# BLIND SOURCE EXTRACTION FOR HANDS-FREE SPEECH RECOGNITION BASED ON WIENER FILTERING AND ICA-BASED NOISE ESTIMATION

*Yu Takahashi, Keiichi Osako, Hiroshi Saruwatari, Kiyohiro Shikano*

Nara Institute of Science and Technology, Nara 630-0192, Japan

## ABSTRACT

In this paper, we proposed a new blind speech extraction method consisting of Wiener filtering and noise estimation based on independent component analysis (ICA). First, we provide both theoretically and experimental investigations on proficiency of ICA in noise estimation under a non-point-source noise condition. Next, computer simulation and experiment in an actual railway-station environment are conducted, and their results also indicate that ICA is proficient in noise estimation under a non-point-source noise condition. Finally, we newly propose a blind speech extraction method based on Wiener filtering and ICA-based noise estimation, and the effectiveness of the proposed method via speech recognition test in an actual railway-station environment.

***Index Terms***— Speech enhancement, acoustic signal processing, acoustic arrays, unsupervised learning

## 1. INTRODUCTION

Blind source separation (BSS) is an approach to estimate original sources using only information of observed signals. Recently, various BSS methods based on independent component analysis (ICA) [1] have been presented for acoustic-sound separation [2, 3]. Indeed, the conventional ICA could work particularly in speech-speech mixing, i.e., all sources can be regarded as point sources, but such a mixing condition is very rare and unrealistic; real noises are often widespread sources. In this paper, we mainly deal with generalized noise that cannot be regarded as a point source. Moreover, we assume this noise to be nonstationary noise that arises in many acoustical environments; however, traditional methods, e.g., adaptive beamformer (ABF) could not treat this noise well. Although ICA is not influenced by nonstationarity of signals unlike ABF, this is still a very challenging task that conventional ICA-based BSS could hardly address because ICA cannot separate widespread sources.

Related with the performance analysis of ICA, one of the authors has mentioned that ICA-based BSS has an equivalence to parallelly-constructed ABFs [4]. However, this investigation was focused on a separation with a non-singular mixing matrix; thus valid for only point sources. In this paper, first, we give a more detailed interpretation of ICA, namely, the beamformers optimized by ICA become specific beamformers which maximize the signal-to-noise ratio (SNR) in each output (so-called *SNR-maximize beamformers*).

Next, we clarify what kinds of beamformers are optimized by ICA under a non-point-source noise condition. We find out that the beamformer which enhances a target speech signal becomes a delay-and-sum (DS) beamformer [5] and the beamformer which picks up a noise signal becomes a null beamformer (NBF) [3] in that environment. In general, the noise reduction performance of DS beamformer is very limited, but that of NBF is markedly high. Therefore, it can be expected that ICA's extraction performance of target speech is not so good rather than that of noise estimation. This imbalanced property is also confirmed by the computer simulation and experiment in an actual railway-station environment. Based on the above mentioned fact, we have proposed blind spatial subtraction array (BSSA) which utilizes ICA as noise estimator [6]. In BSSA, source extraction is achieved by subtracting the power spectrum of the estimated noise signal via ICA from the noisy observations. However, BSSA causes artificial distortion because BSSA is noise reduction method based on spectral subtraction. Such artificial distortion significantly deteriorate the speech recognition result. Therefore, in this paper, we newly introduce an Wiener filtering instead of spectral subtraction in BSSA. In the proposed, it can be expected that the distortion of a target speech signal is decreased because oversubtraction like a spectral subtraction is not performed.

Finally, we compare the conventional BSSA, the conventional ICA which simply uses ICA as a direct target estimator, and the proposed method. In conclusion, speech recognition test in an actual environment reveals the proposed method's superiority to the conventional methods.

## 2. ANALYSIS OF ICA IN WIDESPREAD NOISE

### 2.1. Independent component analysis

We consider an acoustic mixing model where the number of array elements (microphones) is $J$ and the observed signal contains only one target speech signal, which can be regarded as a point source, and an additive noise signal.This additive noise represents noises which cannot be regarded as point sources, e.g., spatially uncorrelated noises, background noises, leakage of reverberation components outside the frame analysis. Hereafter, the observed signal vector in time-frequency domain, $\boldsymbol{x}(f,\tau) = [x_1(f,\tau),\ldots,x_J(f,\tau)]^{\mathrm{T}}$, is given by

$$\boldsymbol{x}(f,\tau) = \boldsymbol{h}(f)s(f,\tau) + \boldsymbol{n}_a(f,\tau), \tag{1}$$

where $f$ is the frequency bin number, $\tau$ is the time index number, $\boldsymbol{h}(f) = [h_1(f),\ldots,h_J(f)]^{\mathrm{T}}$ is a column vector of transfer functions from the target signal component to each microphone, $s(f,\tau)$ is a target speech signal component, and $\boldsymbol{n}_a(f,\tau) = [n_1^{(a)}(f,\tau),\ldots,n_J^{(a)}(f,\tau)]^{\mathrm{T}}$ is a column vector of the additive noise signal. In ICA, we perform signal separation using a complex-valued unmixing matrix $\boldsymbol{W}_{\mathrm{ICA}}(f)$, so that the output signals $\boldsymbol{y}(f,\tau) = [y_s(f,\tau), y_n(f,\tau)]^{\mathrm{T}}$ become mutually independent; this procedure can be represented by [2]

$$\boldsymbol{y}(f,\tau) = [y_s(f,\tau), y_n(f,\tau)]^{\mathrm{T}} = \boldsymbol{W}_{\mathrm{ICA}}(f)\boldsymbol{x}(f,\tau), \tag{2}$$

$$\boldsymbol{W}_{\mathrm{ICA}}^{[p+1]}(f) = \mu\left[\boldsymbol{I} - \langle\boldsymbol{\varphi}(\boldsymbol{y}(f,\tau))\boldsymbol{y}^{\mathrm{H}}(f,\tau)\rangle_\tau\right]\boldsymbol{W}_{\mathrm{ICA}}^{[p]}(f) + \boldsymbol{W}_{\mathrm{ICA}}^{[p]}(f), \tag{3}$$

where $\mu$ is the step-size parameter, $[p]$ is used to express the value of the $p$-th step in iterations, $y_s(f,\tau)$ is the estimated target speech

This work was partly supported by the MEXT e-Society leading project, and the NEDO project for strategic development of advance robotics elemental technologies in Japan.

signal, $y_n(f,\tau)$ is the estimated noise signal, and $\boldsymbol{I}$ is the identity matrix. Besides, $\langle\cdot\rangle_\tau$ denotes a time-averaging operator, $\boldsymbol{M}^{\mathrm{H}}$ denotes conjugate transpose of matrix $\boldsymbol{M}$, and $\varphi(\cdot)$ is an appropriate nonlinear vector function which defined as

$$\varphi(\boldsymbol{y}(f,\tau)) \equiv [\varphi(y_s(f,\tau)), \varphi(y_n(f,\tau))]^{\mathrm{T}}, \tag{4}$$

$$\varphi(y_k(f,\tau)) \equiv \tanh y_k^{(\mathrm{R})}(f,\tau) + i \tanh y_k^{(\mathrm{I})}(f,\tau), \tag{5}$$

where the superscripts (R) and (I) denote the real and imaginary parts, respectively.

### 2.2. SNR-maximize beamformers optimized by ICA

ICA optimizes two beamformers; these can be written as

$$\boldsymbol{W}_{ICA}(f) = [\boldsymbol{g}_s(f), \boldsymbol{g}_n(f)]^{\mathrm{T}}, \tag{6}$$

where $\boldsymbol{g}_s(f) = [g_1^{(s)}(f), \ldots, g_J^{(s)}(f)]^{\mathrm{T}}$ is the coefficient vector of the beamformer to pick up the target speech signal and $\boldsymbol{g}_n(f) = [g_1^{(n)}(f), \ldots, g_J^{(n)}(f)]^{\mathrm{T}}$ is the coefficient vector of the beamformer to pick up the noise. Using Tailor expansion, we can express a factor of the nonlinear vector function of ICA, $\varphi(y_k(f,\tau))$, as

$$\begin{aligned}\varphi(y_k(f,\tau)) &= \tanh y_k^{(\mathrm{R})}(f,\tau) + i \tanh y_k^{(\mathrm{I})}(f,\tau), \\ &= y_k(f,\tau) - \left\{ \frac{\left(y_k^{(\mathrm{R})}(f,\tau)\right)^3}{3} + i\frac{\left(y_k^{(\mathrm{I})}(f,\tau)\right)^3}{3} \right\} + \cdots . \end{aligned} \tag{7}$$

Thus, the calculation part of higher-order correlation in ICA, $\varphi(\boldsymbol{y}(f,\tau))\boldsymbol{y}^{\mathrm{H}}(f,\tau)$, can be decomposed to a second-order correlation matrix and the summation of higher-order correlation matrices of each order. This is shown as

$$\langle\varphi(\boldsymbol{y}(f,\tau))\boldsymbol{y}^{\mathrm{H}}(f,\tau)\rangle_\tau = \langle\boldsymbol{y}(f,\tau)\boldsymbol{y}^{\mathrm{H}}(f,\tau)\rangle_\tau + \Psi(f), \tag{8}$$

where $\Psi(f)$ is a set of higher-order correlation matrices. In ICA, separation filters are optimized so that the all order correlation matrices become diagonal matrices. Then, at least the second-order correlation matrix is diagonalized by ICA. In the following, hence, we prove that ICA optimizes beamformers as SNR-maximize beamformers focusing on only the part of second-order correlation. Then, an absolute value of normalized cross-correlation coefficient (off-diagonal entries) of second-order correlation, $C$, is defined by

$$C = \frac{\left|\langle y_s(f,\tau) y_n^*(f,\tau)\rangle_\tau\right|}{\sqrt{\langle|y_s(f,\tau)|^2\rangle_\tau}\sqrt{\langle|y_n(f,\tau)|^2\rangle_\tau}}, \tag{9}$$

$$y_s(f,\tau) = \hat{s}(f,\tau) + r_s\hat{n}(f,\tau), \tag{10}$$

$$y_n(f,\tau) = \hat{n}(f,\tau) + r_n\hat{s}(f,\tau), \tag{11}$$

where $\hat{s}(f,\tau)$ is a target speech component in ICA's output and $\hat{n}(f,\tau)$ is a noise component in ICA's output, $r_s$ is a coefficient of residual noise component, $r_n$ is a coefficient of target-leakage component, and superscript * represents conjugate complex number. Therefore, SNRs of $y_s(f,\tau)$ and $y_n(f,\tau)$ can be represented by

$$\Gamma_s = \langle\hat{s}^2(f,\tau)\rangle_\tau / (r_s^2\langle\hat{n}^2(f,\tau)\rangle_\tau), \tag{12}$$

$$\Gamma_n = \langle\hat{n}^2(f,\tau)\rangle_\tau / (r_n^2\langle\hat{s}^2(f,\tau)\rangle_\tau), \tag{13}$$

where $\Gamma_s$ is SNR of $y_s(f,\tau)$ and $\Gamma_n$ is SNR of $y_n(f,\tau)$. Using (10), (11), (12) and (13) we can rewrite (9) as

$$C = \frac{\left|1/\sqrt{\Gamma_s} + 1/\sqrt{\Gamma_n}\cdot e^{j(\arg r_n^* - \arg r_s)}\right|}{\sqrt{1+1/\Gamma_s}\sqrt{1+1/\Gamma_n}}, \tag{14}$$

where $\arg r$ represents argument of $r$. Thus, $C$ is a function of only $\Gamma_s$ and $\Gamma_n$. Therefore, the cross-correlation between $y_s(f,\tau)$ and $y_n(f,\tau)$ depends on only the SNRs of beamformers $\boldsymbol{g}_s(f)$ and $\boldsymbol{g}_n(f)$.

Now, we consider $C$ minimization, which is identical with the second-order correlation matrix diagonalization in ICA. When $|\arg r_n^* - \arg r_s| > \pi/2$ where $-\pi < \arg r_s \le \pi$ and $-\pi < \arg r_n^* \le \pi$, it is possible to make $C$ zero or minimization independently of $\Gamma_s$ and $\Gamma_n$. This case is proper to the orthogonalization between $y_s(f,\tau)$ and $y_n(f,\tau)$, which is related to the principal component analysis (PCA) unlike ICA. However, ICA utilizes higher-order cross-correlation to maximize independence among all outputs. This results in the prevention of the orthogonalization of $y_s(f,\tau)$ and $y_n(f,\tau)$; consequently, hereafter we can consider only the case of $|\arg r_n^* - \arg r_s| \le \pi/2$. Then, partial differential of $C^2$ by $\Gamma_s$ is given by

$$\frac{\partial C^2}{\partial\Gamma_s} = \frac{(1-\Gamma_s) + \Gamma_s\sqrt{\Gamma_s\Gamma_n}(1-\Gamma_s)\cdot 2\mathrm{Re}\left[e^{j(\arg r_n^* - \arg r_s)}\right]}{(\Gamma_s+1)^2(\Gamma_n+1)} < 0, \tag{15}$$

where $\Gamma_s > 1$ and $\Gamma_n > 1$. As for the partial differential of $C^2$ by $\Gamma_n$, we can also prove $\partial C^2/\partial\Gamma_n < 0$, where $\Gamma_s > 1$ and $\Gamma_n > 1$ in the same manner. Therefore, $C$ is a monotonically decreasing function of $\Gamma_s$ and $\Gamma_n$. The above-mentioned fact indicates the following in ICA:

- The absolute value of cross-correlation depends on only SNRs of beamformers spanned by each row of an unmixing matrix.
- The absolute value of cross-correlation is a monotonically decreasing function of SNR.
- Therefore, the diagonalization of a second-order correlation matrix leads to SNR maximization.

Thus, we can conclude that ICA, in a parallel manner, optimizes multiple beamformers, i.e., $\boldsymbol{g}_s(f)$ and $\boldsymbol{g}_n(f)$, so that the SNR of the output by each beamformer becomes maximum.

### 2.3. What beamformers are optimized under non-point-source noise condition?

In the previous subsection, it has been proved that ICA optimizes beamformers as SNR-maximize beamformers. In this subsection, we analyze what kind of beamformers are optimized by ICA particularly under a non-point-source noise condition, where we assume the two-source separation problem. The target speech can be regarded as a point source, and the noise source is widespread and spatially uncorrelated noise. First, we focus on the beamformer $\boldsymbol{g}_s(f)$ which picks up the target signal in the environment. It is clear that the desired beamformer is minimum variance distortionless response (MVDR) beamformer [5]. MVDR beamformer is optimized by minimizing the undesired signal's power in noise only interval under the condition that the direction of the target source is known in advance. Thus, MVDR beamformer certainly maximizes the SNR of desired source. Note that we cannot know the true DOA of the target source signal because ICA is *unsupervised* adaptive technique. Thus, MVDR beamformer is expected to be the upper limit of ICA in the presence of non-point-source noises. The beamformer $\boldsymbol{g}_s(f)$ is given by

$$\boldsymbol{g}_s^{\mathrm{T}}(f) = \frac{\boldsymbol{a}(f,\theta_s(f))^{\mathrm{H}}\boldsymbol{R}^{-1}(f)}{\boldsymbol{a}(f,\theta_s(f))^{\mathrm{H}}\boldsymbol{R}^{-1}(f)\boldsymbol{a}(f,\theta_s(f))}, \tag{16}$$

$$\begin{aligned}\boldsymbol{a}(f,\theta_s(f)) = [&\exp(i2\pi(f/M)f_s d_1 \sin\theta_s/c, \\ &\ldots, \exp(i2\pi(f/M)f_s d_J \sin\theta_s/c]^{\mathrm{T}},\end{aligned} \tag{17}$$

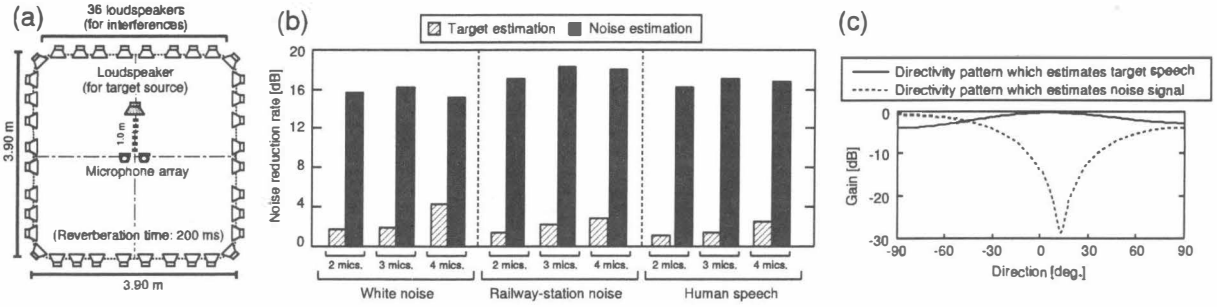


**Fig. 1**. (a) Layout of reverberant room used in simulation, (b) separation results in simulation, and (c) typical directivity patterns under non-point-source condition shaped by ICA at 2 kHz and two-element array.

where $\boldsymbol{a}(f,\theta_s(f))$ is a steering vector, $\theta_s(f)$ is the direction of the target speech, $M$ is the DFT size, $f_s$ is the sampling frequency, and $\boldsymbol{R}(f) = \langle \boldsymbol{n}_a(f,\tau)\boldsymbol{n}_a^{\mathrm{H}}(f,\tau)\rangle_\tau$ is the correlation matrix of $\boldsymbol{n}_a(f,\tau)$. Note that $\theta_s(f)$ is the function of frequency because the direction-of-arrival (DOA) of the source varies in each frequency subband under reverberant condition. Although the correlation is often not diagonalized in lower frequency subbands [5], e.g., diffuse noise, we approximate that the correlation matrix is diagonalized in whole frequency subbands. Then, regarding the power of noise signal as approximately $\delta^2(f)$, the correlation matrix results in $\boldsymbol{R}(f) = \delta^2(f)\cdot\boldsymbol{I}$. Therefore, the inverse of correlation matrix $\boldsymbol{R}^{-1}(f) = \boldsymbol{I}/\delta^2(f)$. Moreover, $\boldsymbol{a}(f,\theta_s(f))^{\mathrm{H}}\boldsymbol{a}(f,\theta_s(f)) = J$. Thus, (16) can be rewritten as

$$\boldsymbol{g}_s(f) = \frac{1}{J}[\exp(-i2\pi(f/M)f_s d_1 \sin\theta_s(f)/c), \ldots, \exp(-i2\pi(f/M)f_s d_J \sin\theta_s(f)/c)]^{\mathrm{T}}. \quad (18)$$

This filter $\boldsymbol{g}_s(f)$ is approximately equal to DS beamformer [5]. Note that the filter $\boldsymbol{g}_s(f)$ is not simple DS beamformer but *reverberation-adapted DS beamformer* because it is optimized for $\theta_s(f)$ in each frequency bin. It is well-known that the noise-reduction performance of the DS beamformer optimized by ICA under a non-point-source noise condition is proportional to $10\log_{10} J$ [dB]; this performance is not so good.

Next, we consider the other beamformer $\boldsymbol{g}_n(f)$ which picks up the noise source. The task of picking up the noise source equals suppressing the target speech arriving from $\theta_s(f)$. Generally, SNR-maximize beamformer for suppressing the target speech signal is likely to become NBF. For instance, NBF for two-element array can be defined by,

$$\boldsymbol{g}_n(f) = [\exp(-i2\pi(f/M)f_s d_1 \sin\theta_s(f)/c), -\exp(-i2\pi(f/M)f_s d_2 \sin\theta_s(f)/c)]^{\mathrm{T}}\cdot\sigma(f), \quad (19)$$

where $\sigma(f)$ is a gain compensation constant. This filter surely satisfies $\boldsymbol{g}_n^{\mathrm{T}}(f)\cdot\boldsymbol{a}(f,\theta_s(f)) = 0$. Thus, this filter steers a directional null against the target speech signal with few number of elements (at least two elements). Moreover, the undesired-signal-reduction performance of NBF is quite high, and this performance does not depend on the number of microphone elements. Also, note that the filter $\boldsymbol{g}_n(f)$ is not simple NBF because it is optimized for $\theta_s(f)$ in each frequency bin respectively. Overall, the performance of enhancing the target speech is very low and that of estimating noise source is high.

### 2.4. Computer simulation

We conduct computer simulations in the reverberant room where the reverberation time is 200 ms to confirm the performance of ICA under a non-point-source noise condition. We used the following 8 kHz-sampled signals as ICA's input; the original target speech (3 seconds) convoluted with impulse responses that were recorded in an actual environment, and to which three types of noise from 36 loudspeakers were added (see Fig. 1(a)). The three types of noise are an independent Gaussian noise, an actually recorded railway-station noise, and interference speech by 36 people. We use 12 speakers (6 males and 6 females) as sources of the original target speech, and the input SNR of test data is set to 0 dB. We use a two-, three-, or four-element microphone array with an interelement spacing of 4.3 cm.

The simulation results are shown in Figs. 1(b) and 1(c). Figure 1(b) shows the result for the average noise reduction rate (NRR) [3] of all the target speakers. NRR is defined as the output SNR in dB minus the input SNR in dB. From this result, we can see an imbalance performance between the target speech estimation and the noise estimation in every noise case; the performance of the target speech estimation is significantly poor, but that of noise estimation is very high. This result is consistent with the theory previously stated. Moreover, Fig. 1(c) shows directivity patterns shaped by the beamformers optimized by ICA in the simulation. It is clearly indicated that the beamformer $\boldsymbol{g}_s(f)$ that picks up the target speech resembles the DS beamformer, and the beamformer $\boldsymbol{g}_n(f)$ that picks up the noise becomes NBF. From these results, we confirm that the previously stated theory, i.e., the beamformers optimized by ICA under a non-point-source noise condition are DS and NBF, is valid.

## 3. PROPOSED METHOD

### 3.1. Overview

As clearly shown in Sects. 2.3 and 2.4, ICA is proficient in noise estimation rather than in target-speech estimation. Thus, we cannot use ICA as a target estimation directly under a non-point-source noise condition. However, we can still use ICA as a noise estimator. In our previous work, we have proposed BSSA [6] algorithm. BSSA comprises ICA-based noise estimator, and noise reduction is achieved by spectral subtraction. However, original BSSA suffers from large distortion of a target speech, which is mainly due to nonlinear artifact such as *musical noise* Therefore, speech recognition results are often damaged by the distortion. In this paper, we newly introduce an Wiener-filter-like method instead of spectral subtraction in BSSA architecture. Figure 2 shows the block diagram of the proposed method. In the proposed method, it can be expected that the distortion of a target speech signal is mitigated because oversubtraction like a spectral subtraction is not performed.

### 3.2. Signal processing in proposed method

The proposed method consists of two paths; a primary path which is DS-based target speech enhancer, and a reference path which is ICA-based noise estimator. Finally, we obtain the target speech extracted

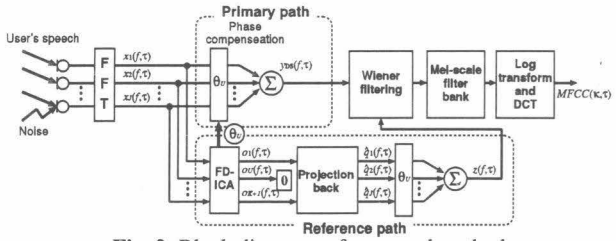


**Fig. 2**. Block diagrams of proposed method.

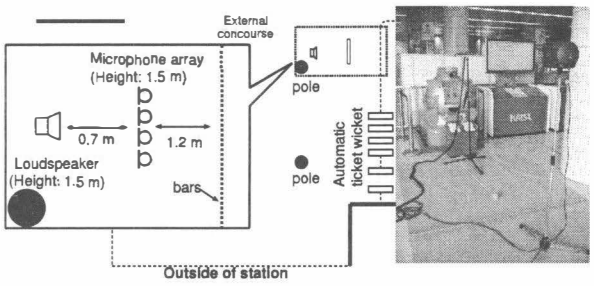


**Fig. 3**. Layout of railway-station used in real-recording experiment.

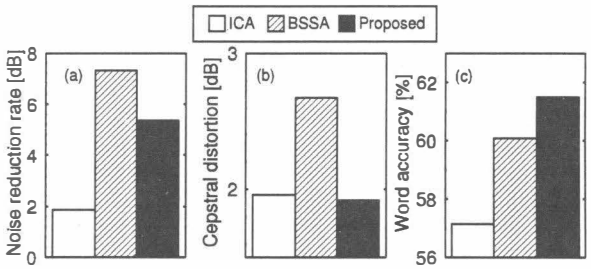


**Fig. 4**. Experimental results of (a) noise reduction rate, (b) cepstral distortion, and (c) speech recognition test, in railway-station environment.

signal based on Wiener filtering. The Wiener filter gain is designed as follows:

$$g_w(f,\tau) = \frac{|y_{DS}(f,\tau)|^2}{|y_{DS}(f,\tau)|^2 + \zeta \cdot |z(f,\tau)|^2}, \quad (20)$$

where $g_w(f,\tau)$ is the Wiener filter gain, $y_{DS}(f,\tau)$ is the output of primary path, $z(f,\tau)$ is the estimated noise via ICA, and $\zeta$ is a gain factor. Finally, we obtain the speech enhanced signal based on Wiener filtering. This procedure can be represented as

$$|y(f,\tau)|^2 = g_w(f,\tau) \cdot |y_{DS}(f,\tau)|^2, \quad (21)$$

where $y(f,\tau)$ is the output of the proposed method.

## 4. EXPERIMENT IN REAL WORLD

### 4.1. Experimental setup

To confirm the effectiveness of the proposed method, we conducted experiments in an actual railway-station environment. Figure 3 shows a layout of the railway-station environment where the reverberation time is about 1000 ms. We used the 46 speakers (200 sentences) as the target speech signal, and noise signal as real-recored noise signal in the environment. The noise in the environment is nonstationary and almost diffuse, and the noise consists of various kinds of interference noise, e.g., background noise, sounds of trains, ticket-vending machines, automatic ticket wickets, foot steps, cars, and wind. A four-element array with the interelement spacing of 2 cm is used. As far as we know, the demonstration in such an actual railway-station, that is very challenging task, is the world's first attempt for ICA study.

### 4.2. Experimental results

First, we would mention an actual separation result by ICA. NRRs of the target estimation are 6.1 dB and 3.8 dB in the noise 1 case and noise 2 case, respectively. Also, NRRs of the noise estimation are 9.6 dB and 14.6 dB in the noise 1 case and noise 2 case, respectively. We can also ascertain the imbalanced performance between target estimation and noise estimation, similarly to the simulation results (see Sect. 2.4), i.e., ICA is proficient in noise estimation.

In the next experiment, we compare the conventional ICA, the conventional BSSA, and the proposed method (Wiener-filter-like method), on the basis of NRR, cepstral distortion, and speech recognition performance. Figure 4(a) shows the results for the average of NRR in whole speakers. From these results, we can see that NRR of the proposed method is inferior to the conventional BSSA, but the cepstral distortion of the proposed method is significantly reduced compared with the conventional BSSA. Finally, we show results of speech recognition, where the extracted sound quality is completely considered, in Fig. 4(c). Speech recognition task is 20 k-word dictation, the acoustic model is phonetic tied mixture [7], we use 260 speakers (150 sentences/speaker) as training data for the acoustic model, and we use Julius [7] 3.5.1 for the speech decoder. From this result, we can conclude that the target-enhancement performance of the proposed method is superior to the conventional BSSA, and ICA which directly estimates the target speech component.

## 5. CONCLUSION

In this paper, first, we revealed that beamformers optimized by ICA become DS beamformer which enhances the target speech signal and NBF which picks up noise signal. Next, computer simulation and experiment in the actual railway-station environment were conducted, and we obtained the separating result in that the performance of enhancing the target signal is poor and that of estimating noise source is very high. Therefore, we realized that ICA is suitable for noise estimator under a non-point-source noise condition. Next, we newly propose the blind source extraction method based on Wiener filtering and ICA-based noise estimation. Finally, it was confirmed that the speech recognition performance of the proposed method overtook those of the conventional ICA, and BSSA.

## 6. REFERENCES

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

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

[3] H. Saruwatari et al., "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Speech and Audio Proc.*, vol.14, no.2, pp.666-678, 2006.

[4] S. Araki et al., "Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures," *EURASIP J. Applied Signal Proc.*, vol.2003, no.11, pp.1157–1166, 2003.

[5] M. Brandstein and D. Ward, "Microphone Arrays: Signal Processing Techniques and Applications," Springer-Verlag, 2001

[6] Y. Takahashi et al., "Blind spatial subtraction array with independent component analysis for hands-free speech recognition," *Proc. of IWAENC*, 2006.

[7] A. Lee et al., "Julius - An open source real-time large vocabulary recognition engine," *Proc. Eurospeech*, pp.1691–1694, 2001.