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

MULTISTAGE ICA FOR BLIND SOURCE SEPARATION OF REAL ACOUSTIC CONVOLUTIVE MIXTURE

Tsuyoki NISHIKAWA [†], *Hiroshi SARUWATARI* [†], *Kiyohiro SHIKANO* [†], *Shoko ARAKI* [‡], *and Shoji MAKINO* [‡]

[†] Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192, JAPAN

[‡] NTT Communication Science Laboratoriess, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 JAPAN

E-Mail: {tsuyo-ni, sawatari, shikano}@is.aist-nara.ac.jp, {shoko, maki}@cslab.kecl.ntt.co.jp

ABSTRACT

We propose a new algorithm for blind source separation (BSS), in which frequency-domain independent component analysis (FDICA) and time-domain ICA (TDICA) are combined to achieve a superior source-separation performance under reverberant conditions. Generally speaking, conventional TDICA fails to separate source signals under heavily reverberant conditions because of the low convergence in the iterative learning of the separation system. On the other hand, the separation performance of conventional FDICA also degrades seriously because the independence assumption of narrow-bin signals collapses when the number of frequency bins increases. In the proposed method, the separated signals of FDICA are regarded as the input signals for TDICA, and we can remove the residual crosstalk components of FDICA by using TDICA. The experimental results obtained under the reverberant condition reveal that the separation performance of the proposed method is superior to those of TDICA- and FDICA-based BSS methods.

1. INTRODUCTION

Blind source separation (BSS) is the approach taken to estimate the original source signals using only the information of the mixed signals observed in each input channel. This technique is applicable to the realization of high-quality hands-free speech recognition system. The BSS methods based on independent component analysis (ICA) [1, 2] can be classified into two groups in terms of the processing domain, i.e., frequency-domain ICA (FDICA) in which the complex-valued separation matrix is calculated in the frequency domain [3, 4], and time-domain ICA (TDICA) in which the separation FIR filter matrix is calculated in the time domain [5, 6]. The recently developed BSS techniques can achieve a good source-separation performance under artificial or short reverberant conditions. However, the performances of these methods under heavily reverberant conditions degrade seriously because of the following problems. (1) In conventional FDICA, the separation performance is saturated before reaching a sufficient performance level because we transform the fullband signals into the narrow-band signals and the independence assumption collapses in each narrow-band [7]. (2) In conventional TDICA, the convergence degrades because the iterative learning rule becomes more complicated as the reverberation increases [6].

In order to resolve the problems, we propose a new BSS algorithm called multistage ICA (MSICA), in which FDICA and TDICA are combined. By using the proposed method, we can achieve a superior separation performance even under heavily reverberant conditions. The results of the signal separation experiments reveal that the separation performance of the proposed algorithm is superior to those of the conventional ICA-based BSS methods.

2. SOUND MIXING MODEL

In general, the observed signals in which multiple source signals are convoluted with room impulse responses are obtained by the following equation:

$$\boldsymbol{x}(t) = \sum_{\tau=0}^{P-1} \boldsymbol{a}(\tau)\boldsymbol{s}(t-\tau), \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t), \cdots, x_K(t)]^{\mathrm{T}}$ is the observed signal vector and $\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}}$ is the source signal vector (see Fig. 1) . K is the number of array elements (microphones) and L is the number of multiple sound sources. In this study, we deal with the case of $K = L = 2$. Also, $\boldsymbol{a}(\tau) = [a_{ij}(\tau)]_{ij}$ ($[\cdot]_{ij}$ denotes the matrix in which ij-th element is $[\cdot]$) is the mixing filter matrix. P is the length of the impulse response which is assumed to be an FIR filter of thousands of taps because we introduce a model to deal with the arrival lags among the elements of the microphone array and the room reverberations.

3. CONVENTIONAL ICA AND PROBLEMS

3.1. FDICA

The conventional BSS based on FDICA is conducted with the following steps: (1) transform the observed fullband signals into the narrow-band signals, (2) optimize the separation matrix in each frequency bin, and (3) reconstruct the fullband separated signal from the narrow-band separated signals. FDICA has the following advantages and disadvantages.

Advantages:

(F1) It is easy to converge the separation filter in iterative ICA learning because we can simplify the convolutive mixture down to simultaneous mixtures by the frequency transform.

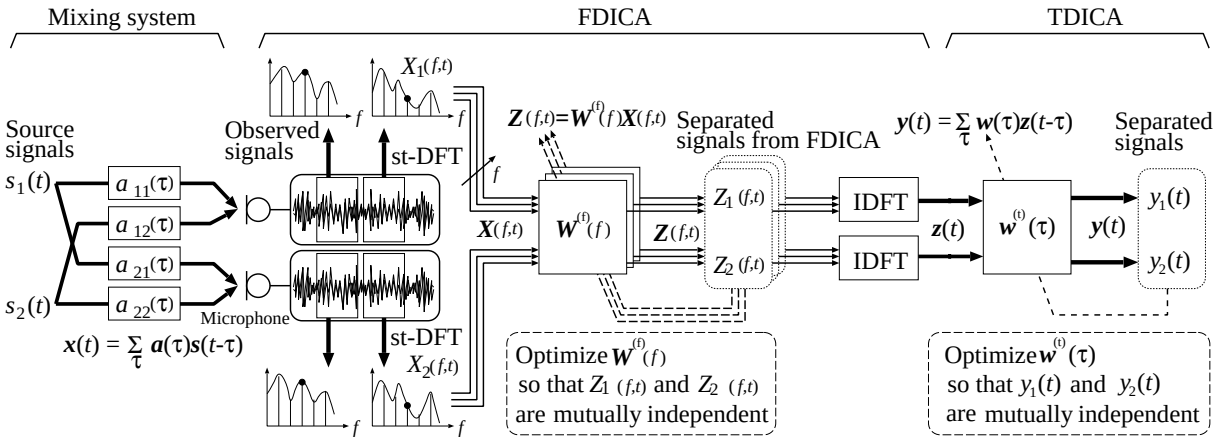

Figure 1: Blind source separation procedure performed in multistage ICA.

Disadvantages:

(F2) The separation performance is saturated before reaching a sufficient performance level because the independence assumption collapses in each narrow-band [7] (see, e.g., Sect. 5.3).

(F3) Permutation among source signals and indeterminacy of each source gain in each frequency bin.

Regarding the disadvantage (F3), various solutions have already been proposed [3, 8, 9, 10]. However, the collapse of the independence assumption, (F2), is a serious and inherent problem, and this prevents us from applying FDICA in a real acoustic environment with a long reverberation.

3.2. TDICA

In the conventional BSS based on TDICA, each element of the separation filter matrix is represented as a FIR filter. We can optimize this separation filter system, by using the fullband observed signals themselves. TDICA has the following advantages and disadvantages.

Advantages:

(T1) We can treat the fullband speech signals where the independence assumption of sources usually holds.

Disadvantages:

(T2) The convergence degrades under reverberant conditions because the iterative rule for FIR filter learning is complicated.

It is known that TDICA works only in the case of mixtures with a short-tap FIR filter, i.e., less than 100 taps. Also, TDICA fails to separate source signals under real acoustic environments because of the disadvantages (T2).

4. PROPOSED METHOD: MSICA

As described above, the conventional ICA methods have some disadvantages. However, note that the advantages and disadvantages of FDICA and TDICA are mutually complementary, i.e., (F2) can be resolved by (T1), and (T2) can be resolved by (F1). Hence, in order to resolve the disadvantages, we propose a new algorithm, MSICA, in which FDICA and TDICA are combined (see Fig. 1).

MSICA is conducted with the following steps. In the first stage, we perform FDICA to separate the source signals to some extent with the high-stability advantages (F1) of FDICA. In the second stage, we regard the separated signals of FDICA as the input signals for TDICA, and we remove the residual crosstalk components of FDICA by using TDICA. Finally, we regard the output signals of TDICA as the resultant separated signals. MSICA can achieve a high stability and a separation performance superior to that of conventional FDICA and TDICA. In the following sections, we describe details of the ICA-learning rules for each stage.

4.1. First-Stage ICA: FDICA

In the first-stage ICA, we introduce the fast-convergence FDICA proposed by one of the authors [4]. We perform the signal separation procedure as described below (see FDICA in Fig. 1).

In FDICA, first, the short-time analysis of observed signals $\boldsymbol{x}(t)$ is conducted by frame-by-frame discrete Fourier transform (DFT). By plotting the spectral values in a frequency bin of each microphone input frame by frame, we consider them as a time series. Hereafter, we designate the time series as $\boldsymbol{X}(f,t) = [X_1(f,t), \cdots, X_K(f,t)]^{\mathrm{T}}$. Next, we perform signal separation using the complex-valued inverse of the mixing matrix, $\boldsymbol{W}^{(\mathrm{f})}(f)$, so that the L time-series output $\boldsymbol{Z}(f,t) = [Z_1^{(\mathrm{f})}(f,t), \cdots, Z_L^{(\mathrm{f})}(f,t)]^{\mathrm{T}}$ becomes mutually independent; this procedure can be given as

$$\boldsymbol{Z}(f,t) \;=\; \boldsymbol{W}^{(\mathrm{f})}(f)\boldsymbol{X}(f,t). \tag{2}$$

We perform this procedure with respect to all frequency bins. Finally, by applying the inverse DFT and the overlap-add technique to the separated time series $\boldsymbol{Z}(f,t)$, we reconstruct the resultant source signals in the time domain, $\boldsymbol{z}(t)$.

In conventional FDICA , the optimal $\boldsymbol{W}^{(\mathrm{f})}(f)$ is obtained by the following iterative equation [3]:

$$\begin{aligned}\boldsymbol{W}_{i+1}^{(\mathrm{f})}(f) \;=\; & \boldsymbol{W}_i^{(\mathrm{f})}(f) + \alpha \Big[\mathrm{diag}\Big(\big\langle \boldsymbol{\Phi}(\boldsymbol{Z}(f,t))\boldsymbol{Z}^{\mathrm{H}}(f,t)\big\rangle_t\Big) \\ & - \big\langle \boldsymbol{\Phi}(\boldsymbol{Z}(f,t))\boldsymbol{Z}^{\mathrm{H}}(f,t)\big\rangle_t\Big]\boldsymbol{W}_i^{(\mathrm{f})}(f),\end{aligned} \tag{3}$$

where $\langle\cdot\rangle_t$ denotes the time-averaging operator, i is used to express the value of the i-th step in the iterations, and α is the step-size parameter. Also, we define the nonlinear vector function $\boldsymbol{\Phi}(\cdot)$ as

$$\boldsymbol{\Phi}(\boldsymbol{Z}(f,t)) \equiv [\Phi(Z_1(f,t)),\cdots,\Phi(Z_L(f,t))]^{\mathrm{T}}, \quad (4)$$

$$\Phi(Z_l(f,t)) \equiv \tanh(Z_l^{(\mathrm{R})}(f,t))+j\cdot\tanh(Z_l^{(\mathrm{I})}(f,t)), \quad (5)$$

where $\mathrm{Re}[Z_l(f,t)]$ and $\mathrm{Im}[Z_l(f,t)]$ are the real and imaginary parts of $Z_l(f,t)$, respectively.

4.2. Second-Stage ICA: TDICA

The output signals from TDICA (i.e., the separated signals of MSICA) can be given as

$$\boldsymbol{y}(t)=\sum_{\tau=0}^{Q-1}\boldsymbol{w}^{(\mathrm{t})}(\tau)\boldsymbol{z}(t-\tau), \quad (6)$$

where $\boldsymbol{y}(t)=[y_1(t),\cdots,y_L(t)]^{\mathrm{T}}$ is the resultant separated signal vector of MSICA, $\boldsymbol{w}^{(\mathrm{t})}(\tau)=[w_{ij}^{(\mathrm{t})}(\tau)]_{ij}$ is the separation filter matrix, and Q is the length of the separation filter. The selection of TDICA is an important issue because the quality of resultant separated signals is determined by TDICA. In this study, we introduce three TDICA algorithms, i.e.,

TDICA1: TDICA based on simultaneous decorrelation of nonstationary signals,

TDICA2: TDICA based on combination of TDICA1 and time-delayed decorrelation approach, and

TDICA3 [12]: TDICA based on minimization of Kullbuck-Leibler divergence (KLD),

and compare these TDICA algorithms.

First, we drive the TDICA1 algorithm. This optimizes the separation filter by minimizing the nonnegative cost function which takes the minimum value only when the second-order cross-correlation becomes zero if the source signals are nonstationary. The cost function is defined as follows (this cost function has been already proposed by Kawamoto et al. [11], however their derivation of learning rule includes *mathematical error*):

$$Q\big(\boldsymbol{w}^{(\mathrm{t})}(\tau)\big)=\frac{1}{2B}\sum_{b=1}^{B}\left\{\log\left(\frac{\det\mathrm{diag}\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\big\rangle_t^{(b)}}{\det\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\big\rangle_t^{(b)}}\right)\right\}, \quad (7)$$

where B is the number of local analysis blocks and $\langle\cdot\rangle_t^{(b)}$ denotes the time-averaging operator for the b-th local analysis block. Equation (7) becomes zero only when $y_i(t)$ and $y_j(t)$ are uncorrelated for all of the local analysis blocks. The *correct* iterative equation to minimize $Q\big(\boldsymbol{w}^{(\mathrm{t})}(\tau)\big)$ is given by [6]

[TDICA1]

$$\begin{aligned}\boldsymbol{w}_{i+1}^{(\mathrm{t1})}(\tau) = \boldsymbol{w}_i^{(\mathrm{t1})}(\tau)+\frac{\beta}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{&\Big(\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\big\rangle_t^{(b)}\Big)^{-1}\\ &\cdot\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\big\rangle_t^{(b)}-\Big(\mathrm{diag}\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\big\rangle_t^{(b)}\Big)^{-1}\\ &\cdot\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\big\rangle_t^{(b)}\Big\}\boldsymbol{w}_i^{(\mathrm{t1})}(d), \quad (8)\end{aligned}$$

where β is the step-size parameter. Since the Eq. (8) evaluates only off-diagonal of $\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\rangle_t^{(b)}$, we confirmed that the iterative equation of Eq. (8) could not achieve a superior separation performance under the reverberant condition (see Sect. 5.4). Namely, the source separation is not achieved by only using nonstationarity of signals. Therefore we use not only nonstationarity of signals but also time-delayed decorrelation approach. We expand Eq. (8) to the following equation to evaluate the off-diagonal of $\langle\boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t^{(b)}$ for all time delays $\tau-d$:

[TDICA2]

$$\begin{aligned}&\boldsymbol{w}_{i+1}^{(\mathrm{t2})}(\tau)\\ &=\boldsymbol{w}_i^{(\mathrm{t2})}(\tau)+\frac{\beta}{B}\sum_{b=1}^{B}\sum_{d=0}^{Q-1}\Big\{\Big(\mathrm{diag}\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\big\rangle_t^{(b)}\Big)^{-1}\\ &\quad\cdot\mathrm{diag}\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\big\rangle_t^{(b)}-\Big(\mathrm{diag}\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t)^{\mathrm{T}}\big\rangle_t^{(b)}\Big)^{-1}\\ &\quad\cdot\big\langle\boldsymbol{y}(t)\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\big\rangle_t^{(b)}\Big\}\boldsymbol{w}_i^{(\mathrm{t2})}(d), \quad (9)\end{aligned}$$

Choi proposed the TDICA algorithm which optimizes the separation filter by minimizing the KLD between the joint probability density function and the marginal probability density function of the separated signals [12]. The KLD is given by

$$KLD\big(\boldsymbol{w}^{(\mathrm{t})}(\tau)\big)=\int p(\boldsymbol{y}(t))\log\frac{p(\boldsymbol{y}(t))}{\prod_{l=1}^{L}\prod_{t=0}^{T-1}q(y_l(t))}d\boldsymbol{y}(t), \quad (10)$$

where $p(\cdot)$ is the joint probability density function, $q(\cdot)$ is the marginal probability density function, and T is the length of the separated signals. With a nonholonomic constraint, the iterative equation of the separation filter to minimize the $KLD(\boldsymbol{w}^{(\mathrm{t})}(\tau))$ is given as (hereafter we designate the iterative equation as "TDICA3"):

[TDICA3]

$$\begin{aligned}\boldsymbol{w}_{i+1}^{(\mathrm{t3})}(\tau) &= \boldsymbol{w}_i^{(\mathrm{t3})}(\tau)\\ &\quad+\alpha\sum_{d=0}^{Q-1}\Big\{\mathrm{diag}\Big(\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t\Big)\\ &\quad-\langle\boldsymbol{\phi}(\boldsymbol{y}(t))\boldsymbol{y}(t-\tau+d)^{\mathrm{T}}\rangle_t\Big\}\boldsymbol{w}_i^{(\mathrm{t3})}(d). \quad (11)\end{aligned}$$

where we define the nonlinear vector function $\boldsymbol{\phi}(\cdot)$ as

$$\boldsymbol{\phi}(\boldsymbol{y}(t))\equiv\big[\tanh(y_1(t)),\cdots,\tanh(y_L(t))\big]^{\mathrm{T}}. \quad (12)$$

In this study, we compare the **MSICA1**, **MSICA2**, and **MSICA3** in which FDICA followed by **TDICA1**, **TDICA2**, and **TDICA2**, respectively.

5. EXPERIMENTS AND RESULTS

5.1. Experimental Setup

A two-element array with the interelement spacing of 4 cm is assumed. We determined this interelement spacing by considering that the spacing should be smaller than half of

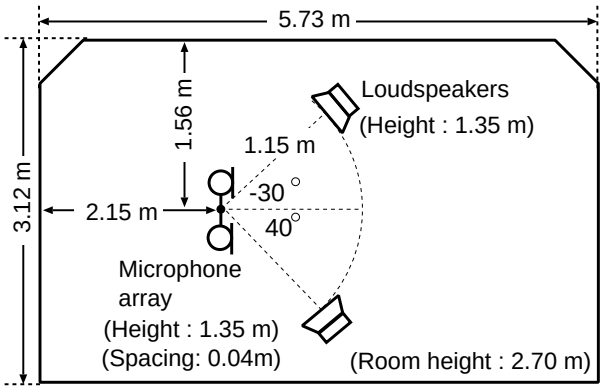

Figure 2: Layout of reverberant room used in experiments.

the minimum wavelength to avoid the spatial aliasing effect; it corresponds to 8.5/2 cm in 8 kHz sampling. The speech signals are assumed to arrive from two directions, -30° and 40°. The distance between the microphone array and the loudspeakers is 1.15 m (see Fig. 2). Two kinds of sentences, those spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research, are used as the original speech samples. The sampling frequency is 8 kHz and the length of speech is limited to within 3 seconds. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. In these experiments, we use the following signals as the source signals: the original speech convolved with the impulse responses specified by the reverberation time of 300 ms. This corresponds to 2400-taps FIR filter in 8 kHz. The impulse responses are recorded in a variable reverberation time room as shown in Fig. 2.

5.2. Objective Evaluation Score

Noise reduction rate (NRR), defined as the output signal-to-noise ratio (SNR) in dB minus input SNR in dB, is used as the objective evaluation score in this experiment. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise. The NRR is defined as

$$\mathrm{NRR} \equiv \frac{1}{2}\sum_{l=1}^{2}\Big(\mathrm{SNR}_l^{(\mathrm{O})} - \mathrm{SNR}_l^{(\mathrm{I})}\Big), \tag{13}$$

$$\mathrm{SNR}_l^{(\mathrm{O})} = 10\log_{10}\frac{\sum_f |H_{ll}(f)S_l(f)|^2}{\sum_f |H_{ln}(f)S_n(f)|^2}, \tag{14}$$

$$\mathrm{SNR}_l^{(\mathrm{I})} = 10\log_{10}\frac{\sum_f |A_{ll}(f)S_l(f)|^2}{\sum_f |A_{ln}(f)S_n(f)|^2}, \tag{15}$$

where $\mathrm{SNR}_l^{(\mathrm{O})}$ and $\mathrm{SNR}_l^{(\mathrm{I})}$ are the output SNR and the input SNR, respectively, and $l \neq n$. Also, $S_l(f)$ is the frequency-domain representation of the source signal, $s_l(t)$, $H_{ij}(f)$ is the element in the i th row and the j th column of the matrix $\boldsymbol{H}(f) = \boldsymbol{W}^{(\mathrm{MSICA})}(f)\boldsymbol{A}(f)$ where $\boldsymbol{W}^{(\mathrm{MSICA})}(f)$ denotes the entire separation process in MSICA including both FDICA and TDICA and $\boldsymbol{A}(f)$ is the mixing matrix which corresponds to the frequency-domain representation of the room impulse responses described in Sect. 2.

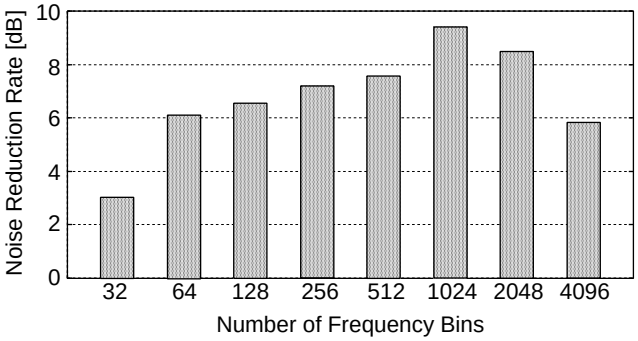

Figure 3: Relation between separation performances and the number of frequency bins in conventional FDICA.

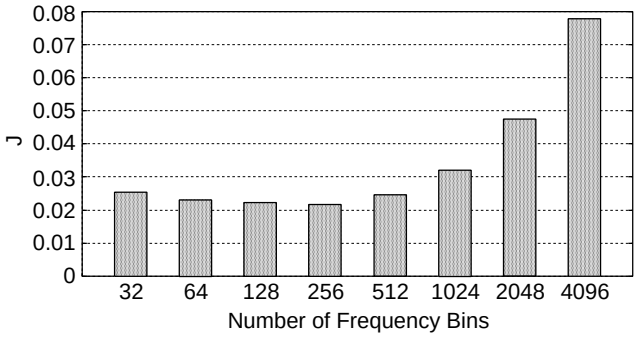

Figure 4: Relation between the number of frequency bins and the value of J defined by Eq. (16), which corresponds to the independence of subband signals.

5.3. Relation between Separation Performance and Number of Frequency Bins in FDICA

In order to confirm the independence problem of narrowband signals in FDICA ((F2) described in Sect. 3.1), we carried out the preliminary experiment under the following analysis conditions. The number of frequency bins (frame length in DFT) is set to be from 32 to 4096, the frame shift is 16 taps, the window function is a Hamming window, the number of iterations in ICA is 30, and the step-size parameter α for iterations is set to be 1.0×10^{-5}.

Figure 3 shows the NRR results for different numbers of frequency bins in FDICA. As shown in Fig. 3, the NRR of FDICA obviously degrades when the number of frequency bins becomes too large, and the separation performance is saturated before reaching a sufficient performance. This is because we transform the fullband signals into the narrowband signals and the independence assumption collapses in each frequency bin, particularly when the number of frequency bins is large.

In order to confirm the fact, we newly define the following objective measure to quantify an independence, and investigate the relation between the number of frequency bins and the independence among subband signals.

$$J = \Big\langle \Big\| \mathrm{diag}\Big(\big\langle \boldsymbol{\Phi}(\boldsymbol{Z}(f,t))\boldsymbol{Z}^{\mathrm{H}}(f,t)\big\rangle_t\Big) - \big\langle \boldsymbol{\Phi}(\boldsymbol{Z}(f,t))\boldsymbol{Z}^{\mathrm{H}}(f,t)\big\rangle_t \Big\| \Big\rangle_f, \tag{16}$$

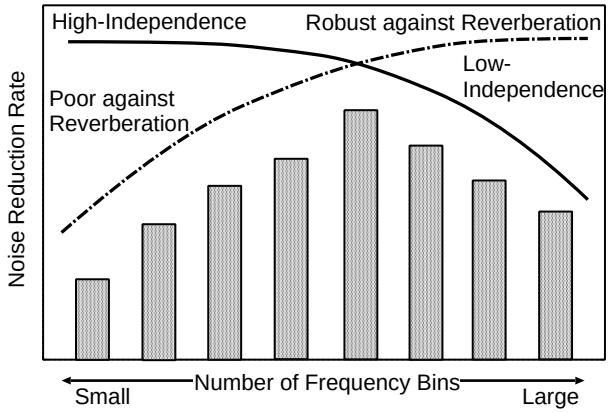

Figure 5: Trade-off relation among the independence of subband signals and robustness against reverberation.

where $\|\cdot\|$ is frobenius norm of matrix. This measure J is a part of the iterative equation (3) and has no dimension. Therefore the absolute value of J is meaningless itself, however, the relative value between the different numbers of frequency bins is important. If narrow-band signals become mutually independent, the measure J becomes zero. Also we can consider that the independence of subband signals is high when J is small. In order to evaluate the independence of real narrow-band speech signals, we carried out the experiment in which the input signal, $\boldsymbol{Y}(f,t)$, in Eq. (16) is regarded as the perfectly separated sources, i.e., original speech samples. Figure 4 shows the relation between the number of frequency bins and the value of J which corresponds to the independence of subband signals. Figure 4 shows that the independence decreases as the number of frequency bins increases, especially when the number of frequency bins is large.

Above-mentioned experimental results clarify the disadvantage that the separation performance is saturated in FDICA because we transform the fullband signals into the narrow-band signals. We should lengthen the separation filter (or FFT length for analysis) when we confront with a long reverberation. In this case, however, the independence of subband signals decreases. Thus, there is a trade-off relation among the independence of subband signals and robustness against reverberation as shown in Figure 5. On the basis of these results, we should cascade another signal processing analysis, e.g., TDICA, with FDICA to obtain the further separation performances.

5.4. Comparison of Separation Performance between MSICAs

We carried out the experiments using MSICA1, MSICA2, and MSICA3 to evaluate the contribution of TDICA1, TDICA2, and TDICA3 for improving the separation performances under reverberant conditions. The analysis conditions of these experiments are as follows: the filter length Q is set to be 2048taps, the maximum number of iterations is 500, and the step-size parameter β for iterations is set to be 5.0×10^{-4} for TDICA1, 1.0×10^{-3} for TDICA2, and 1.0×10^{-6} for TDICA3. As for the local analysis block for

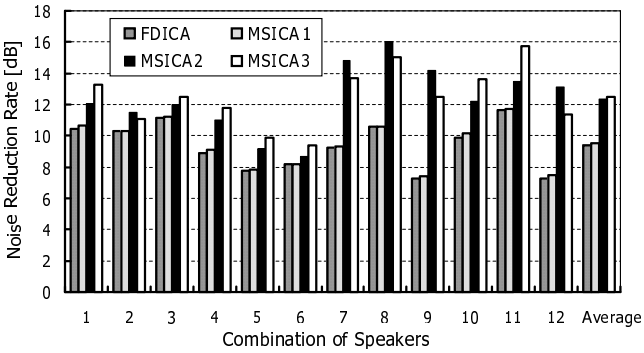

Figure 6: Comparison of separation performance between MSICA1, MSICA2, and MSICA3.

TDICA1 and TDICA2, we divided the signals equally into B parts (B = 1–10). We chose the optimal B and number of iterations for each combination of speaker because the convergence is different for every combination. As for the FDICA part in MSICA, the analysis conditions are the same as those given in Sect. 5.3, except for the number of frequency bins (which is fixed at 1024 bins).

Figure 6 shows the NRR results in the MSICA1, MSICA2 and MSICA3. Figure 6 shows that TDICA1 can not achieve a signal separation under the reverberant condition. Comparing MSICA1 with MSICA2 in Fig. 6, we confirm that TDICA2 can achieve a superior separation performance to TDICA1. These results show that it is necessary to evaluate correlations of different times to achieve a superior performance. TDICA3 can also achieve a superior separation performance to TDICA1. The separation performances of MSICA2 and MSICA3 are superior to that of FDICA and the improvements of the separation performance from FDICA are not so different. Thus, the use of FDICA with TDICA2 or TDICA3 is effective in the proposed MSICA structure.

5.5. Comparison between MSICA2 and MSICA3

In order to compare MSICA2 and MSICA3 in detail, we evaluate the separation performances of MSICA3 and MSICA2 with the various number of local blocks B. In TDICA2 part in MSICA2, the utilization of large B is effective because it can evaluate the nonstationarity of the signals. Figure 7 shows the NRR results in the MSICA3 and MSICA2 with the various B. The separation performance is improved as the B is increased, however, the larger B is not the best for improving the separation performance. In addition, the separation performances of MSICA3 and MSICA2 with the optimal B are not so different. Therefore, MSICA3 is more feasible than MSICA2 because MSICA3 does not require the estimation of the optimal B.

5.6. Discussion on Combination Order in MSICA

As described in the previous section, the combination of FDICA and TDICA can contribute to the improvement of separation. In this combination, the advantage (F1) of FDICA is useful in the initial step of separation procedure

and the advantage (T1) of TDICA is also useful in the later step. Therefore we use FDICA as the first-stage ICA and TDICA as the second-stage ICA. In order to confirm the availability of this combination order, we compare the proposed combination with the combination in which TDICA is used in the first stage and FDICA is used in the second stage (hereafter we designate this swapped combination as "MSICA4").

The experiment of MSICA4 was carried out in the following manner. As for TDICA part in MSICA4, TDICA2 is used, the number of local analysis blocks, B, is fixed at 3, and the filter length is 8 taps improved best the separation performance in TDICA2. This TDICA2 can obtain the NRR of 6.0 dB. As for FDICA part in MSICA4, the analysis conditions are the same as those given in Sect. 5.3, except for the number of frequency bins (which is fixed at 1024 bins). As the result, the NRR of 7.5 dB is obtained in MSICA4, and this performance is better than that of simple TDICA but is poorer than those of MSICA2, MSICA3, and simple FDICA. In MSICA4, the separation performance is still improved by using FDICA in the second stage, however, the separation performance is saturated because of the disadvantage (F2) of FDICA. MSICA4 can not achieve the separation performance of 9.4 dB which corresponds to NRR of simple FDICA. This reason is that FDICA in this paper uses the beamforming technique and the directivity pattern of the array which provide a good initial value of the separation matrix to improve the convergence [4], however, such kind of information is no longer valid in the combination order of MSICA4 because we can not know the effective positions of the array elements after the first-stage TDICA and can not depict the directivity pattern. Thus the separation performance of MSICA4 is almost equal to that of a raw FDICA without the beamforming technique (from [4] we can see the NRR of about 7.5 dB at the 30-iteration point). This fact indicates that the swapped combination order of MSICA4 has no contribution to the improvement of the separation performance, and the proposed combination order of MSICA2 and MSICA3 (FDICA in the first stage and TDICA in the second stage) is essential.

6. CONCLUSION

In this paper, we propose a new algorithm for BSS, in which FDICA and TDICA are combined to achieve a superior source-separation performance under reverberant conditions. In TDICA part in MSICA, we compare (1) TDICA based on simultaneous decorrelation of nonstationary signals, (2) TDICA based on combination of (1) and time-delayed decorrelation approach, and (3) TDICA based on minimization of KLD. The results of the signal separation experiments reveal that the separation performances of (1) and conventional FDICA are not so different and the separation performances of (2) and (3) are superior to that of conventional FDICA. Therefore, the combination of FDICA and (2) or (3) is inherently effective for improving the separation performance. Specifically, the proposed method can improve the SNR by about 3.0 dB over that of FDICA for an average of 12 speaker-combinations.

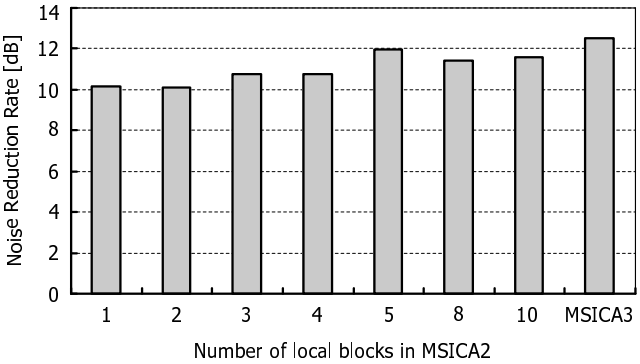

Figure 7: Comparison of separation performance between MSICA3 and MSICA2 with the various B.

7. ACKNOWLEDGEMENT

The authors are grateful to Dr. Mitsuru Kawamoto of Shimane University and Mr. Robert Aichner of NTT. CO., LTD for their greateful discussions. This work was partly supported by CREST (Core Research for Evolutional Science and Technology) of JST (Japan Science and Technology Corporation).

8. REFERENCES

[1] P. Common, "Independent component analysis, a new concept?," *Signal Processing*, vol.36, pp.287–314, 1994.

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

[3] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. of 1998 International Symposium on Nonlinear Theory and Its Application (NOLTA98)*, pp.923–926, Sept. 1998.

[4] H. Saruwatari, T. Kawamura, and K. Shikano, "Blind source separation for speech based on fast-convergence algorithm with ICA and beamforming," *Proc. Eurospeech2001*, pp. 2603–2606, Sept. 2001.

[5] T. W. Lee, *Independent component analysis*, Kluwer academic publishers, 1998.

[6] T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA," *Proc. EUSIPCO2002*, vol.II, pp.15–18, Sept. 2002.

[7] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive of speech," *Proc. ICASSP2001*, pp.2737–2740, May 2001.

[8] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP2000*, pp.3140–3143, June 2000.

[9] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, vol.8, no.3, pp.320–327, May 2000.

[10] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," *Proc. ICASSP2001*, pp.2729–2732, May 2001.

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

[12] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," *Proc. ICA99*, pp.371–376, January 1999.