Overdetermined Blind Separation for Real Convolutive Mixtures of Speech Based on Multistage ICA Using Subarray Processing

Tsuyoki NISHIKAWA, Student Member, Hiroshi ABE, Nonmember, Hiroshi SARUWATARI, Kiyohiro SHIKANO, Members, and Atsunobu KAMINUMA, Nonmember

SUMMARY We propose a new algorithm for overdetermined blind source separation (BSS) based on multistage independent component analysis (MSICA). To improve the separation performance, we have proposed MSICA in which frequency-domain ICA and time-domain ICA are cascaded. In the original MSICA, the specific mixing model, where the number of microphones is equal to that of sources, was assumed. However, additional microphones are required to achieve an improved separation performance under reverberant environments. This leads to alternative problems, e.g., a complication of the permutation problem. In order to solve them, we propose a new extended MSICA using subarray processing, where the number of microphones and that of sources are set to be the same in every subarray. The experimental results obtained under the real environment reveal that the separation performance of the proposed MSICA is improved as the number of microphones is increased.

key words: microphone array, blind source separation, independent component analysis, subarray processing, convolutive mixture

1. Introduction

A hands-free speech recognition system is essential for realizing an intuitive, unconstrained, and stress-free human-machine interface. In the system, we cannot detect the user's speech with a high signal-to-noise ratio (SNR) compared with the case in that we use a close-talking microphone such as a headset microphone. Therefore, the speech recognition performance is significantly degraded. One approach for establishing a noise-robust speech recognition system is to enhance the speech signals by using noise reduction techniques [1]. Among the various noise reduction methods, the method of source separation and speech enhancement based on array signal processing, e.g., a microphone array system, is one of the most effective and promising techniques [2].

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 hands-free speech recognition systems. This technique is based on unsupervised adaptive filtering [3], and provides us with extended flexibility in that the source-separation procedure requires no training sequences and no a priori information on the directions of arrival of the sound sources. Many BSS methods based on independent component analysis (ICA) [4], [5] have been proposed [6]–[14] for the acoustic signal separation. However, the performances of these methods degrade particularly seriously under heavily reverberant conditions. In order to improve the separation performance, we have proposed multistage ICA (MSICA) [13], in which frequency-domain ICA (FDICA) [7], [8], [12] and time-domain ICA (TDICA) [6], [9], [10], [13] are cascaded. In this method, first, FDICA finds an approximate solution to separate the sources to a certain extent, and finally TDICA removes the residual crosstalk components arising in FDICA. It has been verified that conventional MSICA can realize a superior source-separation performance to FDICA and TDICA from the experimental results under the real acoustic conditions [13].

In the conventional ICA research, the specific mixing model is often assumed where the number of microphones is equal to that of sources [7], [8], [12], [14]. In the original MSICA, we also assumed this model and performed the source separation. However, additional microphones are required to achieve an improved separation performance because of the existence of the reflection and the reverberation component. In this paper, we set the number of microphones to be larger than that of sources and we extend the conventional MSICA into a new MSICA method using a large microphones. We point out that the following problems arise in the simple extension of MSICA: (1) the permutation problem [7] in FDICA part becomes very complicated, and (2) the solution of FDICA is likely to be trapped within a trivial solution as described in Sect. 5.3.

In this paper, as a robust method against these problems, we propose a new MSICA method using subarray processing, where the number of each subarray's microphones is set to be equal to that of the sources, and the outputs of FDICA performed in every subarray are weighted to be inserted into TDICA. The experimental results obtained under real acoustic conditions reveal that the separation performance of the proposed MSICA is improved over that of an original MSICA as the number of microphones is increased.

The rest of this paper is organized as follows. In Sect. 2, the sound mixing model of the microphone array is explained. In Sect. 3, the conventional MSICA algorithms and their problems are explained. In Sect. 4, the proposed MSICA using the subarray processing is described in detail. In Sect. 5, from the signal-separation experiments, the
problems in the simply extended MSICA are described and the superiority of the proposed subarray technique over the conventional method is shown. Following a discussion on the results of the experiments, we conclude our findings in Sect. 7.

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:

\[ x_k(t) = \sum_{\tau = 0}^{p-1} a_{k\ell}(\tau)s_\ell(t - \tau), \]  

(1)

where \( x_k(t) = [x_1(t), \ldots, x_K(t)]^T \) is the observed signal vector and \( s_\ell(t) = [s_1(t), \ldots, s_L(t)]^T \) is the source signal vector. \( K \) is the number of array elements (microphones), and \( L \) is the number of multiple sound sources (see Fig. 1). Also, \( a_{k\ell}(\tau) = [a_{ij}(\tau)]_j \) is the matrix in which the \( j \)-th element is \( [\cdot]_j \) is the \( K \times L \) mixing filter matrix defined by

\[ a_{k\ell}(\tau) = \begin{bmatrix} a_{1\ell}(\tau) & \cdots & a_{L\ell}(\tau) \\ \vdots & \ddots & \vdots \\ a_{k1}(\tau) & \cdots & a_{kL}(\tau) \end{bmatrix}. \]  

(2)

\( P \) is the length of the impulse response which is assumed to be an FIR-filter of thousands of taps because we deal with the arrival lags among the elements of the microphone array and the room reverberations.

3. Conventional MSICA and Problems

3.1 BSS Algorithm Based on MSICA [13]

Figure 2 shows the procedure of the original MSICA in which FDICA [12] and TDICA [10] are cascaded. In the case of \( K = L \), MSICA is conducted in the following steps.

First, we perform FDICA to separate the source signals to some extent with the advantage of high stability. The output signals \( z_L(t) = [z_1(t), \ldots, z_L(t)]^T \) from FDICA can be given as

\[ z_L(t) = \sum_{\tau = 0}^{Q-1} v_{LL}(\tau)x_L(t - \tau), \]  

(3)

where \( v_{LL}(\tau) = [v_{ij}(\tau)]_j \) is the \( L \times L \) separation filter matrix for FDICA, and \( Q \) is the length of the separation filter of FDICA. In FDICA, we optimize \( v_{LL}(\tau) \) so that the narrow-band output signals are mutually independent at each frequency (see Fig. 3).

Second, we regard the output signals \( z_L(t) \) 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 \( y_L(t) = [y_1(t), \ldots, y_L(t)]^T \) of MSICA can be given as

\[ y_L(t) = \sum_{\tau = 0}^{R-1} w_{LL}(\tau)z_L(t - \tau), \]  

(4)

where \( w_{LL}(\tau) = [w_{ij}(\tau)]_j \) is the separation filter matrix for TDICA, and \( R \) is the length of the separation filter of TDICA. In this procedure, we optimize \( w_{LL}(\tau) \) so that the fullband separated signals are mutually independent (see Fig. 4).

3.2 Simple Extension of Conventional MSICA

In the conventional MSICA, the specific mixing model is

\[ X_L(t) = \sum_{j=1}^{L} a_{ij}(t)X_j(t), \]  

for \( j = 1, \ldots, L \), where \( X_j(t) \) is the original signal and \( a_{ij}(t) \) is the mixing coefficient. The output signals from FDICA are given by

\[ z_L(t) = \sum_{\tau = 0}^{Q-1} v_{LL}(\tau)x_L(t - \tau), \]  

for \( \tau = 0, \ldots, Q-1 \), and the output signals from TDICA are given by

\[ y_L(t) = \sum_{\tau = 0}^{R-1} w_{LL}(\tau)z_L(t - \tau), \]  

for \( \tau = 0, \ldots, R-1 \).
assumed, where the number of microphones is equal to that of sources. However, additional microphones are required to achieve an improved separation performance because of the reflection and the reverberation component. Thus, we should set the number of microphones to be larger than that of sources (i.e., $K > L$), and we extend the conventional MSICA into a new MSICA method by using a large number of microphones. First, as the simple extension of MSICA, we consider the following two methods in the specific case of $K > L$.

**Method 1**

The $K$ output signals are obtained from FDICA and $L$ separated signals are obtained from TDICA:

$$z_k(t) = \sum_{\tau = 0}^{Q-1} w_{kk}(\tau)x_k(t - \tau),$$

$$y_l(t) = \sum_{\tau = 0}^{Q-1} w_{ll}(\tau)z_k(t - \tau).$$

There is a permutation problem [7] of sources in every frequency bin in FDICA. By using recently proposed techniques [11],[15]–[17], we can easily solve the problem only in the case of $K = L$. However, (P1) the permutation problem in FDICA becomes very complicated as the number of microphones is increased. Also, (P2) the discrimination of the output signals corresponding to the true sources is needed because there exist $K - L$ imaginary outputs. Therefore Method 1 is not applicable to separating sources in the real environment.

**Method 2**

The $L$ output signals are obtained from FDICA and the $L$ separated signals are obtained from TDICA:

$$z_l(t) = \sum_{\tau = 0}^{Q-1} w_{ll}(\tau)x_k(t - \tau),$$

$$y_l(t) = \sum_{\tau = 0}^{Q-1} w_{ll}(\tau)z_k(t - \tau).$$

There still exist some problems as follows. (P3) In the iterative learning of FDICA, the solution is likely to be trapped within a trivial solution as described in Sect. 5.3. (P4) We cannot utilize all the information of the observed signals at $K$ microphones in TDICA because the number of the input signals for TDICA is decreased to $L$ by FDICA.

Due to these problems, a new extension algorithm of MSICA which is not affected by (P1)–(P4) is desired to achieve a superior separation performance. Therefore, in the next section we propose a new BSS algorithm based on the extended MSICA using subarray processing.

4. Proposed MSICA Using Subarray Processing

4.1 Source Separation Algorithm

In the proposed extended MSICA, we regard the $K$ observed signals as combinations of the $L(< K)$ observed signals, and we regard this combination as a subarray (see Fig. 5). First, we divide the whole inputs into $K - 1$ subarrays, and we perform FDICA in every subarray. The output signals $z_L^{(o)}(t) = [z_1^{(o)}(t), \ldots, z_L^{(o)}(t)]^T$ from FDICA in the $n$-th subarray can be given as

$$z_L^{(o)}(t) = \sum_{\tau = 0}^{Q-1} \psi_L^{(o)}(\tau)x_L^{(o)}(t - \tau),$$

where $\psi_L^{(o)}(\tau) = [\psi_L^{(o)}(\tau)]_i$ is the separation filter matrix of FDICA in the $n$-th subarray and

$$x_L^{(o)}(t) = [x_n(t), x_{n+1}(t), \ldots, x_{n+L-1}(t)]^T.$$

As the FDICA algorithm for optimization of the separation filter $\psi_L^{(o)}(\tau)$, we introduce the fast-convergence FDICA proposed by one of the authors [12]. In the FDICA, the optimal $\psi_L^{(o)}(\tau)$ is obtained by the following iterative equation [7]:

$$V_L^{(o)}(f)_{i+1} = d \text{diag} \left( \Phi(Z_L^{(o)}(f, m))Z_L^{(o)}(f, m)^H \right)_m$$

\[= \left( \Phi(Z_L^{(o)}(f, m))Z_L^{(o)}(f, m)^H \right)_m V_L^{(o)}(f)_{i} + V_L^{(o)}(f)_{i} \]
where $Z_{l}(f, m)$ and $Z_{l}^{(0)}(f, m)$ are the real and imaginary parts of $Z_{l}(f, m)$, respectively.

Next, we regard all output signals from FDICA in $K-1$ subarrays as the input signals for TDICA, and we remove the residual crosstalk components from FDICAs. The resultant separated signals $y_{L}^{(i)}(t)$ can be given as

$$y_{L}^{(i)}(t) = \sum_{\tau=0}^{K-1} w_{L,LX,K-1}(\tau) z_{LX,K-1}(t-\tau).$$  \hspace{1cm} (14)

where $w_{L,LX,K-1}(\tau)$ is the $L \times (L \times K-1)$ separation filter matrix and

$$z_{LX,K-1}(t) = [z_{1}^{(1)}(t), z_{2}^{(1)}(t), \ldots, z_{K-1}^{(1)}(t),$$
$$z_{1}^{(2)}(t), z_{2}^{(2)}(t), \ldots, z_{K-1}^{(2)}(t), \ldots, z_{1}^{(K-1)}(t), z_{2}^{(K-1)}(t), \ldots, z_{K-1}^{(K-1)}(t)]^T.$$  \hspace{1cm} (15)

In the TDICA, the optimal $w_{L,LX,K-1}(\tau)$ is obtained by the following iterative equation [10]:

$$w_{L,LX,K-1}(\tau)_{i+1} = \beta \sum_{\delta=0}^{R-1} \left[ \text{diag} \left( \langle \phi(y_{L}^{(i)}(t) y_{L}(t-\tau + \delta)^T \rangle \right) ight.$$
$$- \langle \phi(y_{L}^{(i)}(t) y_{L}(t-\tau + \delta)^T \rangle \rangle w_{L,LX,K-1}(\delta)_{i}$$
$$+ w_{L,LX,K-1}(\tau)_{i}]_{[\cdot]}.$$  \hspace{1cm} (16)

where $\beta$ is the step-size parameter and $\langle \cdot \rangle_{i}$ denotes the time-averaging operator.

We can easily solve the permutation problem by using the conventional methods [11, 15, 17] because the number of microphones is equal to that of sources in every subarray. Also, the discrimination of the output signals corresponding to the true sources is not required because the number of output signals from FDICA is equal to that of sources, i.e., there are no imaginary outputs. The separation filter of FDICA is likely to converge on the optimal point, particularly in the case of $K = L$ (see Sect.5.3). Therefore, in the proposed MSICA, the problems (P1)–(P3) described in Sect.3.2 do not arise. In addition, we can utilize the information of all the element of the microphone array in the TDICA because we use the output signals from FDICA in all subarrays with the information from all microphones. Therefore, (P4) is also solved by the proposed MSICA.

### 4.2 Initial Value for TDICA Part in Proposed MSICA

As the initial value of the TDICA part in the proposed MSICA, we introduce the following coefficient:

$$w_{L,LX,K-1}(\tau)_{[0]} = \begin{cases} 
\frac{1}{K-1} \sum_{l=1}^{K-1} \text{IDFT} [\exp (j\omega d_{l})] \in R, \\
\text{if } (l-1) \times K-1 < k \leq l \times K-1, \\
\end{cases}$$

$$\epsilon_{0} = \sum_{\tau=0}^{T} \left| \langle \phi(z_{j}^{(0)}(t)) z_{j}^{(0)}(t-\tau) \rangle_{i} \right|^2.$$  \hspace{1cm} (17)

where IDFT[..] denotes an inverse DFT of .. $T$ is the length of the output signals from FDICA, $\omega$ is an angular frequency, and $d_{l}$ is the phase delay of input signals for TDICA so that the correlation between the input signal $z_{l}^{(0)}$ and $z_{j}^{(0)}$ is maximum. Also, $\gamma$ is the enhancement parameter to weight with the correlation $c_{j}$. $c_{j}$ corresponds to the Frobenius norm of the update term $\{ \}$ in the TDICA algorithm given by Eq.(16), and we estimate the degree of the separation performance by using this value. We introduce this filter (Eq.(17)) as the initial value of the TDICA part in MSICA. If $\gamma = 0$ in Eq.(17), this filter corresponds to a conventional delay-and-sum beamformer. On the other hand, highly separated output signals from specific FDICAs are strongly weighted as the $\gamma$ is increased. We compare the separation performances of the initial value and the proposed MSICA by changing $\gamma$ and the number of microphones.

### 5. Simulation Experiments Using Measured Impulse Responses

#### 5.1 Experimental Setup

A 14-element array with the interelement spacing of 2.83 cm is assumed. The speech signals are assumed to arrive from two directions, $-40^\circ$ and $20^\circ$. In this paper, we assume that the number of sound sources is known in advance. Regarding the estimation of the number of sound sources, many methods are available, e.g., [18], [19]. The distance between the microphone array and the loudspeakers is 2.0 m (see Fig. 6). Figure 7 shows the numbers of the microphone array and the subarray used in the experiments. Two sentences spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research [20] are used.

![Fig.6](image_url) Layout of reverberant room [21] used in experiments.

![Fig.7](image_url) Configuration of the microphone array and the subarray used in experiments.
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 times of 300 ms. We use the impulse responses recorded in a real room selected from the Real World Computing Partnership (RWCP) sound scene database [21]. 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, and (2) since the mixing condition is explicitly measured, we can easily calculate a reliable objective score for evaluating the separation performance as described in Sect. 5.2.

As the analysis conditions, the filter length of FDICA is 1024 taps and the initial value of FDICA is the null beamformer in which the null steered toward ±60°. Also, the number of iterations of FDICA is 150 and that of TDICA is 500.

5.2 Objective Evaluation Score

Noise reduction rate (NRR), defined as the output 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

$$\text{NRR} = \frac{1}{2} \sum_{k=1}^{2} \left[ \frac{\text{SNR}_{\text{Y}} - \text{SNR}_{\text{Y}}^{(i)}}{\text{SNR}_{\text{Y}}^{(i)}} \right],$$

$$\text{SNR}_{Y} = 10 \log_{10} \frac{\sum_{f} \left| |H_{n}(f)|S(f)\right|^2}{\sum_{f} \left| |H_{n}(f)S_{n}(f)|\right|^2},$$

$$\text{SNR}_{Y}^{(i)} = 10 \log_{10} \frac{\sum_{f} \left| |H_{n}(f)|S_{n}(f)\right|^2}{\sum_{f} \left| |H_{n}(f)S_{n}(f)|\right|^2},$$

where SNR\textsubscript{Y} and SNR\textsubscript{Y}\textsuperscript{(i)} are the output SNR and the input SNR, respectively, and i ≠ n. Also, S\textsubscript{n}(f) is the frequency-domain representation of the source signal, s\textsubscript{r}(t), H\textsubscript{n}(f) is the element in the i\textsuperscript{th} row and the j\textsuperscript{th} column of the matrix H(f) = W(f)V(f)A(f) where A(f) is the mixing matrix which corresponds to the frequency-domain representation of the room impulse responses described in Sect. 2, and W(f) is the frequency-domain representation of the separation filter matrix of TDICA, w(τ).

5.3 Problems in Simply Extended MSICA Based on Method 2

In order to visually evaluate the convergence by FDICA of Method 2, we plot the directivity pattern of the separation filter e\textsubscript{L,E}(τ) provided by FDICA of Method 2 (Eq. (7)). Figure 8 shows the directivity pattern for a different number of microphones (K = 2 or 12), where “Filter 1” is extracting source 1, and “Filter 2” is extracting source 2. In Fig. 8(a),

![Fig. 8 Directivity patterns in 1812.5 Hz of the separation filters provided by FDICAs of Method 2 (Eq. (7)) by using (a) two microphones and (b) 12 microphones. The number of sources is two.](image)

the directional nulls of the separation filters given by FDICA steer in the direction of interference when two microphones are used. However, in Fig. 8(b) where 12 microphones are used, the nulls of separation filter 2 steer not only in the direction of interference but also in the target speech direction. Therefore, the output signal from separation filter 2 becomes a zero signal.

In FDICA, the separation filters are updated so that the output signals are mutually independent and the separated signal from FDICA can be generally given as

$$Z(f, m) = a_{f}(f)S_{k}(f, m),$$

where S\textsubscript{k}(f, m) is the source signal in the time-frequency domain and a\textsubscript{f}(f) is the arbitrary complex-valued coefficient. The coefficient a\textsubscript{f}(f) is not determined because we evaluate only the independence between the output signals in FDICA. The coefficient a\textsubscript{f}(f) in Fig. 8(b) becomes approximately zero and the output signal from filter 1 becomes the zero signal. The speech signal and the zero signal are mutually independent and consequently, the independence assumption holds. However, needless to say, this solution is trivial with respect to the separation of source signals. This phenomenon occurs due to the fact that the degree of freedom of the separation filter becomes high when we use many microphones. We can conclude that the separation filter with a low degree of freedom is desirable in FDICA. This is the motivation behind proposing the extended MSICA using subarray processing in which the number of each subarray's microphones is equal to that of sources.

5.4 Separation Results of FDICA and Conventional MSICA in Each Subarray

Figure 9 shows the NRR results of FDICA and the con-
Fig. 9  Comparison of the source-separation performance by FDICA and conventional MSICA in every subarray. For microphone number and subarray number see Fig. 7.

Fig. 10  Comparison of the initial values in the TDICA part of the proposed MSICA for different \( \gamma \) and numbers of microphones.

Fig. 11  Comparison of the proposed MSICA for different \( \gamma \) and numbers of microphones.

Fig. 12  Relationship between the source-separation performance and the number of microphones or filter length of TDICA part.

5.5 Separation Results of Proposed MSICA for Different Initial Values in TDICA Part

In the proposed MSICA using subarray processing, the microphones which are selected symmetrically with respect to the array center (see the black circle in Fig. 7) are used. For example, the "four-element array" consists of microphones \#5, \#6, \#7, and \#8.

Figures 10 and 11 show the NRR results of the initial value and the proposed MSICA for different \( \gamma \) and numbers of microphones. From Fig. 10, the separation performances of the initial value for the proposed MSICA are improved as \( \gamma \) is increased in all microphones. Therefore, the weighting equation (Eq. (17)) with the input signals for TDICA works effectively. The final separation performance is improved as the number of microphones is increased (see Fig. 11). However, the separation performances of the proposed MSICA which are improvements from the initial values using different \( \gamma \) are not very different in all microphones. We can conclude that the proposed MSICA does not depend on the initial value in the TDICA part and we can achieve a superior separation performance by using the information from many microphones.

To assess the efficacy of the proposed MSICA, we compare the proposed MSICA and the conventional array signal processing [2], e.g., delay-and-sum beamformer and null beamformer which are popular techniques used to enhance the speech signal. The NRR in delay-and-sum beamformer is 2.0 dB and the NRR in null beamformer is 3.9 dB. The NRR in the proposed MSICA is more than 10 dB (see Fig. 11), and consequently the proposed MSICA is more effective for separating speech signal and noise signal compared with the conventional array signal processing.

5.6 Relationship between Separation Performance and the Number of Microphones or Filter Length

Figure 12 shows the NRR results of the proposed ex-
tended MSICA for different numbers of microphones or filter lengths of the TDICA part. In Fig. 12, the horizontal axis shows the number of microphones, the vertical axis shows the filter length and the tone shows the separation performance.

We investigate the relationship between the source-separation performance and the number of microphones or the filter length. On observing the horizontal axis in Fig. 12 it is seen that the separation performance is improved as the number of microphones is increased. Moreover, on observing the vertical axis we note that the separation performance is also improved as the filter length is increased. These results show the same tendencies as those for the conventional microphone array processing, e.g., in terms of delay and sum beamformer. However, in the proposed MSICA, huge amounts of calculations are required. The increase in the number of microphones corresponds to an increase in the number of FDICAs. Therefore, as a future work, we should propose the MSICA with an effective subarray structure.

6. Illustrative Experiment with Real Recordings

6.1 Conditions for Experiment

In this section, the BSS experiment is performed using actual devices in a real acoustic environment. The experiment was carried out in an ordinary room, which has the reverberation time of 200 msec, as shown in Fig. 13. A 14-element array with interelement spacing of 2.1 cm is used. The source signals arrive from two directions, $-60^\circ$ and $30^\circ$; a loudspeaker is placed on the right-hand side ($30^\circ$) to sound a target female speech, and a tower-type personal computer (PC) is placed on the left-hand side ($-60^\circ$) as an interference (noise) sound generator.

As the analysis conditions for these experiments, the sampling frequency is 16 kHz, the filter length of FDICA and TDICA are 2048 taps and the initial value of FDICA is the null beamformer in which the null steered toward $-30^\circ$ and $50^\circ$. Also, the number of iterations of FDICA is 200, that of TDICA is 200, and $\gamma = 0.0$.

The level of background noise, which is not the PC noise but an ambient noise, and the target speech level measured at the array origin, were 37 dB(A) and 54 dB(A), respectively. The levels of the target speech and the PC noise are almost even. It also should be mentioned that all of the experimental apparatus may include possible sensor noise, environment noise, and/or nonlinear error which is produced in, for example, amplifiers.

6.2 Results

Figure 14 shows NRR results in the proposed MSICA. Note that we only depict the NRR with regard to the target speech in this figure because we consider the PC noise as an uninteresting and hence undesired sound source. The results reveal that the separation performance is also improved as the number of microphones is increased, like the simulation results in Sect. 5. This indicates encouraging evidence for the feasibility of the proposed algorithm for a real-world application such as a robust hands-free speech communication system.

7. Conclusion

In this paper, we proposed a MSICA, by setting the number of microphones to be larger than that of sources to achieve an improved separation performance. In the FDICA part in the simple extension of MSICA, the use of additional microphones led to alternative problems: the solution is likely to be trapped within a trivial solution and the permutation problem in FDICA becomes very complicated. In order to solve these problems, we proposed a new extended MSICA using subarray processing, where the number of microphones and that of sources are set to be the same in every subarray. The experimental results obtained under real acoustic environmental conditions reveal that the separation performance of the proposed MSICA is improved as the number of microphones is increased.

Acknowledgement

This work was partly supported by NISSAN MOTOR CO., LTD. in Japan and Core Research for Evolutional Science and Technology (CREST) Program ‘Advanced Media Tech-
ology for Everyday Living” of Japan Science and Technology Agency (JST).

References


Tsuyoki Nishikawa was born in Mie, Japan on 1978. He received the B.E. degree in electronic system and information engineering from Kinki University in 2000 and received the M.E. degree in information and science from Nara Institute of Science and Technology (NAIST) in 2002. He is now a Ph.D. student at Graduate School of Information Science, NAIST. His research interests include acoustics signal processing, sensor array processing, and blind source separation. He received TELECOM System Technology Award and TELECOM System Technology Student Award from the Telecommunications Advancement Foundation in 2004. He is a member of the the Acoustical Society of Japan.

Hiroshi Abe was born in Aichi, Japan on 1979. He received the B.E. degree in division of mathematics from Hokkaido University in 2002. He is now a master student at Graduate School of Information Science, Nara Institute of Science and Technology (NAIST). His research interests include array signal processing and blind source separation. He is a member of the the Acoustical Society of Japan.

Hiroshi Saruwatari was born in Nagoya, Japan, on July 27, 1967. He received the B.E., M.E. and Ph.D. degrees in electrical engineering from Nagoya University, Nagoya, Japan, in 1991, 1993 and 2000, respectively. He joined Intelligent Systems Laboratory, SECOM CO., LTD., Mitaka, Tokyo, Japan, in 1993, where he engaged in the research and development on the ultrasonic array system for the acoustic imaging. He is currently an associate professor of Graduate School of Information Science, Nara Institute of Science and Technology. His research interests include array signal processing, blind source separation, and sound field reproduction. He received the Paper Award from IEICE in 2001 and TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004. He is a member of the IEEE and the Acoustical Society of Japan.
Kiyuhiro Shikano received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980, respectively. He is currently a professor of Nara Institute of Science and Technology (NAIST), where he is directing speech and acoustics laboratory. His major research areas are speech recognition, multi-modal dialog system, speech enhancement, adaptive microphone array, and acoustic field reproduction. From 1972, he had been working at NTT Laboratories, where he had been engaged in speech recognition research. During 1990–1993, he was the executive research scientist at NTT Human Interface Laboratories, where he supervised the research of speech recognition and speech coding. During 1986–1990, he was the Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories, where he was directing speech recognition and speech synthesis research. During 1984–1986, he was a visiting scientist in Carnegie Mellon University, where he was working on distance measures, speaker adaptation, and statistical language modeling. He received the Yonezawa Prize from IEICE in 1975, the Signal Processing Society 1990 Senior Award from IEEE in 1991, the Technical Development Award from ASJ in 1994, IPSJ Yamashita SIG Research Award in 2000, and Paper Award from the Virtual Reality Society of Japan in 2001. He is a member of the Information Processing Society of Japan, the Acoustical Society of Japan (ASJ), Japan VR Society, the Institute of Electrical and Electronics Engineers (IEEE), and International Speech Communication Society.

Atsunobu Kaminuma received the B.E. degree from Akita University in 1993, received the M.E. degree from Japan Advanced Institute of Science and Technology in 1995, and received the Ph.D. degree from Nara Institute of Science and Technology in 2000. In 2000, he joined Nissan Research Center of NISSAN MOTOR CO., LTD. His research interests include sound/speech inter-face system. He received Paper Award from the Virtual Reality Society of Japan in 2001. He is a member of the the Acoustical Society of Japan and the Virtual Reality Society of Japan.