First International Workshop on Independent Component Analysis and Blind Signal Separation ICA'99, Jan. 11–15, 1999, Aussois, France.

# PERFORMANCE COMPARISON OF COMBINED BLIND/NON-BLIND SOURCE SEPARATION ALGORITHMS

*Marcel Joho and Heinz Mathis*

Signal and Information Processing Laboratory,
Swiss Federal Institute of Technology Zurich, Switzerland
*joho@isi.ee.ethz.ch, tel.:+41 (0) 1 632 27 71, mathis@isi.ee.ethz.ch, tel.:+41 (0) 1 632 36 20*

## ABSTRACT

Source separation is becoming increasingly important in acoustical applications for spatial filtering. In the absence of any known source signals (blind case), a blind update equation similar to the natural gradient method [1] is presented, a derivative of which can be used in the case of known references (non-blind case). If some, but not all, source signals are known, blind-only algorithms are suboptimal, since some available information is not exploited. To overcome this problem, non-blind separation techniques can be incorporated. For the instantaneous mixing case (no time delays, no convolution), two different ways of combining blind and non-blind source separation methods are shown, namely an echo canceller-type and an equalizer-like approach. Simulations allow a comparison of the convergence time of both structures versus the convergence time of the blind-only case and clearly demonstrate the benefit of using combined blind/non-blind separation techniques.

## 1. INTRODUCTION

*Blind separation*, *blind deconvolution* and the combination of both, *multichannel blind deconvolution*, are tasks that have to be carried out in an increasing number of applications, particularly in acoustics and communications. The structure of the separation algorithm can be understood as a single-layer neural network, whose training algorithms depend partially on the kind of signals to be separated. Lately, these tasks have been approached using information-theoretic methods [2] or, more generally, higher-order statistics [3]. Fast convergence can be achieved using the natural [1] or relative [4] gradient. In this paper we will deal with the separation problem only (no deconvolution), i.e., the signals $\mathbf{x} = [x_1, \ldots, x_M]^T$ to be processed are linear mixtures of the original source signals $\mathbf{s} = [s_1, \ldots, s_{M_s}]^T$ weighted by scalars

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

$M_s$ denotes the total number of sources and $M$ the number of sensors, respectively. For the blind-only case we assume $M = M_s$. Later we will see that $M < M_s$ is possible, if some source signals are known. The dependency of the signals (but not of the mixing process) on the sampling time $t$ is indicated in (1) but will be omitted in the following for better legibility.

Fig. 1 shows the setup of a system for blind source separation. The source signals $\mathbf{s}$ are mixed by the mixing matrix $\mathbf{A}$ and recovery is attempted by a blind adaptive algorithm adjusting the coefficients of the separation matrix $\mathbf{W}$. The output of the algorithm is therefore

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

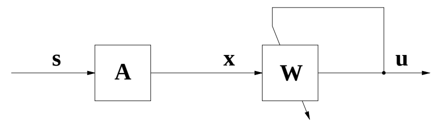


Figure 1: ($\mathcal{A}$1) Blind source separation.

In order to separate the signals well, $\mathbf{WA} = \mathbf{P}$ should become close to a scaled permutation matrix. A possible update equation for the separation matrix $\mathbf{W}$ results from the minimization of the mutual information of the output signals [2] and applying the natural gradient [1]

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

where $\mu$ is the step-size, $\mathbf{I}$ the identity matrix. $\mathbf{g}(\mathbf{u})$ is known as the *Bussgang nonlinearity* [5] or the *score function* and depends on the *pdf* of the source signals $\mathbf{s}$. For super-Gaussian sources such as speech the nonlinearity used is a sigmoid function such as tanh(.) or sign(.). *Blind* means that no components other than $\mathbf{x}$, $\mathbf{W}$ and $\mathbf{u}$ are involved in the update equation (3) or in the separation process $\mathbf{u} = \mathbf{W}\mathbf{x}$.

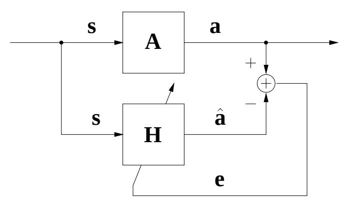


Figure 2: ($\mathcal{A}$2) Echo canceller-like architecture to separate signals with known references.

In the case when the original source signals are known, the problem of blind separation degenerates into an easier non-blind separation task which can be approached in two different ways. In fact, since the source signals are already known in this case, we are actually interested in a good estimate of $\mathbf{A}$ or $\mathbf{A}^{-1}$, respectively. By trying to identify the mixing process just like in an *echo canceller* we get an estimate of $\mathbf{A}$, let's say $\mathbf{H}$, which can then be inverted (provided matrix $\mathbf{A}$ is well conditioned and is of full rank). A diagram of the echo canceller-like architecture can be viewed in Fig. 2. The update equation for the matrix $\mathbf{H}$ can be derived again using a stochastic gradient method, yielding the MIMO LMS

$$\mathbf{H}_{t+1} = \mathbf{H}_t + \eta \mathbf{e}\mathbf{s}^T, \tag{4}$$

with the step-size $\eta$ and the error signal being

$$\mathbf{e} = \mathbf{a} - \mathbf{H}\mathbf{s}. \quad (5)$$

Not only is convergence much faster than in the blind case, as might be expected, but the update equation is computationally easier on the grounds of not containing any nonlinearities and matrix multiplications.

Alternatively, the separation matrix can be directly adapted by an LMS algorithm as can be seen in Fig. 3. The update equation has a form similar to (4)

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu \mathbf{e}\mathbf{x}^T, \quad (6)$$

with the error signal being

$$\mathbf{e} = \mathbf{s} - \mathbf{u}. \quad (7)$$

Note that in contrast to the ubiquitous LMS equalizer, $\mathbf{W}$ in (6) is not a vector containing coefficients of a time-domain filter but a matrix denoting the influence of the input signals $\mathbf{x}$ on the output signals $\mathbf{u}$. There is no temporal memory or delay involved. Yet,

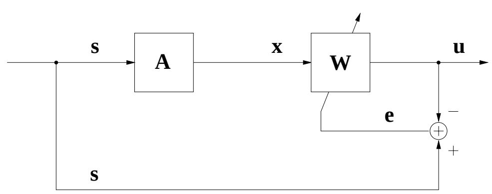


Figure 3: ($\mathcal{A}3$) Equalizer-like architecture to separate signals with known references.

the block structure resembles that of an *equalizer*, hence we will refer to it by that name hereafter. This algorithm shows acceptable convergence if the mixing matrix is far from singular. Variations of (6), e.g., normalized LMS algorithms are possible and tend to reduce convergence time. Still, the algorithm is inferior to the echo canceller $\mathcal{A}2$ in that its convergence time is higher and its performance suffers when the mixing matrix is close to singular.

## 2. IMPROVED UPDATE EQUATIONS

Since $\mathbf{W}$ is the inverse of $\hat{\mathbf{A}}$ and hence of $\mathbf{H}$, a much faster update equation for $\mathcal{A}3$ can be found by applying the *matrix-inversion lemma* to (4)

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu(\mathbf{s}-\mathbf{u})\mathbf{s}^T \left[\hat{\mathbf{R}}_{\mathbf{SS}} - \mu(\mathbf{s}-\mathbf{u})\mathbf{s}^T\right]^{-1} \mathbf{W}_t, \quad (8)$$

which replaces (6). $\hat{\mathbf{R}}_{\mathbf{SS}}$ is an estimate of the source cross-correlation matrix $\mathrm{E}[\mathbf{s}\cdot\mathbf{s}^T]$, which is an identity matrix if the sources are all independent with power $\sigma_{s_i}^2 = 1$. After this modification, $\mathcal{A}3$ shows roughly the same convergence behavior as $\mathcal{A}2$. Using the *Bussgang property* [5], (8) can be further adapted to the blind case and results in

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu(\mathbf{I} - \mathbf{g}(\mathbf{u})\mathbf{u}^T)\left[\mathbf{I} - \mu(\mathbf{I} - \mathbf{g}(\mathbf{u})\mathbf{u}^T)\right]^{-1} \mathbf{W}_t. \quad (9)$$

Here, of course, the source signals are not accessible, so that by the scaling invariance property of blind algorithms $\hat{\mathbf{R}}_{\mathbf{SS}}$ is replaced by the identity matrix. When $\mathbf{W}_t$ is close to the final solution $\mathbf{W}_\infty$, (9) approaches the natural gradient learning algorithm given in (3).

Unfortunately, in a practical situation, the reference signals are usually not provided and a blind approach has to be adopted. However, some of the source signals may be known, e.g., in a situation with feedback. In acoustical applications (e.g., teleconferencing) it may occur that the output signal of a loudspeaker leaks into several microphones placed loosely in front of a group of speakers in the same room. The speakers' signals have to be separated blindly, but for the loudspeaker it would certainly be of some advantage to use the reference signal of the loudspeaker, which is of course well accessible. Hence, algorithms combining both blind and non-blind source separation are needed and will be presented in the sequel.

## 3. COMBINED SEPARATION

Schobben/Sommen [6] have pointed out the necessity of combining echo cancelling and blind separation to be both effective in time varying environments and computationally efficient. They based their joint implementation of the Multi-Channel Acoustical Echo Canceller (MC-AEC) on second-order statistics, which might decorrelate signals but does not always make them completely independent [7]. To make signals truly independent, either of the algorithms $\mathcal{A}2$ and $\mathcal{A}3$ might be combined with the blind separation algorithm.

In the first case ($\mathcal{A}4$) we simply put a blind separation block after the echo-canceller circuit $\mathcal{A}2$ (see Fig. 4). Since only $M_\mathrm{r} = M_\mathrm{s} - M_\mathrm{b}$ source signals are known, the matrix $\mathbf{H}$ is no longer square and is now multiplied by a vector $\mathbf{s}_\mathrm{r}$ consisting of the $M_\mathrm{r}$ components of the source vector $\mathbf{s}$ whose references are given. Whereas $\mathbf{A}$ still is an $M \times M_\mathrm{s}$ matrix, $\mathbf{H}$ is now an $M \times M_\mathrm{r}$ matrix. The update equation (4) and the error equation (5) are still valid if $\mathbf{s}$ and $\mathbf{e}$ are replaced by $\mathbf{s}_\mathrm{r}$ and $\mathbf{e}_\mathrm{b}$, respectively. The blind separation task is then eased off by the cancellation of $M_\mathrm{r}$ signals with known references. Note that the number of sensors equals the number of unknown sources, $M = M_\mathrm{b}$. The signal vector $\mathbf{x}_\mathrm{b}$ after the cancellation contains components of the unknown sources only. Therefore the blind separation part has to separate only the remaining $M_\mathrm{b}$ unknown signals. The same blind algorithms (3) or (9) with the $M \times M$ square matrix $\mathbf{W}$ can then be applied.

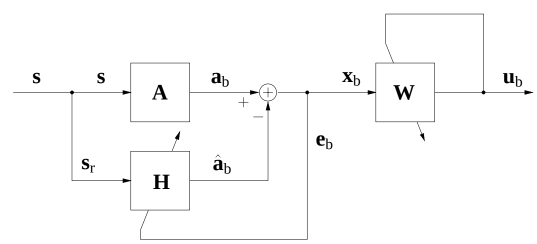


Figure 4: ($\mathcal{A}4$) Combined echo canceller and blind source separation.

The second choice ($\mathcal{A}5$) of combining blind and non-blind separation is obtained by starting with the separation structures $\mathcal{A}1$ and $\mathcal{A}3$ and combining them in a way that can be seen in Fig. 5. In this case $M = M_\mathrm{s}$. The synthesis of equations (8) and (9) leads to

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \mu(\mathbf{R}_{\mathbf{SS}} - \mathbf{R}_{\mathbf{US}})\left[\hat{\mathbf{R}}_{\mathbf{SS}} - \mu(\mathbf{R}_{\mathbf{SS}} - \mathbf{R}_{\mathbf{US}})\right]^{-1} \mathbf{W}_t. \quad (10)$$

Without loss of generality, the elements of the signal vector $\mathbf{s}$ might be arranged, such that its first $M_\mathrm{b}$ components are the unknown source signals and the remaining $M_\mathrm{r}$ elements contain the signals

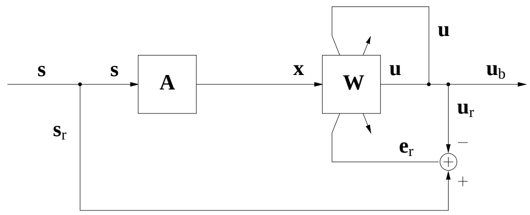


Figure 5: ($\mathcal{A}$5) Combined equalizer and blind source separation.

with known references, i.e. $\mathbf{s} = [\mathbf{s}_\mathrm{b}^T \ \mathbf{s}_\mathrm{r}^T]^T$. Likewise $\mathbf{u} = [\mathbf{u}_\mathrm{b}^T \ \mathbf{u}_\mathrm{r}^T]^T$. This simplifies the representation of the matrices and allows block notation. $\hat{\mathbf{R}}_{\mathbf{SS}}$ consists partly of an identity matrix and partly of an estimate of the cross-correlation matrix of the known sources as in (8)

$$\hat{\mathbf{R}}_{\mathbf{SS}} = \begin{bmatrix} \mathbf{I}_{M_\mathrm{b}} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_{\mathbf{s}_\mathrm{r}\mathbf{s}_\mathrm{r}} \end{bmatrix}. \qquad (11)$$

If the known signals are mutually uncorrelated, $\hat{\mathbf{R}}_{\mathbf{SS}}$ is a diagonal matrix and simplifies the matrix inversion in (10) near convergence. $\mathbf{R}_{\mathbf{SS}}$ and $\mathbf{R}_{\mathbf{US}}$ contain running or instantaneous estimates of correlation terms. For $\sigma_{u_1}^2 = \ldots = \sigma_{u_{M_\mathrm{b}}}^2 = 1$ they show the following block structure

$$\mathbf{R}_{\mathbf{SS}} = \begin{bmatrix} \mathbf{I}_{M_\mathrm{b}} & \mathbf{0} \\ \mathbf{0} & \mathbf{s}_\mathrm{r} \cdot \mathbf{s}_\mathrm{r}^T \end{bmatrix} \qquad (12)$$

and

$$\mathbf{R}_{\mathbf{US}} = \begin{bmatrix} \mathbf{g}(\mathbf{u}_\mathrm{b}) \cdot \mathbf{u}_\mathrm{b}^T & \mathbf{u}_\mathrm{b} \cdot \mathbf{s}_\mathrm{r}^T \\ \mathbf{u}_\mathrm{r} \cdot \mathbf{u}_\mathrm{b}^T & \mathbf{u}_\mathrm{r} \cdot \mathbf{s}_\mathrm{r}^T \end{bmatrix}. \qquad (13)$$

In contrast to $\mathbf{R}_{\mathbf{SS}}$ and $\mathbf{R}_{\mathbf{US}}$, which basically contain blocks of outer products of two vectors for a given time instant $t$ and therefore are not of full rank, $\hat{\mathbf{R}}_{\mathbf{SS}}$ has to be non-singular, i.e., its rank must be $M_\mathrm{s}$, otherwise the inverse of $[\hat{\mathbf{R}}_{\mathbf{SS}} - \mu(\mathbf{R}_{\mathbf{SS}} - \mathbf{R}_{\mathbf{US}})]$ in (10) does not exist.

Finally, both structures, $\mathcal{A}$4 and $\mathcal{A}$5, might be combined, yielding a configuration as shown in Fig. 6. Here some of the known sources are cancelled by the echo canceller and some are separated by the equalizer. This case is given for the sake of completeness and is not considered further.

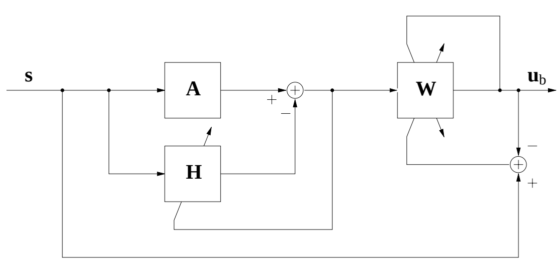


Figure 6: ($\mathcal{A}$0) General setup of the proposed methods.

## 4. PERFORMANCE COMPARISON

### 4.1. Performance measure

In order to find a meaningful measure to judge the separation progress during convergence, a scalar number is needed that describes the average degree of the residual mixing of the separated sources at the output of the network. Amari *et al.* [1] define the performance index as a ratio of the sum of absolute values of unwanted sources leaking through and the absolute value of the desired source. We define our performance measure based on the power of signals rather than on absolute values. Ideally, the scaled permutation matrix in (2),

$$\mathbf{P} = \mathbf{WA} = [p_{ij}] = \begin{bmatrix} \mathbf{P}_\mathrm{b}^{M_\mathrm{b} \times M_\mathrm{s}} \\ \mathbf{P}_\mathrm{r}^{M_\mathrm{r} \times M_\mathrm{s}} \end{bmatrix}, \qquad (14)$$

would have one non-zero entry in each row ($\mathbf{P}$ may happen to be diagonal dominant). Since our real interest here is focused on the estimate $\mathbf{u}_\mathrm{b} = \hat{\mathbf{s}}_\mathrm{b}$, we will concentrate on $\mathbf{P}_\mathrm{b}$. $p_{ij}^2$ is the power transfer from source $s_j$ to the output $u_i$. In each row there will be one dominant entry ($\max_k p_{ik}^2$), indicating the power transfer of the separated source $s_k$. The other entries of the same row of $\mathbf{P}$, which end up being non-zero in a realistic case, if squared and summed, reveal the power of other sources leaking through to the particular output, if all sources have equal power. Vice versa, by squaring the subdominant entries of the same column of $\mathbf{P}$, we get the power of one source leaking through to different outputs. For the mixed blind/non-blind case the row-wise observation of the permutation matrix is more meaningful, so that we define a performance measure as

$$J(\mathbf{P}_\mathrm{b}^{M_\mathrm{b} \times M_\mathrm{s}}) = \sqrt{\frac{1}{M_\mathrm{b}} \sum_{i=1}^{M_\mathrm{b}} \frac{\sum_{j=1, j \neq k}^{M_\mathrm{s}} p_{ij}^2}{\max_k p_{ik}^2}} = \sqrt{\frac{1}{M_\mathrm{b}} \left( \sum_{i=1}^{M_\mathrm{b}} \frac{\sum_{j=1}^{M_\mathrm{s}} p_{ij}^2}{\max_k p_{ik}^2} \right) - 1}. \qquad (15)$$

Note that $J(\mathbf{P}_\mathrm{b})$ is the inverse of the average signal-to-interference ratio (SIR) at the separated output signals $\mathbf{u}_\mathrm{b}$ for independent sources with equal power. Of course, $J(\mathbf{P}_\mathrm{b})$ is available in a simulation environment only. In practical situations, the true matrix $\mathbf{A}$ and therefore the matrix $\mathbf{P}$ are unknown.

### 4.2. Simulation results

For the simulations in this section, the following parameters were set. All source signals were independent Laplace-distributed random signals with a normalized power of $\sigma_{s_i}^2 = 1$. Choosing the nonlinearity $\mathbf{g}(\mathbf{u}) = 2\tanh(\mathbf{u})$ leads to $\sigma_{u_i}^2 \approx 1$ or, equivalently, the singular-values of $\mathbf{P}$, $\lambda_i$, are all around 1 after settling. In order to avoid special cases and be more representative instead, the matrix $\mathbf{A}$ is chosen such that its singular-values $\lambda_i$ are logarithmically distributed and its condition number is $\chi(\mathbf{A}) = \lambda_{\max}/\lambda_{\min} = 100$. The initial value for the separation matrix is chosen as $\mathbf{W}_0 = \mathbf{I}$. Block processing allows faster implementation. In our case, a block length of $L = 16$ samples has been chosen. The desired final separation performance $J_\infty(\mathbf{P}_\mathrm{b})$ determines the step-size $\mu$ and has been found by simulations. Table 1 illustrates this relationship for the first simulation, $\mathcal{A}$1 with (9). In Fig. 7a) the dependency of the separation performance on the step-size for ten unknown sources can clearly be observed. The curves are averages over

| $J_\infty(\mathbf{P}_\mathrm{b})$ | $-10$ dB | $-15$ dB | $-20$ dB |
|---|---|---|---|
| step-size $\mu$ | 0.27 | 0.12 | 0.045 |

Table 1: Step-size $\mu$ of equation (9) for $\mathcal{A}$1 as a function of the final separation performance $J_\infty(\mathbf{P}_\mathrm{b})$.

50 typical runs with different input signals and mixing matrices. Fig. 7b) shows the singular-value decomposition (SVD) of $\mathbf{P}_\mathrm{b}$ for $J_\infty(\mathbf{P}_\mathrm{b}) = -20\,\mathrm{dB}$, which reflects the progress of the decorrelation performance of the permutation matrix $\mathbf{P}_\mathrm{b}$. Because of

$$\mathbf{P}_0 = \mathbf{W}_0 \cdot \mathbf{A} = \mathbf{A}, \tag{16}$$

where the zero denotes the initial value, at the beginning of the plot we clearly see the singular-value spread of $\mathbf{A}$. Very quickly

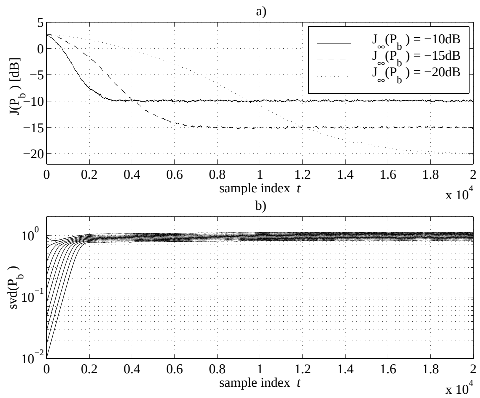


Figure 7: Performance of the blind separation algorithm $\mathcal{A}1$ for $M_\mathrm{s} = 10$ sources. a) Performance measure $J(\mathbf{P}_\mathrm{b})$ as described in section 4.1. b) Singular-value decomposition (SVD) of the permutation matrix $\mathbf{P}_\mathrm{b}$ for $J_\infty(\mathbf{P}_\mathrm{b}) = -20\,\mathrm{dB}$.

the singular values converge to one, indicating that the decorrelation process happens much faster than the separation process, which makes the output signals *mutually independent*. It is due to this behavior that in the blind case the natural gradient algorithm shows about the same convergence time as the proposed algorithm described in (9), because the advantage of the latter only holds as long as the output signals are still correlated.

Next, the results of the combined blind/non-blind algorithms can be compared by the help of Fig. 8. For the nonlinearity we used $\mathbf{g}(\mathbf{u}) = 1.5 \cdot \mathrm{sign}(\mathbf{u})$. The blind-only cases for five and ten channels have been included for reference. For the cases $\mathcal{A}4$ and $\mathcal{A}5$

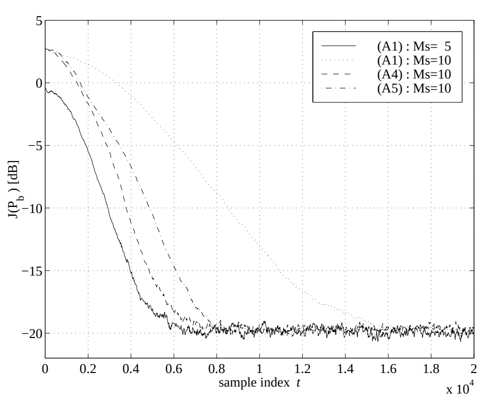


Figure 8: Performance of the combined blind/non-blind separation algorithms averaged over 30 runs. The step-sizes $\mu$ are individually adjusted so as to result in the same separation degree $J_\infty(\mathbf{P}_\mathrm{b}) = -20\,\mathrm{dB}$: $\mathcal{A}1$: $\mu = 0.065$, $\mu = 0.03$, $\mathcal{A}4$: $\mu = 0.055$, $\eta = 0.001$, $\mathcal{A}5$: $\mu = 0.035$.

we assumed that five of the source signals are known, the other five have to be blindly separated. As expected, both algorithms $\mathcal{A}4$ and $\mathcal{A}5$ lie between the two curves for blind separation (for five and ten sources, resp.). Important to note is that they lie much closer to the performance of $\mathcal{A}1$ with only five unknown sources than with ten unknown sources. This clearly indicates that applying combined blind/non-blind separation algorithms dramatically improves convergence times. The difference in the initial separation performance of around 3 dB reflects the fact that in the first case the number of interfering sources is about half of the number for the other three cases.

## 5. CONCLUSIONS

Improved update equations for blind and non-blind source separation have been derived. For the case where some of the source signals are known, two different methods for combining blind and non-blind source separation have been outlined, one of which incorporates an echo canceller-like approach, the other an equalizer structure. Either structure speeds up the separation of the remaining unknown sources considerably. Both approaches can be extended to the multichannel blind deconvolution case where the elements of matrix $\mathbf{A}$ are filter polynomials by using FIR-matrix algebra [8]. In this case the mixing matrix $\mathbf{A}(z)$ as well as the separation matrix $\mathbf{W}(z)$ contain filter polynomials as their matrix elements.

## ACKNOWLEDGEMENT

The authors would like to thank Russell Lambert and Te-Won Lee for helpful discussions.

## 6. REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 752–763, 1996.

[2] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[3] T-W. Lee, M. Girolami, A.J. Bell, and T.J. Sejnowski, "A unifying information-theoretic framework for independent component analysis," *International Journal on Mathematical and Computer Modeling*, in press, 1998.

[4] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[5] R.H. Lambert and A.J. Bell, "Blind separation of multiple speakers in a multipath environment," in *International Conference on Acoustics, Speech & Signal Processing*, Munich, Germany, April 21–24, 1997, pp. 423–426.

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

[7] C. Jutten and J. Hérault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, July 1991.

[8] R.H. Lambert, *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*, Ph.D. thesis, University of Southern California, 1996.