BLIND SOURCE SEPARATION FOR CONVOLUTIVE MIXTURES EXPLOITING NONGAUSSIANITY, NONWHITENESS, AND NONSTATIONARITY

Herbert Buchner, Robert Aichner, Walter Kellermann

Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg
Cauerstr. 7, D-91058 Erlangen, Germany
{buchner,aichner,wk}@LNT.de

ABSTRACT

Generally, there are three types of approaches for blind source separation (BSS) on time series: exploitation of the nonwhiteness, the nonstationarity, and the nongaussianity of the source signals. While methods utilizing the first two properties are usually based on second order statistics (SOS), one needs higher order statistics (HOS) to take into account nongaussianity. In this paper, we combine all these three fundamental approaches (the three 'Non's') for convolutive mixtures to one generic framework, the TRINICON algorithm ('Triple-N ICA for convolutive mixtures'). This is done by introducing an appropriate matrix formulation, combined with the use of multivariate probability densities for considering the time-dependencies of the source signals. It can be shown that our previously introduced generic SOS algorithm follows from the TRINICON as the optimum SOS algorithm. For the general HOS case, we introduce an efficient solution using models for correlated spherically invariant random processes (SIRPs) which are very well suited for a number of signals including speech. In this paper, we consider exclusively time-domain algorithms, but the framework can be extended to the frequency domain.

1. INTRODUCTION

The problem of separating convolutive mixtures of unknown time series arises in several application domains, a prominent example being the so-called cocktail party problem, where we want to recover the speech signals of multiple speakers who are simultaneously talking in a room. The room may be very reverberant due to reflections on the walls, i.e., the original source signals $s_q(n)$, $q = 1, \ldots, Q$ of our separation problem are filtered by a multiple input and multiple output (MIMO) system before they are picked up by the sensors. In the following, we assume that the number Q of source signals $s_q(n)$ equals the number of sensor signals $x_p(n)$, $p = 1, \ldots, P$ (Fig. 1). An M-tap mixing system is thus described by

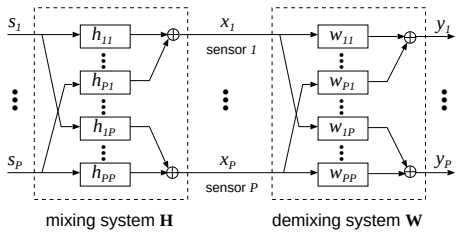

Fig. 1. Linear MIMO model for BSS.

$$x_p(n) = \sum_{q=1}^{P} \sum_{\kappa=0}^{M-1} h_{qp}(\kappa) s_q(n-\kappa), \tag{1}$$

where $h_{qp}(\kappa)$, $\kappa = 0, \ldots, M-1$ denote the coefficients of the filter from the q-th source to the p-th sensor.

In BSS, we are interested in finding a corresponding demixing system according to Fig. 1, where the output signals $y_q(n)$, $q = 1, \ldots, P$ are described by

$$y_q(n) = \sum_{p=1}^{P} \sum_{\kappa=0}^{L-1} w_{pq}(\kappa) x_p(n-\kappa). \tag{2}$$

It can be shown (see, e.g., [1]) that ideally, the MIMO demixing system coefficients $w_{pq}(\kappa)$ can in fact reconstruct the sources up to an unknown permutation and an unknown filtering of the individual signals, where L should be chosen at least equal to M.

Different approaches exist to blindly estimate the P^2L MIMO coefficients $w_{pq}(\kappa)$ by utilizing one of the following properties [1]:

(i) Nonwhiteness property by simultaneous diagonalization of output correlation matrices over multiple time-lags, e.g., [2]

(ii) Nonstationarity property by simultaneous diagonalization of short-time output correlation matrices at different time intervals, e.g., [3]-[5]

(iii) Nongaussianity property using higher order statistics for independent component analysis (ICA), e.g., [6]-[10]

Although it is commonly believed that each one of these properties is sufficient for separation, it has recently been demonstrated for (i) and (ii) using SOS that in practical scenarios, the combination of these criteria can lead to improved performance [11, 12]. This contribution is a further generalization of our previous work [13] combining all three properties into an efficient algorithm. The resulting TRINICON algorithm can thus cope with any kind of source signals (except stationary white Gaussian noise signals).

2. GENERIC HOS-BASED BSS ALGORITHM FOR CONVOLUTIVE MIXTURES

2.1. Matrix notation for convolutive mixtures

To derive an algorithm for block processing of convolutive mixtures, we first need to reformulate the convolution (2) in the fol-

lowing matrix form:

$$\mathbf{y}(m,j) = \mathbf{x}(m,j)\mathbf{W}(m), \tag{3}$$

where m denotes the block index, and $j = 0, \cdots, N-1$ is a time-shift index within a block of length N, and

$$\begin{aligned}
\mathbf{x}(m,j) &= [\mathbf{x}_1(m,j), \ldots, \mathbf{x}_P(m,j)], && (4)\\
\mathbf{y}(m,j) &= [\mathbf{y}_1(m,j), \ldots, \mathbf{y}_P(m,j)], && (5)\\
\mathbf{W}(m) &= \begin{bmatrix} \mathbf{W}_{11}(m) & \cdots & \mathbf{W}_{1P}(m) \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1}(m) & \cdots & \mathbf{W}_{PP}(m) \end{bmatrix}, && (6)\\
\mathbf{x}_p(m,j) &= [x_p(mL+j), \ldots, x_p(mL-2L+1+j)] && (7)\\
\mathbf{y}_q(m,j) &= [y_q(mL+j), \ldots, y_q(mL-D+1+j)] && (8)\\
&= \sum_{p=1}^{P} \mathbf{x}_p(m,j)\mathbf{W}_{pq}(m). && (9)
\end{aligned}$$

D in (8) denotes the number of lags taken into account (c.f. property (i)) as shown below. $\mathbf{W}_{pq}(m)$ denotes a Sylvester matrix that contains all coefficients of the respective filter:

$$\mathbf{W}_{pq}(m) = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{10}$$

2.2. Cost function and algorithm derivation

A generic SOS algorithm for convolutive mixtures has been derived rigorously from a cost function that explicitly contains correlation matrices that include several time-lags (c.f. property (i)) under the assumption of short-time stationarity (c.f. property (ii)) [13]. For property (iii), higher order statistics have to be considered. Higher-order approaches for BSS can be divided into three classes [1]: maximum likelihood (ML) estimation, minimization of the mutual information (MMI) among the output signals, and maximization of the entropy (ME/'infomax'). Although all of these HOS approaches lead to similar update rules, MMI can be regarded as the most general one [9].

Based on a generalization of Shannon's mutual information [14], we now define the following cost function taking into account all three fundamental signal properties (i)-(iii):

$$\begin{aligned}
\mathcal{J}(m) &= -\sum_{i=0}^{m} \beta(i,m) \\
&\quad \cdot \frac{1}{N} \sum_{j=0}^{N-1} \{\log\left(\hat{p}_D(\mathbf{y}_1(i,j)) \cdot \ldots \cdot \hat{p}_D(\mathbf{y}_P(i,j))\right) \\
&\quad - \log\left(\hat{p}_{PD}(\mathbf{y}_1(i,j)\Lambda_1, \ldots, \mathbf{y}_P(i,j)\Lambda_P)\right)\}, \qquad (11)
\end{aligned}$$

where $\hat{p}_D(\cdot)$ and $\hat{p}_{PD}(\cdot)$ are estimated or assumed *multivariate* probability density functions (pdfs) of dimensions D and PD, respectively. Furthermore, D is the memory length, i.e., the number of time-lags to model the nonwhiteness of the P signals as above. Note also that the sequence of these pdf estimates completely describes any multichannel stochastic process with the assumption of short-time stationarity over length-N blocks (this assumption is reasonable for many real-world signals such as speech). The expectation operator of the mutual information [14] is replaced in (11) by short-time averages within these blocks. β is a window function that is normalized according to $\sum_{i=0}^{m} \beta(i,m) = 1$ which allows off-line, on-line, and block-online implementations of the algorithms (e.g., $\beta(i,m) = (1-\lambda)\lambda^{m-i}$ leads to an efficient on-line version allowing for tracking in time-varying environments [15]). The matrices Λ_p, $p = 1, \ldots, P$ represent filters on the output signals to further improve the convergence for nonstationary sources by removing magnitude constraints on the output signals (so that the demixing matrix remains in its so-called 'equivalence class' containing all possible solutions due to the fundamental scaling indeterminacy of BSS) [10].

It can be shown (after a tedious but straightforward derivation) that by taking the *natural gradient* [7] of $\mathcal{J}(m)$ with respect to the demixing filter matrix $\mathbf{W}(m)$ [13],

$$\Delta\mathbf{W} \propto \mathbf{W}\mathbf{W}^H \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}, \tag{12}$$

we obtain the following generic TRINICON update rule:

$$\begin{aligned}
\mathbf{W}(m) &= \mathbf{W}(m-1) - \mu\Delta\mathbf{W}(m), && (13)\\
\Delta\mathbf{W}(m) &= -\sum_{i=0}^{m} \beta(i,m) \sum_{j=0}^{N-1} \mathbf{W}(i) \\
&\quad \cdot \Big(\mathbf{y}^H(i,j)\Phi(\mathbf{y}(i,j)) - \Lambda(i,j)\Big) && (14)
\end{aligned}$$

with the *multivariate score function*

$$\Phi(\mathbf{y}(i,j)) = \left[\frac{\frac{\partial \hat{p}_D(\mathbf{y}_1(i,j))}{\partial \mathbf{y}_1(i,j)}}{\hat{p}_D(\mathbf{y}_1(i,j))}, \ldots, \frac{\frac{\partial \hat{p}_D(\mathbf{y}_P(i,j))}{\partial \mathbf{y}_P(i,j)}}{\hat{p}_D(\mathbf{y}_P(i,j))}\right]. \tag{15}$$

The constraint matrix $\Lambda(i,j)$ is composed from the filters $\Lambda_p(i,j)$. For $\Lambda(i,j) = \mathbf{I}$ we obtain the so-called holonomic algorithm, and $\Lambda(i,j) = \mathrm{bdiag}\{\mathbf{y}^H(i,j)\Phi(\mathbf{y}(i,j))\}$ yields the corresponding generalization of the nonholonomic [10] algorithm with improved convergence characteristics for nonstationary sources. Here, the bdiag operator sets all channel-wise cross-terms to zero.

3. RELATION TO THE GENERIC SOS-BASED BSS ALGORITHM

In [13] we presented a generic SOS BSS approach for convolutive mixtures in the time-domain and frequency-domains based on Oppenheim's inequality. Several popular algorithms for convolutive mixtures such as [5, 12] have turned out to be approximations of this framework. Here, we consider only the time domain for simplicity. The update in the case of two sources and two sensors reads

$$\begin{aligned}
\Delta\mathbf{W}(m) &= \sum_{i=0}^{m} \beta(i,m)\mathbf{W}(i) \\
&\quad \cdot \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12}(i)\mathbf{R}_{22}^{-1}(i) \\ \mathbf{R}_{21}(i)\mathbf{R}_{11}^{-1}(i) & \mathbf{0} \end{bmatrix}, \qquad (16)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{R}_{pq}(i) &= \mathbf{Y}_p^H(i)\mathbf{Y}_q(i), && (17)\\
\mathbf{Y}_q(i) &= [\mathbf{y}_q^T(i,0), \ldots, \mathbf{y}_q^T(i,N-1)]^T. && (18)
\end{aligned}$$

It can be shown that this generic SOS-based BSS follows from the TRINICON algorithm by assuming multivariate Gaussian pdfs

$$\hat{p}_D(\mathbf{y}_p(i,j)) = \frac{1}{\sqrt{2\pi^D \det(\mathbf{R}_{pp}(i))}} e^{-\frac{1}{2}\mathbf{y}_p(i,j)\mathbf{R}_{pp}^{-1}(i)\mathbf{y}_p^H(i,j)}. \tag{19}$$

As a result, we may now draw the important conclusion that the algorithm in [13] is in fact the optimum SOS algorithm for convolutive mixtures in the sense of minimum mutual information or ML, which also implies asymptotic Fisher-efficiency [1]. Another interesting finding is that the SOS BSS algorithm turns out to be nonholonomic for both, $\Lambda(i,j) = \mathbf{I}$, and $\Lambda(i,j) = \mathrm{bdiag}\{\mathbf{y}^H(i,j)\Phi(\mathbf{y}(i,j))\}$ confirming its good performance for speech sources.

4. GENERIC BSS INCORPORATING MODELS FOR NON-STATIONARY AND CORRELATED SPHERICALLY INVARIANT RANDOM PROCESSES

The update rule (14) provides a very general basis for BSS of convolutive mixtures. However, to apply it in a real-world scenario, an appropriate multivariate score function (15) has to be determined, i.e., we have to handle P high-dimensional multivariate pdfs $\hat{p}_D(\mathbf{y}_p(i,j))$, $p = 1, \ldots, P$. In general, this is a very challenging task, as it includes all corresponding higher-order cumulants (including time lags).

Moreover, we want to retain the inherent normalization property of the generic SOS BSS [13], as shown by the inverses of the lagged autocorrelation matrices in (16), and also the nonholonomic form of the update.

Fortunately, there is an efficient solution for these problems by assuming so-called spherically invariant random processes (SIRPs). These models are representative for a wide class of stochastic processes. It has been shown that speech signals in particular can very accurately be represented by SIRPs [17].

4.1. Spherically invariant random processes

One of the great advantages arising from the SIRP model is that multivariate pdfs can be derived analytically (c.f. Sect. 4.3) from the corresponding univariate probability density function together with the (lagged) correlation matrices.

The correlation matrices can be estimated from the data as in the case of the generic SOS BSS, while for the univariate pdf, we can assume one of the well-known functions for speech signals, e.g., the Laplacian density.

The general form of correlated SIRPs of D-th order is given with a properly chosen function $f_D(\cdot)$ by [17]

$$\hat{p}_D(\mathbf{y}_p(i,j)) = \frac{1}{\sqrt{\pi^D \det(\mathbf{R}_{pp}(i))}} f_D\left(\mathbf{y}_p(i,j)\mathbf{R}_{pp}^{-1}(i)\mathbf{y}_p^H(i,j)\right) \tag{20}$$

As the best known example, the multivariate Gaussian can be viewed as a special case of the class of SIRPs. To calculate the score function for SIRPs in general, we employ the chain rule [16]

$$\frac{\frac{\partial \hat{p}_D(\mathbf{y}_p(i,j))}{\partial \mathbf{y}_p(i,j)}}{\hat{p}_D(\mathbf{y}_p(i,j))} = \underbrace{\left[\frac{1}{f_D(s)}\frac{\partial f_D(s)}{\partial s}\right]}_{:=\phi_D(s)} \mathbf{y}_p(i,j)\mathbf{R}_{pp}^{-1}(i), \tag{21}$$

where $s = \mathbf{y}_p\mathbf{R}_{pp}^{-1}\mathbf{y}_p^H$. For convenience, we call the scalar function $\phi_D(s)$ the *SIRP score*.

4.2. Incorporation of SIRPs into the generic BSS

Having derived the multivariate score function for the SIRP model (21), we can now introduce it into the generic TRINICON update equation (14). In the 2-by-2 case, this leads to the following expression for the nonholonomic TRINICON-SIRP update:

$$\begin{aligned}\Delta\mathbf{W}(m) &= \sum_{i=0}^{m} \beta(i,m)\mathbf{W}(i) \\ &\quad \cdot \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{R}}_{12}(i)\mathbf{R}_{22}^{-1}(i) \\ \tilde{\mathbf{R}}_{21}(i)\mathbf{R}_{11}^{-1}(i) & \mathbf{0} \end{bmatrix},\end{aligned} \tag{22}$$

where the modified matrices $\tilde{\mathbf{R}}_{pq}$, $p \neq q$ are given by

$$\begin{aligned}\tilde{\mathbf{R}}_{pq}(i) &= -\sum_{j=0}^{N-1} \phi_D\left(\mathbf{y}_q(i,j)\mathbf{R}_{qq}^{-1}(i)\mathbf{y}_q^H(i,j)\right) \\ &\quad \cdot \mathbf{y}_p^H(i,j)\mathbf{y}_q(i,j) \qquad (23) \\ &= \mathbf{Y}_p^H(i)\tilde{\Lambda}_q(i)\mathbf{Y}_q(i), \qquad (24)\end{aligned}$$

$$\tilde{\Lambda}_q(i) = -\phi_D\left(\mathrm{diag}\left(\mathbf{Y}_q(i)\mathbf{R}_{qq}^{-1}(i)\mathbf{Y}_q^H(i)\right)\right), \tag{25}$$

$$\phi_D(s) = \frac{f_D'(s)}{f_D(s)}. \tag{26}$$

The SIRP score function in (25) is applied element-wise to the matrix in its argument.

From the update equation (22), we see that the inherent normalization of the generic SOS algorithm is retained with the SIRP model, and from (25) we see again the close relation to the SOS algorithm, which is obtained by setting $\tilde{\Lambda}_q(i) = \mathbf{I}$.

4.3. Calculation of optimum SIRP score functions for separation of convolutive mixtures

To derive a TRINICON-SIRP realization using (26) we need an analytical expression of the multivariate pdf (20). As noted above, for SIRPs, this expression can actually be derived from the univariate pdf [17].

To achieve this, a key observation is that a pdf of a certain order must be a marginal density of a higer-order pdf. Thus, they are related by an integral transformation. A simple and elegant way to carry out this transformation is to employ Meijer's G-functions. Most elementary and non-elementary mathematical functions can be expressed as G-functions (for their definition and specific notation, see, e.g., [17]). For practical purposes, symbolic calculations with G-functions can be made using mathematical software, e.g., Maple. As shown in [17] the procedure is as follows:

1.) Express the assumed univariate pdf as a G-Function. Most of the known models can be expressed in the following form:

$$p_1(y) = A \cdot G_{p\,q}^{m\,n}\left(\lambda y^2 \,\middle|\, \begin{matrix} a_p \\ b_q \end{matrix}\right).$$

For example, the laplacian pdf is given by $b_1 = 0$, $b_2 = 1/2$, $A = 1/\sqrt{2\pi}$, $\lambda = 1/2$.

2.) From this expression, the corresponding multivariate pdf (20) can be directly given by the following rule:

$$f_D(s) = \sqrt{\pi} A s^{(1-D)/2} \cdot G_{p+1\,q+1}^{m+1\,n}\left(\lambda s \,\middle|\, \begin{matrix} a_p, & 0 \\ \frac{D-1}{2}, & b_q \end{matrix}\right).$$

3.) Convert the multivariate pdf back to known functions (e.g., using a mathematical software).

Following this procedure, we obtain, e.g., as the *optimum SIRP score for a univariate Laplacian pdf*:

$$\phi_D(s) = \frac{1}{D - \sqrt{2s}\frac{K_{D/2+1}(\sqrt{2s})}{K_{D/2}(\sqrt{2s})}}, \qquad (27)$$

where $K_\nu(\cdot)$ denotes the ν-th order modified Bessel function of the second kind.

5. SIMULATION RESULTS

We conducted our experiments using speech signals from the TIMIT database convolved with impulse responses of a real room with reverberation time $T_{60} \approx 150$ ms. Note that the reverberation time (or filter length) is not very critical here due to the inherent normalization property discussed in Sect. 4 and [13]. The sampling rate was $f_s = 16$ kHz. We used a two-element microphone array with an inter-element spacing of 16 cm. We consider in this paper only time-domain adaptation algorithms. For the filter adaptation (offline) we used both, the generic SOS algorithm in the time-domain [13], and the proposed generic HOS algorithm with SIRP model from the Laplacian pdf. We chose the following parameters: $L = 512$, $N = 1024$, $D = 512$ (note that N has to be chosen greater than D to get improved estimates in the HOS case). To evaluate the performance, as shown in Fig. 2 we used the signal-to-interference ratio (SIR), defined as the ratio of the signal power of the target signal to the signal power from the jammer signal. For Fig. 2, the stepsizes have been maximized up to the stability margin.

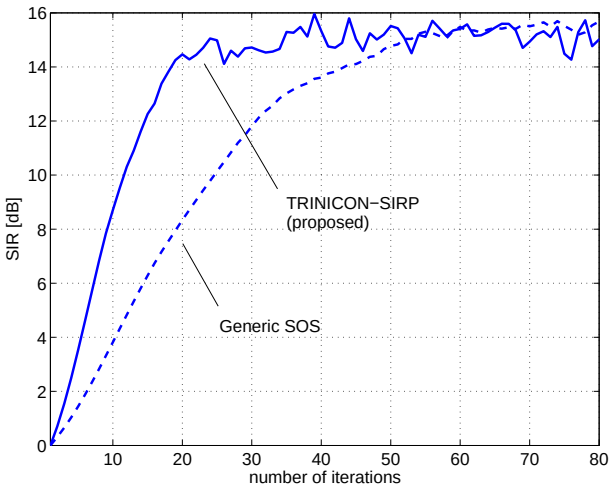

Fig. 2. Simulation results for two sensors and $L = 512$.

6. CONCLUSIONS

We presented a generic algorithm exploiting all three fundamental statistical source properties for BSS, and an efficient solution incorporating a model for spherically invariant random processes. In order to obtain computationally efficient real-time versions which have not been considered in this paper, a corresponding frequency-domain formulation can be derived as in [13].

7. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley & Sons, Inc., New York, 2001.

[2] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, pp. 3634-3636, 1994.

[3] E. Weinstein, M. Feder, and A. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE Trans. on Speech and Audio Processing*, vol 1, no. 4, pp. 405-413, Oct. 1993.

[4] H.-C. Wu and J. C. Principe,"Simultaneous diagonalization in the frequency domain (SDIF) for source separation," in *Proc. ICA*, pp. 245-250, 1999.

[5] C.L. Fancourt and L. Parra, "The coherence function in blind source separation of convolutive mixtures of non-stationary signals," in Proc. *Int. Workshop on Neural Networks for Signal Processing*, 2001.

[6] A.J. Bell and T.J. Seijnowski, "An information-maximization approach for blind signal separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

[7] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in neural information processing systems*, 8, Cambridge, MA, MIT Press, 1996, pp. 757-763.

[8] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009-2025, Oct. 1998.

[9] H.H. Yang and S. Amari, "Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information," *Neural Computation*, vol. 9, pp. 1457-1482, 1997.

[10] S. Amari, T.-P. Chen, and A. Cichocki, "Nonholonomic orthogonal learning algorithms for blind source separation," *Neural Computation*, vol. 12, no. 6, pp. 1463-1484, 2000.

[11] T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA for blind source sepoaration," in *Proc. European Signal Processing Conference*, vol. 2, pp. 15-18, Sep. 2002.

[12] R. Aichner et al., "Time-domain blind source separation of non-stationary convolved signals with utilization of geometric beamforming," in Proc. *Int. Workshop on Neural Networks for Signal Processing*, Martigny, Switzerland, 2002.

[13] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," *Proc. Int. Symposium on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, Apr. 2003.

[14] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley & Sons, New York, 1991.

[15] R. Aichner et al., "On-line time-domain blind source separation of nonstationary convolved signals," *Proc. Int. Symposium on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, Apr. 2003.

[16] D.A. Harville, *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag, New York, 1997.

[17] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing* vol. 12, pp. 119-141, 1987.