PAPER

Evaluation of separation and dereverberation performance in frequency domain blind source separation

Ryo Mukai*, Shoko Araki, Hiroshi Sawada and Shoji Makino

NTT Communication Science Laboratories, NTT Corporation,
2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619–0237 Japan

Abstract: In this paper, we propose a new method for evaluating the separation and dereverberation performance of a convolutive blind source separation (BSS) system, and investigate a separating system obtained by employing frequency domain BSS based on independent component analysis (ICA). As a result, we reveal the acoustical characteristics of the frequency domain BSS for convolutive mixture of speech signals. We show that the separating system removes the direct sound of a jammer signal even when the frame length is relatively short, and it also reduces the reverberation of the jammer according to the frame length. We also confirm that the reverberation of the target is not reduced. Moreover, we propose a technique, suggested by the experimental results, for improving the quality of the separated signals by removing pre-echo noise.

1. INTRODUCTION

Blind source separation (BSS) is a technique for estimating original source signals using only observed mixtures of signals. The BSS of audio signals has a wide range of applications including noise robust speech recognition, hands-free telecommunication systems and high-quality hearing aid equipment.

Independent component analysis (ICA) is a typical BSS method that is effective for instantaneous (non-convolutive) mixtures [1–3], and many attempts have been made to apply BSS to signals mixed in convolutive environments [4–8]. However, it has also been pointed out that an adequate performance level cannot be obtained in environments with long reverberation where the filter lengths of the mixing and separating systems are of the order of thousands or higher [9–11].

Although there have been a large number of studies of BSS, little attention has been paid to the physical and acoustical characteristics of the separating system, since ICA was originally a statistical method. It is useful to investigate the acoustical nature of BSS with a view to improving separation performance, and some researchers have begun to investigate and utilize its physical properties.

The effect of early reflection and reverberation on BSS is investigated in [12]. The relationship between BSS and beamformers is utilized to achieve a better performance in [13,14]. Still only a few physical characteristics is known.

In this paper, we propose a method for analyzing a separating system obtained by frequency domain BSS for convolutive mixture of speech signals from an acoustical viewpoint. We measure the impulse responses of straight and cross paths of the separating system, and examine the separation performance in detail. We focus our attention on the power of (1) the direct sound of the target signal, (2) the reverberation of the target signal, (3) the direct sound of the jammer signal, and (4) the reverberation of the jammer signal, and evaluate each power separately.

As a result, we reveal the acoustical characteristics of frequency domain BSS based on ICA. We show that the separating system removes the direct sound of the jammer signal even when the frame length is relatively short, and it also reduces the reverberation of the jammer according to the frame length. We also confirm that the reverberation of the target is not reduced as expected. Moreover, we propose a technique, suggested by the experimental results, for improving the quality of the separated signals by removing pre-echo noise.

The inability to achieve target dereverberation can be predicted and explained by the ambiguity of a linear

---

*e-mail: ryo@cslab.kecl.ntt.co.jp

transformation of BSS [15,16]. However other characteristics were unknown until we revealed them by means of experiments. Our results also provided experimental confirmation of the equivalence between BSS and adaptive beamformers (ABF) which is discussed in [17,18].

The paper is organized as follows. In the next section, we summarize the algorithm of frequency domain BSS for convolutive mixtures. In Section 3, we define performance measurement factors using impulse responses and their power. Section 4 presents experimental results obtained using speech signals. In Section 5, we discuss the acoustical characteristics of the BSS system obtained by frequency domain ICA. Section 6 concludes this paper.

2. BSS OF CONVOLUTIVE MIXTURES

2.1. Frequency Domain ICA

When the source signals are $s_i(t)\,(1 \leq i \leq N)$, the signals observed by microphone j are $x_j(t)\,(1 \leq j \leq M)$, and the separated signals are $y_k(t)\,(1 \leq k \leq N)$, the BSS model can be described by the following equations:

$$x_j(t) = \sum_{i=1}^{N} (h_{ji} * s_i)(t) \tag{1}$$

$$y_k(t) = \sum_{j=1}^{M} (w_{kj} * x_j)(t) \tag{2}$$

where h_{ji} is the impulse response from source i to microphone j, w_{kj} is the coefficient when we assume that a separating system is used as FIR filters, and $*$ denotes the convolution operator.

A convolutive mixture in the time domain corresponds to instantaneous mixtures in the frequency domain. Therefore, we can apply an ordinary ICA algorithm in the frequency domain to solve a BSS problem in a reverberant environment. Using a short-time discrete Fourier transform for (1), the model is approximated as:

$$X(\omega, n) = H(\omega)S(\omega, n), \tag{3}$$

where ω denotes the frequency, and n represents the frame index. The separating process can be formulated in each frequency bin ω as:

$$Y(\omega, n) = W(\omega)X(\omega, n) \tag{4}$$

$$= W(\omega)H(\omega)S(\omega, n), \tag{5}$$

where $S(\omega, n) = [S_1(\omega, n), \ldots, S_N(\omega, n)]^T$ is the source signal in frequency bin ω, $X(\omega, n) = [X_1(\omega, n), \ldots, X_M(\omega, n)]^T$ denotes the observed signals, $Y(\omega, n) = [Y_1(\omega, n), \ldots, Y_N(\omega, n)]^T$ is the estimated source signal, and $W(\omega)$ represents the separating matrix. $W(\omega)$ is determined so that $Y_i(\omega, n)$ and $Y_j(\omega, n)$ become mutually independent ($i \neq j$).

To calculate the separating matrix W, we use an optimization algorithm based on the minimization of the mutual information of Y. The optimal W is obtained by the following iterative equation using the natural gradient approach [19]:

$$W_{i+1} = W_i + \mu[I - \langle \Phi(Y)Y^H \rangle]W_i, \tag{6}$$

where i is an index for the iteration, I is an identity matrix, μ is a step size parameter, $\langle \cdot \rangle$ denotes the averaging operator, and $\Phi(\cdot)$ is a nonlinear function. Because the signals are complex valued in the frequency domain, we use a polar-coordinated based nonlinear function [20]:

$$\Phi(Y) = \tanh(g \cdot \mathrm{abs}(Y))e^{j\arg(Y)}, \tag{7}$$

where g is a gain parameter to control the nonlinearity.

The above calculations are carried out separately for each frequency.

2.2. Permutation and Scaling Problem

Once we have solved ICA for all frequencies, we need to solve the permutation and scaling problems. Since we are handling signals with complex values, the scaling factors are also complex values. Thus the scaling can be divided into phase scaling and amplitude scaling.

We use the method described in [13] to solve the permutation and phase scaling problems. When we consider a separating system as a microphone array, we can write directivity patterns for every frequency bin. The permutation problem is solved so that the null directions are aligned. We can estimate the directions of the source signals from the aligned directivity patterns, and the phase scaling problem is solved so that the phase response of the estimated source direction becomes zero.

The amplitude scaling problem is solved by the method reported in [21] with a slight modification. We calculate the inverse of the separating matrices $W(\omega)^{-1}$, and decide the scaling factors so that the norms of each column of $W(\omega)^{-1}$ become uniform.

In the following sections, we consider a two-input, two-output convolutive BSS problem, i.e., $N = M = 2$ (Fig. 1).

3. EVALUATION METHOD

The BSS performance is usually evaluated by the signal to interference ratio (SIR), which is defined as the ratio of a
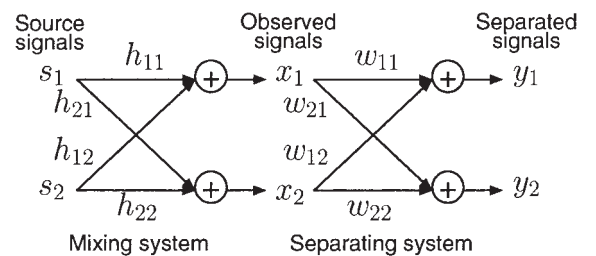

Fig. 1 Model of BSS system.

target-originated signal to a jammer-originated signal [9,10,22]. This measure is reasonable for evaluating the separation performance, but unsuitable for evaluating the dereverberation performance because of its inability to distinguish the direct sound and reverberation. Since we wish to know the components of separated signals in detail, i.e., the direct sound and reverberation of the target and jammer, we adopt the following procedure.

 (i) We estimate the separating matrix $W(\omega)$ for each frequency.
 (ii) By using IFFT, we transform the frequency domain separating matrix $W(\omega)$ to time domain separating filter $w_{ij}(t)$.
(iii) We measure four impulse responses, from s_1 to y_1, s_1 to y_2, s_2 to y_1, and s_2 to y_2.
(iv) We investigate the four impulse responses in detail and compare them with the responses of a null beamformer (NBF).

3.1. Definitions of Performance Measurement Factors

We evaluate the performance of the separating system in the time domain. We consider a separated signal y_1, target signal s_1, and jammer signal s_2. When the target s_1 is an impulse $\delta(t)$ and the jammer $s_2 = 0$, we represent the observed signal x_1 as x_{1s1} [Fig. 2(a)], and y_1 as y_{1s1} [Fig. 2(b)]. Similarly, when $s_1 = 0$ and $s_2 = \delta(t)$, we represent x_1 as x_{1s2}, and y_1 as y_{1s2} [Fig. 2(c)]. x_{1s1} is an impulse response from s_1 to x_1 caused by the mixing system H, and y_{1s1} is an impulse response from s_1 to y_1 caused by the whole system $W \cdot H$. These values are calculated by using h_{ij} and w_{ij} as follows.

$$x_{1s1} = h_{11} \tag{8}$$

$$x_{1s2} = h_{12} \tag{9}$$

$$y_{1s1} = w_{11} * h_{11} + w_{12} * h_{21} \tag{10}$$
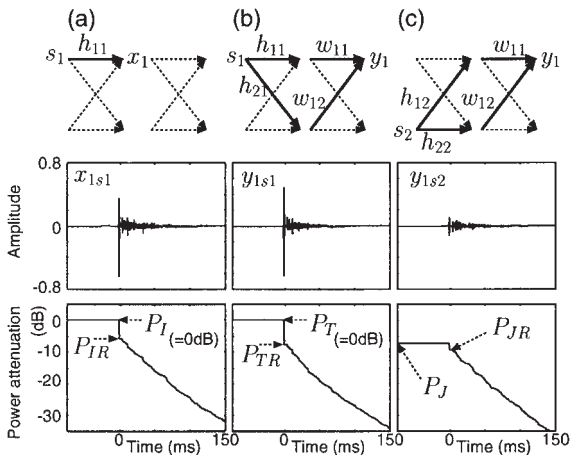
Fig. 2 Impulse responses of (a) observed target, (b) straight path and (c) cross path, and their power attenuation curves.

$$y_{1s2} = w_{11} * h_{12} + w_{12} * h_{22} \tag{11}$$

In terms of source separation, we can consider y_{1s1} to be the direct and reverberant sound of target s_1, and y_{1s2} to be the remaining sound of jammer s_2.

To simplify the evaluation, we normalize h_{ji} so that the power of the observed signals x_{1s1} and x_{1s2} is equal to 0 dB, and employ the following definitions.

- P_I: the power of x_{1s1}. (normalized to 0 dB)
- P_T: the power of y_{1s1}. (scaled to 0 dB)
- P_J: the power of y_{1s2}.
- P_IR: the power of the reverberant sound in x_{1s1},
- P_TR: the power of the reverberant sound in y_{1s1},
- P_JR: the power of the reverberant sound in y_{1s2}.

These are calculated as follows:

$$P_\mathrm{I} = \sum_{t=-\infty}^{\infty} |x_{1s1}(t)|^2, \quad P_\mathrm{IR} = \sum_{t=\tau}^{\infty} |x_{1s1}(t)|^2 \tag{12}$$

$$P_\mathrm{T} = \sum_{t=-\infty}^{\infty} |y_{1s1}(t)|^2, \quad P_\mathrm{TR} = \sum_{t=\tau}^{\infty} |y_{1s1}(t)|^2 \tag{13}$$

$$P_\mathrm{J} = \sum_{t=-\infty}^{\infty} |y_{1s2}(t)|^2, \quad P_\mathrm{JR} = \sum_{t=\tau}^{\infty} |y_{1s2}(t)|^2 \tag{14}$$

where τ is an appropriate time between the arrival of the direct sound and the reverberant sound, which is determined according to experimental conditions. Usually, the impulse response in a real room consists of direct sound, early reflection and reverberation. There are a delay and a level difference between the direct sound and the first reflection due to the difference between their paths. We consider that the early reflections are included in the reverberation, thus τ is set to the time just before the arrival of the first reflective sound. These measurement factors can be indicated on the the power attenuation curves in Fig. 2

We also define the reduction ratio of the reverberation of target signal R_T and the reduction ratio of jammer signal R_J as follows

$$R_\mathrm{T} = -10 \log \frac{P_\mathrm{TR}}{P_\mathrm{IR}} \; (\mathrm{dB}) \tag{15}$$

$$R_\mathrm{J} = -10 \log \frac{P_\mathrm{J}}{P_\mathrm{T}} = -10 \log P_\mathrm{J} \; (\mathrm{dB}). \tag{16}$$

For the analysis of the residual jammer signal in the following section, we also define the power of the early part of the residual jammer signal as follows:

$$P_\mathrm{JE} = \sum_{-\infty < t < \tau} |y_{1s2}(t)|^2$$
$$= P_\mathrm{J} - P_\mathrm{JR}, \tag{17}$$

which is the total power of the remaining direct sound and the non-causal noise called pre-echo.
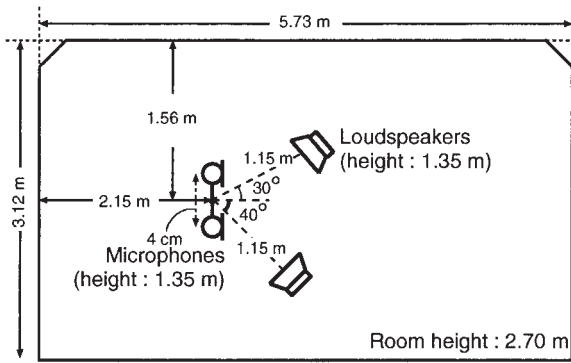
Fig. 3 Layout of the room used for the experiments. Reverberation time $T_{rev} = 300$ ms.

Fig. 4 Target and jammer impulse responses of NBF and ICA ($T = 2,048$, $T_{rev} = 300$ ms).

4. EXPERIMENTS

In order to examine what is separated by a separating system based on ICA, and what remains as noise, we investigated the impulse responses of a system. In frequency domain BSS, it has been confirmed that the separation performance changes according to the length of the frame [10], therefore we chose the frame length as a parameter. In our experiment, the length of the filters of the separating system was the same as the frame length.

4.1. Experimental Conditions

The layout of the room we used to measure the impulse responses of the mixing system H is shown in Fig. 3. The reverberation time of the room T_{rev} was 300 ms, which corresponds to an impulse response of 2,400 taps at an 8 kHz sampling rate. We used a two-element array with an inter-element spacing of 4 cm, which corresponds to almost the half-wave length of the Nyquist frequency 4 kHz. This spacing makes it possible to avoid spatial aliasing over the whole frequency range. The speech signals arrived from two directions, i.e., $-30°$ and $40°$. The contribution of the direct sound of h_{11} and h_{21} was 5.7 dB, and that of h_{12} and h_{22} was 6.6 dB.

We carried out experiments with 42 signal source combinations using seven sentences spoken by two male and two female speakers selected from the ASJ continuous speech corpus. Each of these mixed speech signals were six seconds long. We used the entire six seconds of the mixed data for learning according to (6).

In these experiments, we changed the frame length T from 4 to 8,192 and investigated the performance under each condition. The sampling rate was 8 kHz, and the analysis window was a Hanning window. The frame shift S was $T/4$.

The step size parameter μ affects the convergence of the separating system W and the separation performance. We varied μ from 0.01 to 0.5, and selected the best step size for each frame size T. We used $\mu = 0.1$ for $T \leq 1,024$,
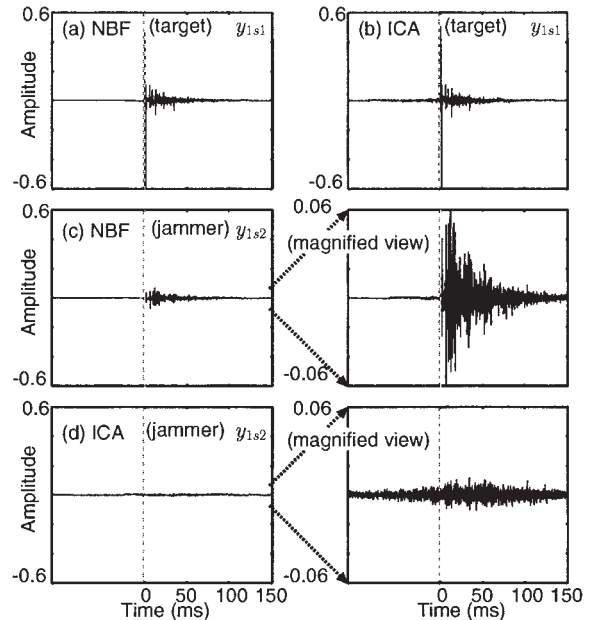
$\mu = 0.05$ for $T = 2,048$, and $\mu = 0.02$ for $T \geq 4,096$. The gain parameter g is ∞. This means that the nonlinear function $\Phi(Y) = e^{j\arg(Y)}$. The number of iterations for (6) is 100.

4.2. Experimental Results

Figure 4(a) and (c) show examples of impulse responses y_{1s1} and y_{1s2} of the separating system obtained by a null beamformer (NBF) that forms a steep null directivity pattern towards a jammer on the assumption that the jammer's direction is known. Figure 4(b) and (d) are results obtained by ICA.

For the target signal, we can see that the reverberation passes the system in both cases (NBF and ICA) in Fig. 4(a) and (b). Figure 4(c) shows that the direct sound of the jammer is eliminated, but the reverberation is not eliminated by the NBF, as expected. By contrast, Fig. 4(d) indicates that ICA not only eliminates the direct sound, but also reduces the reverberation of the jammer. However, we can also observe the appearance of non-causal noise. We discuss this in the following section.

Figure 5 shows the relationship between the frame length T and the reduction ratios R_T and R_J defined by (15) and (16). Figure 5(a) shows results provided by ICA. For the sake of comparison, the NBF performance is shown in Fig. 5(b).

Note that these results are measured by using the power of the impulse responses, and differ from the noise reduction rate (NRR) [10] measured by using a speech signal with a highly colored spectrum. Our results indicate larger values than the NRR.
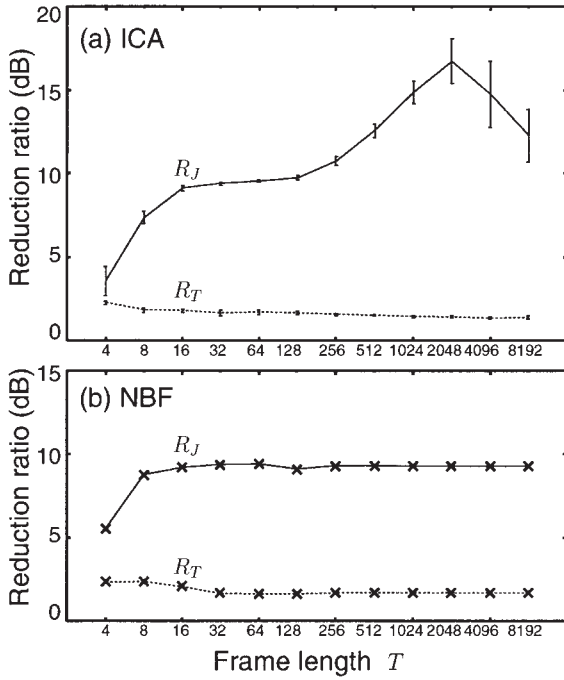
Fig. 5 Relationship between frame length and reduction ratios ($T_{rev} = 300\,ms$). The target reduction ratio R_T and the jammer reduction ratio are defined by (15) and (16).

Fig. 6 Analysis of the power of the residual jammer signal ($T_{rev} = 300\,ms$). P_J is the power of the residual jammer signal, which is decomposed into the early part P_{JE} and the residual reverberation P_{JR}.

Figure 6 shows the power of the residual jammer signal P_J and its decomposition into the early part P_{JE} and the residual reverberation P_{JR}. We used $\tau = 2\,ms$ (16 taps). Figure 6(a) is the result of ICA and Fig. 6(b) is the result of NBF.
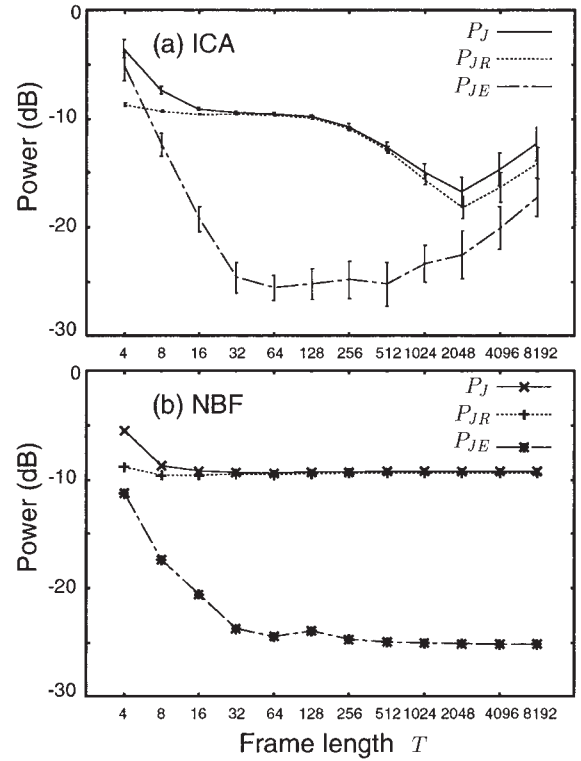
5. DISCUSSION

5.1. Jammer Reduction and Target Dereverberation

First, we discuss the jammer reduction ratio R_J in Fig. 5. When $T \leq 128$, the reduction performance of BSS is as poor as that of NBF, and when $256 \leq T \leq 2,048$, the reduction ratio increases. When $T = 2,048$, the average R_J value is 16.7 dB. This is much greater than the contribution of the direct sound (6.6 dB). This means that the separating system obtained by ICA can reduce not only the direct sound of the jammer but also its reverberant sound. However, as we describe later, the reverberation is not eliminated completely.

By contrast, the reduction ratio of the reverberation of target R_T is low, and does not vary throughout the entire frame length T. This means that dereverberation was not achieved for the target signal. From these results, we can conclude that W is not the approximation of the inverse system of H, but a filter that can eliminate the jammer signal.

It has been pointed out that early reflections of the jammer signal are removed by BSS [12]. We obtained a slightly stronger result that not only the early reflections

but also the reverberation of the jammer signal is reduced to some degree. The reason for this is that frequency domain BSS works like two sets of frequency domain adaptive microphone arrays, i.e., adaptive beamformers (ABF), which adapt to minimize the jammer signal including reverberation. Figure 7 shows the directivity patterns of NBF, the separating system obtained by ICA, and ABF. ICA and ABF provide duller directivity than NBF, thus they can remove not only the direct sound of the jammer but also its reverberation. A theoretical discussion of the relationship between frequency domain BSS using second order statistics and ABF can be found in [17].

5.2. Analysis of Residual Jammer Signal

Next, we examine the power of the residual jammer signal P_J in detail. Figure 6 shows the power of the residual jammer signal P_J and its decomposition to P_{JR} and P_{JE}. P_{JR} is the contribution of the residual reverberation. P_{JE} is the contribution of the early part of the residual jammer signal, which consists of the pre-echo noise and the residual direct sound of the jammer signal.

Figure 8 shows the jammer impulse responses for various frame lengths. From Figs. 6 and 8, we can derive following consideration.

When the frame length is too short, the direct sound cannot be removed by the separating system, and it
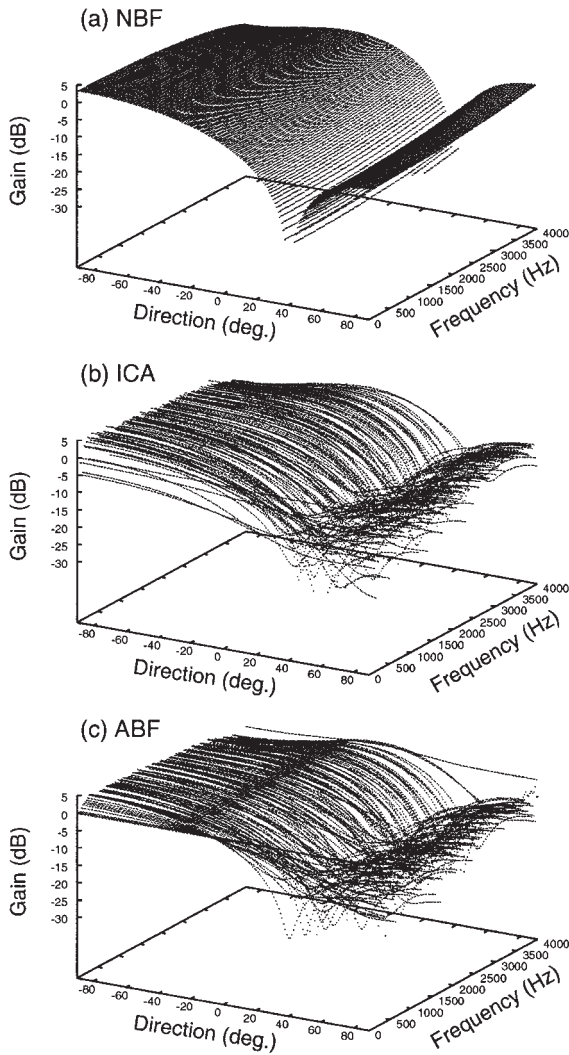
Fig. 7 Directivity patterns of (a) NBF, (b) separating system obtained by ICA, and (c)ABF ($T = 2,048$, $T_{\text{rev}} = 300\,\text{ms}$).

Fig. 8 Jammer impulse responses of BSS system.
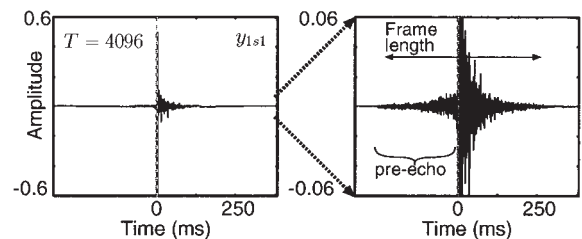
Fig. 9 Target impulse response of BSS system damaged by pre-echo noise.

degrades the separation performance. When $64 \leq T \leq 1,024$, the direct sound is removed well, however the filter is much shorter than the length of the reverberation; accordingly, any reverberation that is longer than the filter cannot be reduced at all.

As the frame length becomes longer, the number of coefficients to be estimated increases while the number of data samples for learning in each frequency bin decreases, since we are using finite length data of 6 seconds. For example, when $T = 8,192$ and $S = T/4$, each frequency bin has only 24 data samples and this is insufficient for the statistical estimation. As a result, the amount of estimation errors increases.

Moreover, the pre-echo noise grows, and this causes the performance to decline. Furthermore, the noise damages the target signal when T is too long (Fig. 9). Thus, there is a tradeoff between coverage of reverberation and estimation errors.
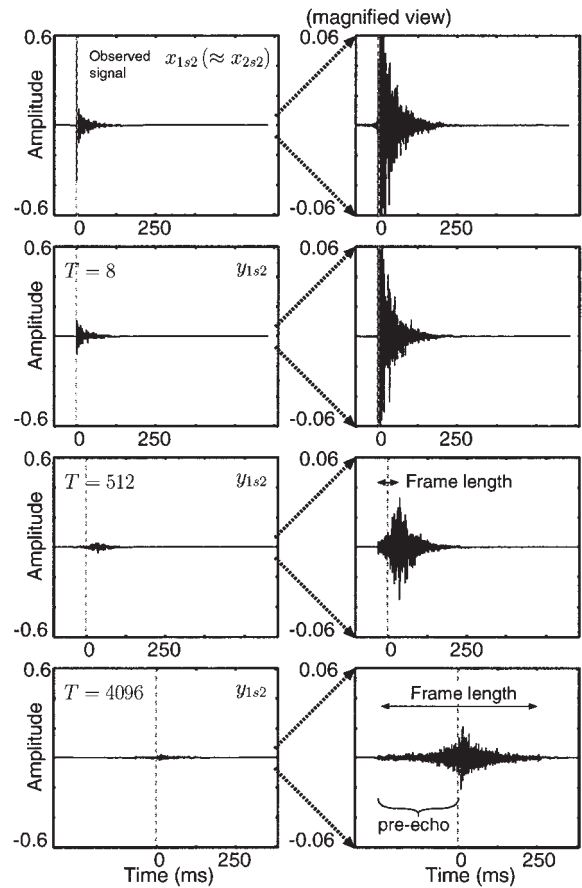
5.3. Eliminating Pre-echo Noise

The pre-echo noise is mainly caused by the estimation errors in the non-causal part of the unmixing filters. Therefore, if we allow a delay in proportion to the frame length (half of the frame length in the above experiments), the noise increases as the frame length become longer. The pre-echo noise can be removed by eliminating the non-causal part of the separating filters. However, this prevents the separating system from removing the direct sound of the jammer signal (Fig. 10), and degrades the separation performance.

One solution is to limit the non-causal part of the separating filters to an appropriate length instead of
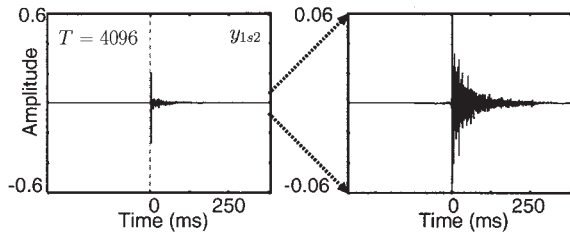
Fig. 10 Jammer impulse response of BSS system when non-causal part of the filter is completely eliminated.
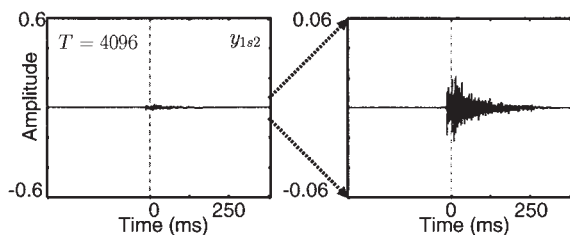
Fig. 11 Jammer impulse response of BSS system when non-causal part of the filter is limited to 128 taps.

eliminating it completely. The lower bound of the filter length, which can remove direct sound, is much shorter than the filter length yielding the best separation performance. Based on P_{JE} in Fig. 6, the proper length of the non-causal part is around 128 taps (corresponding to 16 ms) under our experimental conditions. Figure 11 shows the jammer impulse response when the non-causal part of the filter is limited to 128 taps. The pre-echo noise and the direct sound are both removed. Although we just cut off the separating filters with a rectangular window in this experiment, we did not observed any deterioration on the auditory quality of the separated signals in our preliminary listening test. From a psychoacoustic viewpoint, the pre-echo of 16 ms is relatively allowable, because a human hearing system has premasking effect of about 20 ms [23].

5.4. Applicability of Evaluation Method

We have applied our evaluation method to a BSS algorithm with a different non-linear function Φ and different scaling method in [11], and have confirmed a similar result to that reported in this paper. We have also carried out experiments with changing the locations of source signals, and have obtained similar results.

BSS with target dereverberation is still an open problem, and many algorithms will be designed in the future. Some of them may realize a deconvolution of the target signal, and then their characteristics about the target dereverberation will be different from our results. However, our evaluation method can be used for investigating the characteristics of their performance.

Our method is not limited to frequency domain BSS,

but it can be applied to other kinds of BSS algorithms, including time domain BSS. It is helpful in investigating the acoustical characteristics of BSS in a reverberant environment.

6. CONCLUSION

We investigated the performance of a separating system obtained by frequency domain BSS based on ICA using the impulse responses of target and jammer signals. As a result, we found the acoustical characteristics of the separating system. The separating system eliminates the direct sound of the jammer signal even when the frame length is relatively short, and it also reduces the reverberation according to the frame length, while the reverberation of the target is not reduced. The performance of the target dereverberation does not depend on the frame length and is as poor as that of an NBF. Our results provide experimental confirmation of the equivalence between BSS and ABF.

The jammer reduction performance improves as the frame length increases. However, an overly long frame length has a detrimental effect on the performance due to accumulating errors, and the separated signals are damaged by pre-echo noise. To reduce the pre-echo noise, it is effective to shorten the non-causal part of the separating filters.

Our analysis method is useful for investigating the acoustical characteristics of BSS in a reverberant environment.

REFERENCES

[1] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, **7**, 1129–1159 (1995).

[2] S. Haykin, Ed., *Unsupervised Adaptive Filtering* (John Wiley & Sons, 2000).

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

[4] T. W. Lee, A. J. Bell and R. Orglmeister, "Blind source separation of real world signals," *Neural Networks*, **4**, 2129–2134 (1997).

[5] J. Xi and J. P. Reilly, "Blind separation and restoration of signals mixed in convolutive environment," *Proc. ICASSP '97*, pp. 1327–1330 (1997).

[6] M. Kawamoto, A. Barros, A. Mansour, K. Matsuoka and N. Ohnishi, "Real world blind separation of convolved non-stationary signals," *Proc. Int. Conf. Independent Component Analysis and Blind Source Separation (ICA '99)*, pp. 347–352 (1999).

[7] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, **22**, 21–34 (1998).

[8] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, **8**,

320–327 (2000).

[9] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," *Proc. ICASSP '00*, pp. 1041–1044 (2000).

[10] S. Araki, S. Makino, T. Nishikawa and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," *Proc. ICASSP '01*, pp. 2737–2740 (2001).

[11] R. Mukai, S. Araki and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment," *Proc. Eurospeech '01*, pp. 2599–2602 (2001).

[12] F. Asano, S. Ikeda, M. Ogawa, H. Asoh and N. Kitawaki, "Blind source separation in reflective sound fields," *Proc. Int. Workshop Hands-Free Speech Communication* (*HSC '01*), pp. 51–54 (2001).

[13] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP '00*, pp. 3140–3143 (2000).

[14] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *Proc. NNSP '01*, pp.273–282 (2001).

[15] K. Torkkola, "Blind separation of convolved sources based on information maximization," *Proc. NNSP '96*, pp.423–432 (1996).

[16] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation* (*ICA '01*), pp. 722–727 (2001).

[17] S. Araki, S. Makino, R. Mukai and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," *Proc. Eurospeech '01*, pp. 2595–2598 (2001).

[18] M. Knaak and D. Filbert, "Acoustical semi-blind source separation for machine monitoring," *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation* (*ICA '01*), pp. 361–366 (2001).

[19] S. Amari, A. Cichocki and H.H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems 8* (The MIT Press, 1996), pp.757–763.

[20] H. Sawada, R. Mukai, S. Araki and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," *Proc. ICASSP '02*, pp.1001–1004 (2002).

[21] F. Asano and S. Ikeda, "Evaluation and real-time implementation of blind source separation system using time-delayed decorrelation," *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation* (*ICA '00*), pp. 411–415 (2000).

[22] D. Schobben, K. Torkkola and P. Smaragdis, "Evaluation of blind signal separation methods," *Proc. Int. Conf. Independent Component Analysis and Blind Source Separation* (*ICA '99*), pp. 261–266 (1999).

[23] E. Zwicker and H. Fastl, *Psychoacoustics Facts and Models* (Springer-Verlag, 1990).

Ryo Mukai received the B.S. and the M.S. degrees in information science from the University of Tokyo, Japan, in 1990 and 1992, respectively. He joined NTT in 1992. From 1992 to 2000, he was engaged in research and development of processor architecture for network service systems and distributed network systems. Since 2000, he has been with NTT Communication Science Laboratories, where he is engaged in research of blind source separation. His current research interests include digital signal processing and its applications. He is a member of the IEEE, ACM, the ASJ, IEICE, and the IPSJ.

Shoko Araki received the B.E. and the M.E. degrees in mathematical engineering and information physics from the University of Tokyo, Japan, in 1998 and 2000, respectively. Her research interests include array signal processing, blind source separation applied to speech signals, and auditory scene analysis. She is a member of the IEEE and the ASJ.

Hiroshi Sawada received the B.E., M.E. and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993 and 2001, respectively. In 1993, he joined NTT Communication Science Laboratories. From 1993 to 2000, he was engaged in research on the computer aided design of digital systems, logic synthesis, and computer architecture. Since 2000, he has been engaged in research on signal processing and blind source separation for convolutive mixtures using independent component analysis. He is a member of the IEEE and the IEICE.

Shoji Makino received the B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1979, 1981, and 1993, respectively. He joined NTT in 1981. He is now an Executive Manager at the NTT Communication Science Laboratories. His research interests include blind source separation of convolutive mixtures of speech, acoustic signal processing, and adaptive filtering and its applications. He received the Paper Award of the IEICE in 2002, the Paper Award of the ASJ in 2002, the Achievement Award of the IEICE in 1997, and the Outstanding Technological Development Award of the ASJ in 1995. He is the author or co-author of more than 170 articles in journals and conference proceedings and has been responsible for more than 140 patents. He is a member of the Conference Board of the IEEE SP Society and an Associate Editor of the IEEE Transactions on Speech and Audio Processing. He is a member of the Technical Committee on Audio and Electroacoustics as well as Speech of the IEEE SP Society. Dr. Makino is a Senior Member of the IEEE, a member of the ASJ, and the IEICE.