

STABLE LEARNING ALGORITHM FOR BLIND SEPARATION OF TEMPORALLY CORRELATED SIGNALS COMBINING MULTISTAGE ICA AND LINEAR PREDICTION

Tsuyoki NISHIKAWA Hiroshi SARUWATARI Kiyohiro SHIKANO

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192, JAPAN
E-mail: {tsuyo-ni, sawatari, shikano}@is.aist-nara.ac.jp

ABSTRACT

We newly propose a stable algorithm for blind source separation (BSS) combining multistage ICA (MSICA) and linear prediction. The MSICA is the method previously proposed by the authors, in which frequency-domain ICA (FDICA) for a rough separation is followed by time-domain ICA (TDICA) to remove residual crosstalk. For temporally correlated signals, we must use TDICA with a nonholonomic constraint to avoid the decorrelation effect from the holonomic constraint. However, the stability cannot be guaranteed in the nonholonomic case. To solve the problem, the linear predictors estimated from the roughly separated signals by FDICA are inserted before the holonomic TDICA as a prewhitening processing, and the dewhitening is performed after TDICA. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the pre/dewhitening processing prevents the decorrelation. The experiments in a reverberant room reveal that the algorithm results in higher stability and separation performance.

1. INTRODUCTION

Blind source separation (BSS) is an approach for estimating original source signals only from the information of the mixed signals observed in each input channel. This technique is applicable to high-quality hand-free speech recognition systems. Many BSS methods based on independent component analysis (ICA) [1] have been proposed [2, 3, 4] for the acoustic signal separation. However, the performances of these methods degrade seriously, especially under heavily reverberant conditions.

In order to improve the separation performance, we have proposed an efficient BSS algorithm called multistage ICA (MSICA), in which frequency-domain ICA (FDICA) [3, 5] and time-domain ICA (TDICA) [2, 6] are combined [7]. In this method, first, FDICA can find an approximate solution to separate the sources to a certain extent, and finally TDICA can remove the residual crosstalk components from FDICA. Therefore, the improvement of TDICA is a primary issue because the quality of resultant separated signals is determined by TDICA. In this paper, we discuss the stability of the TDICA algorithm, and newly propose a stable algorithm combining MSICA and linear prediction for temporally correlated signals, e.g., speech signals. First, the following points are explicitly noted: (1) The stability of learning in conventional TDICA with a holonomic constraint [6] is highly acceptable. However, the method cannot work well for speech signals due to the deconvolution property; i.e., the separated speech is harmfully distorted by the whitening process. (2) To decrease the whitening effect, TDICA with a nonholonomic constraint has been proposed [8]. This method, however, includes the inherent drawback that

the stability of learning cannot be guaranteed. In order to solve both problems simultaneously, we propose the novel approach in which the linear predictors estimated from the roughly separated source signals by FDICA are inserted before TDICA with a holonomic constraint as a prewhitening processing (after TDICA, the dewhitening is also performed). The stability of the learning in TDICA can be guaranteed by the holonomic constraint, and it is still possible to separate the temporally correlated signals because the pre/dewhitening processing prevents the ICA from performing the decorrelation. The experimental results under a reverberant condition reveal that the proposed algorithm results in the higher stability and higher separation performance than the conventional MSICA.

2. SOUND MIXING MODEL OF MICROPHONE ARRAY

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

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

where $\mathbf{x}(t) = [x_1(t), \dots, x_K(t)]^T$ is the observed signal vector and $\mathbf{s}(t) = [s_1(t), \dots, s_L(t)]^T$ 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, $\mathbf{a}(\tau) = [a_{ij}(\tau)]_{ij}$ ($[\cdot]_{ij}$ denotes the matrix in which ij -th element is $[\cdot]_{ij}$) 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. Conventional FDICA and TDICA

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.

Disadvantages:

- (F2) The separation performance is saturated before reaching a sufficient performance level because the independence assumption collapses in each narrow-band [9].
- (F3) Permutation among source signals and indeterminacy of each source gain in each subband.

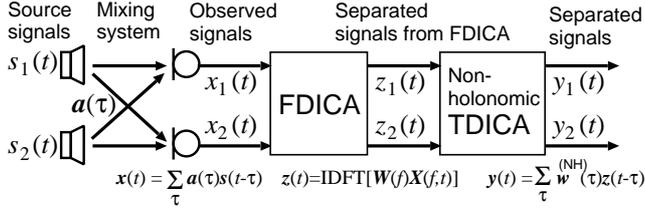


Fig. 1. Blind source separation procedure performed in the original MSICA [7]. In this system, the NH-TDICA is introduced to separate the temporally correlated signals by utilizing the flexibility of the nonholonomic constraint.

Regarding the disadvantage (F3), various solutions have already been proposed [3, 10, 4, 11]. 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.

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).

3.2. BSS Algorithm Based on MSICA [7]

As described above, the conventional ICA methods have some disadvantages. However, 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 have proposed a new algorithm, MSICA, in which FDICA and TDICA are cascaded (see Fig. 1). MSICA is conducted in the following steps. In the first stage, we perform FDICA to separate the source signals to some extent with the advantage of high stability. In the second stage, we regard the separated signals $\mathbf{z}(t)$ from FDICA as the input signals for TDICA, and we can remove the residual crosstalk components of FDICA by using TDICA. Finally, we regard the output signals from TDICA as the resultant separated signals. The separated signals of MSICA can be given as

$$\mathbf{y}(t) = \sum_{\tau=0}^{Q-1} \mathbf{w}(\tau)\mathbf{z}(t-\tau), \quad (2)$$

where $\mathbf{y}(t) = [y_1(t), \dots, y_L(t)]^T$ is the resultant separated signal vector of MSICA and $\mathbf{z}(t) = [z_1(t), \dots, z_L(t)]^T$ is the input signal vector for the TDICA part in MSICA (i.e., the output signals from FDICA). Also, $\mathbf{w}(\tau) = [w_{ij}(\tau)]_{ij}$ is the separation filter matrix, and Q is the length of the separation filter. In this procedure, we optimize $\mathbf{w}(\tau)$ so that the separated signals are mutually independent.

The selection of TDICA is an important issue because the quality of resultant separated signals is determined by TDICA. We have two choices for TDICA algorithms; TDICA with a holonomic constraint [6] and TDICA with a nonholonomic constraint [8]. In the next section, detailed explanations for each algorithm and their problems are described.

3.3. Conventional TDICA with Holonomic Constraint

Amari proposed the TDICA algorithm which optimizes the separation filter by minimizing the Kullback-Leibler divergence (KLD) between the joint probability density function and the marginal probability density function of the separated signals [6]. The iterative equation of the separation filter $\mathbf{w}^{(\text{H})}(\tau)$ to minimize the KLD is given as (hereafter we designate the iterative equation as ‘‘H-TDICA’’):

[H-TDICA]

$$\mathbf{w}_{i+1}^{(\text{H})}(\tau) = \mathbf{w}_i^{(\text{H})}(\tau) + \alpha \sum_{d=0}^{Q-1} \left\{ \mathbf{I}\delta(\tau-d) - \langle \phi(\mathbf{y}(t))\mathbf{y}(t-\tau+d)^T \rangle_t \right\} \mathbf{w}_i^{(\text{H})}(d), \quad (3)$$

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, α is the step-size parameter and \mathbf{I} is the identity matrix. $\delta(\tau)$ is Dirac delta function, where $\delta(0) = 1$ and $\delta(n) = 0$ ($n \neq 0$). Also, we define the nonlinear vector function $\phi(\mathbf{y}(t)) \equiv [\tanh(y_1(t)), \dots, \tanh(y_L(t))]^T$.

3.4. Conventional TDICA with Nonholonomic Constraint

The H-TDICA forces the separated signals to have the characteristic that their higher-order autocorrelation is $\delta(\tau)$, i.e., the signals are temporally decorrelated. This performance might have a negative influence on the source separation. In order to solve the problem, Choi proposed a modified TDICA algorithm with a non-holonomic constraint [8]. In this algorithm, the constraint for the diagonal component of $\{\cdot\}$ part in Eq. (3), i.e., the higher-order autocorrelation of separated signals, is set to be arbitrary. The iterative equation of the separation filter $\mathbf{w}^{(\text{NH})}(\tau)$ is given as (hereafter we designate the iterative equation as ‘‘NH-TDICA’’):

[NH-TDICA]

$$\mathbf{w}_{i+1}^{(\text{NH})}(\tau) = \mathbf{w}_i^{(\text{NH})}(\tau) + \alpha \sum_{d=0}^{Q-1} \left\{ \text{diag} \left(\langle \phi(\mathbf{y}(t))\mathbf{y}(t-\tau+d)^T \rangle_t \right) - \langle \phi(\mathbf{y}(t))\mathbf{y}(t-\tau+d)^T \rangle_t \right\} \mathbf{w}_i^{(\text{NH})}(d). \quad (4)$$

We have also introduced Eq. (4) in the original MSICA [7] to separate the mixed speech which corresponds to the temporally correlated signal by utilizing the flexibility of the nonholonomic constraint.

3.5. Problems in Conventional TDICAs

The advantage and disadvantage of conventional TDICAs can be summarized as follows. (1) The stability of learning in H-TDICA is satisfactory. However, the method cannot work well for speech signals due to the deconvolution property; i.e., the separated speech is harmfully distorted by the whitening process. (2) On the other hand, NH-TDICA possibly performs no deconvolution, i.e., NH-TDICA is applicable to speech signals. This method, however,

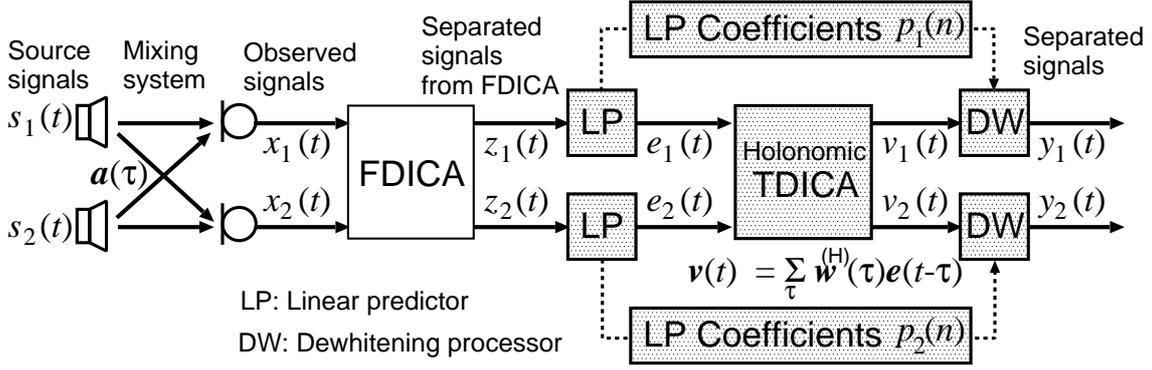


Fig. 2. Blind source separation procedure performed in the proposed algorithm combining MSICA and linear prediction. In this system, the stability of the learning in TDICA can be guaranteed by the holonomic constraint, and it is still possible to separate the temporally correlated signals because the pre/dewhitening processing prevents the ICA from performing the decorrelation.

includes the inherent drawback that the stability of learning cannot be guaranteed as described in Sect. 5.4. Thus, the separation of temporally correlated signals such as speech cannot be achieved only using the conventional TDICAs.

4. PROPOSED ALGORITHM COMBINING MSICA AND LINEAR PREDICTION

This section describes a new stable algorithm combining the linear prediction technique with an original MSICA (see Fig. 1). In the proposed algorithm, the linear predictors estimated from the roughly separated source signals by FDICA are inserted before TDICA with a holonomic constraint as a prewhitening processing (see Fig. 2). After TDICA, the dewhitening is also performed. The stability of the learning in TDICA can be guaranteed by the holonomic constraint, and it is still possible to separate the temporally correlated signals because the pre/dewhitening processing prevents the ICA from performing the decorrelation. The detailed process using the proposed algorithm is as follows.

[STEP 1. FDICA]

FDICA is performed to separate sound sources to some extent. For example, the typical separation performance in FDICA is 9.4 dB under the condition that the reverberation time is 300 ms [5, 7]. The separation filter of FDICA has spectrally flat characteristics in the direction of each sound source [5]. From this, we can estimate the approximate spectra of the sources blindly.

[STEP 2. Prewhitening by Linear Prediction]

In the linear prediction, the auto-regressive model of the generation process of the output signals from FDICA is given as

$$z_l(t) = - \sum_{n=1}^N p_l(n) z_l(t-n) + e_l(t) \quad (l = 1, \dots, L), \quad (5)$$

where $p_l(n)$ is a linear prediction coefficient for the l -th input signal, $e_l(t)$ is the input signal of this model, and N is the order of the linear prediction coefficient. The linear prediction coefficient is obtained by calculating the following Yule-Walker's simultaneous equations:

$$\begin{bmatrix} r_l(0) & \cdots & r_l(N-1) \\ \vdots & \ddots & \vdots \\ r_l(N-1) & \cdots & r_l(0) \end{bmatrix} \begin{bmatrix} p_l(1) \\ \vdots \\ p_l(N) \end{bmatrix} = - \begin{bmatrix} r_l(1) \\ \vdots \\ r_l(N) \end{bmatrix}, \quad (6)$$

where $r_l(\tau) = \langle z_l(t) z_l(t-\tau) \rangle_t$ is the autocorrelation of $z_l(t)$. Solving Eq. (6) basically involves the inversion of the matrix on the left-hand side. However, this can be simplified because the matrix is Toeplitz, i.e., all elements on each superdiagonal are equal and all elements on each subdiagonal are equal. Based on this property, we can easily and efficiently determine $p_l(n)$ by using Levinson-Durbin's recursive algorithm [12].

The whitened signal $e_l(t)$ is obtained by convolving the linear prediction coefficient $p_l(n)$ with $z_l(t)$ as

$$e_l(t) = \sum_{n=0}^N p_l(n) z_l(t-n). \quad (7)$$

[STEP 3. TDICA with Holonomic Constraint]

H-TDICA is performed with whitened signals. The output signals of H-TDICA can be given as

$$\mathbf{v}(t) = \sum_{\tau=0}^{Q-1} \mathbf{w}^{(H)}(\tau) e(t-\tau), \quad (8)$$

where $\mathbf{v}(t) = [v_1(t), \dots, v_L(t)]^T$ is the separated signal vector of H-TDICA, and $\mathbf{e}(t) = [e_1(t), \dots, e_L(t)]^T$ is the input signal vector whitened by the linear prediction for the H-TDICA part in MSICA. We optimize $\mathbf{w}^{(H)}(\tau)$ by H-TDICA:

$$\mathbf{w}_{i+1}^{(H)}(\tau) = \mathbf{w}_i^{(H)}(\tau) + \alpha \sum_{d=0}^{Q-1} \left\{ \mathbf{I} \delta(\tau-d) - \langle \phi(\mathbf{v}(t)) \mathbf{v}(t-\tau+d)^T \rangle_t \right\} \mathbf{w}_i^{(H)}(d). \quad (9)$$

[STEP 4. Dewhitening]

The dewhitening process is performed by using the linear prediction coefficients $p_l(n)$ obtained in STEP 2. The resultant separated signals $y_l(t)$ can be obtained by the following IIR filtering:

$$y_l(t) = - \sum_{n=1}^N p_l(n) y_l(t-n) + v_l(t) \quad (l = 1, \dots, L). \quad (10)$$

Note that the stability of the filtering is guaranteed because $p_l(n)$ is calculated from Levinson-Durbin's algorithm [12].

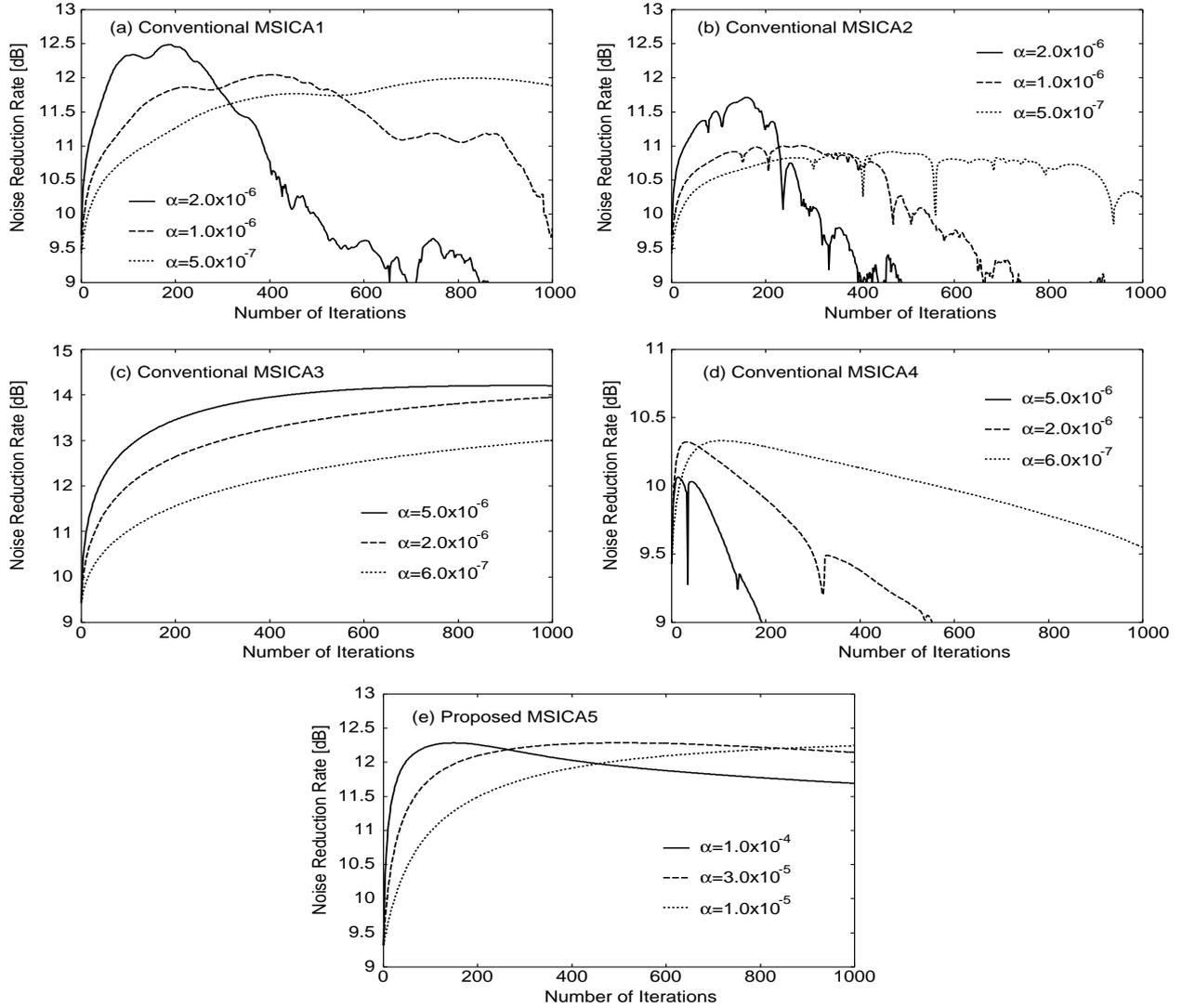


Fig. 3. Comparison of the noise reduction rates in (a) conventional MSICA1: FDICA is followed by NH-TDICA, (b) conventional MSICA2: FDICA is followed by NH-TDICA with spectral compensation, (c) conventional MSICA3: FDICA is followed by H-TDICA, (d) conventional MSICA4: FDICA is followed by H-TDICA with spectral compensation, and (e) proposed MSICA5: FDICA is followed by the proposed method combining H-TDICA and linear prediction.

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 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. Two kinds of sentences, those spoken by two male and two female speakers 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 con-

involved with the impulse responses specified by the reverberation times (RT) of 300 ms. The impulse responses are recorded in a variable reverberation time room. These sound data which are artificially convolved with the real impulse responses have the following advantages: (1) we can use the realistic mixture model of two sources and neglect the effect of background noise, (2) since the mixing condition is explicitly measured, we can easily calculate a reliable objective score to evaluate the separation performance as described in Sect. 5.3.

5.2. Postprocessing for Spectral Compensation

In order to compare the various ICAs fairly, we perform postprocessing for the spectral compensation of the separated signals in this experiment. This processing is based on the utilization of the inverse of the separation filter matrix for the normalization of gain

[3]. In this method, the following operation is performed:

$$\tilde{Y}_l(f) = \left[\mathbf{W}^{(\text{TD})}(f)^{-1} [0, \dots, 0, Y_l(f), 0, \dots, 0]^T \right]_l, \quad (11)$$

where $Y_l(f)$ denotes the frequency-domain component of the l -th estimated source signal by the TDICA part in MSICA, $\tilde{Y}_l(f)$ denotes the l -th resultant separated source signal in the frequency domain after the spectral compensation ($[\cdot]_l$ denotes the l -th element of the argument), and $\mathbf{W}^{(\text{TD})}(f)$ is the frequency-domain representation of the separation filter matrix in the TDICA part, $\mathbf{w}^{(\text{H})}(\tau)$ or $\mathbf{w}^{(\text{NH})}(\tau)$. After the operation, we can obtain the spectrally compensated output signals in the time domain by applying the inverse DFT. In this experiment, we use the DFT and the inverse DFT of 2^{18} points.

By using $\mathbf{W}^{(\text{TD})}(f)^{-1}$, the gain arbitrariness vanishes in the separation procedure. However, this procedure often fails and yields harmful results for signal reconstruction, particularly when the condition number of $\mathbf{W}^{(\text{TD})}(f)$ is large because the invertibility of $\mathbf{W}^{(\text{TD})}(f)$ cannot be guaranteed.

5.3. 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

$$\text{NRR} \equiv \frac{1}{2} \sum_{l=1}^2 \left(\text{SNR}_l^{(\text{O})} - \text{SNR}_l^{(\text{I})} \right), \quad (12)$$

$$\text{SNR}_l^{(\text{O})} = 10 \log_{10} \frac{\sum_f |H_{ll}(f) S_l(f)|^2}{\sum_f |H_{ln}(f) S_n(f)|^2}, \quad (13)$$

$$\text{SNR}_l^{(\text{I})} = 10 \log_{10} \frac{\sum_f |A_{ll}(f) S_l(f)|^2}{\sum_f |A_{ln}(f) S_n(f)|^2}, \quad (14)$$

where $\text{SNR}_l^{(\text{O})}$ and $\text{SNR}_l^{(\text{I})}$ 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 $\mathbf{H}(f) = \mathbf{W}^{(\text{MSICA})}(f) \mathbf{A}(f)$ where $\mathbf{W}^{(\text{MSICA})}(f)$ denotes the entire separation process in MSICA including both FDICA and TDICA (even with the spectral compensation described in the previous section), and $\mathbf{A}(f)$ is the mixing matrix which corresponds to the frequency-domain representation of the room impulse responses described in Sect. 2.

5.4. Experimental Results and Discussion

In this study, we compare the following MSICAs:

MSICA1: FDICA is followed by NH-TDICA,

MSICA2: FDICA is followed by NH-TDICA with spectral compensation,

MSICA3: FDICA is followed by H-TDICA,

MSICA4: FDICA is followed by H-TDICA with spectral compensation, and

MSICA5: FDICA is followed by the proposed method combining H-TDICA and linear prediction.

The length of the separation filters, $\mathbf{w}^{(\text{H})}(\tau)$ or $\mathbf{w}^{(\text{NH})}(\tau)$, is 2048. In the proposed algorithm, the order N in the linear predictor is 1024.

Figures 3(a)–(e) show the NRR results of MSICA1–MSICA5 for different iteration points. These values were averages of all of the combinations with respect to speakers and source directions. The step-size parameters are chosen independently for each of the NH-TDICA, H-TDICA, and the proposed algorithm so that the NRR scores at the early iterations are almost the same in Figs. 3(a)–(e). From these results, the following are revealed. (1) In the conventional MSICA1 and MSICA2 in which the NH-TDICA is used, the behavior of the NRR is not monotonic and there are remarkably consistent deteriorations, even when the step-size parameter is changed. (2) In the proposed algorithm, MSICA5, there are no deteriorations of NRRs. Therefore, the separation performances are almost completely retained during all of the iterations.

Regarding the separation performance of MSICA3 and MSICA4 in which the H-TDICA is used, the following are revealed. (1) The separation performance of MSICA3 is obviously superior to that of the proposed MSICA5. (2) However, its effective separation performance, i.e., the performance of MSICA4, is inferior to that of MSICA5. We speculate that the *specious* performance in MSICA3 is due to the exceeding emphasis of high-frequency components by the whitening effect of H-TDICA. Figure 4 shows the typical long-time averaged spectra of the separated signals obtained by MSICA3 and MSICA4. From these results, we can confirm the spectral distortion in MSICA3. In general, the separation in the high-frequency region is easier than that in low-frequency region [13, 14] because the reverberation is shorter as the frequency increases [15]. Thus, MSICA3 gains the improvement of the NRR only in the high-frequency region, and consequently we can conclude that MSICA3 is useless for separating the speech signals from the practical viewpoint.

In order to confirm the convergence of each MSICA learning, we evaluate the frobenius norms of $\{\cdot\}$ parts on the right-hand side in Eqs. (4) and (9), which are defined as

$$FN^{(\text{NH})} = \frac{1}{Q^2} \sum_{\tau=0}^{Q-1} \sum_{d=0}^{Q-1} \left\| \text{diag} \left(\langle \phi(\mathbf{y}(t)) \mathbf{y}(t - \tau + d)^T \rangle_t \right) - \langle \phi(\mathbf{y}(t)) \mathbf{y}(t - \tau + d)^T \rangle_t \right\|, \quad (15)$$

$$FN^{(\text{H})} = \frac{1}{Q^2} \sum_{\tau=0}^{Q-1} \sum_{d=0}^{Q-1} \left\| \mathbf{I} \delta(\tau - d) - \langle \phi(\mathbf{v}(t)) \mathbf{v}(t - \tau + d)^T \rangle_t \right\|. \quad (16)$$

Figures 5(a) and (b) show $FN^{(\text{NH})}$ of the conventional MSICA1 and $FN^{(\text{H})}$ of the proposed MSICA5. These scores correspond to the stability of the iterative learning; it should be monotonically decreased. As shown in these figures, the conventional ICA loses its stability under the nonholonomic constraint. However, the proposed method can converge in every situation and consequently, we can conclude that the proposed algorithm is effective for improving the stability of the learning.

6. CONCLUSION

We newly proposed a stable algorithm for BSS combining MSICA and linear prediction. In the proposed algorithm, the linear predictors estimated from the roughly separated signals by FDICA are

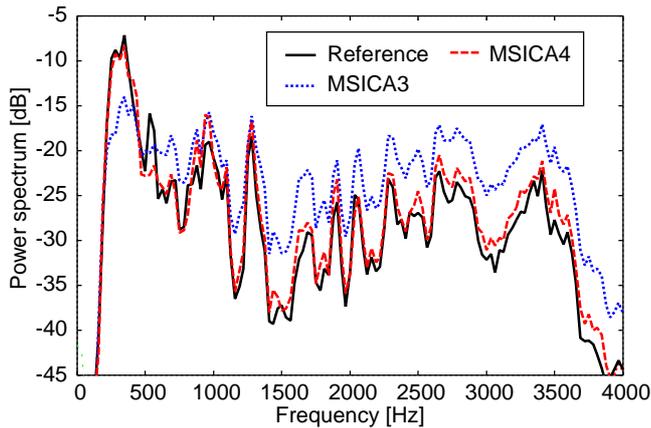


Fig. 4. Comparison of the power spectra in conventional MSICA3: FDICA is followed by H-TDICA, conventional MSICA4: FDICA is followed by H-TDICA with spectral compensation, and reference signal ($s_1(t)$ component recorded at the microphone).

inserted before the holonomic TDICA as a prewhitening processing, and the dewhitening is performed after TDICA. The stability of the proposed algorithm can be guaranteed by the holonomic constraint, and the pre/dewhitening processing prevents the decorrelation. The experimental results under a reverberant condition revealed that the proposed algorithm results in the higher stability and higher separation performance, compared with the conventional MSICA including H-TDICA or NH-TDICA.

Further study on the robustness against the background noise remains an open problem for the future.

7. ACKNOWLEDGEMENT

The authors are grateful to Dr. Shoji Makino and Ms. Shoko Araki of NTT. CO., LTD for their grateful 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] T. W. Lee, *Independent component analysis*, Kluwer academic publishers, 1998.
- [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] L. Parra and C. Spence, "Convulsive blind separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, vol.8, no.3, pp.320–327, May 2000.
- [5] 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.
- [6] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proc. SPAWC97*, pp.101–104, April 1997.
- [7] 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.

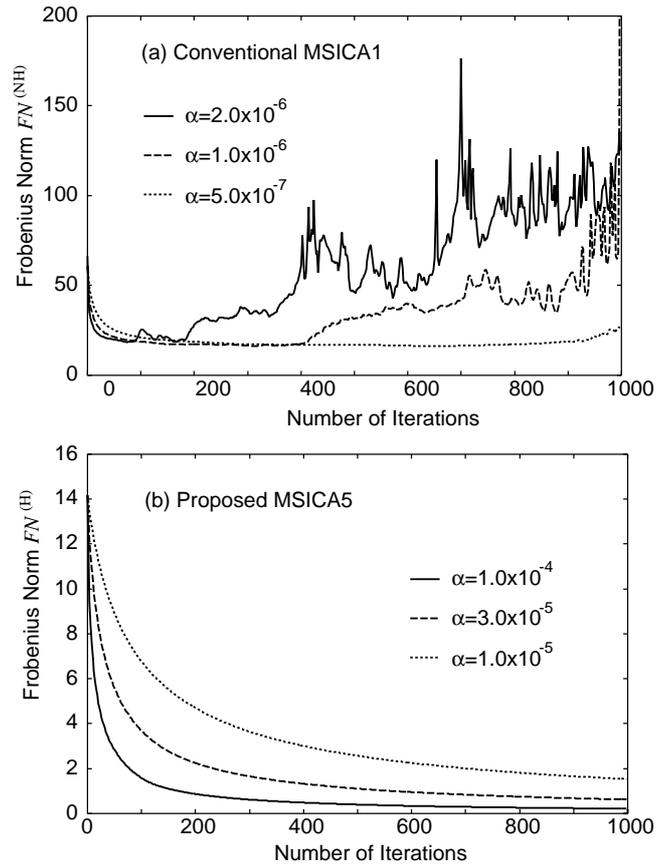


Fig. 5. Comparison of Frobenius norms Eqs. (15), (16) in (a) conventional MSICA1: FDICA is followed by NH-TDICA and (b) proposed MSICA5: FDICA is followed by the proposed method combining H-TDICA and linear prediction.

- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes (Fourth Ed.)*, McGraw-Hill Series in Electrical and Computer Engineering, New York, 2002.
- [13] T. Nishikawa, T. Kawamura, H. Saruwatari, and K. Shikano, "Overdetermined source separation with blind beamformer," *The 2000 Autumn Meeting of the ASJ*, pp.447–448, Sept. 2000 (in Japanese).
- [14] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain ICA blind source separation of non-stationary convoluted signals by utilizing geometric beamforming," *Proc. IEEE International Workshop on Neural Networks for Signal Processing*, pp.445–454, Sept. 2002.
- [15] H. Kuttruff, *Room acoustics (Fourth Ed.)*, Spon Press, London, 2000.