New analytical calculation and estimation of TDOA for underdetermined BSS in noisy environments

Takuro Maruyama*†, Shoko Araki†, Tomohiro Nakatani†,
Shigeki Miyabe*, Takeshi Yamada*, Shoji Makino* and Atsushi Nakamura†
* Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennoudai, Tsukuba-shi, Ibaraki 305-8573, Japan
† NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
Email: maruyama@mmlab.cs.tsukuba.ac.jp

***Abstract*—We have proposed a new algorithm for sparseness-based underdetermined blind source separation (BSS) that can cope with diffused noise environments. This algorithm includes a technique for estimating the time-difference-of-arrival (TDOA) parameter separately in individual frequency bins for each source. In this paper, we propose methods that integrate the frequency-bin-wise TDOA parameter to estimate the TDOA of each source. The accuracy of TDOA estimation with the proposed approach is shown experimentally in comparison with a conventional approach. The separation performance and calculation time of the proposed approach is also examined.**

I. Introduction

Blind Source Separation (BSS) has been intensively investigated because the problem setting matches the real environment very well. With overdetermined BSS cases, source separation can be performed satisfactorily, especially in a clean environment, for example by using Independent Component Analysis (ICA). To be able to handle a more realistic situation, however, we must consider the underdetermined case, where there are fewer sensors than sources.

Many methods have been proposed for underdetermined BSS [1], [2], [3], [4], [5]. They employ the sparseness characteristics of the source signals as a clue for BSS, and are based on the clustering of the source location information, such as time differences of arrival (TDOAs). However, most such methods tend to be weak in the presence of reverberation and background noise.

Recently, Izumi et al. [6] proposed a BSS method that can overcome this problem. The method assumes a diffuse noise environment to make the estimation robust, and solves the TDOA clustering based on the Expectation-Maximization (EM) algorithm. However drawback with this method is that it cannot analytically update the TDOA parameters. As a consequence, this method requires a time-consuming exhaustive search for the TDOA update.

We have proposed a new efficient BSS algorithm to remove the need for such an exhaustive search [7] within the same robust BSS framework proposed in [6]. The new update rule has eliminated the need for the exhaustive search of TDOA, and reduced the computational load. The new update rule provides the TDOA estimation results for each frequency bin.

Such frequency-wise TDOA estimation is sufficient for BSS, however, it is insufficient for TDOA estimation of each source. In concrete terms, we have to determine one common TDOA for each source from the frequency-bin-wise TDOA estimated in individual frequency bins for the source.

So in this paper, we propose methods for estimating the TDOA of source signals from the estimated TDOA parameters at each frequency bin. We provide experimental results showing the accuracy of TDOA estimation with the proposed methods in comparison with that of the conventional method. We also report experimental results for the separation performance and calculation time of the proposed methods in comparison with those of the conventional method.

II. CONVENTIONAL METHOD

This section outlines the method proposed by Izumi et al. [6] and the problem with it. Let $\mathbf{x}_{f,t} = [x_{f,t,L}, x_{f,t,R}]^T$ be signals observed by two microphones represented in the time-frequency domain. Where f, t are the index of the time-frequency slots. If we assume that source signals are sufficiently sparse such that only one source signal is active at each time-frequency point, and each source signal is transferred as a plane wave, $\mathbf{x}_{f,t}$ can be written as

$$\begin{bmatrix} x_{f,t,L} \\ x_{f,t,R} \end{bmatrix} = \begin{bmatrix} 1 \\ e^{j2\pi f\delta_k} \end{bmatrix} s_{f,t,k} + \begin{bmatrix} n_{f,t,L} \\ n_{f,t,R} \end{bmatrix} \quad (1)$$

$$\mathbf{x}_{f,t} = \mathbf{b}_{f,k} s_{f,t,k} + \mathbf{N}_{f,t} \quad (2)$$

, where k is the index of the source, $s_{f,t,k}$ is the complex spectrum of the source signal that is active at a time-frequency slot, $\mathbf{b}_{f,k}$ is the transfer function from the source to the microphones (δ_k is the TDOA between two microphones), and $\mathbf{n}_{f,t}$ is the observation error, which includes reverberation and background noise and is assumed to be uncorrelated with the source signals.

We assume that $\mathbf{N}_{f,t}$ is time-invariant and follows a Gaussian distribution with a zero mean and a covariance matrix $\sigma_f^2 \mathbf{V}_f$ where σ_f^2 is the noise power, and $\mathbf{V}$ is given as follows for the diffused noise

$$\mathbf{V}_f = \begin{bmatrix} 1 & \mathrm{sinc}(2\pi fD/c) \\ \mathrm{sinc}(2\pi fD/c) & 1 \end{bmatrix} \quad (3)$$

. Here c is the velocity of sound, and D represents the distance between the two microphones. The purpose of the conventional method [6] and this paper is to estimate the source signals $s_{f,t,k}$ solely from the mixed observation $\mathbf{x}_{f,t}$.

The likelihood function for the observation $\mathbf{x}_{f,t}$ is

$$p(\mathbf{x}_{f,t}|k,\theta) = \frac{1}{\pi^2\sigma_f^4|\mathbf{V}_f|}\exp\left(-\frac{1}{\sigma_f^2}\mathbf{N}_{f,t,k}^H\mathbf{V}_f^{-1}\mathbf{N}_{f,t,k}\right) \quad (4)$$

.

Let $\theta = \{\sigma_f^2, \delta_k, s_{f,t,k}\}$ be the parameter set. The log likelihood function is

$$L = \sum_t\sum_f \log\sum_k p(\mathbf{x}_{f,t}|k,\theta)p(k|\theta) \quad (5)$$

, where $p(k|\theta)$ is the mixing weight ($\sum_k p(k|\theta) = 1$), and k is an index of a source that is assumed to be dominant at a time frequency point according to the sparseness characteristics assumption.

In [6], this log likelihood is maximized with the EM algorithm. The parameters to be estimated are $\theta = \{\sigma_f^2, \delta_k, s_{f,t,k}\}$ where k is the hidden variable and the auxiliary function in this problem is

$$\begin{aligned} Q(\theta|\theta') &= \mathrm{E}\left[\log p(\mathbf{x}_{f,t};\theta)|\theta'\right] \\ &= \sum_t\sum_f\sum_k m_{k,f,t}\log p(\mathbf{x}_{f,t}|k,\theta)p(k|\theta). \end{aligned} \quad (6)$$

. Here, time-frequency mask $m_{k,f,t}$ is the posterior probability that source k is active at a time-frequency slot, and θ' is the parameter set obtained by the previous iteration.

Moreover, $m_{k,f,t}$ functions as a soft mask for separating the k-th source in the EM algorithm, and after the convergence of the EM algorithm we can estimate the separated source signal $y_{f,t,k}$ as the expectation using the estimated parameters and $m_{f,t,k}$ given by

$$y_{f,t,k} = m_{f,t,k}\frac{\mathbf{b}_{f,k}^H\mathbf{V}_f^{-1}\mathbf{x}_{f,t}}{\mathbf{b}_{f,k}^H\mathbf{V}_f^{-1}\mathbf{b}_{f,k}} \quad (7)$$

.

In this paper, the E step updates $m_{k,f,t}$ and the auxiliary function, and the M step updates the other parameters.

The time-frequency mask $m_{k,f,t}$ is updated by

$$m_{f,t,k} = p(k|\mathbf{x}_{f,t,k},\theta') = \frac{p(k|\theta')p(\mathbf{x}_{f,t}|k,\theta')}{\sum_k' p(k'|\theta')p(\mathbf{x}_{f,t}|k',\theta')} \quad (8)$$

.

The parameters σ_f^2 and $s_{f,t,k}$ are estimated by differentiating the auxiliary function with respect to each parameter, and setting them at zero,

$$\sigma_f^2 = \frac{1}{T}\sum_t\sum_k m_{f,t,k}\mathbf{N}_{f,t,k}^H\mathbf{V}_f^{-1}\mathbf{N}_{f,t,k} \quad (9)$$

$$s_{f,t,k} = \frac{\mathbf{b}_{f,k}^H\mathbf{V}_f^{-1}\mathbf{x}_{f,t}}{\mathbf{b}_{f,k}^H\mathbf{V}_f^{-1}\mathbf{b}_{f,k}}, \quad (10)$$

and the mixing weight $p(k|\theta)$ (where $\sum_k p(k|\theta) = 1$) is calculated by

$$p(k|\theta) = \frac{1}{TF}\sum_t\sum_f m_{f,t,k} \quad (11)$$

, where T and F are the numbers of time frames and frequency bins, respectively.

In [6], as δ_k cannot be solved analytically, the update is performed by calculating $Q(\theta|\theta')$ for all the discretized δ_k and selecting δ_k that gives the maximum of Q

$$\delta_k = \mathrm{argmax}_{\delta_k} Q(\theta|\theta') \quad (12)$$

.

This update rule has two problems. One is that we cannot obtain an exact, optimal update of δ_k but only its discretized approximation in each iteration. The other problem is that this exhaustive search requires a large computational cost. To overcome these problems, we derive an analytical update rule for estimating the TDOA parameter δ_k.

III. PROPOSED METHOD

In this section, we provide an analytical update rule for the TDOA parameter δ_k.

A. Calculation of TDOA parameter

The right-hand side of the likelihood function (4) is denoted by using the components of the vectors $\mathbf{x}$, $\mathbf{b}$ and the matrix $\mathbf{V}$

$$\begin{aligned} &\mathbf{N}_{f,t,k}^H\mathbf{V}_f^{-1}\mathbf{N}_{f,t,k} \\ =\ &\left[n_{f,t,L}^* n_{f,t,R}^*\right]\frac{1}{\sigma_f^2(1-\phi_f^2)}\begin{bmatrix} 1 & -\phi_f \\ -\phi_f & 1 \end{bmatrix}\begin{bmatrix} n_{f,t,L} \\ n_{f,t,R} \end{bmatrix} \\ =\ &\frac{1}{\sigma_f^2(1-\phi_f^2)}\big\{|n_{f,t,L}|^2 + |n_{f,t,R}|^2 \\ &-\phi_f n_{f,t,R}^* n_{f,t,L} - \phi_f N_{f,t,R}N_{f,t,L}^*\big\}, \end{aligned} \quad (13)$$

where ϕ_f denotes $\phi_f = \mathrm{sinc}(2\pi fD/c)$.

From (1)

$$\begin{bmatrix} n_{f,t,L} \\ n_{f,t,R} \end{bmatrix} = \begin{bmatrix} x_{f,t,L} \\ x_{f,t,R} \end{bmatrix} - \begin{bmatrix} 1 \\ \beta_{f,k} \end{bmatrix} s_{f,t,k} \quad (14)$$

,

where $\beta_{f,k}$ denotes $\beta_{f,k} = e^{j2\pi f\delta_k}$. By substituting (14) into (13), we obtain (15)

$$
\begin{aligned}
& \mathbf{N}_{f,t,k}^H \mathbf{V}_f^{-1} \mathbf{N}_{f,t,k} \\
= & \frac{1}{\sigma_f^2(1-\phi_f^2)} \{ |x_{f,t,L} - s_{f,t,k}|^2 + |x_{f,t,R} - \beta_{f,k} s_{f,t,k}|^2 \\
& -\phi_f (x_{f,t,R} - \beta_{f,k} s_{f,t,k})^* (x_{f,t,L} - s_{f,t,k}) \\
& -\phi_f (x_{f,t,R} - \beta_{f,k} s_{f,t,k}) (x_{f,t,L} - s_{f,t,k})^* \} \\
= & \frac{1}{\sigma_f^2(1-\phi_f^2)} \{ s_{f,t,k}^* \beta_{f,k}^* \beta_{f,k} s_{f,t,k} \\
& -\beta_{f,k}^* s_{f,t,k}^* [x_{f,t,R} - \phi_f (x_{f,t,L} - s_{f,t,k})] \\
& -\beta_{f,k} s_{f,t,k} [x_{f,t,R} - \phi_f (x_{f,t,L} - s_{f,t,k})]^* \\
& +|x_{f,t,R}|^2 - \phi_f x_{f,t,R}^* (x_{f,t,L} - s_{f,t,k}) \\
& -\phi_f x_{f,t,R} (x_{f,t,L} - s_{f,t,k})^* + |x_{f,t,L} - s_{f,t,k}|^2 \} \\
= & \frac{1}{\sigma_f^2(1-\phi_f^2)} \{ \Big(\xi_{f,t,k} - \beta_{f,k} s_{f,t,k}\Big)^* \Big(\xi_{f,t,k} - \beta_{f,k} s_{f,t,k}\Big) \\
& +(1-\phi_f^2) |x_{f,t,L} - s_{f,t,k}|^2 \}, \qquad (15)
\end{aligned}
$$

where

$$
\xi_{f,t,k} = [x_{f,t,R} - \phi_f (x_{f,t,L} - s_{f,t,k})]. \qquad (16)
$$

Furthermore, the clause containing the above-mentioned $\beta_{f,k}$ can be rewritten as

$$
\begin{aligned}
& \Big(\xi - \beta_{f,k} s_{f,t,k}\Big)^* \Big(\xi - \beta_{f,k} s_{f,t,k}\Big) \\
= & |\xi_{f,t,k}|^2 + |s_{f,t,k}|^2 \\
& -\xi_{f,t,k}^* s_{f,t,k} e^{j2\pi f \delta_k} - \xi_{f,t,k} s_{f,t,k}^* e^{-j2\pi f \delta_k} \\
= & |\xi_{f,t,k}|^2 + |s_{f,t,k}|^2 \\
& -2|\xi_{f,t,k}||s_{f,t,k}| \cos(\psi_{S_k} - \psi_{\xi_k} - 2\pi f \delta_k), \qquad (17)
\end{aligned}
$$

where ψ_{S_k} and ψ_{ξ_k} represent the phases of $s_{f,t,k}$ and $\xi_{f,t,k}$, respectively (i.e, $s_{f,t,k} = |s_{f,t,k}| e^{j\psi_{S_k}}$, $\xi_{f,t,k} = |\xi_{f,t,k}| e^{j\psi_{\xi_k}}$) If $\phi_f = 0$, $\psi_{S_k} - \psi_{\xi_k}$ is the phase difference between two microphones.

By using (15) and (17), the likelihood function (4) can be rewritten as

$$
\begin{aligned}
p(\mathbf{x}_{f,t}|k,\theta) = & \frac{1}{2\pi\sqrt{\sigma_f^2 |\mathbf{V}_f|}} \exp(C) \\
& \cdot \exp\Big(\frac{|\xi_{f,t,k}||s_{f,t,k}|}{\sigma_f^2(1-\phi^2)} \cos(\psi_s - \psi_\xi - 2\pi f \delta_{f,k}) \Big) \qquad (18)
\end{aligned}
$$

, where C is independent of $\delta_{f,k}$.

The last term,

$$
\exp\Big(\frac{|\xi_{f,t,k}||s_{f,t,k}|}{\sigma_f^2(1-\phi^2)} \cos(\psi_s - \psi_\xi - 2\pi f \delta_{f,k}) \Big) \qquad (19)
$$

has the shape of the von Mises distribution [8]

$$
g(x|\kappa,\mu) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)} \qquad (20)
$$

, where $-\pi < x \leq \pi$, μ is the mean of the distribution ($-\pi < \mu \leq \pi$), $\kappa > 0$ is a concentration parameter, and $I_0(x)$ is a modified Bessel function of the first kind and order zero. Thus, (19) indicates that the phase difference $\psi_s - \psi_\xi \approx \arg(x_{f,t,R}) - \arg(x_{f,t,L})$ follows a von Mises distribution whose mean value corresponds to the frequency-dependent TDOA $\mu = 2\pi f \delta_{f,k}$, and the concentration parameter is the signal to noise ratio (SNR) related [1] value $\kappa = \frac{|\xi_{f,t,k}||s_{f,t,k}|}{\sigma_f^2(1-\phi_f^2)}$.

Therefore, we can derive the update rule for δ_k using a method similar to a maximum likelihood estimation of the mixture model of the von Mises distribution. However, because the cosine part of (19) depends on the frequency f, we have to derive the update rule for $\delta_{f,k}$ at each frequency, which is different from the previous frequency independent update rule (12). By substituting (19) into (6) and setting $\frac{\partial Q}{\partial \delta_{f,k}} = 0$, the update rule becomes

$$
2\pi f \delta_{f,k} = \arctan \frac{\sum_t m_{f,t,k} |\xi_{f,t,k}||s_{f,t,k}| \sin(\psi_{\xi_k} - \psi_{S_k})}{\sum_t m_{f,t,k} |\xi_{f,t,k}||s_{f,t,k}| \cos(\psi_{\xi_k} - \psi_{S_k})} \qquad (21)
$$

.

It should be noted that the function $\arctan(x)$ is unique only if $-\pi/2 < x < \pi/2$. However, $2\pi f \delta_{f,k}$ can fall in the $-\pi$ to π range. Therefore, when $|x| \geq \pi/2$, we have to modify the estimated value by checking the inflection point of the auxiliary function. To accomplish this, we calculate the second order differential of the auxiliary function, and modify the values as follows

- If $\delta_{f,k} < 0$, $\frac{\partial^2 Q}{\partial(\delta_{f,k}^2)} \geq 0$, then $2\pi f \delta_{f,k} \leftarrow 2\pi f \delta_{f,k} + \pi$
- If $\delta_{f,k} > 0$, $\frac{\partial^2 Q}{\partial(\delta_{f,k}^2)} \geq 0$, then $2\pi f \delta_{f,k} \leftarrow 2\pi f \delta_{f,k} - \pi$
- Otherwise, we do not modify $2\pi f \delta_{f,k}$

In summary, the proposed method estimates the parameters σ_f^2, $s_{f,t,k}$ and $m_{f,t,k}$ in the same ways as described in Section 2, and the frequency-dependent TDOA parameter $\delta_{f,k}$ is calculated with the update rule (21).

B. Estimation of TDOA parameters

We described how to compute the TDOA parameters using the method proposed in [6]. However, the TDOAs are estimated independently in individual frequency bins by this method. We propose the following three methods for estimating the frequency-independent TDOAs.

- PROPOSED METHOD A
  A presumed $\delta_{f,k}$ is used as an independent TDOA for every frequency bin.
- PROPOSED METHOD B
  An average value of $\delta_{f,k}$ over all the frequency bins is used as a TDOA estimate for the source indexed by k. (This means $\delta_k = \frac{1}{F}\sum_f \delta_{f,k}$)
  Let this be PROPOSED METHOD B.
- PROPOSED METHOD C
  Figure 1 shows an example set of estimated TDOAs obtained by the method proposed in [6].
  The solid lines show the value of the estimated TDOA, and the dashed lines show the values of the correct

[1] When we assume that $\phi_f = 0$, $\frac{|\xi_{f,t,k}||s_{f,t,k}|}{\sigma_f^2(1-\phi^2)} = \frac{|s_{f,t,k}|^2}{\sigma_f^2}$, which is the SNR.

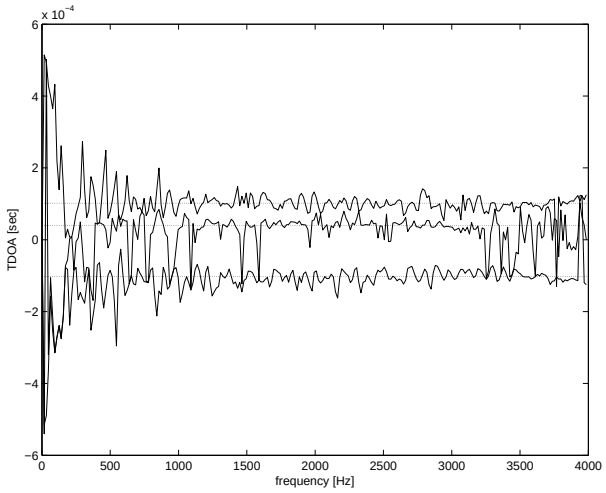

Fig. 1. Sample of calculated $\delta_{f,k}$

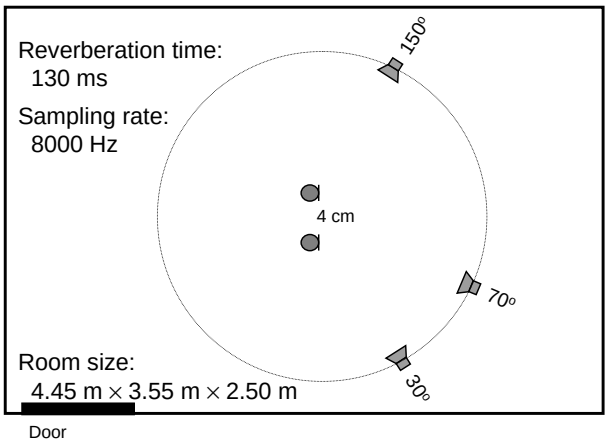

Fig. 2. Experimental setup

TDOAs. Figure 1 shows that the estimation error is very large in low frequency bins. Moreover, many of the estimated TDOAs in high frequency bins also include many errors. So, in PROPOSED METHOD C, the range of f to be used for the estimation is restricted. An average value of $\delta_{f,k}(200 \text{ Hz} < f < 3000 \text{ Hz})$ is adopted and it is used as a TDOA estimate of each source signal. (This means $\delta_k = \frac{1}{3000-200}\sum_{200<f<3000}\delta_{f,k}$)
Let this be PROPOSED METHOD C.

C. Separate source signals

The proposed separation method employs the procedure described in section II, except that (12) is replaced by (21).

IV. EXPERIMENTS

A. Experimental conditions

We performed experiments with measured impulse responses in a room with a reverberation time of 130 ms. The experimental setup is shown in Fig. 2. We used two microphones, whose spacing was 4 cm. The number of sources K was $K = 2$ ($70°$ and $150°$), or $K = 3$ ($30°, 70°$ and $150°$). Observed signals are made by convolving the measured room impulse responses and 5-second English speech signals sampled at 8 kHz. The frame size and frame shift for STFT were 64 ms and 16 ms, respectively.

In our experiments we used a Gaussian noise with zero mean and a covariance matrix of $\sigma_f^2\mathbf{V}_f$, where $\mathbf{V}_f$ was given by (3), and σ_f was determined so that the SNR with respect to source 1 had a preset value.

We compared the direction of arrival (DOA) estimation accuracy and computational times of the conventional method and PROPOSED METHODS A, B and C. For a discrete search of the DOA parameters δ_k with the conventional method, we compared the Q functions of the $0°$ to $180°$ range in increments of $1°$ to find the optimum according to Eq, (12).

B. DOA estimation accuracy

We compared DOA estimation errors. For each K, 10 speaker combinations were tested and the results were averaged. Table I shows the errors of the angular estimation. We can see from table I that PROPOSED METHOD C was the best error of the proposed methods. However, the conventional method was comparable to or even better than all the proposed methods.

To investigate the influence of noise power and reverberation on the TDOA estimation we also compared the estimation errors of the conventional and proposed methods by changing the SNR and the reverberation time. Table II shows the influence of the noise on the TDOA estimation performance. The number of sources K was 3 ($30°, 70°$ and $150°$), and the other conditions were the same as those in Section IV-A. By adjusting the power of the noise σ_f^2, the SNR was set at 27.5 dB (= Table IV. (b)), 20 dB (Table II. (a)), and 10 dB (Table II. (b)).

Table III shows the influence of the reverberation time on the TDOA estimation performance. The number of sources K was 3 ($30°, 100°$ and $135°$), and the other conditions were the same as those in Section 4.1. The reverberation time was set at 300 ms.

We can see from Tables I, II and III that PROPOSED METHOD C was the best of the three proposed methods. However, the conventional method was comparable to or even better than all the proposed methods.

C. Separation performance and computational time

We compared the separation performance and computational time of the conventional method and the proposed methods A, B and C. The separation performance was evaluated in terms of the signal to interference-plus-noise ratio (SINR) and the signal-to-distortion ratio (SDR) [2]. For each K, 10 speaker combinations were tested and the results were averaged.

Figure 3 shows example spectrograms of the signals separated by the different methods. The figure shows that the source signal was successfully separated from the observed

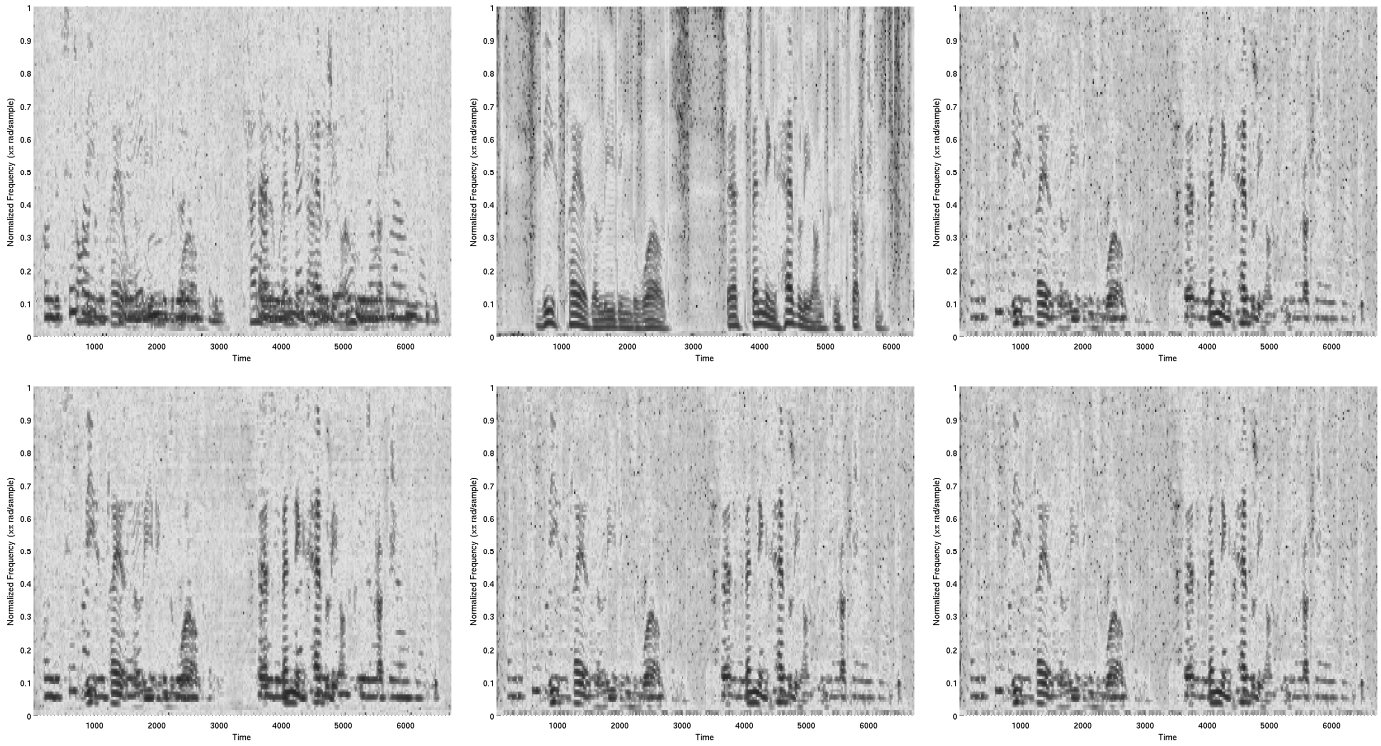

Fig. 3. Signal spectrogram. Top left: observed signal, top center: one example of source image, top right: separated signal with conventional method, bottom left: separated signal with proposed method A, bottom center: separated signal with proposed method B, bottom right: separated signal with proposed method C.

TABLE I
ANGULAR ESTIMATION ERROR

(a) $K = 2$ (70° and 150°)

Method	Source1	Source2
Conventional	1.1[deg]	3.0[deg]
Proposed A	2.2[deg]	6.1[deg]
Proposed B	3.9[deg]	3.0[deg]
Proposed C	1.3[deg]	0.6[deg]

(b) $K = 3$ ($30^\circ, 70^\circ$ and 150°)

Method	Source1	Source2	Source3
Conventional	2.8[deg]	1.5[deg]	3.5[deg]
Proposed A	12.1[deg]	6.3[deg]	7.7[deg]
Proposed B	4.1[deg]	1.9[deg]	6.4[deg]
Proposed C	5.6[deg]	2.4[deg]	1.6[deg]

TABLE II
DOA ESTIMATION ACCURACY UNDER SEVERAL SNR CONDITIONS

(a) $SNR = 20$ [dB]

Method	Source1	Source2	Source3
Conventional	1.7[deg]	0.3[deg]	4.2[deg]
Proposed A	12.2[deg]	6.2[deg]	7.2[deg]
Proposed B	5.0[deg]	0.7[deg]	6.4[deg]
Proposed C	4.9[deg]	2.2[deg]	3.2[deg]

(b) $SNR = 10$ [dB]

Method	Source1	Source2	Source3
Conventional	0.1[deg]	1.7[deg]	5.0[deg]
Proposed A	13.0[deg]	8.9[deg]	7.4[deg]
Proposed B	5.3[deg]	9.5[deg]	6.4[deg]
Proposed C	4.6[deg]	0.4[deg]	3.4[deg]

TABLE III
DOA ESTIMATION UNDER A REVERBERATION CONDITION

reverberation time = 300 [ms]

Method	Source1	Source2	Source3
Conventional	25.3[deg]	26.0[deg]	6.8[deg]
Proposed A	11.2[deg]	26.5[deg]	5.6[deg]
Proposed B	26.3[deg]	39.0[deg]	2.8[deg]
Proposed C	30.0[deg]	41.4[deg]	7.6[deg]

speech mixture. Table IV shows the SINR, SDR and computational time obtained in the experiments. We can see from table IV that our proposed method achieved a comparable performance to the conventional approach and greatly reduced the computational time by 1/10 to 1/20. This result shows that our proposed method can successfully separate signals using the estimated TDOA parameter δ_k without using an exhaustive search that is essential with the conventional method.

To investigate the influence of noise power and reverberation on separation performance, we also compared the SINR, SDR, and computational time of the conventional and the proposed methods by changing the SNR and reverberation time.

Table V shows the influence of noise on the separation performance. The number of sources K was 3 ($30^\circ, 70^\circ$ and 150°), and the other conditions were the same as those in Section 4.1. By adjusting the power of the noise σ_f^2, the SNR was set at 27.5 dB (= Table IV (b)), 20 dB (Table V (a)), and 10 dB (Table V (b)).

TABLE IV
SOURCE SEPARATION RESULTS

(a) $K = 2$ (70° and 150°)

Method	SINR	SDR	Calculation Time
Conventional	15.3[dB]	6.7[dB]	398.6[s]
Proposed A	14.8[dB]	7.7[dB]	25.7[s]
Proposed B	15.1[dB]	6.6[dB]	16.5[s]
Proposed C	15.3[dB]	6.8[dB]	16.4[s]

(b) $K = 3$ ($30^\circ, 70^\circ$ and 150°)

Method	SINR	SDR	Calculation Time
Conventional	7.6[dB]	5.4[dB]	451.5[s]
Proposed A	6.7[dB]	5.9[dB]	39.1[s]
Proposed B	7.6[dB]	5.2[dB]	15.0[s]
Proposed C	7.7[dB]	5.4[dB]	14.8[s]

TABLE V
PERFORMANCE UNDER SEVERAL SNR CONDITIONS

(a) $SNR = 20$ [dB]

Method	SINR	SDR	Calculation Time
Conventional	7.6[dB]	5.4[dB]	470.1[s]
Proposed A	6.7[dB]	5.9[dB]	41.0[s]
Proposed B	7.5[dB]	5.3[dB]	17.8[s]
Proposed C	7.6[dB]	5.4[dB]	14.5[s]

(b) $SNR = 10$ [dB]

Method	SINR	SDR	Calculation Time
Conventional	5.3[dB]	5.5[dB]	512.2[s]
Proposed A	6.1[dB]	5.9[dB]	47.1[s]
Proposed B	4.9[dB]	5.3[dB]	31.6[s]
Proposed C	5.3[dB]	5.4[dB]	15.4[s]

Table VI shows the influence of the reverberation on the separation performance. The number of sources K was 3 ($30^\circ, 100^\circ$ and 135°), and the other conditions were the same as those in Section 4.1. The reverberation time was set at 300 ms.

We can see from Tables V and VI that the performance of our proposed method was comparable to that of the conventional method in that if reduced the computational time, regardless of noise and reverberation.

V. CONCLUSION

This paper proposed new methods for estimating the TDOAs of multiple sources for sparseness-based blind source separation in noisy and reverberant environments. The proposed methods determine a unique TDOA for each sound source by taking the average of the TDOAs over different frequency bins, which can be estimated in a computationally very efficient manner by using an analytical update equation. We confirmed that our proposed methods can greatly reduce the computation time without degrading the quality of the separated signals by comparison with the conventional method. However, we also confirmed that the TDOA estimation accuracy of the proposed methods was slightly degraded compared with that of the conventional method. We plan to improve our methods as regards accuracy of TDOA estimation, to evaluate them in real environments, and to introduce a noise model that is more appropriate for reverberation.

TABLE VI
PERFORMANCE UNDER A REVERBERATION CONDITION

reverberation time = 300 [ms]

Method	SINR	SDR	Calculation Time
Conventional	7.0[dB]	2.9[dB]	962.5[s]
Proposed A	6.3[dB]	2.7[dB]	65.9[s]
Proposed B	5.1[dB]	3.0[dB]	24.7[s]
Proposed C	5.9[dB]	2.9[dB]	24.9[s]

REFERENCES

[1] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, Vol. 52, No. 7, pp. 1830-1847, 2004.

[2] S. Araki, H. Sawada, R. Mukai and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," Signal Processing, Vol. 77, No. 8, pp. 1833-1847, 2007.

[3] M. Mandel, D. Ellis and T. Jebara, "An EM algorithm for localizing multiple sound sources in reverberant environments," *Proc. Neural Info. Proc. Sys.*, 2006.

[4] C. Fevotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Transactions on Speech and Audio Processing*, Vol. 14, No. 6, pp. 2174-2188, 2006.

[5] H. Sawada, S. Araki and S. Makino, "A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures," *Proc. WASPAA2007.* , pp. 139-142, 2007.

[6] Y. Izumi, N. Ono, and S. Sagayama, "Sparseness-based 2ch BSS using the EM algorithm in reverberant environment," in *Proc. WASPAA2007*, pp. 147-150, 2007.

[7] T. Maruyama, S. Araki, T. Nakatani, S. Miyabe, T. Yamada, S. Makino and A. Nakamura, "New analytical update rule for TDOA inference for underdetermind BSS in noisy environments" *Proc. ICASSP2012*, pp. 269-272, 2012,

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2008