Combined Blind/Nonblind Source Separation Based on the Natural Gradient

Marcel Joho, Heinz Mathis, and George S. Moschytz

***Abstract*—It is a known fact that blind algorithms have convergence times of an order of magnitude longer than their nonblind counterparts. However, as shown in this letter, the knowledge of a subset of signals can greatly accelerate the convergence of blind source separation. The convergence behavior of the proposed algorithm is compared with the blind-only case.**

***Index Terms*—Adaptive blind source separation, blind signal processing, natural gradient learning algorithm, semi-blind learning algorithm, teleconferencing, virtual sensors.**

I. INTRODUCTION

THE CLASSICAL blind separation problem of an instantaneous mixture of independent source signals can be formulated as follows. The sensor signals $\mathbf{x} = [x_1, \ldots, x_{M_s}]^T$ to be processed are linear mixtures of the original source signals $\mathbf{s} = [s_1, \ldots, s_{M_s}]^T$ weighted by scalars

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

i.e., the source signals $\mathbf{s}$ are mixed by the mixing matrix $\mathbf{A}$. Neither $\mathbf{s}$ nor $\mathbf{A}$ are known. Separation can be achieved through a blind adaptive algorithm adjusting the coefficients of the separation matrix $\mathbf{W}$. The output of the algorithm is therefore

$$\mathbf{u} = [u_1, \ldots, u_{M_s}]^T = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{P}\mathbf{s}. \tag{2}$$

In order to separate the signals effectively, $\mathbf{WA} = \mathbf{P}$ should approximate a scaled permutation matrix. A possible update equation for the separation matrix $\mathbf{W}$ that shows good convergence properties was derived in [1] based on the *natural gradient*

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu \left(\mathbf{I} - \mathbf{g}(\mathbf{u})\mathbf{u}^T\right) \mathbf{W}_t \tag{3}$$

$$= \mathbf{W}_t + \mu \Delta \mathbf{W}_t \tag{4}$$

where μ is the step size and $\mathbf{I}$ the identity matrix. $\mathbf{g}(.)$ is a nonlinearity, whose choice depends on the probability density function (pdf) of the source signals $\mathbf{s}$. In certain applications, some of the source signals of a mixture are known, e.g., in systems with feedback. For example, in acoustical applications (e.g., teleconferencing), the output signal of a loudspeaker may leak into several microphones that are placed loosely in front of a group of speakers in the same room, shown in Fig. 1. Although the speakers' signals have to be separated blindly, speaker separation can be accelerated if the known reference signal of the loudspeaker is incorporated into the blind algorithm.

Manuscript received February 28, 2000. This work was supported in part by the Swiss Federal Institute of Technology, ETH Zurich, Switzerland. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Messer-Yaron.

M. Joho is with Phonak Inc., Champaign, IL 61820 USA (e-mail: joho@ieee.org).

H. Mathis and G. S. Moschytz are with the Signal and Information Processing Laboratory, Swiss Federal Institute of Technology, ETH, Zurich, Switzerland (e-mail: mathis@isi.ee.ethz.ch; moschytz@isi.ee.ethz.ch).

Publisher Item Identifier S 1070-9908(01)06118-1.

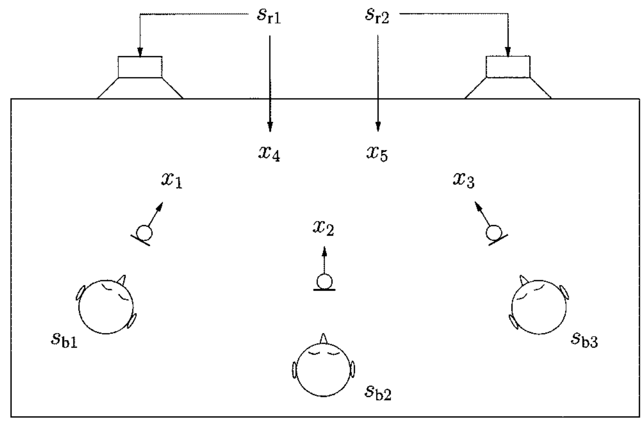

Fig. 1. Teleconferencing setup. Some signals are unknown (speakers) and some signals are known (loudspeaker signals). The loudspeaker signals s_{r1} and s_{r2} are directly conveyed to the virtual sensors x_4 and x_5, respectively.

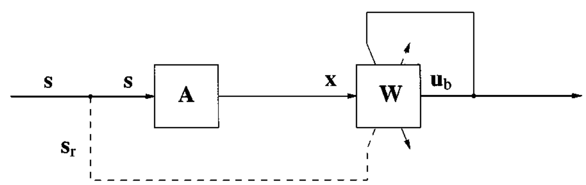

Fig. 2. Basic architecture for a combined blind/nonblind separation algorithm. The dashed line shows the path with the subset of known signals. Discarding this path leads to the more familiar blind-only case.

Algorithms exploiting information of partially known signals were introduced in [2] and [3]. Two approaches were presented. The first one simply combines an echo-canceller (in a spatial sense) with a following blind stage. The second one uses an equalizer-like structure, in which again, additional information of known source signals is used to speed up convergence. In this letter, we propose a new algorithm for solving the combined blind/nonblind signal separation problem for an instantaneous mixing system. Possible methods to extend the algorithm to cope with a convolutive mixing system are described in [4].

II. NEW APPROACH

Fig. 2 shows the structure of the proposed approach. Without loss of generality, we can rearrange the subsets of M_b unknown and M_r known source signals such as to express them in the form $\mathbf{s} = [\mathbf{s}_b^T \quad \mathbf{s}_r^T]^T$, where the index b indicates "unknown"

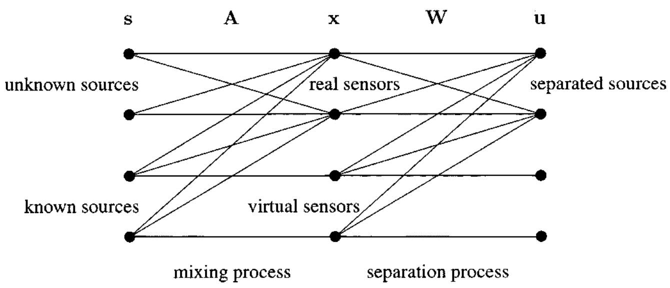

Fig. 3. Incorporation of the known sources as virtual sensors ($M_r = 2, M_b = 2$).

(blind) and the index r indicates "known" (reference), respectively. Likewise, $\mathbf{u} = [\mathbf{u}_b^T \;\; \mathbf{u}_r^T]^T$. This simplifies the representation of the matrices and allows block matrix notation. Although $\mathbf{u}_r$ could be estimated blindly using (3), we already know that $\mathbf{u}_r = \mathbf{s}_r$ with $\mathbf{s}_r$ being the known subset of source signals. In the blind-only case, $M_s = M_b + M_r$ different mixtures of the source signals are required in order to completely separate all of them. But since M_r source signals are known, we can set up M_r virtual sensors, whose signals are the known source signals (see Fig. 3). The remaining M_b real sensors contain different linear mixtures of all source signals (including the known ones). For the sensor signals, we can therefore write

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_b \\ \mathbf{s}_r \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{s}_b \\ \mathbf{s}_r \end{bmatrix}. \tag{5}$$

The structure of the mixing matrix $\mathbf{A}$ can now be represented in block form, revealing the dependence on the two subsets of signals. $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$ describe the influence of the unknown and known source signals to the real sensor signals, respectively. The zero and the identity submatrix result from the introduction of the M_r virtual sensors. The separation matrix $\mathbf{W}$ can now also be written as a simplified block matrix, so that

$$\begin{aligned} \mathbf{u} = \begin{bmatrix} \mathbf{u}_b \\ \mathbf{s}_r \end{bmatrix} &= \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{s}_b \\ \mathbf{s}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{W}_{11}\mathbf{A}_{11} & \mathbf{W}_{11}\mathbf{A}_{12} + \mathbf{W}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{s} \\ &= \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{s}. \end{aligned} \tag{6}$$

The lower part of $\mathbf{W}$ is now directly obtained as $[\mathbf{0} \;\; \mathbf{I}]$ without adaptation, making the calculation of $[\Delta\mathbf{W}_{21} \quad \Delta\mathbf{W}_{22}]$ in (4) redundant. According to (3), only the upper subblocks $[\mathbf{W}_{11} \;\; \mathbf{W}_{12}]$ of $\mathbf{W}$ need to be updated, i.e.,

$$\begin{aligned} &[\Delta\mathbf{W}_{11} \;\; \Delta\mathbf{W}_{12}] \\ &\quad = \left[\left(\mathbf{I} - \mathbf{g}(\mathbf{u}_b)\mathbf{u}_b^T\right)\mathbf{W}_{11} \;\;\; \left(\mathbf{I} - \mathbf{g}(\mathbf{u}_b)\mathbf{u}_b^T\right)\mathbf{W}_{12} - \mathbf{g}(\mathbf{u}_b)\mathbf{s}_r^T\right] \\ &\quad = \left[\mathbf{I} - \mathbf{g}(\mathbf{u}_b)\mathbf{u}_b^T\right][\mathbf{W}_{11} \;\; \mathbf{W}_{12}] - \left[\mathbf{0} \;\; \mathbf{g}(\mathbf{u}_b)\mathbf{s}_r^T\right]. \end{aligned} \tag{7}$$

The first term in (7) has exactly the same form as in (3), where no source signals are known. The second term only influences the updating of $\mathbf{W}_{12}$, which describes the signal flow between the virtual sensors $\mathbf{x}_r = \mathbf{s}_r$ and the output $\mathbf{u}_b$. The term $\mathbf{g}(\mathbf{u}_b)\mathbf{s}_r^T$ is a measure of the dependence between the known source signals $\mathbf{s}_r$ and the blindly recovered signals $\mathbf{u}_b$. During adaptation, the expectation of this term is reduced causing $\mathbf{u}_b$ to become a linear mixture of only unknown source signals $\mathbf{s}_b$. The first term in (7) then performs the actual separation, such that the output signals in $\mathbf{u}_b$ become independent.

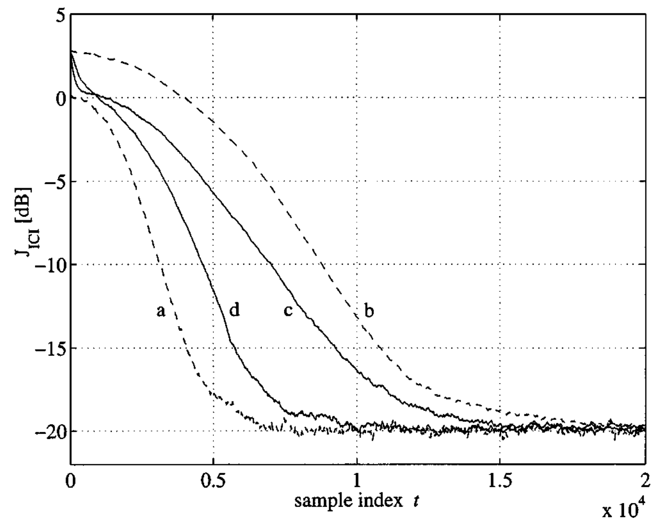

Fig. 4. Separation performance of the combined blind/nonblind separation algorithm averaged over 50 runs. (a) $M_s = 5$, $M_r = 0$, $M_b = 5$, (blind only), and $\mu = 0.004$. (b) $M_s = 10$, $M_r = 0$, $M_b = 10$, (blind only), and $\mu = 0.002$. (c) $M_s = 10$, $M_r = 5$, $M_b = 5$, and $\mu_1 = \mu_2 = 0.002$. (d) $M_s = 10$, $M_r = 5$, $M_b = 5$, $\mu_1 = 0.003$, and $\mu_2 = 0.001$.

Since the two terms in (7) have different tasks regarding the updating of the separation matrix $\mathbf{W} = [\mathbf{W}_{11} \;\; \mathbf{W}_{12}]$, different step sizes can be assigned to them. The resulting update equation is

$$\mathbf{W}_{t+1} = \mathbf{W} + \mu_1 \left[\mathbf{I} - \mathbf{g}(\mathbf{u}_b)\mathbf{u}_b^T\right]\mathbf{W} - \mu_2 \left[\mathbf{0} \;\; \mathbf{g}(\mathbf{u}_b)\mathbf{s}_r^T\right]. \tag{8}$$

In order to find a meaningful measure to judge the separation progress during convergence, a scalar measure is needed that describes the average degree of residual mixing. Such a measure, the so-called *interchannel interference* (ICI) was described in [5]

$$J_{\text{ICI}}(\mathbf{P}_b) = \frac{1}{M_b} \sum_{i=1}^{M_b} \frac{\sum_{j=1}^{M_s} p_{ij}^2 - \max_j p_{ij}^2}{\max_j p_{ij}^2} \tag{9}$$

where p_{ij} are the elements of $\mathbf{P}_b = [\mathbf{P}_{11} \;\; \mathbf{P}_{12}]$, see also (6). Of course, $J_{\text{ICI}}(\mathbf{P}_b)$ is available in a simulation environment only. In practical situations, the true matrix $\mathbf{A}$ and therefore, the matrix $\mathbf{P}_b$ is unknown.

For the following simulation using Laplacian distributed source signals, $\mathbf{g}(.) = \sqrt{2} \cdot \text{sign}(.)$ was chosen as the nonlinearity. The block length of the algorithm was $L = 16$. The step sizes were individually adjusted so as to result in the same final separation degree of $J_{\text{ICI}_\infty}(\mathbf{P}_b) = -20$ dB. Fig. 4 shows the performance of the algorithm (8) for five known and five unknown source signals (solid curves). It is compared with the blind-only algorithm (3) for five and ten unknown source signals (dashed curves), respectively. The individual adjustment of the step sizes in (8) introduces a further degree of freedom and allows tuning for optimized convergence speed without using explicit step-size control. In the example given, curve (c) has been obtained using the same step size for both terms in (8). If, as shown by curve (d), the step sizes are tuned further, faster convergence can be achieved for the same separation degree.

It can now clearly be observed that the additional knowledge of some source signals is worthwhile, bringing the curve much closer to curve (a), which can be regarded as a lower bound. Besides reduced computational complexity and faster convergence, fewer sensor signals are required in the case where some of the source signals are known. Simulations have also shown that a mixture in which all known source signals plus one unknown source signal have Gaussian distributions is still separable with the proposed algorithm.

REFERENCES

[1] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Processing Syst.*, vol. 8, pp. 757–763, 1996.

[2] D. W. E. Schobben and P. C. W. Sommen, "Transparent communication," in *IEEE Benelux Signal Processing Chapter Symp.*, Leuven, Belgium, Mar. 26–27, 1998, pp. 171–174.

[3] M. Joho and H. Mathis, "Performance comparison of combined blind/nonblind source separation algorithms," in *Proc. Int. Conf. Independent Component Analysis and Blind Signal Separation (ICA)*, Aussois, France, Jan. 11–15, 1999, pp. 139–142.

[4] M. Joho, "A systematic approach to adaptive algorithms for multichannel system identification, inverse modeling, and blind identification," Ph.D. dissertation, ETH, Zürich, Switzerland, Dec. 2000.

[5] M. Joho, H. Mathis, and G. S. Moschytz, "An FFT-based algorithm for multichannel blind deconvolution," in *IEEE Int. Symp. Circuits and Systems (ISCAS)*, Orlando, FL, May 30–June 2 1999, pp. III-203–206.