad (6)$$
$$\text{Sub-Gaussian :} \quad \kappa_4 < 0 \quad f(y) = ay^3 \qquad (7)$$
$$a = \frac{1}{\kappa_4 + 3} \qquad (8)$$
$$\text{Super-Gaussian :} \quad \kappa_4 > 0 \quad f(y) = b\tanh y \qquad (9)$$
$$b = \frac{1}{E[y\tanh y]} \qquad (10)$$

a and b are scaling factors used to adjust the output power.

3. NONLINEAR FUNCTIONS CONTROLLED BY KURTOSIS

3.1. Two Nonlinear Functions Switched by Polarity of Kurtosis

The nonlinear functions discussed in sec2.3 can be used in an online manner. They can be switched taking the polarity of κ_4 into account. κ_4 is iteratively estimated by repeating the following equations.

$$\bar{y}(n) = (1-\alpha)\bar{y}(n-1) + \alpha y(n) \qquad (11)$$
$$\kappa(n) = (1-\alpha)\kappa(n-1) + \alpha(y(n) - \bar{y}(n))^4 \qquad (12)$$
$$\sigma^2(n) = (1-\alpha)\sigma^2(n-1) + \alpha(y(n) - \bar{y}(n))^2 \qquad (13)$$
$$\kappa_4(n) = \frac{\kappa(n)}{\sigma^4(n)} - 3 \qquad (14)$$
$$0 < \alpha \ll 1 \qquad (15)$$

3.2. Continuously Controlled Nonlinear Function by Kurtosis

A nonlinear function is formed as a linear combination of the nonlinear functions given in sec2.3. Their scaling factor are controlled by kurtosis [19].

$$f(y) = a\tanh y + (1-a)\frac{y^3}{4} \qquad (16)$$

$f(y)$ can cover a wide range of kurtosis. The other nonlinear function is fixed as folows:

$$g(y) = y \quad \text{fixed} \qquad (17)$$

In Eq.(16), a is controlled by the kurtosis, which is estimated by Eqs.(11) through (15). The pdf $p(y)$ of the output is derived from the stability condition

$$f(y) = -\frac{dp(y)/dy}{p(y)} \tag{18}$$

Using Eq.(16), $p(y)$ becomes

$$p(y) = e^{-\left[a(\log\cosh y + 0.25) + (1-a)\left(\frac{y^4}{16} + 0.45\right)\right]} \tag{19}$$

0.25 and 0.45 are used to normalize $p(y)$. Using this expression, the relation between a and κ_4 is numerically calculated, and is approximated by

$$a(n) = \frac{1 - e^{-2.5\kappa_4(n) - 2.1}}{1 + e^{-2.5\kappa_4(n) - 2.1}} \tag{20}$$

$a(n)$ is updated at each sample n.

Finally, the nonlinear function is controlled as

$$f(y(n)) = a(n)\tanh y + (1 - a(n))\frac{y^3(n)}{4} \tag{21}$$

$$g(y(n)) = y(n) \quad \text{fixed} \tag{22}$$

4. NONLINEAR FUNCTION CONTROLLED BY OUTPUT PDF

Since the nonlinear function described in sec3.2 is a combination of two kinds of functions, its form is rather limited. Furthermore, the corresponding pdf is limited to a single Gaussian function, which is not enough to approximate the output pdf.

In this paper, this idea is further extended by directly using the output pdf $p(y)$ for adjusting the nonlinear functions. $p(y)$ is iteratively estimated sample by sample. The nonlinear function $f(y)$ and the pdf $p(y)$ are also related by the stability condition Eq.(18).

$p(y)$ is formed as a mixture Gaussian model in order to express more general distribution.

$$p(y) = \sum_{i=1}^{M} C_i \phi(y; \mu_i, \sigma_i^2) \tag{23}$$

$$\phi(y; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left(\frac{-(y-\mu_i)^2}{2\sigma_i^2}\right) \tag{24}$$

The mean μ_i and the variance σ_i^2 are estimated through the following adaptive mixture (AM) algorithm [21].

$$g_i = \frac{C_i \phi_i(y(n+1))}{p(y(n+1))} \tag{25}$$

$$C_i(n+1) = C_i(n) + \frac{1}{v}(g_i - C_i(n)) \tag{26}$$

$$\mu_i(n+1) = \mu_i(n) + \frac{g_i}{v}(y(n+1) - \mu_i(n)) \tag{27}$$

$$\sigma_i^2(n+1) = \sigma_i^2(n) + \frac{g_i}{v + g_i - 1} \times \left(\frac{v}{v+g_i}(y(n+1) - \mu_i(n))^2 - \sigma_i(n)\right) \tag{28}$$

The above update is carried out for $i = 1 \sim M$. The parameters g, C, μ, σ^2 requires the initial guess. Numerical data will be shown in simulation.

Using $p(y)$ estimated by Eq.(23) and its derivative, shown below, $f(y)$ is obtained by Eq.(18) sample by sample.

$$\frac{dp(y)}{dy} = \sum_{i=1}^{M} C_i \frac{d\phi(y; \mu_i, \sigma_i^2)}{dy} \tag{29}$$

$$\frac{d\phi(y; \mu_i, \sigma_i^2)}{dy} = \frac{1}{\sqrt{2\pi\sigma_i^2}} \frac{-2(y-\mu_i)}{2\sigma_i^2} \times exp\left(\frac{-(y-\mu_i)^2}{2\sigma_i^2}\right) \tag{30}$$

5. SIMULATION AND DISCUSSIONS

5.1. Kurtosis, PDF and Nonlinear Functions

One example of the kurtosis of music signal is shown in Fig.2. It takes negative and positive values, and dynamically changes.

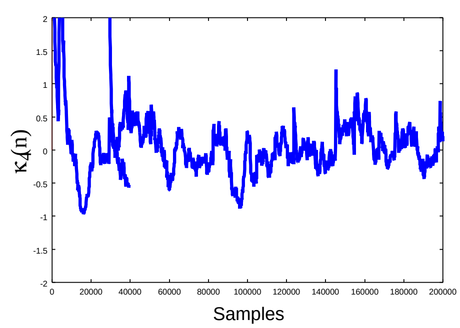

Fig. 2. One example of kurtosis of music signal.

Figure 3 shows the nonlinear function given by Eq.(16), where a is changed from 0 to 1.

Figure 4 shows the histogram of music signals and its estimated pdfs. In the upper and the lower figures, they are estimated at 35,000 samples and 50,000 samples, respectively. The pdf, which is estimated by using the kurtosis,

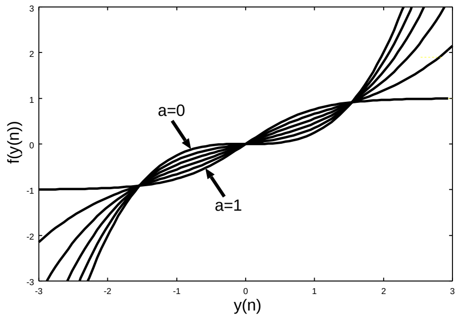

Fig. 3. Nonlinear function given by Eq.(16).

is obtained by solving the differential equation Eq.(18) for $p(y)$ using $f(y)$ controlled by the kurtosis. In estimating the pdf by the iterative method shown in sec4, the following parameters are used. From these results, the pdf estimated by using the kurtosis is not enough.

$$v = 500 \sim 800 \quad (31)$$

$$g_i(1) = 0.1 \quad (32)$$

$$\mu_i(1) = -0.8 \sim 1.0 \quad (33)$$

$$\sigma_i^2(1) = 0.05 \quad (34)$$

$$m = 10 \quad (35)$$

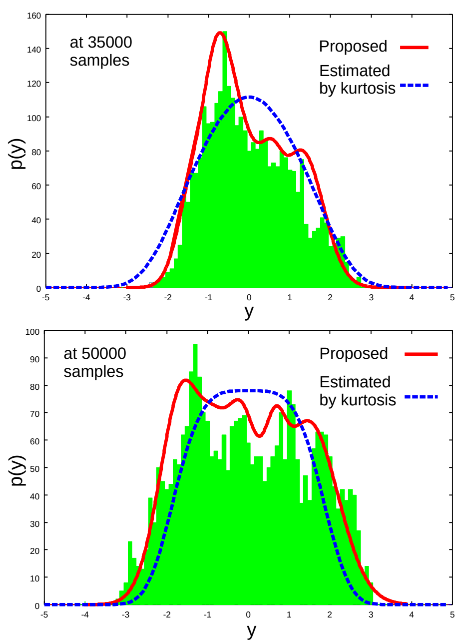

Fig. 4. Examples of histogram and pdf of music signal. pdf is estimated by using kurtosis and iterative method (Proposed).

Furthermore, the nonlinear functions $f(y)$ calculated by using the kurtosis and the pdf are shown in Fig.5. Both nonlinear functions are somewhat different.

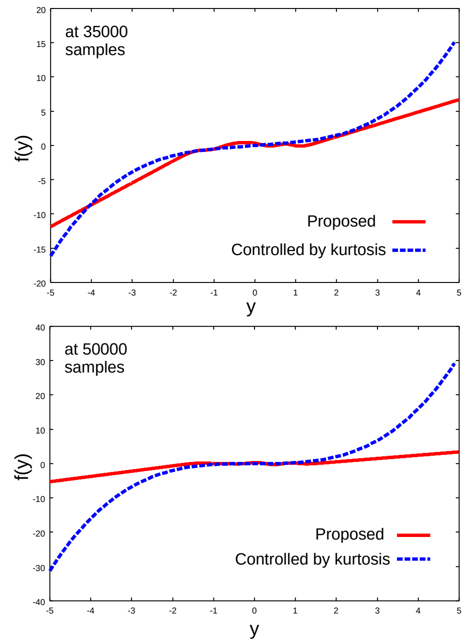

Fig. 5. Nolinear functions controlled by kurtosis and pdf (Proposed) in Fig.4

.

5.2. Blind Source Separation Performance

Separation performance is evaluated by the following SNR.

$$SNR = 10log\frac{\Sigma_{i,j\in\Omega_1}p_{ij}^2}{\Sigma_{i,j\in\Omega_2}p_{ij}^2} \quad (36)$$

p_{ij} is the elements of $\boldsymbol{P}$ in Eq.(5). Ω_1 includes the signal sources and Ω_2 includes the interference components. Examples of the mixing matrix for 3-channel and 5-channel are shown here.

$$A_{3ch} = \begin{bmatrix} 1.0 & 0.6 & 0.5 \\ 0.3 & 1.0 & 0.7 \\ 0.4 & 0.5 & 1.0 \end{bmatrix} \quad (37)$$

$$A_{5ch} = \begin{bmatrix} 1.0 & 0.3 & 0.6 & 0.7 & 0.3 \\ 0.5 & 1.0 & 0.4 & 0.6 & 0.4 \\ 0.6 & 0.5 & 1.0 & 0.5 & 0.4 \\ 0.3 & 0.6 & 0.5 & 1.0 & 0.5 \\ 0.4 & 0.6 & 0.7 & 0.3 & 1.0 \end{bmatrix} \quad (38)$$

Figure 6 shows SNR of blind source separation of music signals using the three kinds of methods. The upper figure shows 2-channel case and the followings are 3-channel, 4-channel and 5-channel, in this order. In all cases, the proposed nonlinear function controlled by the output pdf is superior to the others.

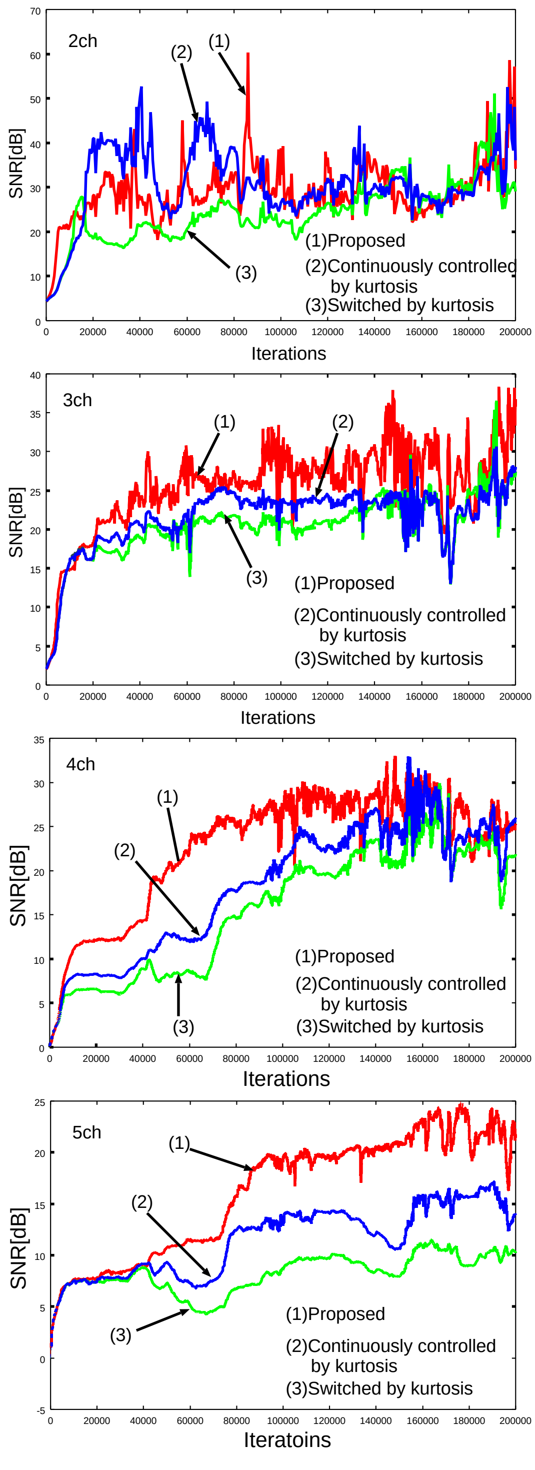

Fig. 6. SNR of blind source separation of music signals by three kinds of methods. From top to bottom, 2-ch, 3-ch, 4-ch and 5-ch cases are arranged in this order.

5.3. Number of Gaussian Functions

The above simulations were carried out by using $M = 10$, that is ten Gaussian functions $\phi(y;\mu_i,\sigma_i^2)$ are used in Eq.(23). Effect of the number of the Gaussian functions is evaluated. The simulation results in the 5-channel case are shown in Fig.7. The number in these figure indicate the number of Gaussian functions. SNR obtained by using 3∼10 Gaussian functions are almost the same. From this result, 3 Gaussian functions are enough in the proposed method. The conventional nonlinear function, continuously controlled by the kurtosis, equivalently uses only a single Gaussian function. Therefore, by extending the number of the Gaussian functions used in the pdf, separation performance can be effectively improved.

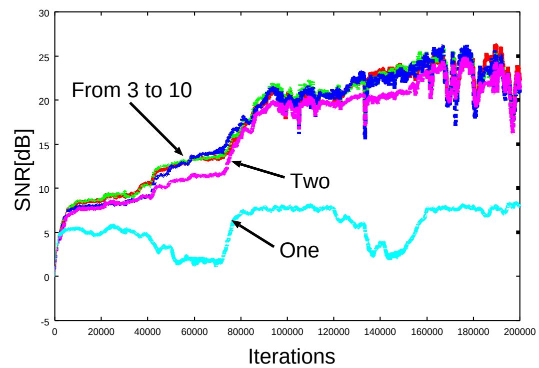

Fig. 7. Effect of number of Gaussian functions on separation performance in 5-channel case.

5.4. Effect of Mixing Matrix

The mixing process is evaluated by the following SNR.

$$SNR_{mix} = 10log\frac{\Sigma_{i,j\in\Omega_1} a_{ij}^2}{\Sigma_{i,j\in\Omega_2} a_{ij}^2} \quad (40)$$

a_{ij} is the elements of $\boldsymbol{A}$ in Eq.(3). Ω_1 includes the main path and Ω_2 includes the interference paths. SNR_{mix} can express dificulty of separation. Furthermore, separation performance is highly dependent on the variance of the components in $\boldsymbol{A}$. a_{ij} in Ω_1 are fixed to 1 and the others are changed. The standard deviation of a_{ij} in Ω_2 is evaluated. The separation performance with different standard deviation are simulated for 5-channel case, and are shown in Fig.8. The proposed method is not affected by the variance, while the other methods are affected. Especially, the switching method is not useful for this kind of difficult separation problem.

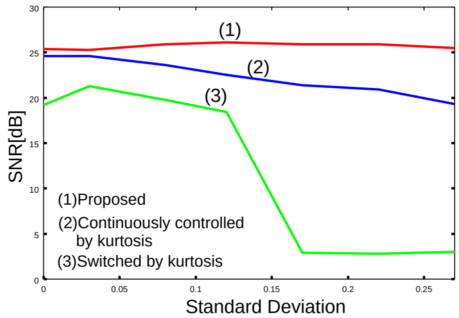

Fig. 8. Separation performance with different standard deviation of mixing matrix in 5-channel case.

6. CONCLUTIONS

In blind source separation, the nonlinear functions are used in updating the coefficients of the separation block. In this paper, an adaptive nonlinear function has been proposed. The nonlinear function is adjusted by using the output pdf following the stability condition. The output pdf, which is expressed by the mixture Gaussian form, is iteratively estimated. Simulation results using 2$\sim$5-channel music signals have been shown. The proposed method can achieve high separation performance compared with the conventional adaptive nonlinear functions, including the switching method and the continuous control method by the kurtosis. In the proposed method, 3 Gaussian functions in the mixture model are enough in expressing the output pdf.

7. REFERENCES

[1] C.Jutten, J.Herault and A.Guerin, ”IIN.C.A: An independent components analyzer based on an adaptive neuromimetic network”, in: J.Degmongeot, T.Herve, V.Raille and C.Roche, eds., Artificial Intelligence and Cognitive Science, Manhester Univ. Press, 1988.

[2] C.Jutten and J.Herault, ”Analog implementation of permanent unsupervised learning algorithm”, Proc. NATO Advanced Research Workshop on Neurocomputing, Les Arcs, France, pp.145-152, Feb.27-March 3, 1989.

[3] P.Comon, ”Separation of stochastic process whose linear mixtures observed”, Proc. ONR- NSF-IEEE Workshop on Higher Spectral Analysis Vail, pp.174-179, June 28-30, 1989.

[4] P.Comon, ”Separation of sources using higher-order cumulants”, SPIE Conference, Vol.1152, Advanced Algorithms and Architectures for Signal Processing IV, San Diego, pp. 170-181, August 6-11, 1989.

[5] J.F.Cardoso, ”Eigen structure of the 4th order cumulant tensor with application to the blind source separation problem”, ICASSP Proc. pp. 2655-1658.

[6] C.Jutten and J. Herault, ”Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture”, Signal Proc. 24, pp.1-10, 1991.

[7] P.Comon, C.Jutten and J.Herault, ”Blind separation of sources, Part II: Problems statement”, Signal Proc. 24, pp.11-20, 1991.

[8] E.Sorouchyari, ”Blind separation of sources, Part III: Stability analysis”, Signal Proc., 24, pp.21-29, 1991.

[9] H.L.Nguyen Thi and C.Jutten, ”Blind source separation for convolutive mixtures”, Signal Processing, vol.45, no.2, pp.209–229, March 1995.

[10] K.Nakayama, A.Hirano and M.Nitta, ”A constraint learning algorithm for blind source separation”, Proc. IJCNN’2000, pp.24-27, July, 2000.

[11] S.Amari, A.Cichocki and H.H.Yang, ”Recurrent neural networks for blind separation of sourdes”, Proc. NOLTA-95, Las Vegas, pp.37-42, 1995.

[12] A.Cichocki, S.Amari, M.Adachi, W.Kasprzak, ”Self-adaptive neural networks for blind separation of sources”, Proc. ISCAS’96, Atlanta, pp.157-161, 1996.

[13] S.Amari, T.Chen and A.Cichocki, ”Stability analysis of learning algorithms for blind source separation”, Neural Networks, vol.10, no.8, pp.1345-1351, 1997.

[14] J.F.Cardoso, ”Informax and maximum likelihood for source separation”, IEEE Signal Processing Letter, vol.4, no.4, pp.112-114, Apr. 1997.

[15] A.Mansour and C.Jutten, ”What should we say about the kurtosis”, IEEE Signal Processing Letters, vol.6, no.12, pp.321-322, Dec. 1999.

[16] H.Mathis and S.C.Douglas, ”On optimal and universal nonlinearities for blind signal separation”, IEEE Proc. ICASSP’01, MULT-P3.3, May 2001.

[17] K.Nakayama, A.Hirano and T.Sakai, ”A pair-channel learning algorithm with constraints for multi-channel blind separation”, Proc. IJCNN’01, July 2001.

[18] L.Xu, C.C.Cheung and S.Amari, ”Learned parametric mixture based ICA algorithm”, Neurocomputing 22, pp.69-80, 1998.

[19] K.Nakayama, A.Hirano and T.Sakai, ”An adaptive nonlinear function controlled by kurtosis for blind source separation”, Proc. IJCNN’2002, pp.1234-1239, May. 2002.

[20] K.Nakayama, A.Hirano and A.Horita, ”A learning algorithm for convolutive blind source separation with transmission delay constraint”, Proc. IJCNN’2002, pp.1287-1292, May 2002.

[21] Carey E.Priebe,”Adaptive mixtures”, Journal of the American Statistical Association, vol.89, no.427, pp.796-806, Sept. 1994.