BLIND SOURCE SEPARATION BASED ON BINAURAL ICA

Tomoya TAKATANI, Tsuyoki NISHIKAWA, and Hiroshi SARUWATARI

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

ABSTRACT

We newly propose a novel blind separation framework for binaural acoustic signals based on the extended ICA algorithm, Binaural ICA (BICA). The BICA consists of multiple ICAs and fidelity controller, and each ICA runs in parallel under the control of the fidelity of the whole separation system. The BICA can separate the mixed signals into not monaural source signals but binaurally-heard signals of independent sources. Thus, the separated signals of BICA can maintain spatial qualities of each sound source. In order to evaluate its effectiveness, separation experiments are carried out under a reverberant condition. The experimental results reveal that (1) the signal separation performance of the proposed BICA is the same as that of the conventional ICA-based method, and (2) the spatial quality of the separated sound in BICA is remarkably superior to that of the conventional method, especially for the fidelity of the sound reproduction.

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. This technique can be applicable to the high-quality hands-free telecommunication systems. In the recent works for the BSS based on the independent component analysis (ICA) [1], various methods have been proposed to deal with a separation of acoustical sounds which correspond to convolutive mixture case [2, 3, 4]. However, the conventional ICA-based BSS approaches are basically for extracting each of independent sound sources as a *monaural* signal, and consequently they have a serious drawback that the separated sounds cannot maintain the information about directivity, localization, and any spatial qualities of each sound source. This prevents any BSS methods from being applied to the binaural signal processing and high-fidelity sound reproduction system.

In this paper, we propose a new blind separation framework for binaural acoustic signals based on the extended ICA algorithm, Binaural ICA (BICA). In the scenario of BICA, the unknown multiple source signals which are mixed through unknown acoustical transmission channels are observed at the microphones, and these signals can be separated into not monaural source signals but binaurally-heard signals of independent sources. Thus, the separated signals of BICA can maintain the spatial quality of each sound source.

In order to evaluate its effectiveness, separation experiments are carried out under a reverberant condition. The experimental results reveal that (1) the signal separation performance of the proposed BICA is the same as that of the conventional ICA, and (2) the sound quality of the separated signals in BICA is remarkably superior to that of the conventional ICA, especially for the spatial quality and the fidelity of the sound reproduction.

2. MIXING PROCESS AND CONVENTIONAL BSS

2.1. Mixing process

In this study, the number of array elements (microphones) is K and the number of multiple sound sources is L, where we deal with the case of $K = L = 2$. In general, the observed signals in which multiple source signals are mixed linearly are expressed as the following equations:

$$\boldsymbol{x}(t) = \sum_{n=0}^{N-1} \boldsymbol{a}(n)\boldsymbol{s}(t-n) = \boldsymbol{A}(z)\boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{s}(t)$ is the source signal vector, $\boldsymbol{x}(t)$ is the observed signal vector, $\boldsymbol{a}(n)$ is the mixing matrix with the length of N, and $\boldsymbol{A}(z)$ is the z-transform of $\boldsymbol{a}(n)$; these are given as

$$\boldsymbol{s}(t) = [s_1(t), \cdots, s_L(t)]^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{x}(t) = [x_1(t), \cdots, x_K(t)]^{\mathrm{T}}, \tag{3}$$

$$\boldsymbol{a}(n) = \begin{bmatrix} a_{11}(n) & \cdots & a_{1L}(n) \\ \vdots & \ddots & \vdots \\ a_{K1}(n) & \cdots & a_{KL}(n) \end{bmatrix}, \tag{4}$$

$$\boldsymbol{A}(z) = \sum_{n=0}^{N-1} \boldsymbol{a}(n)z^{-n} = \left[\sum_{n=0}^{N-1} a_{ij}(n)z^{-n}\right]_{ij}, \tag{5}$$

where z^{-1} is used as the unit-delay operator, i.e., $z^{-n} \cdot x(t) = x(t-n)$, a_{kl} is the impulse response between k-th microphone and l-th sound source, and $[X]_{ij}$ denotes the matrix which includes the element X in the i-th row and the j-th column.

2.2. Conventional ICA-based BSS method

As the BSS method, we consider the time-domain ICA (TDICA), in which each element of the separation matrix is represented as an FIR filter. In the TDICA, we optimize the separation matrix by only using the fullband observed signals without subband processing (see Fig. 1). The separated signal vector $\boldsymbol{y}(t) = [y_1(t), \cdots, y_L(t)]^{\mathrm{T}}$ is expressed as the following equation:

$$\begin{aligned} \boldsymbol{y}(t) &= \sum_{n=0}^{D-1} \boldsymbol{w}(n)\boldsymbol{x}(t-n) = \boldsymbol{W}(z)\boldsymbol{x}(t) \\ &= \boldsymbol{W}(z)\boldsymbol{A}(z)\boldsymbol{s}(t), \end{aligned} \tag{6}$$

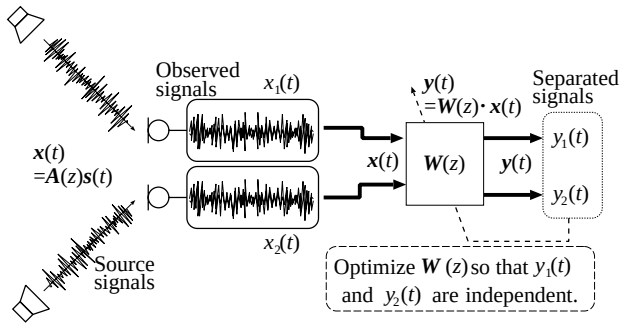


Fig. 1. Configuration of conventional TDICA.

where $\boldsymbol{w}(n)$ is the separation matrix, $\boldsymbol{W}(z)$ is the z-transform of $\boldsymbol{w}(n)$, and D is the filter length of $\boldsymbol{w}(n)$. In our study, separation matrix is optimized by minimizing Kullbuck-Leibler divergence between the joint probability density function (PDF) of $\boldsymbol{y}(t)$ and the product of marginal PDFs of $y_l(t)$. The iterative learning rule is given by [4]

$$\begin{aligned}\boldsymbol{w}^{[j+1]}(n) &= \boldsymbol{w}^{[j]}(n) \\ &\quad - \alpha \sum_{d=0}^{D-1} \left[\{\text{off-diag}\ \langle \boldsymbol{\varphi}(\boldsymbol{y}^{[j]}(t))\boldsymbol{y}^{[j]}(t-n+d)^{\mathrm{T}}\rangle_t\}\right. \\ &\quad \left. \cdot\, \boldsymbol{w}^{[j]}(d)\right], \end{aligned} \tag{7}$$

where α is the step-size parameter, the superscript $[j]$ is used to express the value of the j-th step in the iterations, $\langle\cdot\rangle_t$ denotes the time-averaging operator, and off-diag$\boldsymbol{W}(z)$ is the operation to set every diagonal element of matrix $W(z)$ to be zero. Also, we define the nonlinear vector function $\boldsymbol{\varphi}(\cdot)$ as

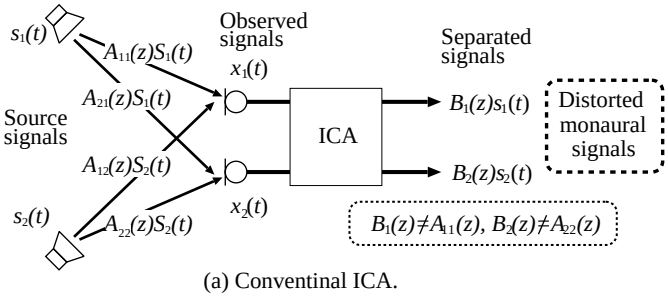
$$\boldsymbol{\varphi}(\boldsymbol{y}(t)) = [\tanh(y_1(t)), \cdots, \tanh(y_L(t))]^{\mathrm{T}}. \tag{8}$$

2.3. Problems in conventional ICA

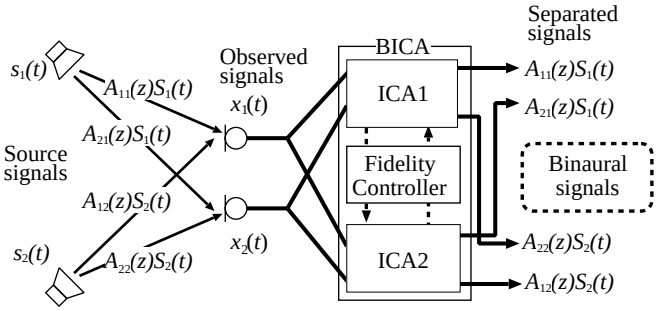
The conventional ICA is basically for extracting each of independent sound sources as a monaural signal. In addition, the quality of the separated sound cannot be guaranteed, i.e., the separated signals are possible to include any spectral distortions because the modified separated signals convolved with arbitrary linear filters are still mutually independent (see Fig. 2(a)). Therefore, the conventional ICA has a serious drawback that the separated sounds cannot maintain the information about directivity, localization, and any spatial qualities of each sound source. To resolve the problem only on the sound quality, modified ICA based on Minimal Distortion Principle has been proposed by Matsuoka et al. [5]. However, this method is valid for only monaural outputs, and the fidelity of the output signals as binaural sounds cannot be guaranteed.

3. PROPOSED ALGORITHM; BINAURAL ICA

In order to resolve the above-mentioned problems essentially, we propose a new blind separation method for binaural acoustic signals based on BICA. The BICA consists of multiple ICA parts and Fidelity Controller, and each ICA runs in parallel under the control



(a) Conventinal ICA.



(b) Proposed BICA.

Fig. 2. Input and output relations in (a) conventional ICA and (b) proposed BICA.

of the fidelity of the whole separation system (see Fig. 2(b)). The separated signals of BICA are defined as the following equations:

$$\begin{aligned}\boldsymbol{y}_{\mathrm{ICA1}}(t) &= [y_1^{(1)}(t),\ y_2^{(2)}(t)]^{\mathrm{T}} = \sum_{n=0}^{D-1} \boldsymbol{w}_{\mathrm{ICA1}}(n)\boldsymbol{x}(t-n) \\ &= \boldsymbol{W}_{\mathrm{ICA1}}(z)\boldsymbol{x}(t), \end{aligned} \tag{9}$$

$$\begin{aligned}\boldsymbol{y}_{\mathrm{ICA2}}(t) &= [y_2^{(1)}(t),\ y_1^{(2)}(t)]^{\mathrm{T}} = \sum_{n=0}^{D-1} \boldsymbol{w}_{\mathrm{ICA2}}(n)\boldsymbol{x}(t-n) \\ &= \boldsymbol{W}_{\mathrm{ICA2}}(z)\boldsymbol{x}(t), \end{aligned} \tag{10}$$

where $y_m^{(k)}(t)$ is the separated signal which extracts the source signal $s_m(t)$ from the observed signal $x_k(t)$. In this case, $y_1^{(1)}(t)$ and $y_1^{(2)}(t)$ are regarded as the binaural components which correspond to $s_1(t)$, and $y_2^{(1)}(t)$ and $y_2^{(2)}(t)$ are regarded as the binaural components which correspond to $s_2(t)$.

As for the fidelity controller, we newly introduce the following cost function to be minimized,

$$\mathrm{E}\left[\| \boldsymbol{y}_{\mathrm{ICA1}}(t) + \boldsymbol{y}_{\mathrm{ICA2}}(t) - \boldsymbol{x}(t - D/2) \|^2\right], \tag{11}$$

where $\| \boldsymbol{x} \|$ is Euclidean norm of vector $\boldsymbol{x}$. If we obtain the independent sound sources from Eqs. (9) and (10), and simultaneously minimize the Eq. (11) to be zero, then we can obtain the appropriate separated signals maintaining their binaural properties. To achieve this, the natural gradient [4] of the cost function with respect to $\boldsymbol{w}_{\mathrm{ICA1}}(n)$ and $\boldsymbol{w}_{\mathrm{ICA2}}(n)$ should be added in the iterative learning rule of separation filter given by Eq. (7); thus the new iterative algorithm of BICA is given by

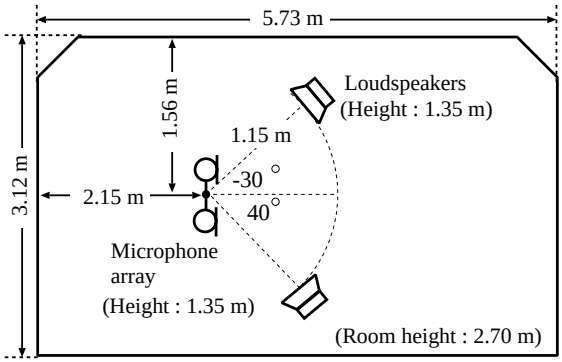


Fig. 3. Layout of reverberant room used in experiments.

$$
\begin{aligned}
&\boldsymbol{w}_{\mathrm{ICA1}}^{[j+1]}(n) \\
&= \boldsymbol{w}_{\mathrm{ICA1}}^{[j]}(n) \\
&\quad - \alpha \sum_{d=0}^{D-1} \big[\{\text{off-diag } \langle \boldsymbol{\varphi}(\boldsymbol{y}_{\mathrm{ICA1}}^{[j]}(t))\boldsymbol{y}_{\mathrm{ICA1}}^{[j]}(t-n+d)^{\mathrm{T}}\rangle_t \\
&\quad + \beta\langle(\boldsymbol{y}_{\mathrm{ICA1}}^{[j]}(t) + \boldsymbol{y}_{\mathrm{ICA2}}^{[j]}(t) - \boldsymbol{x}(t-D/2)) \\
&\quad \cdot \boldsymbol{y}_{\mathrm{ICA1}}^{[j]}(t-n+d)^{\mathrm{T}}\rangle_t\} \cdot \boldsymbol{w}_{\mathrm{ICA1}}^{[j]}(d)\big], \qquad (12) \\
&\boldsymbol{w}_{\mathrm{ICA2}}^{[j+1]}(n) \\
&= \boldsymbol{w}_{\mathrm{ICA2}}^{[j]}(n) \\
&\quad - \alpha \sum_{d=0}^{D-1} \big[\{\text{off-diag } \langle \boldsymbol{\varphi}(\boldsymbol{y}_{\mathrm{ICA2}}^{[j]}(t))\boldsymbol{y}_{\mathrm{ICA2}}^{[j]}(t-n+d)^{\mathrm{T}}\rangle_t \\
&\quad + \beta\langle(\boldsymbol{y}_{\mathrm{ICA1}}^{[j]}(t) + \boldsymbol{y}_{\mathrm{ICA2}}^{[j]}(t) - \boldsymbol{x}(t-D/2)) \\
&\quad \cdot \boldsymbol{y}_{\mathrm{ICA2}}^{[j]}(t-n+d)^{\mathrm{T}}\rangle_t\} \cdot \boldsymbol{w}_{\mathrm{ICA2}}^{[j]}(d)\big], \qquad (13)
\end{aligned}
$$

where α and β are the step-size parameters; α is for the control of the total update quantity and β is for the fidelity control.

4. EXPERIMENT

4.1. Conditions for experiment

As the preliminary study on the proposed BICA, we carried out the source separation experiment using a simple microphone array, neglecting the effect of the head-related transfer function (HRTF). A two-element array with interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, -30° and 40° (see Fig. 3). The distance between microphone array and loudspeaker is 1.15 m. 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. Using these sentences, we obtain 6 combinations. The sampling frequency is 8 kHz and the length of speech is limited in 3 seconds. The reverberation time of the impulse responses recorded in the experimental room is 150 ms. The step-size parameter α is changed from 5.0×10^{-8} to 1.0×10^{-6} and β is changed from 2.0×10^{-10} to 4.9×10^{-10} to search the optima which minimize Eq. (11). The length of $\boldsymbol{w}(n)$ is 512, and the initial value is Null-Beamformer [3] whose directional null is steered to $\pm 60^\circ$. The number of iterations in ICA is 5000. As for the conventional ICA for comparison, we used Eqs. (12) and (13) in the case of $\beta = 0$.

4.2. Objective evaluation score

In this experiment, three objective evaluation scores are defined. First, *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 of separation performance, where we don't care the distortion of the separated signal. The SNRs are calculated under the assumption that the suppressed speech signal is regarded as noise. The NRR is defined as

$$
\begin{aligned}
\mathrm{NRR} &\equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{k=1}^{2}\big(\mathrm{OSNR}_l^{(\mathrm{ICA}k)} - \mathrm{ISNR}_l^{(\mathrm{ICA}k)}\big), \qquad (14) \\
\mathrm{OSNR}_l^{(\mathrm{ICA1})} &= 10\log_{10}\frac{\sum_t \mid H_{ll}^{\mathrm{ICA1}}(z)S_l(t)\mid^2}{\sum_t \mid H_{ln}^{\mathrm{ICA1}}(z)S_n(t)\mid^2}, \\
\mathrm{ISNR}_l^{(\mathrm{ICA1})} &= 10\log_{10}\frac{\sum_t \mid A_{ll}(z)S_l(t)\mid^2}{\sum_t \mid A_{ln}(z)S_n(t)\mid^2}, \\
\mathrm{OSNR}_l^{(\mathrm{ICA2})} &= 10\log_{10}\frac{\sum_t \mid H_{ln}^{\mathrm{ICA2}}(z)S_n(t)\mid^2}{\sum_t \mid H_{ll}^{\mathrm{ICA2}}(z)S_l(t)\mid^2}, \\
\mathrm{ISNR}_l^{(\mathrm{ICA2})} &= 10\log_{10}\frac{\sum_t \mid A_{ln}(z)S_n(t)\mid^2}{\sum_t \mid A_{ll}(z)S_l(t)\mid^2},
\end{aligned}
$$

where $\mathrm{OSNR}_l^{(\mathrm{ICA}k)}$ and $\mathrm{ISNR}_l^{(\mathrm{ICA}k)}$ are the output SNR and the input SNR for ICAk, respectively, and $l \neq n$. Also, $H_{ij}^{\mathrm{ICA}k}(z)$ is the element in the i-th row and the j-th column of the matrix $\boldsymbol{H}^{\mathrm{ICA}k}(z) = \boldsymbol{W}_{\mathrm{ICA}k}(z)\boldsymbol{A}(z)$. Secondly, in order to evaluate the sound quality of the separated signal, the *Sound Quality* (SQ) is defined as the following equation.

$$
\begin{aligned}
\mathrm{SQ} &\equiv \frac{1}{4}\sum_{l=1}^{2}\sum_{n=1}^{2}\mathrm{SQ}_{y_l^{(n)}}, \qquad (15) \\
\mathrm{SQ}_{y_1^{(1)}} &= 10\log_{10}\frac{\sum_t \mid A_{11}(z)S_1(t)\mid^2}{\sum_t \mid A_{11}(z)S_1(t) - H_{11}^{\mathrm{ICA1}}(z)S_1(t)\mid^2}, \\
\mathrm{SQ}_{y_2^{(1)}} &= 10\log_{10}\frac{\sum_t \mid A_{12}(z)S_2(t)\mid^2}{\sum_t \mid A_{12}(z)S_2(t) - H_{12}^{\mathrm{ICA2}}(z)S_2(t)\mid^2}, \\
\mathrm{SQ}_{y_1^{(2)}} &= 10\log_{10}\frac{\sum_t \mid A_{21}(z)S_1(t)\mid^2}{\sum_t \mid A_{21}(z)S_1(t) - H_{21}^{\mathrm{ICA2}}(z)S_1(t)\mid^2}, \\
\mathrm{SQ}_{y_2^{(2)}} &= 10\log_{10}\frac{\sum_t \mid A_{22}(z)S_2(t)\mid^2}{\sum_t \mid A_{22}(z)S_2(t) - H_{22}^{\mathrm{ICA1}}(z)S_2(t)\mid^2},
\end{aligned}
$$

where $\mathrm{SQ}_{y_l^{(n)}}$ is the sound quality of separated signal $y_l^{(n)}$. The last evaluation score is *Fidelity* (F). It is defined as the following equation,

$$
\mathrm{F} = \frac{\mathrm{E}[\| \boldsymbol{x}(t) \|^2]}{\mathrm{E}[\|\boldsymbol{y}_{\mathrm{ICA1}}(t) + \boldsymbol{y}_{\mathrm{ICA2}}(t) - \boldsymbol{x}(t-D/2) \|^2]}. \qquad (16)
$$

4.3. Results and discussion

Figure 4 (a) shows the result of NRR for different speaker combination. The bars on the right of this figure correspond to the averaged results of them. In the averaged scores, the deterioration of NRR in BICA is 0.2 dB compared with the conventional ICA. From this results, it is revealed that the signal separation performance of the proposed BICA is almost the same as that of the conventional ICA-based method.

On the other hand, Figs. 4 (b) and (c) show the results of SQ and F for different speaker combination. The bars on the right of

each figure correspond to the averaged results of them. In the averaged scores, compared with the conventional ICA, the improvement of SQ is 3.3 dB, and the improvement of F is 31.8 dB. From these results, it is revealed that the sound quality of the separated signals in BICA is remarkably superior to that of the conventional method, especially for the spatial quality and the fidelity of the sound reproduction.

The whole of the results indicates the followings. (1) In BICA, the addition of fidelity controller is effective to compensate the spatial quality of the separated binaural signals. (2) There is no deterioration in the separation performance (NRR) even with the additional compensation of sound quality in BICA. Therefore, we can conclude that the proposed BICA can be applicable to the binaural signal processing and high-fidelity sound reproduction system.

5. CONCLUSION

We newly propose a novel blind separation framework for binaural acoustic signals based on the extended ICA algorithm, Binaural ICA (BICA). BICA is the algorithm to separate the mixed signals into not monaural source signals but binaurally-heard signals of independent sources without loss of their spatial qualities. In order to evaluate its effectiveness, separation experiments are carried out using 2 microphones and 2 sources under the condition that the reverberation time is set to be 150 ms. The experimental results reveal that (1) the signal separation performance of the proposed BICA is the same as that of the conventional ICA-based method, and (2) the spatial quality of the separated sound in BICA is remarkably superior to that of the conventional method, especially for the fidelity of the sound reproduction. Therefore, we can conclude that the proposed BICA can be applicable to the binaural signal processing and high-fidelity sound reproduction system. The further experiment with HRTF is an open problem.

6. ACKNOWLEDGEMENT

The authors are grateful to Dr. Makino, Miss Araki of NTT. CO., LTD, and Dr. Matsuoka of Kyusyu Institute of Technology for their discussions. This work was partly supported by CREST (Core Research for Evolutional Science and Technology) in Japan.

7. REFERENCES

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

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol.22,pp.21–34,1998.

[3] 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.

[4] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*,pp.371–376, 1999.

[5] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation", *Proc. International Conference on independent Component Analysis and Blind Signal Separation*, pp.722–727, Dec. 2001.

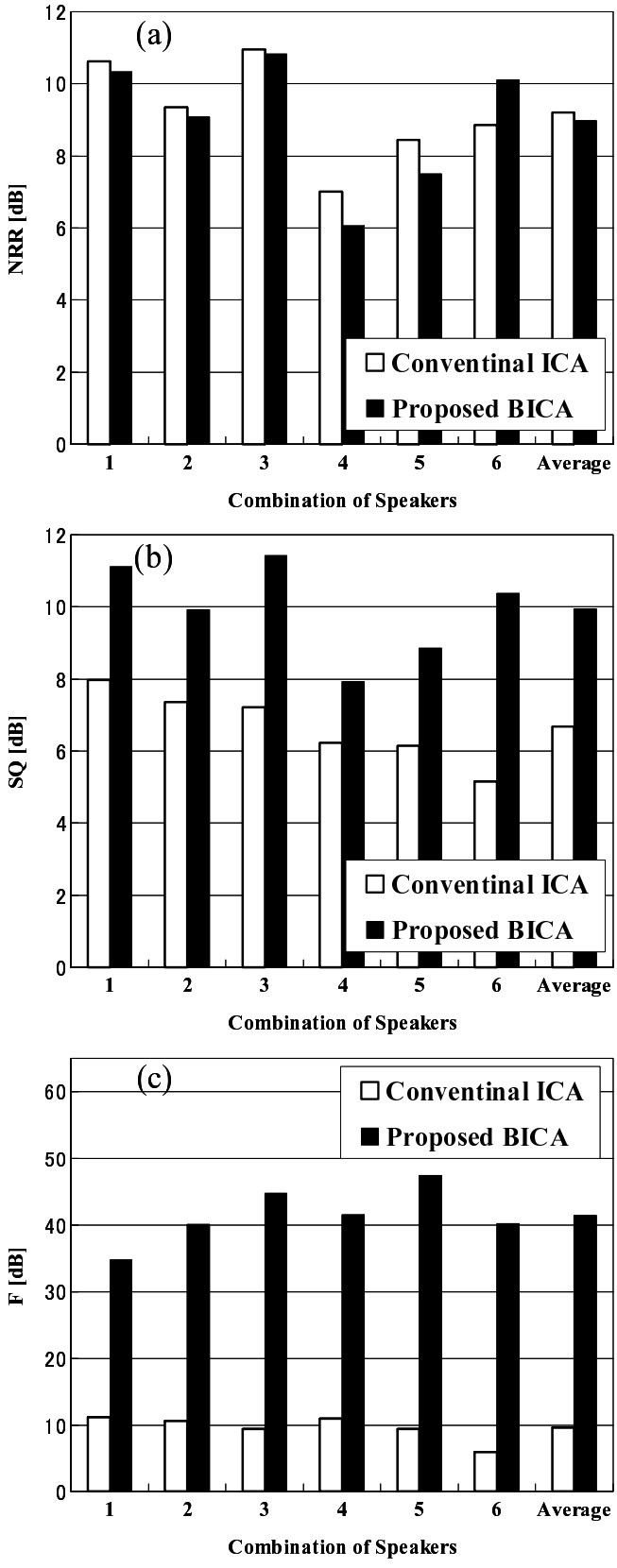


Fig. 4. Results of (a) NRR, (b) SQ, and (c) F.