EFFICIENT BLIND SOURCE SEPARATION COMBINING CLOSED-FORM SECOND-ORDER ICA AND NONCLOSED-FORM HIGHER-ORDER ICA

Kentaro Tachibana, Hiroshi Saruwatari, Yoshimitsu Mori, Shigeki Miyabe, Kiyohiro Shikano[†]*, Akira Tanaka*[‡]

[†]Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, JAPAN
[‡]Hokkaido University, Kita-14, Nishi-9, Kita-ku, Sapporo, 060-0814, JAPAN

ABSTRACT

In this paper, first, we propose a computational-cost efficient blind source separation combining closed-form 2nd-order independent component analysis (ICA) and nonclosed-form higher-order ICA. The closed-form solution of the 2nd-order ICA has been recently presented by one of the authors. This finding motivates us to combine the closed-form 2nd-order ICA and higher-order ICA, where the preceding closed-form ICA produces a good initial value and the following higher-order ICA updates the separation filters from the advantageous status. Secondly, we utilize the proposed architecture to address an essential question that which type of statistics is more beneficial to ICA among non-stationarity and non-Gaussianity. This can be conducted owing to the attractive property that the closed-form ICA can provide a good estimate of the theoretical upper limitation of the separation performance among 2nd-order ICAs without suffering from poor-convergence problems. Experimental results reveal that the non-Gaussianity-based ICA can outperform the non-stationarity-based ICA.

Index Terms— Separation, speech enhancement, acoustic arrays, acoustic signal processing, adaptive signal processing

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. Basically BSS is classified into unsupervised filtering technique, and in that the source-separation procedure requires no training sequences and no a priori information on the directions-of-arrival of the sound sources. Owing to the attractive features of BSS, much attention has been paid to BSS in many fields of signal processing such as speech enhancement.

In recent researches of BSS based on independent component analysis (ICA), various methods have been proposed to tackle acoustic-sound separation [1]–[6] which is referred to as convolutive mixing problem. This paper also addresses the BSS problem under reverberant conditions which often arise in many practical audio applications. Generally speaking, almost all the algorithms in ICA, e.g., 2nd-order ICA [2, 3, 5, 6] and higher-order ICA [1, 4] are conducted through nonclosed-form, in other words, *iterative*, optimization, where the separation filters are improved along with the gradient of an appropriate cost function. However, this property often leads to the difficult problem of the poor and slow convergence [7]. In addition, the latency in the convergence prevents ICA-based BSS from being applicable to real-time processing.

In this paper, first, we newly propose an efficient BSS method combining closed-form 2nd-order ICA and nonclosed-form higher-order ICA. The closed-form solution of the 2nd-order ICA has been recently presented by one of the authors [8]. This mathematical contribution yields an idea of combining the closed-form 2nd-order ICA and the higher-order ICA, where the preceding closed-form ICA can produce a good initial value and the following higher-order ICA can update the separation filters from the advantageous status.

We thank KOBE STEEL LTD. for funding and collaboration.

Secondly, based on the above-mentioned structure, we address an essential question that which cost function is better among non-stationarity (on 2nd-order ICA) and non-Gaussianity (on higher-order ICA). This can be conducted using the proposed method's attractive property that the closed-form ICA approximately shows the theoretical upper limitation of the separation performance among 2nd-order ICAs without suffering from poor-convergence problems. The evaluation of the separation performance in the proposed combination easily indicates the winner of non-stationarity vs. non-Gaussianity in ICA.

2. MIXING PROCESS AND CONVENTIONAL ICA

In this study, the number of microphones is K and the number of multiple sound sources is L, where we deal with the case of $K = L$.

Multiple mixed signals are observed at the microphone array, and these signals are converted into discrete-time series via an A/D converter. By applying the short-time discrete-time Fourier transform framewisely, we can express the observed signals, in which multiple source signals are linearly mixed, as follows in the time-frequency domain:

$$\boldsymbol{x}(f,t) = \boldsymbol{A}(f)\boldsymbol{s}(f,t), \tag{1}$$

where $\boldsymbol{x}(f,t) = [x_1(f,t), \cdots, x_K(f,t)]^{\mathrm{T}}$ is the observed signal vector, and $\boldsymbol{s}(f,t) = [s_1(f,t), \cdots, s_L(f,t)]^{\mathrm{T}}$ is the source signal vector. Also, $\boldsymbol{A}(f)$ is the mixing matrix which is complex-valued because we introduce a model to deal with the relative time delays among the microphones and room reverberations.

Next, we perform signal separation using the complex-valued unmixing matrix $\boldsymbol{W}(f)$, so that the L time-series output $\boldsymbol{y}(f,t) = [y_1(f,t), \cdots, y_L(f,t)]^{\mathrm{T}}$ becomes mutually independent; this procedure can be given as

$$\boldsymbol{y}(f,t) = \boldsymbol{W}(f)\boldsymbol{x}(f,t). \tag{2}$$

We perform this procedure with respect to all frequency bins.

The optimal $\boldsymbol{W}(f)$ is obtained by many types of ICAs, where several cost functions are used to measure the independence among sources. The most popular statistics used in the cost functions are *non-stationarity* and *non-Gaussianity*. For example, the conventional 2nd-order ICA utilizes non-stationarity of sources. The optimization can be achieved by minimizing, e.g., the following function

[3]:

$$J_{\mathrm{SO}}(\boldsymbol{W}(f)) = \sum_i ||\boldsymbol{W}(f)\boldsymbol{R}_{t_i}(f)\boldsymbol{W}(f)^{\mathrm{H}} - \mathrm{diag}[\boldsymbol{W}(f)\boldsymbol{R}_{t_i}(f)\boldsymbol{W}(f)^{\mathrm{H}}]||^2, \quad (3)$$

where superscript H represents a conjugate transposition, $\boldsymbol{R}_{t_i}(f)$ ($i = 1, 2, ...$) are the cross-correlation matrices of the input $\boldsymbol{x}(f,t)$, which are calculated around the multiple time indices t_i, and $\mathrm{diag}[\cdot]$ is the operation for setting every off-diagonal element to zero. The minimization of $J_{\mathrm{SO}}(\boldsymbol{W}(f))$ yields simultaneous diagonalization (decorrelation) of the correlation matrix of $\boldsymbol{y}(f,t)$.

In the typical higher-order ICA, Kullback-Leibler divergence between the joint probability density function (PDF) of $\boldsymbol{y}(f,t)$ and the product of marginal PDFs of $y_l(f,t)$ is used for the cost function to be minimized as

$$J_{\mathrm{HO}}(\boldsymbol{W}(f)) = \int p(\boldsymbol{y}(f,t)) \log \frac{p(\boldsymbol{y}(f,t))}{\prod_{l=1}^{L}\prod_{t=0}^{T-1} p(y_l(f,t))} d\boldsymbol{y}(f,t), \quad (4)$$

where $p(y_l(f,t))$ is the marginal PDF of $y_l(f,t)$, $p(\boldsymbol{y}(f,t))$ is the joint PDF of $\boldsymbol{y}(f,t)$. This cost function is highly relevant to higher-order statistics of the sources and non-Gaussianity.

In general, for both 2nd- and higher-order ICAs, the optimization procedures can be conducted via nonclosed-form (i.e., iterative) optimization in which $\boldsymbol{W}(f)$ is updated along with the negative direction of gradient of $J_{\mathrm{SO}}(\boldsymbol{W}(f))$ or $J_{\mathrm{HO}}(\boldsymbol{W}(f))$. Therefore it has an inherent disadvantage in that there is difficulty with the poor and slow convergence of nonlinear optimization, particularly when we are confronted with very complex convolutive mixtures and unfortunately set a bad initial value. Furthermore, ordinary ICA-based BSS algorithms require huge computational complexities. The disadvantages reduce the applicability of the approach to the general audio applications which often need real-time processing.

3. PROPOSED METHOD

3.1. Motivation

In a previous study, closed-form solution of the 2nd-order ICA was proposed by one of the authors [8], who showed that simple algebraic calculations enable the separation of mixed signals without iterative filter updating. This finding has motivated us to combine the closed-form 2nd-order ICA and higher-order ICA, where the computational cost is considerably reduced (see Sect. 3.4). Moreover, our strategy provides a good tool for an insight into the essential question that which cost function is better among non-stationarity and non-Gaussianity (see Sect. 3.5). Hereinafter we describe the detailed algorithm.

3.2. First stage: closed-form 2nd-order ICA

In the original reference [8], the principle of the closed-form 2nd-order ICA was derived, especially from the mathematical point of view. This subsection briefly describes the overview of signal processing in the closed-form ICA. The strict proofs of the theorem will be omitted due to the limitation of the current manuscript's space.

First, we obtain the correlation matrices with different time points as

$$\boldsymbol{R}_{t_i}(f) = \langle \boldsymbol{x}(f,t)\boldsymbol{x}(f,t)^{\mathrm{H}} \rangle_{t \in t_i}, \quad (5)$$

where $\langle\cdot\rangle_{t\in t_i}$ denotes the time-averaging operator over specific time duration t_i, and $i = 1, 2, ...$ represent indices of time-averaging block.

Next, we apply the singular value decomposition (SVD) to a superposition of $\boldsymbol{R}_{t_i}(f)$, which is represented as

$$\sum_i \boldsymbol{R}_{t_i}(f) = \boldsymbol{U}(f)\mathrm{diag}(\lambda_1, \lambda_2, ...)\boldsymbol{U}(f)^{\mathrm{H}}, \quad (6)$$

where λ_k are the eigenvalues, $\mathrm{diag}(\lambda_1, ...)$ denotes the diagonal matrix which includes the eigenvalues, and $\boldsymbol{U}(f)$ is the matrix consisting of the eigenvectors. Then we obtain a full-rank decomposition for pseudo-inverse of $\sum_i \boldsymbol{R}_{t_i}(f)$ as follows

$$\Big[\sum_i \boldsymbol{R}_{t_i}(f)\Big]^{+} = \boldsymbol{L}(f)\boldsymbol{L}(f)^{\mathrm{H}}, \quad (7)$$

$$\boldsymbol{L}(f) = \boldsymbol{U}(f)\mathrm{diag}(1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, ...). \quad (8)$$

It can be proved [8] that if the covariance of the sources $\boldsymbol{s}(f,t)$ in t_i is negligible, every $\boldsymbol{L}(f)^{\mathrm{H}}\boldsymbol{R}_{t_i}(f)\boldsymbol{L}(f)$ for any i shares the same eigenvectors, and this is given via SVD form as

$$\boldsymbol{L}(f)^{\mathrm{H}}\boldsymbol{R}_{t_i}(f)\boldsymbol{L}(f) = \boldsymbol{T}(f)\mathrm{diag}(\sigma_1(t_i), \sigma_2(t_i), ...)\boldsymbol{T}(f)^{\mathrm{H}}, \quad (9)$$

where $\sigma_k(t_i)$ are the eigenvalues for a specific time block t_i, and $\boldsymbol{T}(f)$ denotes the matrix consisting of shared eigenvectors which are *independent* of time-block index i. Therefore, for any i, the simultaneous diagonalization of $\boldsymbol{R}_{t_i}(f)$ can be achieved as follows;

$$\boldsymbol{T}(f)^{\mathrm{H}}\boldsymbol{L}(f)^{\mathrm{H}}\boldsymbol{R}_{t_i}(f)\boldsymbol{L}(f)\boldsymbol{T}(f) = \mathrm{diag}(\sigma_1(t_i), \sigma_2(t_i), ...), \quad (10)$$

and this means that the optimal separation filter matrix in the 2nd-order sense is given by

$$\boldsymbol{W}_{\mathrm{SO}}(f) = (\boldsymbol{L}(f)\boldsymbol{T}(f))^{\mathrm{H}}. \quad (11)$$

Note that, for the calculation of $\boldsymbol{T}(f)$ in Eq. (9), it is sufficient for us to only apply a single SVD to an *arbitrary* single time-block t_i because of the eigenvector-sharing property.

It is worth mentioning that Molgedey et al. have shown the closed-form solution only for the case that the number of correlation matrix blocks is up to 2 [9]. In contrast, the algorithm [8] used in the proposed method is the first generalized closed-form solution which can be applicable even to the case of $i > 2$.

3.3. Second stage: nonclosed-form higher-order ICA

The separation filter matrix $\boldsymbol{W}_{\mathrm{SO}}(f)$ obtained by 2nd-order ICA often provides insufficient source-separation performance. To polish up the separation filter matrix and gain the further performance, we propose to combine the nonclosed-form higher-order ICA after the 2nd-order ICA. This strategy regards the separation filter matrix $\boldsymbol{W}_{\mathrm{SO}}(f)$ as an initial value for higher-order ICA's iterative learning. The higher-order ICA is conducted by the following manner;

$$\boldsymbol{W}^{[0]}(f) = \boldsymbol{W}_{\mathrm{SO}}(f), \quad (12)$$

$$\boldsymbol{W}^{[j+1]}(f) = \eta\left[\boldsymbol{I} - \left\langle \boldsymbol{\Phi}(\boldsymbol{y}(f,t))\boldsymbol{y}^{\mathrm{H}}(f,t)\right\rangle_t\right]\boldsymbol{W}^{[j]}(f) + \boldsymbol{W}^{[j]}(f), \quad (13)$$

where superscript $[j]$ represents the number of iterations, $\boldsymbol{I}$ is the identity matrix, $\langle\cdot\rangle_t$ denotes the time-averaging operator over whole time indices, and $\boldsymbol{\Phi}(\cdot)$ is the appropriate nonlinear vector function, e.g., [**?**, 10] (we use [10] in this paper).

In general, the higher-order ICA suffers from an problem of the poor and slow convergence of nonlinear optimization. In the proposed method, however, the preceding closed-form 2nd-order ICA can give a better initial state for the higher-order ICA, and the proposed combination mitigates the drawbacks on the poor convergence.

3.4. Computational-cost efficiency of proposed method

In the first stage, the closed-form 2nd-order ICA mainly requires the following computations.

Calculation of correlation matrices: The computations for obtaining $\boldsymbol{R}_{t_i}(f)$ result in, e.g., more than hundred multiplications-accumulations to deal with the observed signal of several seconds.

Calculation of $\boldsymbol{L}(f)$: To obtain $\boldsymbol{L}(f)$ in Eq. (8), a single SVD should be performed as in Eq. (6), where the computational load is $O(K^3)$ (K corresponds to the dimension of $\boldsymbol{L}(f)$).

Calculation of $\boldsymbol{T}(f)$: The matrix $\boldsymbol{T}(f)$ in Eq. (11) needs one more SVD in Eq. (9) with the computational load of $O(K^3)$.

In summary, overall amount of computations in the closed-form 2nd-order ICA approximately depends on the cost of obtaining $\boldsymbol{R}_{t_i}(f)$ because the calculations of $\boldsymbol{L}(f)$ and $\boldsymbol{T}(f)$ are relatively negligible when K is small, e.g., 2 or 3. In addition, it should be mentioned that the whole computations in the closed-form solution are almost the same as *those for 1 or 2 iterations in the higher-order ICA*, and thus almost all the computational resources can be dedicated to the higher-order ICA part in the second stage. Furthermore, the computational complexities can be totally reduced because the good initialization by the closed-form ICA saves the number of iterations in the following higher-order ICA's updating.

3.5. As a judging tool for non-stationarity vs. non-Gaussianity

Another contribution of the closed-form 2nd-order ICA is concerned with a comparison on non-stationarity and non-Gaussianity. Owing to Eq. (10), $\boldsymbol{W}_{\rm SO}(f)$ given by Eq. (11) can diagonalize each correlation matrix when the covariance of $\boldsymbol{s}(f,t)$ in t_i is negligible. Consequently under such a condition, the cost function defined by Eq. (3) is minimized to be zero, i.e, the following relation holds;

$$J_{\rm so}(\boldsymbol{W}_{\rm SO}(f)) \quad = \quad 0. \tag{14}$$

This remains us that the closed-form solution Eq. (11) gives a good estimate of the *theoretical upper limitation* of the separation performance among the 2nd-order ICAs based on source non-stationarity. Note that there are no affections from poor-convergence and local-minimum problems which often arise in the conventional nonclosed-form (iterative) method. Therefore, by seeing the results of the first stage and the possible performance increase/decrease by the second stage, we can put a period to the discussion on non-stationarity vs. non-Gaussianity in ICA, i.e., the increase implies the superiority of non-Gaussianity.

4. EXPERIMENTS AND DISCUSSIONS

4.1. Experimental conditions

To evaluate the efficacy of the proposed method, we carried out sound-separation experiments in a real reverberant room illustrated in Fig. 1, where two sources and two directional microphones (stereo-microphone) are set. The reverberation time in this room is 200 ms. Two speech signals are assumed to arrive from different directions, θ_1 and θ_2, where we prepare three kinds of source direction patterns as follows; $(\theta_1, \theta_2) = (-90^\circ, -10^\circ)$, $(-10^\circ, 0^\circ)$, or $(30^\circ, 60^\circ)$. We used the speech signals spoken by two male and two female speakers as the source samples, and we generated 6 combinations of speakers. The sampling frequency is 8 kHz and the length of each sound sample is limited to 3 s. The DFT size is 1024, and the frame shift length is 256. The block size for calculation of each $\boldsymbol{R}_{t_i}(f)$ is set to 1.5 s in the closed-form 2nd-order ICA part.

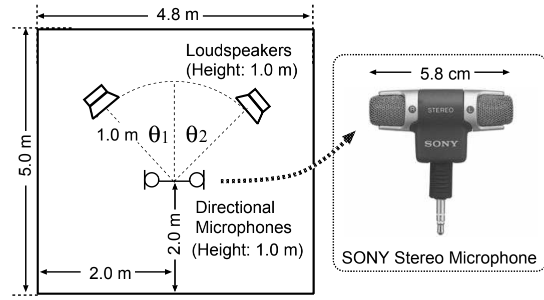

Fig. 1. Layout of reverberant room used in experiments.

4.2. Evaluation of computational-cost efficiency

In order to compare the proposed method with several conventional ICAs, we prepare two higher-order ICAs with different initial filter matrix $\boldsymbol{W}^{[0]}(f)$ as (A) a matrix which has entries of random complex value, and (B) identity matrix. The step-size parameter η in the higher-order ICA is fixed to 0.1 throughout the experiments.

Noise reduction rate (NRR) [4], defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is used as the objective indication of separation performance. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise. Figures 2–4 show the convergence curves of NRR under different speaker allocations. As for the proposed method, we plot the results only in the higher-order ICA part. These scores are the averages of 6 speaker combinations.

From the results, we first confirm that the closed-form 2nd-order ICA can score the NRRs of 8–10 dB (see the point of *Number of iterations = 0*) regardless of the speaker directions. This consistent and tolerable performance is very attractive if we take into account the low computational cost. Application of the higher-order ICA in the second stage can remarkably improve the separation performance, and the proposed BSS outperforms all of the conventional methods, especially on its convergence time.

4.3. Judge of non-stationarity vs. non-Gaussianity in speech

In this subsection, we compare the 2nd-order ICA (non-stationarity) and higher-order ICA (non-Gaussianity). Here we prepare 3 s (max i = 2) or 36 s (max i = 64) long observed signals to consider the dependence on the data length. The higher-order ICA after the closed-form 2nd-order ICA is conducted as two manners; learning with full-length data, or with 1.5 s data, to equalize the data-size effect of time-block averaging.

As shown in Figures 5–7, the performance of the 2nd-order ICA is slightly improved as the observed data length increases, but it cannot reach the level of the higher-order ICA at all. Although the experimental results are very limited number of evidences, we can speculate that non-Gaussianity is more beneficial than non-stationarity in ICA for speech signals inherently. This result is also consistent with previous *all-iteration-type* ICAs' results (see, e.g., [6]; unfortunately these methods still suffer from local-minima problem). As far as we know, our new comparison and conclusion are the world's first appearances because we derive it via closed-form method, and this can help the further investigations on this topic.

5. CONCLUSION

First, we proposed a new efficient BSS method combining closed-form 2nd-order ICA and nonclosed-form higher-order ICA, where the preceding closed-form ICA can provide a good initial value and the following higher-order ICA can update the separation filters from the advantageous status. This enables us to reduce the computational

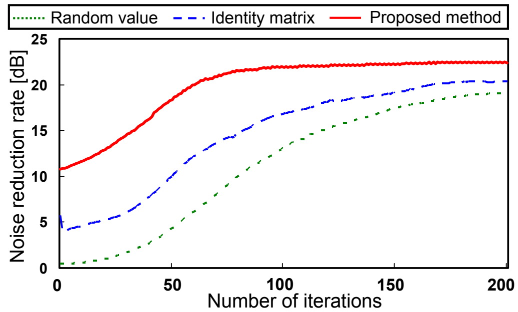

Fig. 2. NRR convergence for $(\theta_1, \theta_2) = (-90^\circ, -10^\circ)$.

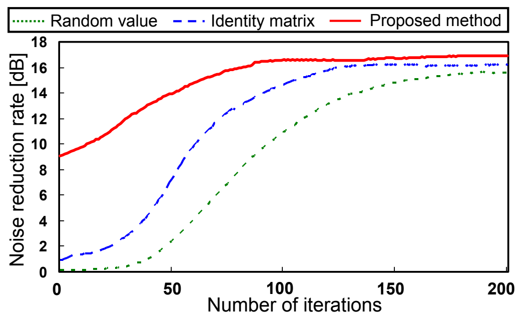

Fig. 3. NRR convergence for $(\theta_1, \theta_2) = (-10^\circ, 0^\circ)$.

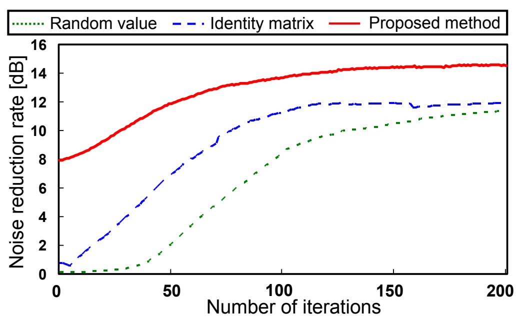

Fig. 4. NRR convergence for $(\theta_1, \theta_2) = (30^\circ, 60^\circ)$.

complexities without deteriorating the separation performance. Secondly, using the proposed method, we compare two types of ICAs with non-stationarity and non-Gaussianity. Experimental results reveal that the performances of the 2nd-order ICA are inferior to those of the higher-order ICA.

6. REFERENCES

[1] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. NOLTA*, pp.923–926, 1998.

[2] S. Ikeda and N. Murata, "A method of blind source separation based on temporal structure of signals," *Proc. ICONIP*, pp.737–742, 1998.

[3] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech & Audio Process.*, vol.8, pp.320–327, 2000.

[4] H. Saruwatari et al., "Blind source separation combining independent component analysis and beamforming," *EURASIP Journal on Applied Sig. Process.*, vol.2003, pp.1135–1146, 2003.

[5] T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundam.*, vol.E86-A, no.4, pp.846–858, 2003.

[6] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. Speech & Audio Process.*, vol.13, pp.120–134, 2005.

[7] H. Saruwatari et al., "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Speech & Audio Process.*, vol.14, pp.666–678, 2006.

[8] A. Tanaka, H. Imai, and M. Miyakoshi, "Theoretical foundations of second-order-statistics-based blind source separation for non-stationary sources," *Proc. ICASSP*, pp.III-600–III-603, 2006.

[9] L. Molgedey and H. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Phys. Rev. Lett.*, vol.72, no.23, pp.3634–3637, 1994.

[10] H. Sawada et al., "Polar coordinate based nonlinear function for frequency domain blind source separation," *IEICE Trans. Fundam.*, vol.E86-A, no.3, pp.590–596, 2003.

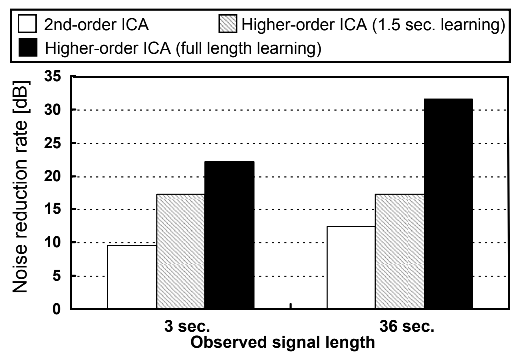

Fig. 5. Results of NRR for $(\theta_1, \theta_2) = (-90^\circ, -10^\circ)$.

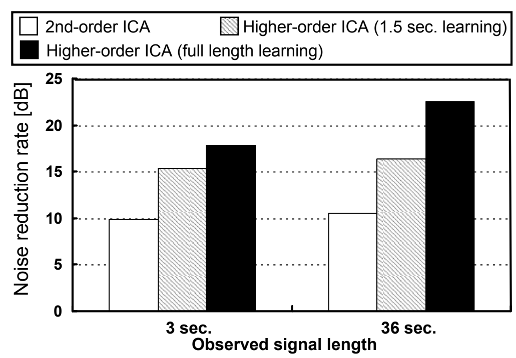

Fig. 6. Results of NRR for $(\theta_1, \theta_2) = (-10^\circ, 0^\circ)$.

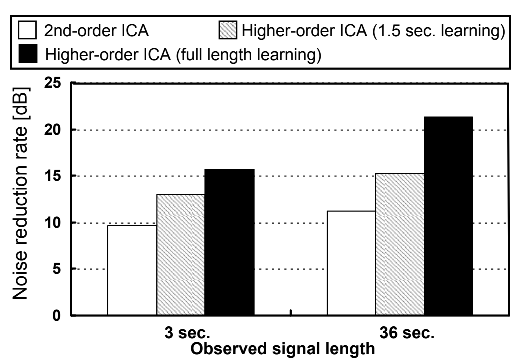

Fig. 7. Results of NRR for $(\theta_1, \theta_2) = (30^\circ, 60^\circ)$.