

PAPER

A Distortion-Free Learning Algorithm for Feedforward Multi-Channel Blind Source Separation

Akihide HORITA^{†a)}, Student Member, Kenji NAKAYAMA^{†b)}, and Akihiro HIRANO^{†c)}, Members

SUMMARY FeedForward (FF-) Blind Source Separation (BSS) systems have some degree of freedom in the solution space. Therefore, signal distortion is likely to occur. First, a criterion for the signal distortion is discussed. Properties of conventional methods proposed to suppress the signal distortion are analyzed. Next, a general condition for complete separation and distortion-free is derived for multi-channel FF-BSS systems. This condition is incorporated in learning algorithms as a distortion-free constraint. Computer simulations using speech signals and stationary colored signals are performed for the conventional methods and for the new learning algorithms employing the proposed distortion-free constraint. The proposed method can well suppress signal distortion, while maintaining a high source separation performance.

key words: blind source separation, signal distortion, convergence, learning algorithm, convolutive mixture

1. Introduction

Signal processing, including noise cancellation, echo cancellation, equalization of transmission lines, estimation and restoration of signals has become a very important research area. These techniques require separation between desired signals and interference. However, information regarding the signals and their interference is insufficient in many cases. Furthermore, their mixing and transmission processes are not well known in advance. In these kinds of situations, blind source separation (BSS) technologies using statistical properties of signal sources have become very important [1]–[5].

In many applications, the mixing processes are convolutive mixtures. Therefore, separation processes require convolutive models. Various methods have been proposed for separating sources in the time domain and the frequency domain. Their separation performance is highly dependent on the signal sources and the transfer functions of the mixture [7]–[10], [14], [15], [17], [18].

BSS learning algorithms make the output signals to be statistically independent. However, these approaches do not always guarantee distortion-free separation. In order to suppress signal distortion, the following methods have been proposed. One of them reduces the distance between the separation block outputs and observations [11]. However,

since the observations include several signal sources, the interference components in the outputs are not well reduced. The other method modifies the separation block outputs by multiplying the inverse of the separation block matrix [13]. However, it is somewhat difficult to achieve high performances in both source separation and signal distortion.

In this paper, first, evaluation measures for signal distortion are discussed. Secondly, conditions for source separation and distortion-free are derived. Based on these conditions, a new learning algorithm with a distortion-free constraint is proposed. Performances of a new learning algorithm in comparison with conventional methods are analyzed through computer simulations.

2. BSS Systems for Convolutive Mixture

2.1 Network Structure and Equations

A block diagram of a feedforward (FF-) BSS system with 2 signal sources and 2 sensors is shown in Fig. 1. The mixing stage has a convolutive structure. In the separation block, $W_{kj}(z)$ are realized by using FIR filters.

Let $s_i(n)$, $x_j(n)$ and $y_k(n)$ be the inverse z-transform of $S_i(z)$, $X_j(z)$ and $Y_k(z)$, respectively. Furthermore, $h_{ji}(l)$ and $w_{kj}(l)$ are the coefficients of $H_{ji}(z)$ and $W_{kj}(z)$, respectively, when they are realized by using FIR filters. We define the following vectors and matrices,

$$\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^T \quad (1)$$

$$\mathbf{x}(n) = [x_1(n), \dots, x_N(n)]^T \quad (2)$$

$$\mathbf{y}(n) = [y_1(n), \dots, y_N(n)]^T \quad (3)$$

$$\mathbf{h}(l) = \begin{bmatrix} h_{11}(l) & \cdots & h_{1N}(l) \\ \vdots & \ddots & \vdots \\ h_{N1}(l) & \cdots & h_{NN}(l) \end{bmatrix} \quad (4)$$

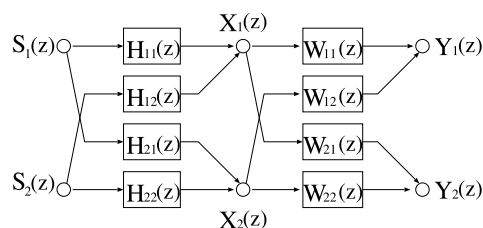


Fig. 1 BSS system with 2 signal sources and 2 sensors.

Manuscript received March 2, 2007.

Manuscript revised June 24, 2007.

Final manuscript received September 10, 2007.

[†]The authors are with the Graduate School of Natural Science and Technology, Kanazawa Univ., Kanazawa-shi, 920-1192 Japan.

a) E-mail: horita@leo.ec.t.kanazawa-u.ac.jp

b) E-mail: nakayama@t.kanazawa-u.ac.jp

c) E-mail: hirano@t.kanazawa-u.ac.jp

DOI: 10.1093/ietfec/e90-a.12.2835

$$\mathbf{w}(n, l) = \begin{bmatrix} w_{11}(n, l) & \cdots & w_{1N}(n, l) \\ \vdots & \ddots & \vdots \\ w_{N1}(n, l) & \cdots & w_{NN}(n, l) \end{bmatrix}. \quad (5)$$

T is the transpose operation of vectors and matrices. l expresses the sample number of the impulse responses in the mixing block and the filter coefficients in the separation block. n indicates the iteration number for updating the separation block $W_{kj}(z)$. Update will be carried out sample by sample. N is the number of channels. In Fig. 1, $N = 2$.

By using these definitions and expressions, the observations $\mathbf{x}(n)$ and the output signals $\mathbf{y}(n)$ are given by

$$\mathbf{x}(n) = \sum_{l=0}^{K_h-1} \mathbf{h}(l)\mathbf{s}(n-l) \quad (6)$$

$$\mathbf{y}(n) = \sum_{l=0}^{K_w-1} \mathbf{w}(n, l)\mathbf{x}(n-l) \quad (7)$$

where K_h is a length of the impulse responses in the mixing block. K_w is the number of the taps in the FIR filters used in the separation block.

In the z -domain, the above equations can be expressed by

$$\mathbf{X}(z) = \mathbf{H}(z)\mathbf{S}(z) \quad (8)$$

$$\mathbf{Y}(z) = \mathbf{W}(z)\mathbf{X}(z) \quad (9)$$

$$\mathbf{S}(z) = [S_1(z), \dots, S_N(z)]^T \quad (10)$$

$$\mathbf{X}(z) = [X_1(z), \dots, X_N(z)]^T \quad (11)$$

$$\mathbf{Y}(z) = [Y_1(z), \dots, Y_N(z)]^T \quad (12)$$

$$\mathbf{H}(z) = \begin{bmatrix} H_{11}(z) & \cdots & H_{1N}(z) \\ \vdots & \ddots & \vdots \\ H_{N1}(z) & \cdots & H_{NN}(z) \end{bmatrix} \quad (13)$$

$$\mathbf{W}(z) = \begin{bmatrix} W_{11}(z) & \cdots & W_{1N}(z) \\ \vdots & \ddots & \vdots \\ W_{N1}(z) & \cdots & W_{NN}(z) \end{bmatrix}. \quad (14)$$

The relation between the signal sources and the outputs is further expressed by

$$\mathbf{Y}(z) = \mathbf{W}(z)\mathbf{H}(z)\mathbf{S}(z) = \mathbf{A}(z)\mathbf{S}(z) \quad (15)$$

$$\mathbf{A}(z) = \begin{bmatrix} A_{11}(z) & \cdots & A_{1N}(z) \\ \vdots & \ddots & \vdots \\ A_{N1}(z) & \cdots & A_{NN}(z) \end{bmatrix}. \quad (16)$$

2.2 Learning Algorithm in Time Domain

A learning algorithm for separating sources based on a natural gradient method using mutual information as a cost function has been proposed [6]. This learning algorithm in the time domain is expressed by

$$\begin{aligned} \mathbf{w}(n+1, l) = & \mathbf{w}(n, l) + \eta \sum_{q=0}^{K_w-1} [\mathbf{I}\delta(l-q) \\ & - \langle \Phi(\mathbf{y}(n))\mathbf{y}^T(n-l+q) \rangle] \mathbf{w}(n, q) \end{aligned} \quad (17)$$

$$\Phi(\mathbf{y}(n)) = [\Phi(y_1(n)), \dots, \Phi(y_N(n))]^T \quad (18)$$

$$\Phi(y_k(n)) = \frac{1 - e^{-y_k(n)}}{1 + e^{-y_k(n)}}. \quad (19)$$

η is a learning rate, \mathbf{I} is the identity matrix, $\langle \rangle$ is an averaging operation and $\delta(n)$ is Dirac's delta function, where $\delta(0) = 1$ and $\delta(n) = 0, n \neq 0$.

2.3 Learning Algorithm in Frequency Domain

Based on the same idea, a learning algorithm in the frequency domain has been proposed in [6], [12], [16].

$$\begin{aligned} \mathbf{W}(r+1, m) = & \mathbf{W}(r, m) \\ & + \eta [\mathbf{I} - \langle \Phi(\mathbf{Y}(r, m))\mathbf{Y}^H(r, m) \rangle] \mathbf{W}(r, m) \end{aligned} \quad (20)$$

$$\Phi(\mathbf{Y}(r, m)) = [\Phi(Y_1(r, m)), \dots, \Phi(Y_N(r, m))]^T \quad (21)$$

$$\Phi(Y_k(r, m)) = \frac{1}{1 + e^{-Y_k^R(r, m)}} + \frac{j}{1 + e^{-Y_k^I(r, m)}}. \quad (22)$$

$\mathbf{Y}^H(r, m)$ means the Hermitian matrix of $\mathbf{Y}(r, m)$. The parameter r is the frame number used in the FFT, and m indicates the frequency point in each frame. $\mathbf{W}(r, m)$ is the weight matrix of the r -th FFT frame and the m -th frequency point. $\mathbf{Y}(r, m)$ is the output of the r -th FFT frame and the m -th frequency point. $Y_k^R(r, m)$ and $Y_k^I(r, m)$ represent the real part and the imaginary part of $Y_k(r, m)$, respectively.

It has been reported that the learning algorithm given in Eq. (20) transforms the outputs into the white signals due to the identity matrix \mathbf{I} . In order to avoid this problem, \mathbf{I} in Eq. (20) is replaced by $\text{diag}(\langle \Phi(\mathbf{Y}(r, m))\mathbf{Y}^H(r, m) \rangle)$ [16] as follows:

$$\begin{aligned} \mathbf{W}(r+1, m) = & \mathbf{W}(r, m) + \eta [\text{diag}(\langle \Phi(\mathbf{Y}(r, m))\mathbf{Y}^H(r, m) \rangle) \\ & - \langle \Phi(\mathbf{Y}(r, m))\mathbf{Y}^H(r, m) \rangle] \mathbf{W}(r, m) \end{aligned} \quad (23)$$

where, $\text{diag}(\mathbf{x})$ is the diagonal matrix of \mathbf{x} . In this paper, the learning algorithms using Eqs. (20) and (23) will be referred to as **FREQ(1)** and **FREQ(2)**, respectively.

3. Evaluation of Signal Distortion

3.1 Criterion for Signal Distortion

One candidate of the criteria for signal distortion may be the signal sources. However, in the learning algorithms of the BSS systems, where the signal sources are assumed to be statistically independent, the separation block is trained so that its outputs become statistically independent. The transfer functions in the mixing block are not estimated. Therefore, it is difficult to measure signal distortion as the difference between the separation block outputs and the signal sources.

In the conventional methods, the signal sources included in the observations are used as the criteria. This means that changes of the frequency responses, caused in the separation block, are treated as the signal distortion [4], [11], [13], [19]–[21]. This paper also follows this criterion. As a result, $H_{ii}(z)S_i(z)$ or $H_{ji}(z)S_i(z)$, $j \neq i$ are taken into account as a criterion for the signal distortion.

3.2 Conditions for Source Separation and Distortion-Free

For simplicity, the BSS system with two sources and two sensors, as shown in Fig. 1, is used. Furthermore, the sources $S_i(z)$ are assumed to be separated at the outputs $Y_i(z)$. This does not lose generality. Considering the criterion of signal distortion as defined in Sect. 3.1, the condition for distortion-free source separation can be expressed as follows:

Source separation : Non diagonal elements of $A(z)$ are zero.

$$A_{12}(z) = W_{11}(z)H_{12}(z) + W_{12}(z)H_{22}(z) = 0 \quad (24)$$

$$A_{21}(z) = W_{21}(z)H_{11}(z) + W_{22}(z)H_{21}(z) = 0 \quad (25)$$

Distortion free : Diagonal elements of $A(z)$ are $H_{ii}(z)$.

$$A_{11}(z) = W_{11}(z)H_{11}(z) + W_{12}(z)H_{21}(z) = H_{11}(z) \quad (26)$$

$$A_{22}(z) = W_{21}(z)H_{12}(z) + W_{22}(z)H_{22}(z) = H_{22}(z) \quad (27)$$

The conventional learning algorithm given by Eqs. (17)–(23) employs only Eqs. (24) and (25) as cost functions. Equations (26) and (27) are not guaranteed to be satisfied.

The number of the equations for source separation, that is no interference, is two. The number of variables, that is $W_{kj}(z)$, is four. Therefore, the solutions for $W_{kj}(z)$ are not unique. There exist some degrees of freedom in $W_{kj}(z)$. Since different speech signals are not statistically independent exactly, the output signals expressed by $A_{ii}(z)S_i(z)$ may be changed so as to be statistically independent to each other by adjusting $W_{kj}(z)$, which satisfy the conditions Eqs. (24) and (25). Thus, the output signals can deviate from the observed signal sources $H_{ii}(z)S_i(z)$, resulting in the signal distortion.

Therefore, by applying Eqs. (17) through (23), it is very likely that signal distortion will occur. This analysis is valid for BSS systems trained in both the time and the frequency domains.

The scaling problem is well known in the BSS. Scaling of the BSS outputs cannot be controlled by the learning algorithms. The separation block is adjusted so that the outputs become statistical independent to each other. There still remains a degree of freedom for scaling by constants. Especially, in the frequency domain BSSs, scaling for different frequency points may be different. This causes distortion in the frequency response, that is ‘signal distortion.’ Thus, in the frequency domain BSSs, ‘signal distortion’ and ‘scaling problem’ can be regarded as the same problem. The conventional learning algorithms given by Eqs. (17) though (23) do not solve the scaling problem.

4. Conventional Methods for Suppressing Signal Distortion

4.1 Minimal Distortion Principle (MDP)

A learning algorithm for reducing signal distortion has been proposed [11]. The cost function of the conventional method described in Sect. 2.2 has been extended by including the distance between the observed signals and the output signals. Therefore, the output signals are forced to approach to the observed signals.

The update equation for the filter coefficients is given by

$$\begin{aligned} \mathbf{w}(n+1, l) = & \mathbf{w}(n, l) + \eta \sum_{q=0}^{K_w-1} [\mathbf{I}\delta(l-q) \\ & - \langle \Phi(\mathbf{y}(n))\mathbf{y}^T(n-l+q) \rangle \\ & - \mu(\mathbf{y}(n) - \mathbf{x}(n))\mathbf{y}^T(n-l+q)]\mathbf{w}(n, q). \end{aligned} \quad (28)$$

Convergence property of this method is analyzed here. If $Y_i(z)$ approach to $X_i(z)$, that is $Y_i(z) = X_i(z)$, then,

$$\begin{aligned} Y_i(z) - X_i(z) = & (A_{ii}(z) - H_{ii}(z))S_i(z) \\ & + (A_{ij}(z) - H_{ij}(z))S_j(z) = 0, \quad i \neq j \end{aligned} \quad (29)$$

Since, $S_i(z)$ and $S_j(z)$ are assumed to be statistically independent, the above equation is satisfied when

$$A_{ii}(z) = H_{ii}(z) \quad (30)$$

$$A_{ij}(z) = H_{ij}(z). \quad (31)$$

According to our discussion from Sect. 3.2, if $A_{ii}(z)$ approach to $H_{ii}(z)$, then distortion-free is guaranteed. Furthermore, in the same section, it was also concluded that the non-diagonal elements of $A(z)$ should be zero in order to achieve source separation. However, here $A_{ij}(z)$ tend to approach to $H_{ij}(z)$. This means the interference components remain in the outputs, resulting in poor source separation. Although weight of the source separation and the signal distortion can be controlled by using the scaling factor μ , it is difficult to achieve good performance in both the source separation and the signal distortion at the same time.

It should be pointed out that the criterion of the signal distortion, that we introduced in this paper, is not the observed signal itself but the signal source included in the observed signals. This is a very important difference between them.

4.2 Signal Distortion Suppression by Using Inverse Separation Matrix

A method for suppressing signal distortion in the frequency domain BSS has been proposed [13]. After the source separation, the separation block outputs are modified by multiplying the inverse of the separation matrix $\mathbf{W}(r, m)$, which

is given by Eqs. (20) or (23). More concretely, the diagonal elements in $\mathbf{W}^{-1}(r, m)$ are used. Thus, the modified output vector is given by

$$\hat{\mathbf{Y}}(r, m) = \text{diag}[\mathbf{W}^{-1}(r, m)]\mathbf{Y}(r, m). \quad (32)$$

This approach is called a projection back (PB) method. In this method, the learning of source separation and suppressing signal distortion are separated.

5. A New Learning Algorithm with Distortion-Free Constraint

5.1 Distortion-Free Constraint

In this subsection, a distortion-free constraint is derived [22]–[25]. The conditions for distortion-free source separation, which are expressed by Eqs. (24) through (27) for the 2 channel case, can be arranged in a matrix form as follows:

$$\mathbf{W}(z)\mathbf{H}(z) = \mathbf{\Lambda}(z) \quad (33)$$

$$\mathbf{\Lambda}(z) = \text{diag}[\mathbf{H}(z)]. \quad (34)$$

Let $\mathbf{\Gamma}(z)$ be a matrix containing the non-diagonal elements of $\mathbf{H}(z)$ as follows:

$$\mathbf{\Gamma}(z) = \mathbf{H}(z) - \mathbf{\Lambda}(z). \quad (35)$$

Substituting $\mathbf{H}(z)$ in Eq. (35) for that in Eq. (33), we obtain

$$\mathbf{W}(z)(\mathbf{\Lambda}(z) + \mathbf{\Gamma}(z)) = \mathbf{\Lambda}(z). \quad (36)$$

Solve this equation for $\mathbf{\Gamma}(z)$

$$\mathbf{\Gamma}(z) = \mathbf{W}^{-1}(z)(\mathbf{I} - \mathbf{W}(z))\mathbf{\Lambda}(z) \quad (37)$$

$$= (\mathbf{W}^{-1}(z) - \mathbf{I})\mathbf{\Lambda}(z). \quad (38)$$

From Eq. (35), it follows

$$\text{diag}[\mathbf{\Gamma}(z)] = \text{diag}[(\mathbf{W}^{-1}(z) - \mathbf{I})\mathbf{\Lambda}(z)] = \mathbf{0}. \quad (39)$$

Since $\mathbf{\Lambda}(z)$ is the diagonal matrix, the above equation can be rewritten as

$$\text{diag}[(\mathbf{W}^{-1}(z) - \mathbf{I})] = \mathbf{0} \quad (40)$$

This condition holds, when the diagonal elements of $\mathbf{W}^{-1}(z)$ are 1. The inverse matrix is generally expressed by

$$\mathbf{W}^{-1}(z) = \frac{\text{adj } \mathbf{W}(z)}{\det \mathbf{W}(z)} \quad (41)$$

where $\text{adj } \mathbf{W}(z)$ is the adjugate matrix of $\mathbf{W}(z)$. Since, the diagonal elements of $\mathbf{W}^{-1}(z)$ equal 1, then

$$\text{diag}[\mathbf{W}^{-1}(z)] = \frac{\text{diag}[\text{adj } \mathbf{W}(z)]}{\det \mathbf{W}(z)} = \mathbf{I}. \quad (42)$$

Letting the j -th diagonal element of $\text{adj } \mathbf{W}(z)$ be $\hat{W}_{jj}(z)$, the above equation can be solved for $\hat{W}_{jj}(z)$ as follows:

$$\frac{\hat{W}_{jj}(z)}{\det \mathbf{W}(z)} = 1 \quad (43)$$

$$\hat{W}_{jj}(z) = \det \mathbf{W}(z) \quad (44)$$

$\hat{W}_{jj}(z)$ is also a cofactor of $\mathbf{W}(z)$. This equation expresses both conditions of ‘Source Separation’ and ‘Distortion-Free’ simultaneously.

5.2 Approximation of Distortion-Free Constraint

Solving Eq. (44) is computationally expensive. Therefore, we introduce an approximation formula for this calculation.

Equation (44) can be rewritten as follows:

$$W_{jj}(z) = 1 + \mathbf{w}_{row,j}^T(z)\mathbf{M}_{jj}^{-1}(z)\mathbf{w}_{col,j}(z) \quad (j = 1, \dots, N). \quad (45)$$

The derivation is given in Appendix A. $\mathbf{M}_{jj}(z)$ is an $(N - 1) \times (N - 1)$ minor matrix, i.e. removing the j -th row and j -th column from $\mathbf{W}(z)$. The vectors $\mathbf{w}_{col,j}(z)$ and $\mathbf{w}_{row,j}(z)$ are:

$$\mathbf{w}_{col,j}(z) = [W_{1j}(z), W_{2j}(z), \dots, W_{Nj}(z)]^T \quad (46)$$

$$\mathbf{w}_{row,j}(z) = [W_{j1}(z), W_{j2}(z), \dots, W_{jN}(z)]^T \quad (47)$$

$\mathbf{w}_{col,j}(z)$ and $\mathbf{w}_{row,j}(z)$ do not include $W_{jj}(z)$. Equation (45) is not an explicit solution, because $W_{kk}(z)$ is included in $\mathbf{M}_{jj}^{-1}(z)$ ($j \neq k$), which is used to calculate $W_{jj}(z)$.

In the time domain, $w_{jj}(n + 1, l)$ are calculated as follows: Since $W_{jj}(z)$ is realized as an FIR filter, whose transfer function is a series of z^{-1} , the right hand side of Eq. (45) is first expanded into a series of z^{-1} . Next, the corresponding coefficients of both hand sides are compared, and the coefficients of $W_{jj}(z)$, which are $w_{jj}(n + 1, l)$, are calculated. However, this series expansion requires a huge number of computations. Therefore, this process is equivalently carried out in the frequency domain in order to save the computational load. $\mathbf{w}(n + 1, l)$, given by Eq. (17), are first transformed into the frequency domain. Next, $W_{jj}(n + 1, m)$, which express samples of $W_{jj}(e^{j\omega T})$, are obtained by using Eq. (45). After that, $W_{jj}(n + 1, m)$ are inversely transformed into $w_{jj}(n + 1, l)$.

5.3 Combination of Learning Algorithm and Constraint in Frequency Domain

The condition for source separation and distortion-free expresses only a relation between $W_{jj}(z)$ and $W_{jk}(z)$, $k \neq j$. The optimal solutions for them cannot be obtained. Therefore, this condition should be combined with the learning algorithms for the source separation given by Eqs. (20) or (23). The basic idea behind the proposed method is as follows: First, $W_{jj}(z)$ and $W_{jk}(z)$ are updated following the original learning algorithm for source separation. Next, $W_{jj}(z)$ are modified following the condition given by Eq. (45).

The proposed learning algorithm is expressed as

Step 1: Update $W_{jj}(r, m)$ and $W_{jk}(r, m)$ following Eqs. (20) or (23), resulting in $W_{jj}(r + 1, m)$ and $W_{jk}(r + 1, m)$.

Step 2: Calculate $W_{jj}(r + 1, m)$ by Eq. (45). Here, $W_{jj}(r + 1, m)$ is changed to $\tilde{W}_{jj}(r + 1, m)$ for convenience.

$$\begin{aligned} \tilde{W}_{jj}(r + 1, m) &= 1 + \mathbf{w}_{row,j}^T(r + 1, m)\mathbf{M}_{jj}^{-1}(r + 1, m) \\ &\quad \cdot \mathbf{w}_{col,j}(r + 1, m) \end{aligned} \quad (48)$$

Step 3: Modify $W_{jj}(r+1, m)$ following

$$W_{jj}(r+1, m) = (1 - \alpha)W_{jj}(r+1, m) + \alpha\tilde{W}_{jj}(r+1, m) \quad (0 < \alpha \leq 1). \quad (49)$$

$W_{jj}(r+1, m)$ in the left-hand side is the modified version, which will be used in the next frame $r+1$.

In the above procedure, Step 3 is introduced. The reason can be explained as follows: The constraint is derived based on the source separation and the distortion-free. However, at the beginning of the learning process, the sources are not well separated. Therefore, complete replacement of $W_{jj}(r+1, m)$ by using only the constraint may cause some unstable behavior. So, in order to gradually impose the constraint, the parameter α is introduced in our method. α will be optimized by experience.

As described in Sect. 5.2, Eq. (45) is not an explicit solution for all $W_{jj}(r+1, m)$, $j = 1, \dots, N$. However, it can be expected that the update changes of $W_{kk}(z)$ are very small, because a small learning rate is usually applied. Therefore, Eq. (45) can be used to solve $W_{jj}(z)$ by treating the $W_{kk}(z)$ in $\mathbf{M}_{jj}^{-1}(z)$ as constants. Usefulness of this approximation will be examined through simulation.

5.4 Combination of Learning Algorithm and Constraint in Time Domain

The distortion-free constraint can also be applied to BSS systems trained in the time domain.

Step 1: Update $w_{jj}(n, l)$ and $w_{jk}(n, l)$ following Eqs. (17) through (19) resulting in $w_{jj}(n+1, l)$ and $w_{jk}(n+1, l)$.

Step 2: $w_{jj}(n+1, l)$ is calculated by Eq. (45), resulting in $\tilde{w}_{jj}(n+1, m)$.

Step 3: Modify $w_{jj}(n, l)$ following

$$w_{jj}(n+1, l) = (1 - \alpha)w_{jj}(n+1, l) + \alpha\tilde{w}_{jj}(n+1, l) \quad (0 < \alpha \leq 1). \quad (50)$$

$w_{jj}(n+1, l)$ in the left-hand side is the modified version, which will be used in the next sample $n+1$. How to calculate $\tilde{w}_{jj}(n+1, l)$ is described in the last paragraph in Sect. 5.2.

5.5 Learning Algorithm with Distortion-Free Constraint for Two Channels

In case of two channels, the distortion-free constraint defined in Eq. (44) can be solved explicitly. Equation (44) is rewritten by

$$W_{jj}^2(z) - W_{jj}(z) - W_{jk}(z)W_{kj}(z) = 0 \quad (51)$$

$$j = 1, 2, k = 1, 2, j \neq k.$$

This equation is solved for $W_{jj}(z)$ as follows:

$$W_{jj}(z) = \frac{1 \pm \sqrt{1 + 4W_{12}(z)W_{21}(z)}}{2}, \quad j = 1, 2. \quad (52)$$

The above equation includes a \pm sign. Thus, there exist two solutions. When the power of $S_j(z)$ is maximum in $X_j(z)$, $S_j(z)$ are likely to be separated in $Y_j(z)$. When the sensor of $X_j(z)$ locates close to $S_j(z)$, this assumption is likely to be satisfied. In this case, we can select the $+$ sign in this equation.

Therefore, in case of two channels, $\tilde{W}_{jj}(r, m)$ in Eq. (49) is expressed as

$$\tilde{W}_{jj}(r+1, m) = \frac{1 + \sqrt{1 + 4W_{12}(r+1, m)W_{21}(r+1, m)}}{2} \quad (53)$$

Next, time domain implementation is considered. Equation (51) is rewritten as follows:

$$\tilde{W}_{jj}^2(z) - \tilde{W}_{jj}(z) = W_{jk}(z)W_{kj}(z) \quad (54)$$

$$\left(\sum_{l=0}^{K_w-1} \tilde{w}_{jj}(n+1, l)z^{-l} \right)^2 - \sum_{l=0}^{K_w-1} \tilde{w}_{jj}(n+1, l)z^{-l} = \sum_{l_1=0}^{K_w-1} \sum_{l_2=0}^{K_w-1} w_{12}(n+1, l_1)w_{21}(n+1, l_2)z^{-(l_1+l_2)}. \quad (55)$$

By comparing the coefficients of the left-hand side and the right-hand side of the above equation, $\tilde{w}_{jj}(n+1, l)$ in Eq. (50) can be obtained as follows:

$$\tilde{w}_{jj}(n+1, 0) = \frac{1 + \sqrt{1 + 4w_{12}(n+1, 0)w_{21}(n+1, 0)}}{2} \quad (56)$$

$$\tilde{w}_{jj}(n+1, l) = \frac{\sum_{l_1=0}^l w_{12}(n+1, l_1)w_{21}(n+1, l-l_1)}{2\tilde{w}_{jj}(n+1, 0) - 1} - \frac{\sum_{l_2=1}^{l-1} \tilde{w}_{jj}(n+1, l_2)\tilde{w}_{jj}(n+1, l-l_2)}{2\tilde{w}_{jj}(n+1, 0) - 1}, \quad l > 0. \quad (57)$$

Equation (57) is successively calculated starting from $l = 1$ until $l = K_w - 1$. The derivation of Eqs. (56) and (57) is given in Appendix B.

5.6 Comparison between Conventional Methods and Proposed Method

In the MDP method, the source separation is affected, since the output signals are controlled to approach to the observations, even though the signal distortion can be reduced. In the PB method, first the learning process for source separation is carried out. After that, the signal distortion is suppressed by multiplying the diagonal elements of the inverse separation matrix.

On the other hand, in the proposed method, the constraint should be combined with the learning algorithm for the source separation. This is because the constraint is derived from both conditions of the complete separation and the signal distortion-free. The learning process is required to make progress taking both conditions into account simultaneously.

6. Simulations and Discussions

6.1 Learning Methods and Their Abbreviations

In this section, several kinds of learning methods will be compared. They are summarized in Table 1.

As described in Sect. 3.2, **FREQ(1)** and **FREQ(2)** do not solve the scaling problem, and may cause the signal distortion.

In the simulation, $w_{jk}(n, l)$ and $w_{jj}(n, l)$ are updated sample by sample. $W_{jk}(r, m)$ and $W_{jj}(r, m)$ are updated frame by frame. Therefore, the average operation $\langle \cdot \rangle$ in Eqs. (17), (20) and (23) are replaced by the instantaneous values.

Simulations are performed for 2 channel and 3 channel BSS systems. In the 2 channel BSS, **TIME (DF)** uses Eqs. (56) and (57). **FREQ (1+DF)** and **FREQ (2+DF)** use Eq. (53). In any case, the same results are obtained by using Eq. (45). In the 3 channel BSS, Eq. (45) is always applied.

6.2 Simulation Setup

A mixture block, simulating actual acoustic spaces by using 256 tap FIR filters, is applied. Speeches and stationary colored signals, generated by 2nd-order AR models, are used as sources. The FFT size is set to 256 points for training in the frequency domain. FIR filters with 256 taps are used for training in the time domain. The initial guess for the separation blocks are $W_{jj}(z) = 1$ and $W_{kj}(z) = 0, k \neq j$. η in the learning algorithms Eqs. (17), (20), (23) and (28) is set to $1.0 \times 10^{-6} \sim 7.0 \times 10^{-6}$. μ in the MDP algorithm Eq. (28) is $0.1 \sim 0.2$. α in the proposed learning algorithms Eqs. (49) and (50) is optimized by experience, and is set to be 1 and 3.0×10^{-4} for the frequency domain BSS and the time domain BSS, respectively. In the frequency domain BSS, $0 < \alpha \leq 1$ have been also tried, as a result, $\alpha = 1$ can provide the best performance. $\alpha = 1$ means that Step 3 is not used. The difference between them should be more investigated.

6.3 Evaluation Measures for Source Separation and Signal Distortion

Source separation is evaluated by the following two signal-

to-interference ratios SIR_1 and SIR_2 . Here, the sources $S_i(z)$ are assumed to be separated at the outputs $Y_i(z)$. This does not lose generality.

$$\sigma_{s1}^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} |A_{ii}(e^{j\omega}) S_i(e^{j\omega})|^2 d\omega \quad (58)$$

$$\sigma_{i1}^2 = \frac{1}{2\pi} \sum_{k=1}^N \sum_{\substack{i=1 \\ \neq k}}^N \int_{-\pi}^{\pi} |A_{ki}(e^{j\omega}) S_i(e^{j\omega})|^2 d\omega \quad (59)$$

$$SIR_1 = 10 \log_{10} \frac{\sigma_{s1}^2}{\sigma_{i1}^2} \quad (60)$$

$$\sigma_{s2}^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} |A_{ii}(e^{j\omega})|^2 d\omega \quad (61)$$

$$\sigma_{i2}^2 = \frac{1}{2\pi} \sum_{k=1}^N \sum_{\substack{i=1 \\ \neq k}}^N \int_{-\pi}^{\pi} |A_{ki}(e^{j\omega})|^2 d\omega \quad (62)$$

$$SIR_2 = 10 \log_{10} \frac{\sigma_{s2}^2}{\sigma_{i2}^2} \quad (63)$$

σ_{s1}^2 and σ_{s2}^2 are the power of the signal source to be separated, and the related transfer function, respectively. σ_{i1}^2 and σ_{i2}^2 are the power of the interference signal and the related transfer function, respectively.

In evaluating the signal distortion, we assume that $S_i(z)$ are dominant in $X_i(z)$, and $H_{ii}(z)S_i(z)$ can be used as the criteria. Furthermore, $S_i(z)$ are assumed to be separated at $Y_i(z)$. Based on these assumptions, four different measures, as shown below, are used to evaluate signal distortion.

$$\sigma_{d1a}^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} |H_{ii}(e^{j\omega}) S_i(e^{j\omega}) - A_{ii}(e^{j\omega}) S_i(e^{j\omega})|^2 d\omega \quad (64)$$

$$\sigma_{d1b}^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} (|H_{ii}(e^{j\omega}) S_i(e^{j\omega})| - |A_{ii}(e^{j\omega}) S_i(e^{j\omega})|)^2 d\omega \quad (65)$$

$$\sigma_1^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} |H_{ii}(e^{j\omega}) S_i(e^{j\omega})|^2 d\omega \quad (66)$$

$$SD_{1x} = 10 \log_{10} \frac{\sigma_{d1x}^2}{\sigma_1^2}, x = a, b \quad (67)$$

$$\sigma_{d2a}^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} |H_{ii}(e^{j\omega}) - A_{ii}(e^{j\omega})|^2 d\omega \quad (68)$$

$$\sigma_{d2b}^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} (|H_{ii}(e^{j\omega})| - |A_{ii}(e^{j\omega})|)^2 d\omega \quad (69)$$

$$\sigma_2^2 = \frac{1}{2\pi} \sum_{i=1}^N \int_{-\pi}^{\pi} |H_{ii}(e^{j\omega})|^2 d\omega \quad (70)$$

Table 1 Abbreviations of applied learning algorithms.

TIME	Eqs. (17)–(19) [6]
TIME (MDP)	Eq. (28) [11]
TIME (DF)	Eqs. (17)–(19) with new distortion-free constraint Eqs. (45) or (56) and (57)
FREQ(1)	Eqs. (20)–(22) [6], [16]
FREQ(1+PB)	Eqs. (20)–(22) with Eq. (32) [13]
FREQ(1+DF)	Eqs. (20)–(22) with new distortion-free constraint Eqs. (45) or (53)
FREQ(2)	Eqs. (21)–(23) [16]
FREQ(2+PB)	Eqs. (21)–(23) with Eq. (32) [13]
FREQ(2+DF)	Eqs. (21)–(23) with new distortion-free constraint Eqs. (45) or (53)

$$SD_{2x} = 10 \log_{10} \frac{\sigma_{d2x}^2}{\sigma_2^2}, x = a, b \quad (71)$$

Since BSS systems are unable to control amplitude of the output signals, the output signals may differ from the criteria in amplitude. In order to neglect this scaling effect in calculating SD_{1x} and SD_{2x} , the average power of $H_{ii}(e^{j\omega})S_i(z)$, $A_{ii}(e^{j\omega})S_i(z)$, $H_{ii}(e^{j\omega})$, and $A_{ii}(e^{j\omega})$ are normalized to unity.

Since the signal distortion SD_{ix} are evaluated based on the difference between the criteria and the simulation results, the smaller SD_{ix} indicate the lower signal distortion. The signal source separation SIR_i are evaluated by a signal to interference ratio. Therefore, basically speaking, the larger SIR_i assume the higher separation performance. However, when the signal distortion is large, the separated signal includes the distortion components, and SIR_i are not reliable. SIR_i should be evaluated under lower signal distortion.

6.4 Two Channel Speech Signal Sources

Simulation results regarding 2 channel speech signals are summarized in Table 2.

In the time domain implementation, TIME is not good in the signal distortion. SIR_i in TIME, which are higher than the others, have no meaning due to the large signal distortions. TIME (MDP) can improve the signal distortion as expected. However, its SIR_i are not good due to the remaining interference $A_{ij}(z)S_j(z)$ at the output $Y_i(z)$ as discussed in Sect. 4.1.

Similarity between $H_{ij}(z)$ and $A_{ij}(z)$ is evaluated by using SD_{ix} defined by Eqs. (67) through (71), where $H_{ii}(z)$, $A_{ii}(z)$ and $S_i(z)$ are replaced by $H_{ij}(z)$, $A_{ij}(z)$ and $S_j(z)$, respectively. Simulation results are shown in Table 3. TIME (MDP) has the smallest values. This means $A_{ij}(z)$ approaches to $H_{ij}(z)$ in TIME (MDP) and the source separations are degraded. When the learning rate μ in Eq. (28) is decreased, SIR_i may be improved, at the same time SD_{ix} are increased. Thus, it is difficult to achieve high performances in both SIR_i and SD_{ix} simultaneously.

On the other hand, the proposed method TIME (DF) can improve the signal distortion, while maintaining relatively high SIR_i .

In the frequency domain implementation, the conventional method FREQ(1) is not good in both SIR_i and SD_{ix} .

Table 2 Comparison of learning algorithms of BSS systems for 2 channel speech signals.

Methods	SIR_1	SIR_2	SD_{1a}	SD_{1b}	SD_{2a}	SD_{2b}
TIME	12.2	5.56	0.25	-2.94	0.57	-3.82
TIME (MDP)	3.98	2.90	-10.3	-13.6	-8.24	-12.3
TIME (DF)	8.33	4.33	-12.1	-16.2	-15.4	-19.9
FREQ(1)	7.37	2.61	-6.23	-8.67	-2.82	-2.95
FREQ(1+PB)	15.2	6.59	-23.1	-27.1	-16.9	-18.8
FREQ(1+DF)	9.68	6.38	-13.5	-18.1	-15.1	-18.3
FREQ(2)	13.0	12.9	-9.43	-15.1	-10.9	-13.9
FREQ(2+PB)	14.3	11.6	-16.9	-20.6	-16.9	-18.9
FREQ(2+DF)	19.8	11.8	-24.6	-28.1	-17.9	-20.6

FREQ(1+PB) can provide good performances in both SIR_i and SD_{ix} , which are higher than those of the proposed method. Furthermore, FREQ(2) can improve both SIR_i and SD_{ix} compared to FREQ(1). However, its SD_{ix} are not sufficient. The improved conventional method FREQ(2+PB) can improve SD_{ix} from FREQ(2). However, the proposed method FREQ(2+DF) can improve more in both SIR_i and SD_{ix} than FREQ(1+PB) and FREQ(2+PB).

6.5 Two Channel Stationary Colored Signal Sources

2nd-order AR models are used to generate stationary colored signals. The amplitude responses of $H_{ii}(z)$, which are the criteria for the signal distortion, are shown in Fig. 2. The broken lines show amplitude responses of $S_1(z)$ and $S_2(z)$. This is because the frequency bands, where their amplitude responses have large values, are meaningful. The horizontal axis is the frequency ranging from 0 to $f_s/2 = 4000$ Hz, where $f_s = 8000$ Hz is the sampling frequency. The frequency components are mainly located around 1540 Hz and 2460 Hz. SIR_i and SD_i are summarized in Table 4.

In the time domain implementations, the signal distortions in TIME are not good. $|A_{ij}(e^{j\omega})|$ are shown in Fig. 3. $|A_{ii}(e^{j\omega})|$ are very different from the criteria shown in Fig. 2.

Table 3 Similarity between $H_{ij}(z)$ and $A_{ij}(z)$ for 2 channel speech signals.

Methods	SD_{1a}	SD_{1b}	SD_{2a}	SD_{2b}
TIME	2.44	-0.25	2.03	-0.63
TIME (MDP)	-9.36	-10.3	-8.53	-9.39
TIME (DF)	-1.14	-5.00	-5.83	-8.12
FREQ(1)	3.80	-4.87	-0.42	-0.55
FREQ(1+PB)	2.19	-2.41	-2.56	-3.16
FREQ(1+DF)	-0.21	6.51	-4.08	-5.11
FREQ(2)	3.89	-5.57	-0.31	-1.00
FREQ(2+PB)	2.21	-1.92	1.38	0.43
FREQ(2+DF)	3.99	-5.39	-0.03	-0.76

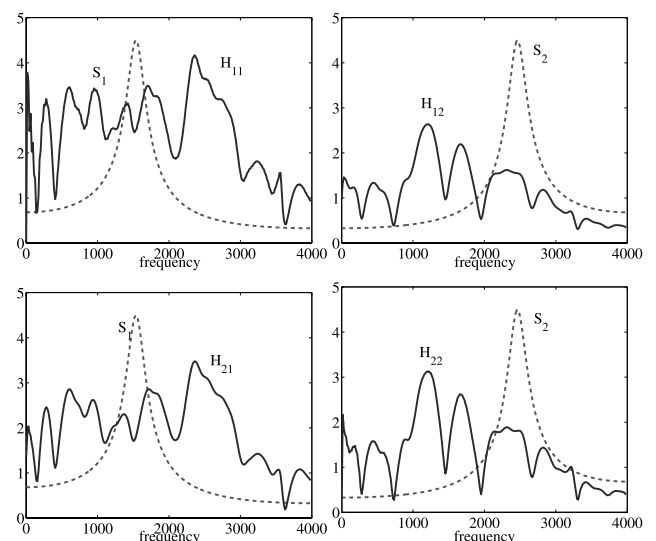


Fig. 2 Amplitude responses of $H_{ij}(z)$ and $S_i(z)$ for colored signals.

They amplify the output signal spectra outside the signal bands. This is due to the degree of freedom in the FF-BSS systems as discussed in Sect. 3.2. The signal distortions can be reduced by TIME (MDP). However, SIR_i are not good as shown in Table 4. The proposed method TIME (DF) can improve both SIR_i and SD_{ix} . $|A_{ij}(e^{j\omega})|$ are shown in Fig. 4. $|A_{ii}(e^{j\omega})|$ are similar to $|H_{ii}(e^{j\omega})|$ in the signal bands, and $|A_{ij}(e^{j\omega})|$ are small in the interference bands. However, in the frequency bands, where the signal source spectra are small, $|A_{ii}(e^{j\omega})|$ and $|A_{ij}(e^{j\omega})|$ are somewhat different from the criteria. Because the learning process is controlled by the output signals, whose spectra are related to the signal source spectra.

In the frequency domain implementations, FREQ (1) is not good in both properties. FREQ(1+PB) can improve SD_{ix} , while maintaining SIR_i lower level. On the other hand, the proposed method FREQ(1+DF) can provide very good performances in both properties.

FREQ(2) can improve SD_{ix} compared with FREQ(1). Even though FREQ(2+PB) and FREQ(2+DF) can improve SD_{ix} slightly, their differences are not so much. In this case, FREQ(1+DF) can provide the best performances in both SIR_i and SD_{ix} .

Table 4 Comparison of learning algorithms of BSS systems for 2 channel stationary colored signals.

Methods	SIR_1	SIR_2	SD_{1a}	SD_{1b}	SD_{2a}	SD_{2b}
TIME	9.49	7.07	-0.28	-3.11	-0.69	-4.99
TIME (MDP)	4.49	2.20	-15.7	-18.8	-13.7	-16.5
TIME (DF)	8.05	4.07	-14.4	-16.6	-10.4	-13.2
FREQ(1)	4.93	4.32	-3.62	-5.03	-3.67	-4.28
FREQ(1+PB)	3.22	1.51	-12.7	-14.3	-7.83	-9.38
FREQ(1+DF)	7.63	4.06	-18.4	-20.8	-10.8	-14.1
FREQ(2)	5.22	2.82	-12.2	-14.3	-7.42	-9.44
FREQ(2+PB)	5.35	2.62	-14.1	-15.8	-8.39	-10.7
FREQ(2+DF)	5.67	2.72	-13.7	-15.7	-8.27	-11.1

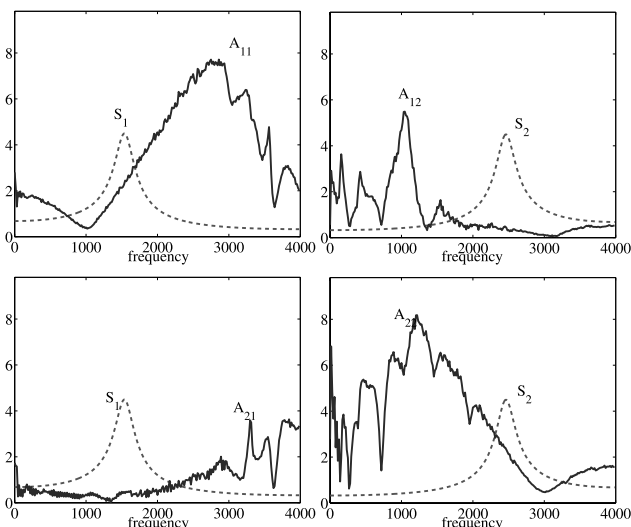


Fig. 3 Amplitude responses $A_{ki}(z)$ obtained by TIME and $S_i(z)$ for colored signals.

6.6 Three Channel Speech Signal Sources

The distortion-free constraint of Eq. (45) is applied for 3 channel signal sources. Simulation results are listed in Table 5. Similar results are obtained as in the previous simulations. FREQ(1+PB) and FREQ(2+PB) can improve SD_{ix} from FREQ(1) and FREQ(2), respectively. FREQ(1+PB) is also better than FREQ(1+DF) as in the 2-channel speech signal. Thus, the conventional method is better than the proposed method in the learning algorithm given by Eq. (20). However, the proposed method FREQ(2+DF) can still provide good performances for all measures.

6.7 Comparison between FREQ (1+DF) and FREQ (2+DF)

Properties of FREQ(1+DF) and FREQ(2+DF), which are recognized from the simulation results, are summarized as follows: FREQ(1+DF) is effective for stationary signals, and FREQ(2+DF) is useful for non-stationary signals. Furthermore, we have confirmed that FREQ(2+DF) can provide good performances for a combination of stationary and non-stationary signals. However, comparison between them

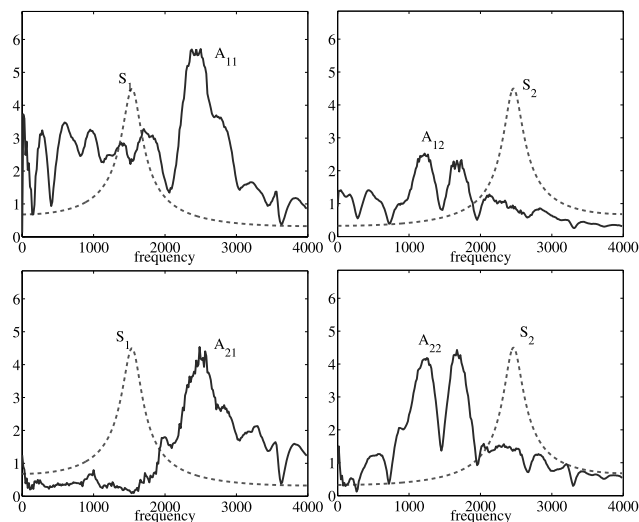


Fig. 4 Amplitude responses $A_{ki}(z)$ obtained by TIME(DF) and $S_i(z)$ for colored signals.

Table 5 Comparison of learning algorithms of BSS systems for 3 channel speech signals.

Methods	SIR_1	SIR_2	SD_{1a}	SD_{1b}	SD_{2a}	SD_{2b}
TIME	13.3	7.58	0.50	-2.48	-0.71	-4.65
TIME (MDP)	6.25	4.36	-8.02	-11.2	-10.2	-14.1
TIME (DF)	8.00	5.24	-14.0	-17.9	-17.1	-20.5
FREQ(1)	11.2	5.32	-10.4	-16.1	-4.39	-4.52
FREQ(1+PB)	15.9	8.61	-19.5	-24.2	-17.7	-19.7
FREQ(1+DF)	8.50	7.36	-14.7	-21.1	-16.3	-18.7
FREQ(2)	17.2	10.3	-13.4	-19.4	-13.9	-16.4
FREQ(2+PB)	16.3	9.68	-17.3	-21.1	-18.0	-19.9
FREQ(2+DF)	16.1	9.56	-24.1	-28.4	-20.1	-21.7

should be more investigated taking spectra of the sources, stationary or non-stationary properties and combinations of different kinds of the sources into account.

6.8 Permutation Problem

In the simulation, we assume that the microphone of $X_i(z)$ are located close to the sources $S_i(z)$. Therefore, the transfer gain from $S_i(z)$ to $X_i(z)$ is larger than those from $S_j(z)$ to $X_j(z)$, $j \neq i$. Furthermore, the initial guess of the separation matrix $W(z)$ are set to be the identity matrix I . Under these conditions, we confirmed ‘Permutation’ did not occur in our simulations. However, since the permutation is highly dependent on the mixing process, possibility of the permutation should be more investigated by changing the conditions on the mixture.

7. Conclusion

In this paper, the condition, which satisfies both complete separation and distortion-free simultaneously for multi-channel FF-BSS systems, is derived. This condition expresses a relation among the filter coefficients of the separation block, and is combined with the learning algorithms as the distortion-free constraint. The new learning algorithms can be applied to the BSS systems implemented in both the time domain and the frequency domain. Compared with the conventional method [13], the proposed method adjusts the separation block so as to satisfy both source separation and signal distortion reduction simultaneously. Simulation results demonstrated one of the proposed methods can provide good performance.

Acknowledgments

The authors are grateful to Dr. Boris Jansen and A.Prof. Eisuke Horita for their fruitful discussions on this work.

A part of this research has been supported by Grant-in-Aid for Scientific Research ((C)17560335).

References

- [1] C. Jutten and J. Herault, “Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture,” *Signal Process.*, vol.24, no.1, pp.1–10, July 1991.
- [2] P. Comon, C. Jutten, and J. Herault, “Blind separation of sources, Part II: Problems statement,” *Signal Process.*, vol.24, no.1, pp.11–20, July 1991.
- [3] E. Sorouchyari, “Blind separation of sources, Part III: Stability analysis,” *Signal Process.*, vol.24, no.1, pp.21–29, July 1991.
- [4] H.L. Nguyen Thi and C. Jutten, “Blind source separation for convolutive mixtures,” *Signal Process.*, vol.45, no.2, pp.209–229, March 1995.
- [5] A. Cichocki, S. Amari, M. Adachi, and W. Kasprzak, “Self-adaptive neural networks for blind separation of sources,” *Proc. ISCAS’96*, pp.157–161, Atlanta, May 1996.
- [6] S. Amari, T. Chen, and A. Cichocki, “Stability analysis of learning algorithms for blind source separation,” *Neural Netw.*, vol.10, no.8, pp.1345–1351, Nov. 1997.
- [7] C. Simon, G. d’Urso, C. Vignat, Ph. Loubaton, and C. Jutten, “On the convolutive mixture source separation by the decorrelation approach,” *IEEE Proc. ICASSP’98*, pp.IV-2109–2112, Seattle, May 1998.
- [8] S. Cruces and L. Castedo, “A Gauss-Newton methods for blind source separation of convolutive mixtures,” *IEEE Proc. ICASSP’98*, pp.IV2093–2096, Seattle, May 1998.
- [9] L. Parra and C. Spence, “Convolutive blind separation of nonstationary source,” *IEEE Trans. Speech Audio Process.*, vol.8, no.3, pp.320–327, May 2000.
- [10] H. Mathis and S.C. Douglas, “On optimal and universal nonlinearities for blind signal separation,” *IEEE Proc. ICASSP’01, MULT-P3.3*, Salt Lake City, UT, USA, May 2001.
- [11] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *Proc. ICA2001*, pp.722–727, San Diego, CA, USA, Dec. 2001.
- [12] I. Kopriva, Z. Devcic, and H. Szu, “An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures,” *IEEE&INNS Proc. IJCNN’01*, pp.424–429, Washington, DC, July 2001.
- [13] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” *Neuro-computing*, vol.41, pp.1–24, Oct. 2001.
- [14] K. Nakayama, A. Hirano, and T. Sakai, “An adaptive nonlinear function controlled by kurtosis for blind source separation,” *IEEE&INNS, Proc. IJCNN2002*, pp.1234–1239, Honolulu, Hawaii, May 2002.
- [15] K. Nakayama, A. Hirano, and A. Horita, “A learning algorithm for convolutive blind source separation with transmission delay constraint,” *IEEE&INNS, Proc. IJCNN2002*, pp.1287–1292, Honolulu, Hawaii, May 2002.
- [16] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, “The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech,” *IEEE Trans. Speech Audio Process.*, vol.11, no.2, pp.109–116, March 2003.
- [17] K. Nakayama, A. Hirano, and A. Horita, “A learning algorithm with adaptive exponential stepsize for blind source separation of convolutive mixtures with reverberations,” *IEEE&INNS, Proc. IJCNN2003* pp.1092–1097, Portland, Oregon, July 2003.
- [18] K. Nakayama, A. Hirano, and T. Sakai, “An adaptive nonlinear function controlled by estimated output pdf for blind source separation,” *Proc. ICA2003*, pp.427–432, Nara, Japan, April 2003.
- [19] K. Nakayama, A. Hirano, and Y. Dejima, “Analysis of signal separation and distortion analysis in feedforward blind source separation for convolutive mixture,” *IEEE Proc. The 47th International Midwest Symposium on Circuits and Systems (MWSCAS2004)*, pp.III 207–III 210, Hiroshima, Japan, Sept. 2004.
- [20] A. Horita, Y. Dejima, K. Nakayama, and A. Hirano, “Comparative study of convergence performance and signal distortion in feedforward BSS and feedback BSS,” *SIP Symposium 2004*, no.B3-2, Yatsugatake, Nov. 2004.
- [21] A. Horita, K. Nakayama, A. Hirano, and Y. Dejima, “Analysis of signal separation and signal distortion in feedforward and feedback blind source separation based on source spectra,” *IEEE&INNS, Proc. IJCNN2005*, pp.1257–1262, Montreal, July-Aug. 2005.
- [22] A. Horita, K. Nakayama, A. Hirano, and Y. Dejima, “Analysis of a learning algorithm with distortion free constraint for convolutive blind source separation in time domain,” *SIP Symposium 2005*, no.C2-1, Kochi, Nov. 2005.
- [23] A. Horita, K. Nakayama, A. Hirano, and Y. Dejima, “A distortion free learning algorithm for feedforward BSS with convolutive mixture and multi-channel signal sources,” *IEICE Technical Report, SIP2006-15*, May 2006.
- [24] A. Horita, K. Nakayama, A. Hirano, and Y. Dejima, “A distortion free learning algorithm for feedforward BSS and its comparative study with feedback BSS,” *IEEE&INNS, Proc. IJCNN2006*, pp.7642–7649, Vancouver, July 2006.

- [25] A. Horita, K. Nakayama, A. Hirano, and Y. Dejima, "A learning algorithm with distortion free constraint and comparative study for feedforward and feedback BSS," Proc. EUSIPCO2006, Florence, Italy, Sept. 2006.

Appendix A: Derivation of Eq.(45)

det $\mathbf{W}(z)$ is given by

$$\det \mathbf{W}(z) = \sum_{k=1}^N W_{jk}(z)(-1)^{j+k} \det \mathbf{M}_{jk}(z). \quad (\text{A} \cdot 1)$$

In general, $\hat{W}_{jj}(z)$ is expressed as

$$\hat{W}_{jj}(z) = (-1)^{2j} \det \mathbf{M}_{jj}(z) = \det \mathbf{M}_{jj}(z). \quad (\text{A} \cdot 2)$$

From Eqs. (44), (A·1) and (A·2), we obtain

$$\det \mathbf{M}_{jj}(z) = \sum_{k=1}^N W_{jk}(z)(-1)^{j+k} \det \mathbf{M}_{jk}(z). \quad (\text{A} \cdot 3)$$

In this equation, $W_{jj}(z)$ is extracted.

$$\begin{aligned} \det \mathbf{M}_{jj}(z)(1 - W_{jj}(z)) \\ = \sum_{\substack{k=1 \\ \neq j}}^N W_{jk}(z)(-1)^{j+k} \det \mathbf{M}_{jk}(z) \end{aligned} \quad (\text{A} \cdot 4)$$

det $\mathbf{M}_{jk}(z)$ is further rewritten as:

$$\det \mathbf{M}_{jk}(z) = \sum_{\substack{l=1 \\ \neq j}}^N W_{lj}(z)\kappa(j, k, l) \det \mathbf{m}_{lj}(z) \quad (\text{A} \cdot 5)$$

$$\kappa(j, k, l) = \begin{cases} (-1)^{l+j} & (l < k) \\ -(-1)^{l+j} & (l \geq k) \end{cases} \quad (\text{A} \cdot 6)$$

where $\mathbf{m}_{lj}(z)$ is an $(N-2) \times (N-2)$ minor matrix, i.e. removing the l -th row, j -th column from $\mathbf{M}_{jk}(z)$. Therefore, the right hand side of Eq. (A·4) is rewritten as:

$$- \sum_{\substack{k=1 \\ \neq j}}^N W_{jk}(z) \sum_{\substack{l=1 \\ \neq j}}^N W_{lj}(z)\kappa(k, k, l) \det \mathbf{m}_{lj}(z) \quad (\text{A} \cdot 7)$$

$$= -\mathbf{w}_{row,j}^T(z) \text{adj } \mathbf{M}_{jj}(z) \mathbf{w}_{col,j}(z). \quad (\text{A} \cdot 8)$$

Finally, this results in the equation given in Eq. (45).

$$W_{jj}(z) = 1 + \mathbf{w}_{row,j}^T(z) \frac{\text{adj } \mathbf{M}_{jj}(z)}{\det \mathbf{M}_{jj}(z)} \mathbf{w}_{col,j}(z) \quad (\text{A} \cdot 9)$$

$$= 1 + \mathbf{w}_{row,j}^T(z) \mathbf{M}_{jj}^{-1}(z) \mathbf{w}_{col,j}(z). \quad (\text{A} \cdot 10)$$

Appendix B: Derivation of Eqs. (56) and (57)

Equation (55) is rewritten as:

$$\begin{aligned} \sum_{l=0}^{K_w-1} \sum_{l'=0}^{K_w-1} \tilde{w}_{jj}(n+1, l) \tilde{w}_{jj}(n+1, l') z^{-(l+l')} \\ - \sum_{l=0}^{K_w-1} \tilde{w}_{jj}(n+1, l) z^{-l} \\ = \sum_{l_1=0}^{K_w-1} \sum_{l_2=0}^{K_w-1} w_{12}(n+1, l_1) w_{21}(n+1, l_2) z^{-(l_1+l_2)}. \end{aligned} \quad (\text{A} \cdot 11)$$

In Eq. (A·11), coefficients of the term z^0 are related by

$$\begin{aligned} \tilde{w}_{jj}^2(n+1, 0) - \tilde{w}_{jj}(n+1, 0) \\ = w_{12}(n+1, 0) w_{21}(n+1, 0). \end{aligned} \quad (\text{A} \cdot 12)$$

This equation is solved for $\tilde{w}_{jj}(n+1, 0)$, then Eq. (56) is obtained.

Furthermore, coefficients of the terms z^l , ($l > 0$) are expressed as follows:

Coefficients of the first term in the left hand side in Eq. (A·11):

$$\sum_{k=0}^l \tilde{w}_{jj}(n+1, k) \tilde{w}_{jj}(n+1, l-k). \quad (\text{A} \cdot 13)$$

Coefficients of the second term in the left hand side in Eq. (A·11):

$$\tilde{w}_{jj}(n+1, l). \quad (\text{A} \cdot 14)$$

Coefficients of the right hand side in Eq. (A·11):

$$\sum_{l_1=0}^l w_{12}(n+1, l_1) w_{21}(n+1, l-l_1). \quad (\text{A} \cdot 15)$$

In Eq.(A·13), $\tilde{w}_{jj}(n+1, l)$ is separated from the summation as follows:

$$\begin{aligned} 2\tilde{w}_{jj}(n+1, 0) \tilde{w}_{jj}(n+1, l) \\ + \sum_{k=1}^{l-1} \tilde{w}_{jj}(n+1, k) \tilde{w}_{jj}(n+1, l-k). \end{aligned} \quad (\text{A} \cdot 16)$$

Therefore, coefficients of the terms z^l , ($l > 0$) in Eq. (A·11) are related by

$$\begin{aligned} (2\tilde{w}_{jj}(n+1, 0) - 1) \tilde{w}_{jj}(n+1, l) \\ + \sum_{k=1}^{l-1} \tilde{w}_{jj}(n+1, k) \tilde{w}_{jj}(n+1, l-k) \\ = \sum_{l_1=0}^l w_{12}(n+1, l_1) w_{21}(n+1, l-l_1). \end{aligned} \quad (\text{A} \cdot 17)$$

Equation (A·17) is solved for $\tilde{w}_{jj}(n+1, l)$, then Eq. (57) is derived.



Akihide Horita was born in Toyama, Japan in 1979. He received a B.E. degree in engineering in 2002 and a M.E. degree in engineering in 2004 both from Kanazawa University. He is currently working towards his Ph.D. degree at Kanazawa University. His research interests include adaptive filters and blind source separation.



Kenji Nakayama received the B.E. and Dr. degrees in electronics engineering from Tokyo Institute of Technology (TIT), Tokyo, Japan, in 1971 and 1983, respectively. From 1971 to 1972 he was engaged in the research on classical network theory in TIT. He was involved in NEC Corporation from 1972 to 1988, where his research subjects were filter design methodology and digital signal processing. He joined the Department of Electrical and Computer Engineering at Kanazawa University, in Aug. 1988. He is

currently a Professor of Graduate School of Natural Science and Technology. He has served as the Director of the International Student Center from April 2003 to March 2005. His current research interests include adaptive signal processing and neural networks. He is a senior member of IEEE and a member of INNS and RISP, Japan.



Akihiro Hirano received the B. Eng., M. Eng. and Dr. Eng. degrees from Kanazawa University, Kanazawa, Japan in 1987, 1989 and 2000, respectively. He was involved in NEC Corporation, Kawasaki, Japan from 1989 to 1998, where he had been a Research Engineer in Research and Development Group. He joined Faculty of Engineering, Kanazawa University in 1998. He is currently an Assistant Professor in Division of Electrical Engineering and Computer Science, Graduate School of Natural Science and Technology, Kanazawa University. He has been engaged in

researches on adaptive signal processing and neural networks. He was awarded the 1995 Academic Encouragement Award by IEICE. Dr. Hirano is a member of IEEE.