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

ON-LINE TIME-DOMAIN BLIND SOURCE SEPARATION OF NONSTATIONARY CONVOLVED SIGNALS

Robert Aichner, Herbert Buchner

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

Shoko Araki, Shoji Makino

NTT Communication Science Laboratories
NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun,
Kyoto 619-0237, Japan
{shoko, maki}@cslab.kecl.ntt.co.jp

ABSTRACT

In this paper we propose a time-domain gradient algorithm that exploits the nonstationarity of observed signals and recovers the original sources by simultaneously decorrelating time-varying second-order statistics. By introducing a generalized weighting factor in our cost function we can formulate an on-line algorithm that can be applied to time-varying multipath mixing systems. A further benefit is the possibility of implementing updates in a recursive manner and thus reduce computational complexity. We show that this method inherently possesses an adaptive step size and hence avoids stability problems. Furthermore we present a new geometric initialization for time-domain gradient algorithms that improves separation performance in strongly reverberant environments. In our experiments we compared the separation performance of the proposed algorithm with those of its off-line counterpart and another multiple-decorrelation-based on-line algorithm in the frequency domain for real-world speech mixtures.

1. INTRODUCTION

Blind source separation (BSS) refers to the problem of recovering signals from several observed linear mixtures. The adjective "blind" stresses the fact that the source signals are not observed and that no information is available on the multi-path mixing process. This lack of *a priori* knowledge about the mixing system is compensated for by a statistically strong but physically plausible assumption of independence. The absence of prior information is actually the strength of the BSS model. Thus BSS has received considerable attention in the recent years, and many algorithms have been proposed [1, 2, 3], mainly in relation to scalar mixing systems.

We focus on BSS for acoustic signals, which is characterized by a convolutive mixing process and therefore is an even more challenging topic. Among the unresolved problems are two of particular interest. Firstly, in strong reverberant environments most of the current algorithms fail to achieve sufficient separation performance. This is due to the fact that most algorithms try to invert the multi-path mixing system by adapting multi-path finite impulse response (FIR) filters. Thus in highly echoic rooms a very large number of filter coefficients need to be identified. Secondly, in real-world scenarios, the multi-path mixing system is usually time-variant because of moving sources, moving sensors or changing environments. This requires a continuous adaptation of the algorithm to the time-varying mixing system with a sufficiently fast convergence of the algorithm. We deal with both problems in this paper and propose a new geometric initialization method to improve the separation under reverberant conditions. The second problem is addressed by introducing a generalized weighting factor to allow us to track time-variant environments.

This paper is organized as follows. In Sec. 2 we present a modified cost function with a generalized weighting factor. We show that the inherent normalization avoids stability problems. Thereafter we derive the natural gradient adaptation for the proposed cost function and present a recursive formulation for an on-line time-domain gradient algorithm. We then introduce a new geometric initialization method. After that we address the whitening problem for time-domain BSS algorithms. In Sec. 3 we compare simulation results for the proposed on-line algorithm with those of its off-line counterpart and another on-line algorithm in the frequency domain.

2. TIME-DOMAIN GRADIENT ALGORITHM

2.1. Convolutive BSS model

In real environments, signals arrive at the sensors with different time delays due to reflections. This scenario is referred to as a multi-path environment and can be described as a finite impulse response (FIR) convolutive mixture:

$$\mathbf{x}(t) = \sum_{k=0}^{P-1} \mathbf{H}(k)\,\mathbf{s}(t-k) \tag{1}$$

where $\mathbf{s}(t) = [s_1(t), ..., s_n(t)]^T$ captures the *n* mutually independent source signals and $\mathbf{x}(t) = [x_1(t), ..., x_m(t)]^T$ are the mixed signals obtained by the microphones. The superscript ^{T} denotes transposition. The mixing system $\mathbf{H}$ is an $m \times n$ matrix consisting of channel impulse responses $h_{ij}(k)$ $(i = 1, ..., m, j = 1, ..., n)$ that are modeled by FIR filters of length *P* with the filter coefficients $k = 0, ..., P-1$.

To obtain the estimated sources $\mathbf{y}(t) = [y_1(t), ..., y_n(t)]^T$, we seek a $n \times m$ matrix $\mathbf{W}$ of FIR filters of length *L* operating on the sensor measurements $\mathbf{x}(t)$ such that the components of the output vector $\mathbf{y}(t)$ are statistically independent:

$$\mathbf{y}(t) = \sum_{k=0}^{L-1} \mathbf{W}(k)\mathbf{x}(t-k) \tag{2}$$

where the unmixing matrix $\mathbf{W}(k)$ is defined as

$$\mathbf{W}(k) = \begin{bmatrix} w_{11}(k) & \cdots & w_{1m}(k) \\ \vdots & \ddots & \vdots \\ w_{n1}(k) & \cdots & w_{nm}(k) \end{bmatrix}. \quad (3)$$

We introduce $\overline{\mathbf{W}}(z)$ as the z-transform of the unmixing filter coefficient $\mathbf{W}(k)$ with $k = 0, ..., L-1$:

$$\overline{\mathbf{W}}(z) = \sum_{k=0}^{L-1} \mathbf{W}(k) z^{-k}, \quad (4)$$

where L denotes the length of the unmixing filter and z^{-1} is used as the unit-delay operator for convenience, i.e., $z^{-k}x(t) = x(t-k)$.

To recover the original source signals from the observed mixtures, BSS uses the mutual independence of the original sources. However, the mutual independence cannot resolve the following two ambiguities:

1. The order of the recovered source signals $\mathbf{y}(t)$ can be arbitrarily permuted as the mutual independence of the sources is unaffected by the permutation.
2. The recovered signals $\mathbf{y}(t)$ may be arbitrarily filtered with a diagonal matrix Λ where each element on the diagonal represents an FIR filter of length *P*.

Thus we can express the recovered signals $\mathbf{y}(t)$ as:

$$\mathbf{y}(t) = \mathbf{P}\Lambda * \mathbf{s}(t) \quad (5)$$

where * denotes convolution.

In the remainder of the paper we consider a two-speaker, two-microphone BSS scenario as shown in Fig. 1

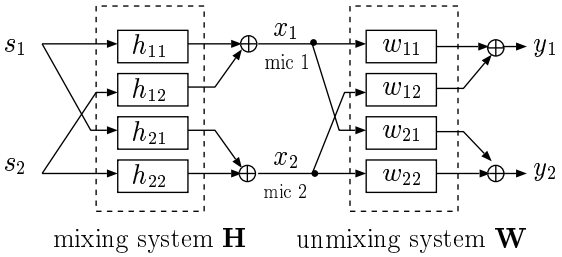


Fig. 1. BSS model.

2.2. Cost function

The assumption of independence causes the correlation matrices of the sources $\mathbf{R}_{ss}(t,k) = E\left[\mathbf{s}(t)\mathbf{s}^T(t+k)\right]$ to become diagonal matrices. Acoustic sources are assumed to be nonstationary, i.e., the auto-correlations of sources change independently over time *t*. Hence the correlation matrix of the outputs $\mathbf{R}_{yy}(t,k) = E\left[\mathbf{y}(t)\mathbf{y}^T(t+k)\right]$ also varies over time *t*. Thus, if we force estimated outputs $\mathbf{y}(t)$ to be uncorrelated at every time *t*, we obtain a much stronger condition than mere decorrelation for a single time instant *t*, and this enables us to separate the sources.

Matsuoka et al. [4] used this principle for separating scalar mixtures. This method was subsequently extended by Kawamoto et al. [5] and used to separate convolutive mixtures. As a measure of uncorrelatedness, we use the cost function proposed in [5] and modify it for use with block processing methods. Therefore we introduce $\mathbf{R}_{yy}^{(b)}(k)$ to represent the correlation matrix of $\mathbf{y}(t)$ in the *b*-th analysis block with time delay k. The correlation matrix defined by $\mathbf{R}_{yy}^{(b)}(k) = E_{(b)}\left[\mathbf{y}(t)\mathbf{y}^T(t+k)\right]$ is calculated using a block processing procedure, where $E_{(b)}[x]$ denotes the time average of *x* for the *b*-th block.

Additionally, with regard to deriving an on-line time-domain gradient algorithm, we introduce a generalized weighting factor $\beta(m,b)$ where *m* and *b* denote block indices.

$$\mathcal{J}_1\left(m, \overline{\mathbf{W}}(z)\right) = \frac{1}{2}\sum_{b=1}^{m}\beta(m,b)\left\{\log\left(\det \operatorname{diag}\mathbf{R}_{yy}^{(b)}(0)\right) - \log\left(\det \mathbf{R}_{yy}^{(b)}(0)\right)\right\} \quad (6)$$

The operator diag $\mathbf{X}$ denotes the diagonal elements of the matrix $\mathbf{X}$.

The segmentation of the observed mixtures into blocks ensures that we are calculating the cross-correlations at multiple times. The non-negative cost function becomes zero only when $y_i(t)$ and $y_j(t)$ are uncorrelated for all the local analysis blocks, i.e., $E_{(b)}[y_i(t)y_j(t)] = 0$ $(i,j = 1,...,n;\ i \neq j,\ \ b = 1,...,m)$. This corresponds to a simultaneous diagonalization of multiple correlation matrices at different times *t*.

2.3. Weighting factor

In adaptive filter signal processing it is customary to introduce a weighting factor $\beta(m,b)$ into the definition of the cost function [6]. The weighting factor, which we normalize so that $\sum_{b=1}^{m}\beta(m,b) = 1$, allows the more recent samples to have greater influence on the output error, allowing the tracking of time-variant acoustic environments. The introduced weighting factor is based on the block indices *m,b*. Special forms of the weighting factor that are commonly used are sliding windows or the exponential

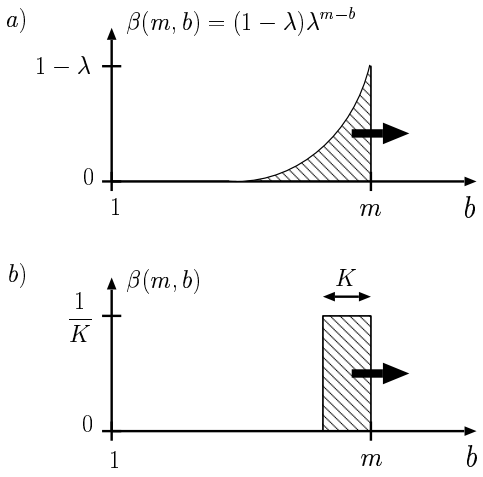


Fig. 2. Weighting factors for adaptive algorithms. (a) exponential forgetting factor, (b) sliding window of length *K*.

weighting factor (forgetting factor) (Fig. 2) defined by

$$\beta(m,b) = (1-\lambda)\lambda^{m-b} \tag{7}$$

where λ is a positive constant close to, but less than 1.

This on-line formulation of algorithms including a generalized weighting factor is also similar to off-line or so-called batch methods. For a batch method, the entire observed signal has to be processed using the sliding window. After the block processing of the entire signal, the output signals are generated by applying the final unmixing filters to the sensor signals. Hence the algorithm can be interpreted as one iteration of a batch method, as we have no continuous separation of the observed signals.

In [7] we presented an off-line algorithm based on the cost function (6) and an additional modification. This approach leads to good results in highly reverberant environments.

In this paper, we consider the on-line counterpart with an exponential forgetting factor, as it can be formulated in a recursive manner that is computationally simple. In the following paragraph we will provide an on-line formulation of the update rule.

2.4. Natural gradient adaptation

We use the natural gradient adaptation, as introduced by Amari [8], which provides an isotropic convergence independent of the mixing matrix **H**, to minimize the cost function:

$$\Delta \mathbf{W}_m(k) \propto -\frac{\partial \mathcal{J}_1\left(m, \overline{\mathbf{W}}(z)\right)}{\partial \mathbf{W}(k)} \overline{\mathbf{W}}^T(z^{-1}) \overline{\mathbf{W}}(z), \tag{8}$$

where the symbol $\propto$ means "proportional to" and $k = 0, ..., L-1$ denotes the filter tap index. The adaptation rule for an off-line algorithm using the natural gradient is derived in [9]. For a cost function including a general weighting factor $\beta(m,b)$ (6) this results in:

$$\begin{aligned} \mathbf{W}_m(k) &= \alpha \sum_{b=1}^{m} \beta(m,b) \left\{ \left(\mathbf{R}_{yy}^{(b)}(0)^{-1} \right)^T \mathbf{R}_{yy}^{(b)}(k) \right. \\ &\quad \left. - \left(\operatorname{diag} \mathbf{R}_{yy}^{(b)}(0) \right)^{-1} \mathbf{R}_{yy}^{(b)}(k) \right\} \overline{\mathbf{W}}_{m-1}(z) \\ &\quad + \mathbf{W}_{m-1}(k) \end{aligned} \tag{9}$$

where α is a step size parameter, *m* denotes the current block and *k* is the filter tap index. Equation (9) converges if the off-diagonal components of $\mathbf{R}_{yy}^{(b)}(0)$ are minimized for all blocks. We confirmed in experiments that if we consider only time-delay $k = 0$, we cannot achieve separation in a strongly reverberant environment. Therefore, we modify (9) to the following equation to evaluate the off-diagonal components of $\mathbf{R}_{yy}^{(b)}(k)$ for all time delays $k = 0, ..., L-1$:

$$\begin{aligned} \mathbf{W}_m(k) &= \alpha \sum_{b=1}^{m} \beta(m,b) \left\{ \left(\operatorname{diag} \mathbf{R}_{yy}^{(b)}(0) \right)^{-1} \right. \\ &\quad \left. \left(\operatorname{diag} \mathbf{R}_{yy}^{(b)}(k) - \mathbf{R}_{yy}^{(b)}(k) \right) \right\} \\ &\quad \overline{\mathbf{W}}_{m-1}(z) + \mathbf{W}_{m-1}(k) \end{aligned} \tag{10}$$

This update equation simultaneously diagonalizes the correlation matrices $\mathbf{R}_{yy}^{(b)}(k)$ for multiple blocks (depending on the shape of $\beta(m,b)$) and additionally decorrelates the cross-correlations at multiple time lags *k*. Hence we are using the nonstationarity and the nonwhiteness of the source signals which lead to improved separation.

With regard to an on-line algorithm, in our update equation (10) we use the exponential forgetting factor defined in (7). This results in

$$\mathbf{W}_m(k) = \alpha \Delta \mathbf{W}_m(k) + \mathbf{W}_{m-1}(k) \tag{11}$$

where

$$\Delta \mathbf{W}_m(k) = \sum_{b=1}^{m} (1-\lambda)\lambda^{m-b} \mathcal{Q}_b(k) \tag{12}$$

$$\begin{aligned} \mathcal{Q}_b(k) &= \left\{ \left(\operatorname{diag} \mathbf{R}_{yy}^{(b)}(0) \right)^{-1} \right. \\ &\quad \left. \left(\operatorname{diag} \mathbf{R}_{yy}^{(b)}(k) - \mathbf{R}_{yy}^{(b)}(k) \right) \right\} \overline{\mathbf{W}}_{m-1}(z) \end{aligned} \tag{13}$$

and $k = 0, ..., L-1$ denotes the filter tap index of the unmixing filter **W**. Equation (12) can be formulated in a recursive manner to reduce the computational complexity. Another benefit is the reduced memory requirements, since only the preceding filter tap matrix has to be saved for the update.

$$\Delta \mathbf{W}_m(k) = \lambda \Delta \mathbf{W}_{m-1}(k) + (1-\lambda)\mathcal{Q}_b(k) \tag{14}$$

Thus, by using the exponential forgetting factor, we obtain an on-line time-domain gradient algorithm that can be applied to time-variant multipath environments.

2.5. Inherent normalization

Another approach for measuring the uncorrelatedness of the estimated output signals is the cost function defined by:

$$\mathcal{J}_2(\mathbf{W}) = \sum_{b,k} \| \mathbf{R}_{yy}^{(b)}(k) - \operatorname{diag} \mathbf{R}_{yy}^{(b)}(k) \|^2 \tag{15}$$

where $\mathbf{R}_{yy}^{(b)}(k) = E_{(b)}[\mathbf{y}(t)\mathbf{y}^T(t+k)]$, the index *b* denotes block processing and $\| \cdot \|^2$ is the Frobenius norm. This cost function exploits the nonstationarity and nonwhiteness of the sources and is widely used (see, e.g., [10]). To achieve fast convergence in this case it is usually necessary to consider second order gradient expressions. A proper Newton-Raphson update requires the inverse of the Hessian. This is both tedious to derive and computationally very demanding. Thus there are many heuristic approaches with which to solve this normalization problem.

The cost function $\mathcal{J}_1$, however, shows an inherent normalization. It can be seen in the update equation (9) that the correlation matrix $\mathbf{R}_{yy}^{(b)}(k)$ is scaled by the inverse of $\mathbf{R}_{yy}^{(b)}(0)^T$ and the inverse of $\operatorname{diag} \mathbf{R}_{yy}^{(b)}(0)$, respectively.

Similarily we can observe that the off-diagonal components of the correlation matrix in the modified update rule (10), which are expressed by $\operatorname{diag} \mathbf{R}_{yy}^{(b)}(k) - \mathbf{R}_{yy}^{(b)}(k)$, are normalized by the inverse of $\operatorname{diag} \mathbf{R}_{yy}^{(b)}(0)$. This means a normalization to the short-time power of the output signals. We observed that this inherent normalization ensures a fast convergence of the algorithm, thus making an adaptive step size becomes unnecessary.

2.6. Utilization of geometric beamforming

The convergence and the separation performance of gradient-based algorithms are greatly influenced by the initial value, especially if we need long FIR filters to model the room impulse responses. To solve this problem we propose a new approach for calculating the initial value by adding geometric information on the positions of the microphones and the assumed positions of the speakers. This method has so far only been applied to frequency-domain methods [11] where the geometric initialization showed superior separation especially for a large number of microphones.

The equivalence of adaptive beamformers and BSS shown in [12] was our motivation to use a beamformer technique for initializing the adaptation algorithm. We assume the sources to be spatially separated so that we can employ a null beamformer with beams that place spatial zeros at the orientations of interfering sources. For our experiments with a two-speaker, two-microphone scenario, we assume two sources with angles of $\theta_i = \pm 60^o$, measured with respect to the normal vector of the microphone array.

We calculate the null beamformer for both target and jammer signal configurations with respect to the angle of the interfering source and the microphone positions. The delays resulting from the null beamformer are then used as cross path filters w_{12} and w_{21}. As these delays are fractions of the sampling time, we have to use a sinc function for interpolation. The filter in the straight path w_{11}, w_{22} is initialized as a unit impulse. The small components that exist in addition to the unit impulse (Fig. 3) are the result of disregarding frequency components smaller than 62.5 Hz. They were disregarded because we cannot calculate a sharp spatial null for low frequencies due to the small microphone spacing of 4 cm. Additionally we incorporate a delay of $\frac{L}{2}$ to access both future and past values of the observed signals. Thus a noncausal filter is formed that allows us to identify non-minimum phase systems. In [7] we showed for a batch algorithm that the new initialization

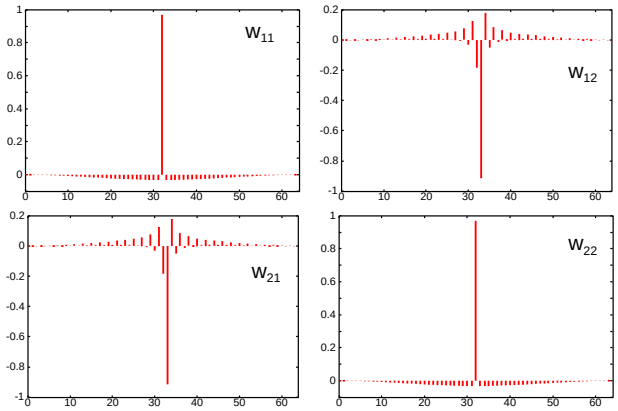


Fig. 3. Initial value $\mathbf{W}$ for an unmixing filter length of $L = 64$ taps.

method improves the separation significantly and that we can estimate long unmixing FIR filters in the time domain that can cover the entire reverberation. As shown in Sec. 3 we could also confirm this experimentally when using the on-line version with an exponential forgetting factor.

2.7. Dewhitening of output signals

In Sec. 2.1 we pointed out that the recovered source signals possess two ambiguities; the signal order may be permuted and the outputs can be arbitrarily filtered. The latter ambiguity causes a change in the output signal spectrum. Higher frequencies are usually emphasized whereas lower frequencies are attentuated and thus the spectrum becomes whitened. The flattened spectrum gives the speech signals an unnatural timbre. Although this ambiguity can arise in both time-domain and frequency-domain BSS, it is seldom addressed in time-domain BSS. To simplify the explanation, here we consider the problem in the frequency domain. Frequency-domain BSS decomposes the convolutive mixing system in multiple scalar mixing systems in each frequency bin. The arbitrary filtering corresponds to an arbitrary scaling in each frequency bin. Before recomposing the signals from each frequency bin, the arbitrary scaling has to be resolved because otherwise no separation is achieved.

To overcome the filtering ambiguity, we apply post-filters to our system. These filters are based on the method for removing the amplitude ambiguity in frequency-domain BSS proposed by Ikeda [13]. To the authors' knowledge, no post-processing method using this principle has been proposed for time-domain BSS.

The general idea is to transfer the separated output signals $\mathbf{y}(t)$ into the frequency domain and then solve the problem of irregular amplitude for each frequency bin. The observed mixtures $\mathbf{X}(\omega)$ are described by $\mathbf{X}(\omega) = \mathbf{H}(\omega)\,\mathbf{S}(\omega)$ and $\mathbf{X}(\omega) = \mathbf{W}^{-1}(\omega)\,\mathbf{Y}(\omega)$. We can assume that, when sources are set at almost the same distance from a microphone array, the amplitudes of all the elements of the mixing filter matrix $\mathbf{H}(\omega)$ are equal, because the attenuations of all observed sound signals are nearly equal due to the small microphone spacing. The inverse of the unmixing matrix $\mathbf{W}(\omega)$ is denoted as $\mathbf{W}^{-1}(\omega) = \left[W_{ij}^{-1}(\omega)\right]$. For the two-speaker, two-microphone scenario we obtain:

$$X_1(\omega) = W_{11}^{-1}(\omega)Y_1(\omega) + W_{12}^{-1}(\omega)Y_2(\omega) \quad (16)$$

$$X_1(\omega) = 1 \cdot S_1(\omega) + 1 \cdot S_2(\omega) \quad (17)$$

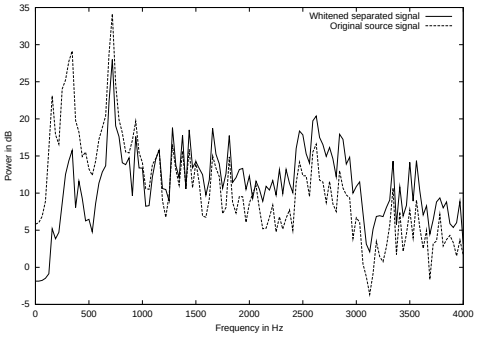
$$X_2(\omega) = W_{21}^{-1}(\omega)Y_1(\omega) + W_{22}^{-1}(\omega)Y_2(\omega) \quad (18)$$

$$X_2(\omega) = 1 \cdot S_1(\omega) + 1 \cdot S_2(\omega) \quad (19)$$

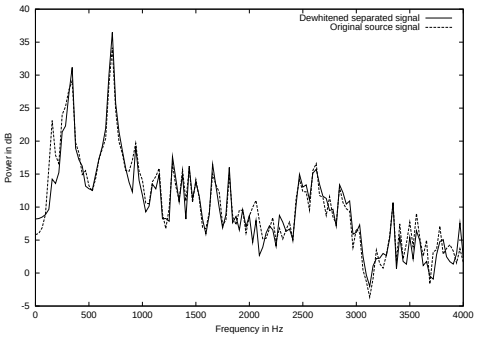
We can now rescale the output signal $Y_1(\omega)$ in each frequency bin by multiplying $Y_1(\omega)$ with the factor $W_{11}^{-1}(\omega)$ so that the amplitude of $X_1(\omega)$ in (16) is equal to the amplitude of $X_1(\omega)$ in (17). The same process is applied to the output signal $Y_2(\omega)$ so that the amplitudes of $X_2(\omega)$ in (18) and (19) are the same. Thus we obtain a dewhitening matrix $\mathbf{V}(\omega)$, which is defined as $\mathbf{V}(\omega) = \operatorname{diag}\mathbf{W}^{-1}(\omega)$. The dewhitened output signals $\tilde{\mathbf{Y}}(\omega)$ are obtained by

$$\tilde{\mathbf{Y}}(\omega) = \mathbf{V}(\omega)\mathbf{Y}(\omega) = \mathbf{V}(\omega)\mathbf{W}(\omega)\mathbf{X}(\omega). \quad (20)$$

The post-processing filter emphasizes low frequencies to restore the original spectral content of the source signals (Fig. 4). The algorithm without dewhitening returns higher SIR values due to the emphasis of high frequencies. The smaller separation capability of our BSS system for low frequencies is not taken into account. Thus the dewhitening process is corrects the higher signal-to-interference ratios. We can observe that the obtained spectrum complies well with the original spectral content.



(a) Whitened spectrum



(b) Dewhitened spectrum

Fig. 4. Spectrum of output signal before (a) and after (b) post-processing.

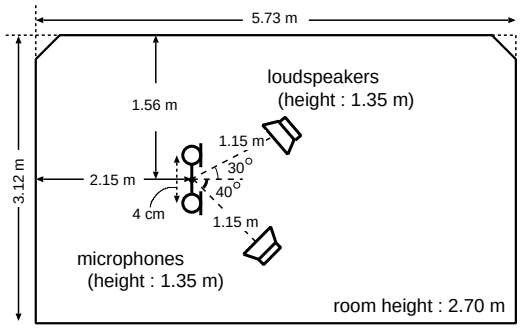


Fig. 5. Layout of the reverberant room used in the experiments.

3. EXPERIMENTS AND RESULTS

3.1. Experimental conditions

We conducted the experiments using speech data convolved with the impulse responses of a real room (Fig. 5), with a reverberation time $T_R = 300$ ms. Since the sampling frequency was 8 kHz, the reverberation time corresponded to a room impulse response of 2400 taps. We used a two-element microphone array with an inter-element spacing of 4 cm. The speech signals arrived from two different directions, -30^o and 40^o. We used two sentences spoken by two male and two female speakers selected from the ASJ continuous speech corpus as source signals [14]. We used six combinations of speakers with a signal length of seven seconds. To evaluate the performance, 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. The SIR was continuously calculated for each block by using the post-processed dewhitened separated output signals.

3.2. Experimental results

We compare the performance of the proposed on-line algorithm with that of its off-line counterpart [7] and with the second on-line decorrelation algorithm in the frequency domain presented in [10].

To compare the two on-line algorithms we set the unmixing filter length at $L = 512$ taps. For the time-domain algorithm we chose a blocklength of 1024 samples with an overlap factor of 4 and used an exponential forgetting factor with $\lambda = 0.9999$. The frequency-domain BSS algorithm also had an FFT length of 1024 with an overlap factor of 4. We applied the new geometric initialization method to both classes, but for the two-source, two-microphone scenario we could only achieve an SIR improvement for the time-domain algorithms.

Due to the filtering ambiguity the output signals of the on-line time-domain algorithm showed a flattened spectrum. The frequency-domain method avoids the scaling ambiguity in each frequency bin by enforcing a unity gain constraint for the diagonal filters. However, we observed that this does not fully prevent the whitening of the spectrum of the separated output signals. Figure 6 shows the average SIR for six different combinations of source signals for the investigated algorithms. We observed that the geometric initialization of the time-domain algorithm leads to fast convergence. Moreover it is seen that the postprocessing method, which restores the original spectral content of the sources, reduces the SIR. This results from the fact that the dewhitening process emphasizes the low frequencies, which exhibit a smaller SIR caused by strong reverberation.

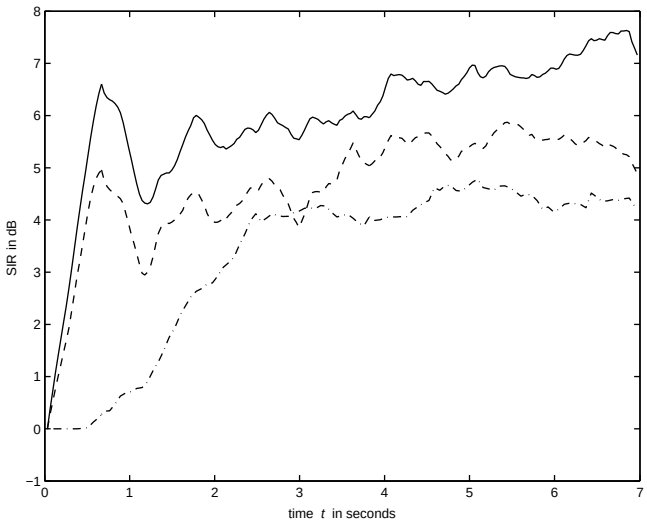


Fig. 6. Average SIR depicted for the time-domain on-line algorithm without (solid line) and with (dashed line) the dewithening postprocessing method. In comparison the frequency-domain on-line algorithm without postprocessing (dash-dotted line) [10]. Reverberation time $T_{60} = 300$ ms.

Figure 6 also shows that the frequency-domain on-line algorithm converges to a lower SIR than with the time-domain method. By contrast, the frequency-domain algorithm is also applicable to strongly reverberant environments where no *a priori* information on the array geometry is available.

The on-line BSS algorithm can exploit the observed signals only once and yet it achieves good separation. The off-line algorithm on the other hand can repeatedly access their values for each iteration. This further improves the separation performance seen in Fig. 7 which shows the SIR for the batch algorithm with and without the postprocessing method. For the off-line algorithm we set the unmixing filter length at $L = 512$ taps, and we chose a blocklength of 1024 samples. This corresponds to the parameters we used for the on-line algorithm.

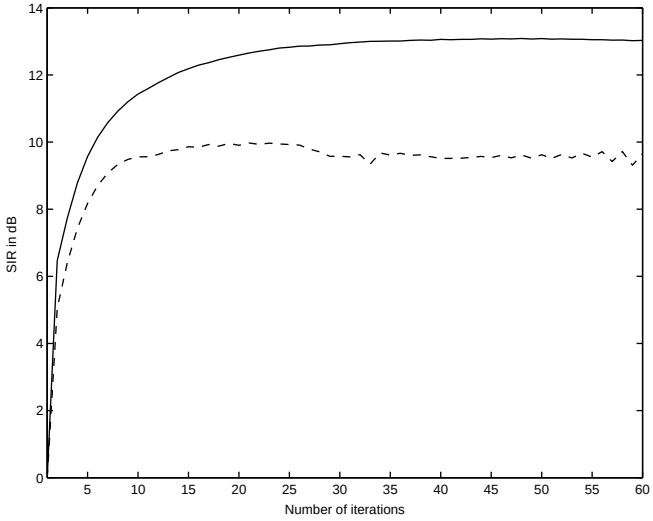


Fig. 7. Average SIR of the off-line time-domain algorithm without (solid line) and with the dewhitening postprocessing method (dashed line). Reverberation time $T_{60} = 300$ ms.

4. CONCLUSION

We proposed a time-domain gradient algorithm with a generalized weighting function. Using an exponential forgetting factor, we introduced a recursive on-line BSS algorithm, which can be applied to time-varying multipath environments. We presented a new geometric initialization for time-domain algorithms, which showed good results for strongly reverberant environments. Our simulation results showed that both on-line and off-line methods are capable of separating real-world speech signals in echoic surroundings.

5. ACKNOWLEDGMENTS

We thank Prof. Walter Kellermann for his continuous encouragement and constructive criticism. We are also grateful to Prof. Hiroshi Saruwatari from Nara Institute of Science and Technology for providing the room impulse responses.

6. REFERENCES

[1] J.-F. Cardoso, "Blind signal separation: statistical principles," in *Proc. IEEE*, Oct. 1998, vol. 9, pp. 2009–2025.

[2] S. Haykin, Ed., *Unsupervised Adaptive Filtering*, Volume 1 Blind Source Separation. John Wiley & Sons, 2000.

[3] A. Hyvaerinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[4] K. Matsuoka, M. Ohya, and M. Kawamoto, "Neural net for blind separation of nonstationary signals," *IEEE Trans. Neural Networks*, vol. 8, no. 3, pp. 411–419, 1995.

[5] M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, vol. 22, pp. 157–171, 1998.

[6] S. Haykin, *Adaptive Filter Theory*, Prentice Hall Inc., Englewood Cliffs, NJ, 1996.

[7] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time-domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," in *Proc. Neural Networks for Signal Processing*, June 2002, pp. 445–454.

[8] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, 1998.

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

[10] L. Parra and C. Spence, "On-line convolutive source separation of non-stationary signals," *Journal of VLSI Signal Processing*, vol. 26, no. 1/2, pp. 39–46, Aug. 2000.

[11] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," in *Proc. Neural Networks for Signal Processing*, 2001, pp. 273–282.

[12] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency-domain blind source separation and frequency-domain adaptive null beamformers," in *Proc. Eurospeech*, Sept. 2001, pp. 2595–2598.

[13] S. Ikeda and N. Murata, "A method of ICA in time-frequency-domain," in *Proc. ICA*, Jan. 1999, pp. 365–371.

[14] T. Kobayashi, S. Itabashi, S. Hayashi, and T. Takezawa, "ASJ continuous speech corpus for research," *J. Acoust. Soc. Jpn.*, vol. 48, no. 12, pp. 888–893, 1992, (in Japanese).