# MULTIPLE ICA-BASED REAL-TIME BLIND SOURCE EXTRACTION APPLIED TO HANDY SIZE MICROPHONE

*T.Hiekata*[1], *T.Morita*[1], *Y.Ikeda*[1], *H.Hashimoto*[1], *R.Zhang*[2], *Y.Takahashi*[3], *H.Saruwatari*[3], *K.Shikano*[3]

[1]Kobe Steel,Ltd., Kobe, 651-2271, Japan
[2]Feng Co.,Ltd., Himeji, 670-0995, Japan
[3]Nara Institute of Science and Technology, Nara, 630-0192, Japan

## ABSTRACT

A new blind source extraction method in widespread noise conditions is proposed, which is based on multiple frequency-domain independent component analysis (FDICA) combining projection back and spectral subtraction. In addition, We implement the proposed method to digital signal processor (DSP) for a more realistic real-time operation, and develop a new blind source extraction (BSE) microphone which can extract a target sound in real-time. In this paper, we illustrate and evaluate the proposed method and BSE microphone. And experimental results reveal that the extraction performance of the proposed method are superior to that of conventional methods, and we show the efficacy of microphone.

***Index Terms***— Real time systems, Acoustic signal processing, Digital signal processors, Acoustic devices, Microphones.

## 1. INTRODUCTION

The real-time extraction of target sound is demanded for many applications, e.g., speech dialogue systems, cellular phones, and car navigation systems. Blind source separation (BSS) is one of the beneficial approach for this purpose because BSS is a flexible approach for estimating original source signals using only information on the mixed signals observed in each input channel. Over the last decade, independent component analysis (ICA) has become one of the most notable candidates of the BSS method for separating and reducing interfering sounds in acoustical signal processing [1, 2, 3, 4, 5].

As mentioned above, the conventional ICA could work especially in speech-speech (or point sources) mixing. However, such a mixing condition is very rare and not realistic, i.e. real noises are often widespread sources. In such a sound mixing condition, it is known that ICA is proficient in noise estimation rather than in target speech estimation [6]. Based on the above-mentioned fact, we propose a new blind source extraction method which is suitable for implementation to DSP, and handy-size device.

In this paper, first, we mainly illustrate the proposed method. Next we mention the real-time BSE implementation

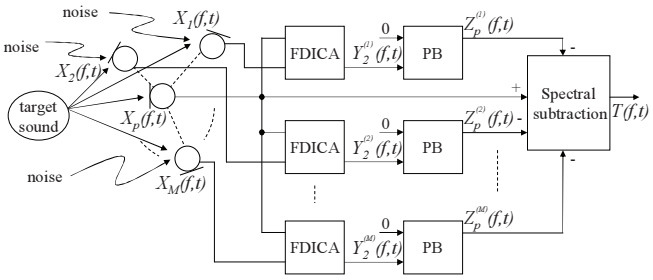


**Fig. 1**. The block diagram of proposed method.

on handy-size hardware. Several recent research studies [4] have addressed the real-time implementation of ICA, but all of them required high-speed personal computers. And also there is no fact and paper that BSE is implemented to a small-size LSI on handy-size hardware. We implement proposed method to DSP on handy-size microphone. Finally, we extensively evaluate proposed method. From our results, we can show the efficacy of proposed method implemented to the microphone.

## 2. PROPOSED BLIND SOURCE EXTRACTION

The block diagram of the proposed method is shown in Fig. 1. The proposed method needs a primary microphone and some reference microphones.

First, we designate the observed signal vector in time-frequency domain as

$$\boldsymbol{X}_m(f,t) = [X_p(f,t), X_m(f,t)]^{\mathrm{T}}, \tag{1}$$

where $\boldsymbol{X}_m(f,t)$ is the observed signal vector, $f$ is the frequency bin, $t$ $(=0,1,2,\dots)$ is time frame index, $m$ $(=0,1,2,\dots,M)$ is reference microphone index, $X_p(f,t)$ is the the observed signal of a primary microphone, and $X_m(f,t)$ is the observed signal of reference microphone $m$.

By using $\boldsymbol{X}_m(f,t)$, the ICA-based noise estimation is performed. We use frequency-domain ICA (FDICA)[1] because it is one of the ICA-based BSS approaches that require fewer computational complexities than other ICA-based BSS

approaches. In FDICA, we perform signal separation using the complex-valued unmixing matrix given as

$$\boldsymbol{W}_m(f) = \begin{bmatrix} W_{11}^{(m)}(f) & \cdots & W_{1K}^{(m)}(f) \\ \vdots & \ddots & \vdots \\ W_{L1}^{(m)}(f) & \cdots & W_{LK}^{(m)}(f) \end{bmatrix}, \quad (2)$$

where K and L are input and output number of FDICA respectively and the proposed algorithm deals with the case of $K = L = 2$, so that the output $\boldsymbol{Y}_m(f,t) = [Y_1^{(m)}(f,t), Y_2^{(m)}(f,t)]^{\mathrm{T}}$ becomes mutually independent; this procedure can be given as

$$\boldsymbol{Y}_m(f,t) = \boldsymbol{W}_m(f)\boldsymbol{X}_m(f,t). \quad (3)$$

We perform this procedure with respect to all frequency bins. The optimal $\boldsymbol{W}_m(f)$ is obtained by, e.g., the following iterative updating equation

$$\begin{aligned} &\boldsymbol{W}_m^{[i+1]}(f) \\ &= \eta\left[\boldsymbol{I} - \left\langle \boldsymbol{\Phi}(\boldsymbol{Y_m}(f,t))\boldsymbol{Y_m}^{\mathrm{H}}(f,t)\right\rangle_t\right]\boldsymbol{W}_m^{[i]}(f) \\ &\quad + \boldsymbol{W}_m^{[i]}(f), \end{aligned} \quad (4)$$

where $\boldsymbol{I}$ is the identity matrix, $\langle\cdot\rangle_t$ denotes the time-averaging operator, $[i]$ is used to express the value of the $i$-th step in the iterations, $\eta$ is the step-size parameter, and $\boldsymbol{\Phi}(\cdot)$ is the appropriate nonlinear vector function. In each FDICA, it is only required to estimate noise component. Thus, the target signal component $Y_N^{(m)}(f,t)$ is removed from the output signal vector $\boldsymbol{Y}_m(f,t)$. This processing can be designated as

$$\boldsymbol{U}_m(f,t) = [0, Y_2^{(m)}(f,t)]^{\mathrm{T}}. \quad (5)$$

Next, we apply the projection back (PB) [7] method to remove the ambiguity of amplitude. This procedure can be represented as

$$\boldsymbol{Z}_m(f,t) = \boldsymbol{M}_m^{+}(f)\boldsymbol{U}_m(f,t), \quad (6)$$

where $\boldsymbol{M}^{+}$ denotes Moore-Penrose pseudo inverse matrix of $\boldsymbol{W}_m$, $\boldsymbol{Z}_m(f,t) = [Z_p^{(m)}(f,t), Z_r^{(m)}(f,t)]^{\mathrm{T}}$ is a m-th estimated noise vector, and $Z_p^{(m)}(f,t)$ and $Z_r^{(m)}(f,t)$ are the m-th estimated noises inputted in a primary microphone and a reference microphone respectively. Note that $\boldsymbol{Z}_m(f,t)$ is the function of the frame number $t$, unlike the constant noise prototype estimated in the traditional spectral subtraction method [8]. Therefore, the proposed method can deal with *nonstationary* noise. Finally, source extraction is achieved by spectral subtraction as follows

$$\begin{aligned} &|T(f,t)| \\ &= \begin{cases} \left\{|X_p(f,t)|^2 - \sum_{m=1}^{M}\alpha_m \cdot |Z_p^{(m)}(f,t)|^2\right\}^{\frac{1}{2}}, \\ (\text{ if } |X_p(f,t)|^2 - \sum_{m=1}^{M}\alpha_m \cdot |Z_p^{(m)}(f,t)|^2 \geq 0\ ) \\ \beta \cdot |X_p(f,t)| \quad (\text{otherwise}), \end{cases} \end{aligned} \quad (7)$$

**Fig. 2**. BSE microphone.

$$T(f,t) = |T(f,t)| \cdot e^{jarg(X_p(f,t))}, \quad (8)$$

where $T(f,t)$ is the final output of the proposed method, $\alpha_m$ is the oversubtraction parameter, and $\beta$ is the flooring parameter. The appropriate setting, e.g., $(\sum_{m=1}^{M}\alpha_m) \simeq 1$ and $\beta \ll 1$, gives an efficient noise reduction.

## 3. REAL-TIME BSE MICROPHONE

### 3.1. Overview

We developed a new BSE microphone which is developed for real-time implementation of BSE. Figure 2 shows a picture of the BSE microphone, and the picture of the internal board is shown in Fig. 3. Also, the main specifications are listed in Table 1. As can be observed, the BSE microphone is one of the world's smallest BSE microphone miniaturized into handy-size hardware. This device is equipped with one primary microphone and two reference microphones, i.e., this device can deal with the case of M = 2 in the proposed method. The reference microphones have directivity in different directions respectively, and the robust estimation of widespread noise can be realized by the configuration of these microphones.The BSE microphone is equipped with a floating-point small-size DSP, and we implemented the proposed algorithm to DSP.

## 4. EXPERIMENTAL EVALUATIONS

### 4.1. Conditions

To grasp the basic behavior of the proposed method implemented to the BSE microphone, we measure the following three scores. First, *noise reduction rate (NRR)* [2], defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is evaluated as the objective indicator of the

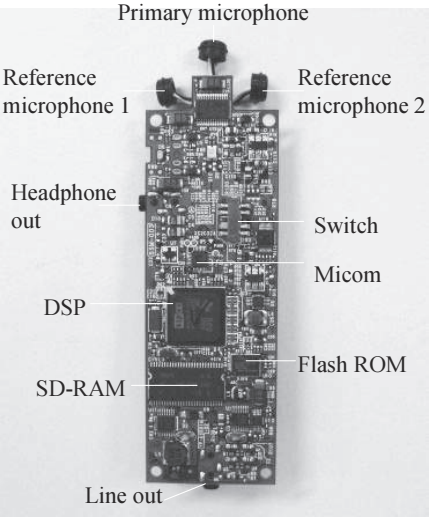


**Fig. 3**. Internal board of BSE microphone.

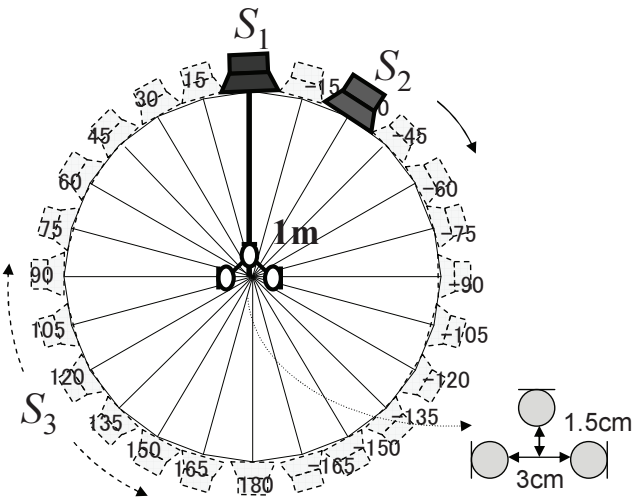


**Fig. 4**. Measurement conditions.

**Table 1**. Specifications of BSE microphone

| | |
|---|---|
| processor | TI-DSP TMS320VC6727 (Clock 300 MHz) |
| input | one primary microphone and two reference microphones |
| output | line out |
| | headphone out |
| terminal | line out |
| | headphone out |
| sampling frequency | 16 kHz or 8 kHz |
| battery | AA battery × 2 |
| memory | Flash ROM: 8 MB (used about 430 KB) |
| | SDRAM: 128 MB (used about 3.9 MB) |
| size | 136 mm (H) × 45 mm (W) × 27 mm (D) |
| weight | 125 g (including battery) |

degree of interference reduction (we do not take into account sound distortion). Secondly, we measure *cepstral distortion (CD)*, which indicates the distance between the spectral envelope of the original source signal and the target component in the separated output (CD does not take into account the degree of interference reduction unlike NRR). Thirdly, we score *PESQ MOS-LQO* (ITU-T Recommendation P.862.1), which is comparable to the *mean opinion score (MOS)* and corresponds to the subjective indicator of sound quality related to both NRR and CD.

Figure 4 shows the measurement conditions. The following real-recorded 16kHz-sampled signals were used in the experiments. The target signal is male user's speech which is talked in front of the microphone (i.e. we fix source 1 in $\theta_1 = 0°$) and 1 m apart from the microphone. As for noise, two noises are added simultaneously. First noise is an interference female speech arranged from $\theta_2 = 0°$ to $\theta_2 = 360°$ at $5°$ intervals. The second noise is music emitted from surrounded around 24 loudspeakers (In Fig. 4, source 3 shows the music). Each music sounds emitted from loudspeakers are played simultaneously but asynchronously. These sound sources are compositions in *Audio/Acoustics Technical CD*, and the length of these signals is limited to 30 s.

### 4.2. Experimental Results

We evaluate two conditions. The average SNRs between source 1 and source 3 are (1) 20dB and (2) 0dB respectively. Figure 5 shows the measurement results under the condition (1), where Fig. 5$(a)$ shows NRR, Fig. 5$(b)$ shows CD, and, Fig. 5$(c)$ shows PESQ MOS-LQO. Figure 6 shows the measurement results under the condition (2), where Fig. 6$(a)$ shows NRR, Fig. 6$(b)$ shows CD, and, Fig. 6$(c)$ shows PESQ MOS-LQO. The x-axis indicates angle of interference speech $S_2$. We compare the following three methods: (A) the conventional spectral subtraction, that is, primary microphone minus reference microphones (3ch SS), (B) conventional 3-channel FDICA (3ch ICA : uses one FDICA), (C) simple combination of conventional 3-channel FDICA and 1-channel spectral subtraction in primary microphone(3ch FDICA + 1ch SS), and (D) proposed method (the ICA part in (B), (C) or (D) uses 3-s-duration buffering for estimating the separation matrix). Figure 5 shows that the performance of the proposed method is similar to ICA-based conventional methods under comparatively noiseless conditions. On the other hand, Figure 6 shows that the proposed method greatly outperform the conventional methods in NRR, and although the proposed method doesn't outperform the ICA-based conventional methods in CD slightly, PESQ MOS-LQO results of proposed method are superior to conventional methods. This indicates that the proposed method can extract a target sound robustly under very noisy and widespread noise conditions.

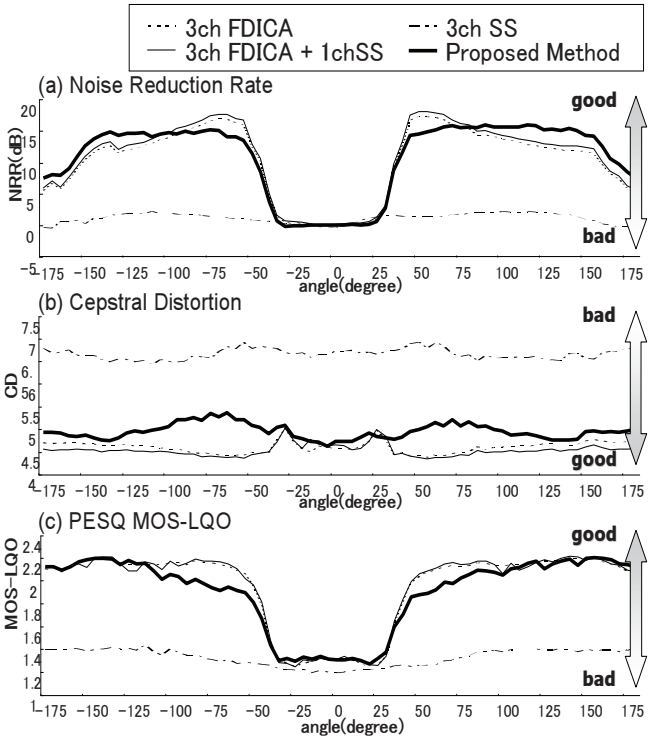


**Fig. 5**. Experimental results under the condition of SNR 20dB: (a) noise reduction rate, (b) cepstral distortion, and (c) PESQ MOS-LQO.

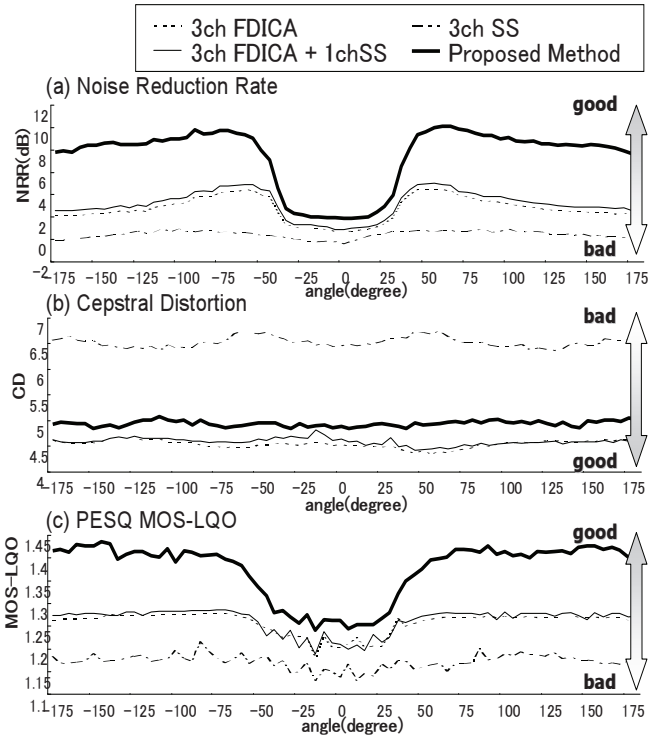


**Fig. 6**. Experimental results under the condition of SNR 0dB: (a) noise reduction rate, (b) cepstral distortion, and (c) PESQ MOS-LQO.

## 5. CONCLUSION

We proposed a novel BSE method in widespread noise conditions, which is based on multiple FDICA combining projection back and subtraction. In addition, We introduced and evaluated a new BSE microphone. Experimental results reveal that the extraction performance of the proposed method are superior to that of conventional methods, especially under very noisy conditions, and we show the efficacy of the BSE microphone. This leads us to expect that the technologies related to the use of the BSE microphone will be adopted in many future applications.

## 6. REFERENCES

[1] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. NOLTA98*, vol.3, pp.923–926, 1998.

[2] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, "Blind source separation combining independent component analysis and beamforming," *EURASIP Journal on Applied Signal Processing*, vol.2003, pp.1135–1146, 2003.

[3] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency domain blind source separation," *IEICE Trans. Fundamentals*, vol.E86-A, no.3, pp.590–596, 2003.

[4] 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 Processing*, vol.13, no.1, pp.120–134, 2005.

[5] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, Y. Ikeda, H. Hashimoto, and T. Morita, "Blind separation of acoustic signals combining SIMO-model-based independent component analysis and binary masking," *EURASIP Journal on Applied Signal Processing*, vol.2006, Article ID 34970, 17 pages, 2006.

[6] Y. Takahashi, T. Takatani, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array with independent component analysis for hands-free speech recognition," *Proc. of IWAENC*, 2006.

[7] S. Ikeda and N. Murata, "A method of ICA in the frequency domain," *Proc. Intern. Workshop on ICA and BSS*, pp.365–371, 1999.

[8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, Signal Proc.*, vol.ASSP-27, no.2, pp.113–120, 1979.