Evaluation of Multistage SIMO-Model-Based Blind Source Separation Combining Frequency-Domain ICA and Time-Domain ICA

Satoshi Ukai¹, Hiroshi Saruwatari¹, Tomoya Takatani¹, Kiyohiro Shikano¹, Ryo Mukai², and Hiroshi Sawada²

¹ Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan
sato-uk@is.aist-nara.ac.jp
² NTT Communication Science Laboratories
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, Japan

Abstract. In this paper, single-input multiple-output (SIMO)-model-based blind source separation (BSS) is addressed, where unknown mixed source signals are detected at the microphones, and these signals can be separated, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. This technique is highly applicable to high-fidelity signal processing such as binaural signal processing. First, we provide an experimental comparison between two kinds of the SIMO-model-based BSS methods, namely, traditional frequency-domain ICA with projection-back processing (FDICA-PB), and SIMO-ICA recently proposed by the authors. Secondly, we propose a new combination technique of the FDICA-PB and SIMO-ICA, which can achieve a more higher separation performance in comparison to two methods. The experimental results reveal that the accuracy of the separated SIMO signals in the simple SIMO-ICA is inferior to that of FDICA-PB, but the proposed combination technique can outperform both simple FDICA-PB and SIMO-ICA.

1 Introduction

Blind source separation (BSS) is the approach taken to estimate original source signals using only the information of the mixed signals observed in each input channel. In recent BSS works based on independent component analysis (ICA), various methods [1-4] have been proposed to deal with the separation of convolutive acoustical-sound mixtures, but these approaches only output each of the independent sound sources as a monaural signal. Accordingly, the separated sounds cannot maintain any spatial qualities of each sound source, e.g., directivity, localization, etc. This prevents any traditional BSS methods from being applied to binaural signal processing [5], or any high-fidelity acoustic signal processing.

In order to solve the problem, we should adopt a new blind separation framework in which Single-Input Multiple-Output (SIMO)-model-based BSS is considered. Here the term "SIMO" represents the specific transmission system in...
which the input is a single source signal and the outputs are its transmitted signals observed at multiple sensors. In the SIMO-model-based separation scenario, unknown multiple source signals which are mixed through unknown acoustical transmission channels are detected at the microphones, and these signals can be separated, not into monaural source signals but into SIMO-model-based signals from independent sources as they are at the microphones. Thus, the SIMO-model-based separated signals can maintain the spatial qualities of each sound source. Obviously the attractive feature is highly applicable to high-fidelity acoustic signal processing.

As an early contribution for SIMO-model-based BSS, Murata et al. have proposed frequency-domain ICA (FDICA) with projection-back processing [1] (hereafter we call it FDICA-PB). Also, we have proposed SIMO-ICA which consists of multiple time-domain ICAs (TDICAs) [6]. Following these methods, inspired by Nishikawa’s multistage ICA approach [3], we are now studying and proposing a combination technique [7] of the FDICA-PB and SIMO-ICA, which can achieve a more higher separation performance with the low computational complexity. Our previous report [7], however, only showed limited and slightly unreliable experimental results in that the reverberation is too short, the number of data sets is little, and there is a lack of consistency on initial value in ICA. To improve them and provide more reliable evidences, this paper mainly describes an experimental evaluation of the proposed combination technique under more realistic conditions, and adds a discussion on the importance of the combination order. The experiments results explicitly reveal that the superiority of the proposed combination technique over the FDICA-PB or SIMO-ICA.

2 Mixing Process

In this study, the number of microphones is $K$ and the number of multiple sound sources is $L$. The observed signals in which multiple source signals are mixed linearly are expressed as

$$
\mathbf{x}(t) = \sum_{n=0}^{N-1} \alpha(n)s(t-n),
$$

where $s(t) = [s_1(t), \cdots, s_L(t)]^T$ is the source signal vector, and $\mathbf{x}(t) = [x_1(t), \cdots, x_K(t)]^T$ is the observed signal vector. Also, $\alpha(n) = [a_{kl}(n)]_{kl}$ is the mixing filter matrix with the length of $N$, $a_{kl}(n)$ is the impulse response between the $k$-th microphone and the $l$-th sound source, and $[\mathbf{X}]_{ij}$ denotes the matrix which includes the element $X$ in the $i$-th row and the $j$-th column. Hereafter, we only deal with the case of $K = L$ in this paper.

3 SIMO-Model-Based BSS 1: Conventional FDICA-PB

In the conventional FDICA-PB, first, the short-time analysis of observed signals 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 $\mathbf{X}(f,t) = [X_1(f,t), \cdots, X_K(f,t)]^T$. 
Next, we perform signal separation using the complex-valued unmixing matrix, \( W(f) = [W_k(f)]_{k} \), so that the \( L \) time-series output \( Y(f,t) = [Y_1(f,t), \cdots, Y_L(f,t)]^T \) becomes mutually independent; this procedure can be given as
\[
Y(f,t) = W(f)X(f,t).
\] (2)

We perform this procedure with respect to all frequency bins. The optimal \( W(f) \) is obtained by, e.g., the following iterative updating:
\[
W^{[i+1]}(f) = \eta \left[ I - \langle \Phi(Y(f,t))Y^H(f,t) \rangle \right] W^{[i]}(f) + W^{[i]}(f),
\] (3)
where \( \langle \cdot \rangle \) denotes the time-averaging operator, \([i]\) is used to express the value of the \( i \)th step in the iterations, and \( \eta \) is the step-size parameter. In our research, we define the nonlinear vector function \( \Phi(\cdot) \) as \( [e^{j \arg(Y_1(f,t))}, \cdots, e^{j \arg(Y_L(f,t))}]^T \), where \( \arg(\cdot) \) represents an operation to take the argument of the complex value [4]. After the iterations, the permutation problem, i.e., indeterminacy in ordering sources, can be solved by [8].

Finally, in order to obtain the SIMO components, the separated signals are projected back onto the microphones by using the inverse of \( W(f) \) [1]. In this method, the following operation is performed.
\[
Y_k^{(l)}(f,t) = \left\{ W(f)^{-1} \left[ 0, \cdots, 0, Y_l(f,t), 0, \cdots, 0 \right]^T \right\}_{k},
\] (4)
where \( Y_k^{(l)}(f,t) \) represents the \( l \)-th resultant separated source signal which is projected back onto the \( k \)-th microphone, and \( \cdot \)\(_k\) denotes the \( k \)-th element of the argument.

The FDICA-PB has the advantage that (F1) this method is very fast and nonsensitive to the initial value in the iterative updating because the calculation of FDICA given by (3) and the projection-back processing given by (4) are simple. There exists, however, the disadvantages that (F2) the inversion of \( W(f) \) often fails and yields harmful results because the invertibility of every \( W(f) \) cannot be guaranteed, and (F3) the circular convolution effect inherent in FDICA is likely to cause the deterioration of the separation performance.

### 4 SIMO-Model-Based BSS 2: SIMO-ICA

The SIMO-ICA [6] consists of \( (L - 1) \) TDICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. The separated signals of the \( l \)-th ICA \( (l = 1, \cdots L - 1) \) in SIMO-ICA are defined by
\[
y_{(ICA_{l})}(t) = [y_{k}^{(ICA_{l})}(t)]_{k} = \sum_{n=0}^{D-1} w_{(ICA_{l})}(n)x(t-n),
\] (5)
where \( w_{(ICA_{l})}(n) = [w_{ij}^{(ICA_{l})}(n)]_{ij} \) is the separation filter matrix in the \( l \)-th ICA, and \( D \) is the filter length.
Regarding the fidelity controller, we calculate the following signal vector, in which all elements are to be mutually independent,

$$y_{\text{ICAL}}(t) = x(t - D/2) - \sum_{i=1}^{L-1} y_{\text{ICAL}}(t).$$  \hspace{1cm} (6)

Hereafter, we regard $y_{\text{ICAL}}(t)$ as an output of a virtual "L-th" ICA. To explicitly show the meaning of the fidelity controller, we rewrite (6) as

$$\sum_{i=1}^{L} y_{\text{ICAL}}(t) = 0.$$  \hspace{1cm} (7)

This equation means a constraint to force the sum of all ICAs' output vectors $\sum_{i=1}^{L} y_{\text{ICAL}}(t)$ to be the sum of all SIMO components $\sum_{n=0}^{N-1} a_{ki}(n) s_{i}(n - t + D/2)$, where $D/2$ is used to deal with nonminimum phase systems. Using (5) and (6), we can obtain the appropriate separated signals and maintain their spatial qualities as follows.

**Theorem:** If the independent sound sources are separated by (5), and simultaneously the signals obtained by (6) are also mutually independent, then the output signals converge on unique solutions, up to the permutation, as

$$y_{\text{ICAL}}(t) = \sum_{n=0}^{L-1} \text{diag} \left[ a(n) P^T_l \right] P_l s(n - t + D/2),$$  \hspace{1cm} (7)

where $P_l \ (l = 1, \ldots, L)$ are exclusively-selected permutation matrices which satisfy

$$\sum_{l=1}^{L} P_l = [1]_{ij}.$$  \hspace{1cm} (8)

Regarding a proof of the theorem, see [6]. Obviously the solutions given by (7) provide necessary and sufficient SIMO components, $\sum_{n=0}^{L-1} a_{ki}(n) s_{i}(n - t + D/2)$, for each $l$-th source.

In order to obtain (7), the natural gradient of Kullback-Leibler divergence of (6) with respect to $w_{\text{ICAL}}(n)$ should be added to the existing TDICA-based iterative learning rule [2] of the separation filter in the $l$-th ICA ($l = 1, \ldots, L-1$). The new iterative algorithm of the $l$-th ICA part ($l = 1, \ldots, L-1$) in SIMO-ICA is given as

$$w^{[i+1]}_{\text{ICAL}}(n) = w^{[i]}_{\text{ICAL}}(n)$$

$$- \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \varphi(y^{[i]}_{\text{ICAL}}(t) y^{[i]}_{\text{ICAL}}(t - n + d)^T \right\rangle_t \right\}$$

$$+ \left\langle \varphi(x(t - D/2) - \sum_{i=1}^{L-1} y^{[i]}_{\text{ICAL}}(t)) \right\rangle_t$$

$$+ \left\langle x(t - n + d - D/2) - \sum_{i=1}^{L-1} y^{[i]}_{\text{ICAL}}(t - n + d)^T \right\rangle_t.$$
Fig. 1. Example of input and output relations in proposed method in the case of 2 sources with 2 microphones.

\[
\left( I \delta(d - \frac{D}{2}) - \sum_{l=1}^{L-1} w_\text{ICA}(d) \right),
\]

where \( \alpha \) is the step-size parameter, \( \delta(n) \) is a delta function, i.e., \( \delta(0) = 1 \) and \( \delta(n) = 0 \) \((n \neq 0)\), and \( \psi(.) \) is the nonlinear vector function, e.g., the \( l \)th element is \( y_l(t)/|y_l(t)| \). Also, the initial values of \( w_\text{ICA}(n) \) for all \( l \) should be different.

The SIMO-ICA has the following advantage and disadvantage. (T1) This method is free from both the circular convolution effect and the invertibility of the separation filter matrix. (T2) Since the SIMO-ICA is based on TDICA which involves more complex calculations than FDICA, the convergence of the SIMO-ICA is very slow, and the sensitivity to the initial settings of separation filter matrices is very high.

### 5 Proposed Combination Technique of FDICA-PB and SIMO-ICA

As described above, two kinds of SIMO-model-based BSS methods have some disadvantages. However, note that the advantages and disadvantages of FDICA-PB and SIMO-ICA are mutually complementary, i.e., (F2) and (F3) can be resolved by (T1), and (T2) can be resolved by (F1). Therefore, we propose a new multistage technique combining FDICA-PB and SIMO-ICA.

The proposed multistage technique is conducted with the following steps (see Fig. 1). In the first step, we perform FDICA to separate the source signals to some extent with the fast- and robust-convergence advantage (F1). After the FDICA, we generate a specific initial value \( w_\text{ICA}(n) \) for SIMO-ICA performed in the next step by using \( W(f) \) obtained from FDICA. This procedure is given by

\[
w_\text{ICA}(n) = \text{IFFT} \left[ \text{diag} \left[ W(f)^{-1} P_t^T \right] P_t W(f) \right],
\]

where \( P_t \) are set to be, e.g., (8), and IFFT[.] represents an inverse DFT with the time shift of \( D/2 \) samples. In the final step, we perform SIMO-ICA (9) to obtain resultant SIMO components with the advantage (T1).

Compared with the simple SIMO-ICA, this combination algorithm is not so sensitive to the initial value of the separation filter because FDICA is used
for estimating the good initial value. Also, this technique has the possibility to provide a more accurate separation result over the simple FDICA because the resultant quality of the output signal is determined by the separation ability of the SIMO-ICA starting from the good initial state.

6 Experiments and Results

6.1 Conditions for Experiment

A two-element array with an interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, $-30^\circ$ and $40^\circ$. The distance between the microphone array and the loudspeakers is 1.15 m. Two kinds of sentences spoken by two male and two female speakers are used as the source speech samples. Using these sentences, we obtain 12 combinations. The sampling frequency is 8 kHz and the length of speech is limited to 7.5 seconds. To simulate the convolutive mixtures, the source signals are convolved with two kinds of impulse responses recorded in the experimental room which has a reverberation time (RT) of 150 ms or 300 ms. The length of the separation filter is set to be 2048 in both FDICA-PB and SIMO-ICA. The initial value in both the methods is null-beamformer whose directional null is steered to $\pm 45^\circ$.

As an objective evaluation score, SIMO-model accuracy (SA) [9] is used to indicate a degree of similarity (mean-squared-error) between the SIMO-model-based BSSs' outputs and the original SIMO-model-based signals ($\sum_{n=0}^{N-1} a_{kl}(n) s_1(n - t + D/2)$).

6.2 Comparison of Conventional Method and Proposed Method

Figure 2 and 3 show the results of SAs for FDICA-PB, SIMO-ICA, and the proposed combination technique in all speaker combinations, for each of reverberation conditions. In the results of the proposed combination technique, there exists a consistent improvement of SA compared with FDICA-PB as well as the simple SIMO-ICA. In RT = 150 ms, the average score of the improvement is 8.3 dB over SIMO-ICA, and is 2.9 dB over FDICA-PB. Also, in RT = 300 ms, the average score of the improvement is 5.1 dB over SIMO-ICA, and is 2.9 dB over FDICA-PB. From these results, we can conclude that the proposed combination technique can assist the SIMO-ICA in improving the separation performance, and successfully achieve the SIMO-model-based BSS under reverberant conditions.

6.3 Discussion on Combination Order

As described in the previous section, the combination of FDICA-PB and SIMO-ICA can contribute to the improvement of separation. In this combination, the advantage of FDICA-PB is useful in the initial step of separation procedure and the advantage of SIMO-ICA is also useful in the later step. Therefore we use FDICA-PB as the first-stage BSS and SIMO-ICA as the second-stage BSS. In order to confirm the availability of this combination order, we compare the proposed combination with another combination in which SIMO-ICA is used in
the first stage and FDICA-PB is used in the second stage (hereafter we designate this combination as "Swapped Combination").

The experiment of Swapped Combination was carried out in the following manner. Regarding SIMO-ICA part in Swapped Combination, parameters of SIMO-ICA part are the same as those of the simple SIMO-ICA in Sect. 6.2. Leaning of SIMO-ICA part is stopped at the peak of SA. As for FDICA-PB part in Swapped Combination, the analysis conditions and the parameter of FDPCA-PB are the same as those of the simple FDICA-PB given in Sect. 6.2.

We show the result of comparison of the simple SIMO-ICA, simple FDICA-PB, proposed combination, and Swapped Combination in Table 1. The result of Table 1 is the average of 12 experiments with different combinations of speakers. The average SA of 11.2 dB is obtained in Swapped Combination, and this performance is still better than that of simple SIMO-ICA and almost the same as that of simple FDICA-PB, but it is poorer than that of the proposed combination. In Swapped Combination, the SA is still improved by using FDICA-PB in the second stage, however, the separation performance is saturated because of the disadvantages (F2) and (F3) of FDICA-PB. This fact indicates that the proposed combination order (FDICA-PB in the first stage and SIMO-ICA in the second stage) is essential and the best.
Table 1. Comparison of SIMO-model accuracy among FDICA-PB, SIMO-ICA, proposed combination (Proposed), and Swapped Combination (Swapped) (unit is dB).

<table>
<thead>
<tr>
<th></th>
<th>SIMO-ICA</th>
<th>FDICA-PB</th>
<th>Proposed</th>
<th>Swapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT = 150 ms</td>
<td>11.3</td>
<td>16.7</td>
<td><strong>19.6</strong></td>
<td>17.1</td>
</tr>
<tr>
<td>RT = 300 ms</td>
<td>8.4</td>
<td>10.6</td>
<td><strong>13.4</strong></td>
<td>11.2</td>
</tr>
</tbody>
</table>

7 Conclusion

In this paper, first, the conventional FDICA-PB and the proposed SIMO-ICA were compared under a reverberant condition to evaluate the feasibility of SIMO-model-based BSS. Secondly, we proposed a new combination technique of FDICA-PB and SIMO-ICA to achieve the more higher separation performance compared with each of two methods. The experimental results revealed that the accuracy of the separated SIMO signals in the simple SIMO-ICA is inferior to that of FDICA-PB under low-quality initial value conditions, but the proposed combination technique of FDICA-PB and SIMO-ICA can outperform both simple FDICA-PB and SIMO-ICA. The average of the improvement was 8.3 dB over SIMO-ICA, and was 2.9 dB over FDICA-PB in RT = 150 ms, and 5.1 dB over SIMO-ICA, and was 2.9 dB over FDICA-PB in RT = 300 ms.

Acknowledgement. This work was partly supported by CREST Program “Advanced Media Technology for Everyday Living” of JST in Japan.

References