Performance analysis of multiple sparse source localization with triangular pyramid microphone array in noise environment

Conference Paper · August 2012

DOI: 10.1109/ICSPCC.2012.6335702

2 authors, including:

Yuexian Zou
Peking University

Performance Analysis of Multiple Sparse Source Localization with Triangular Pyramid Microphone Array in Noise Environment

Mengqi Ren, and Yuexian Zou*, Senior Member, IEEE
Advanced Digital Signal Processing Laboratory
Peking University Shenzhen Graduate School
Shenzhen 518055, China
*Correspondence author: zouyx@pkusz.edu.cn

Abstract—**Direction of arrival (DOA) estimation for multiple speakers using a small microphone array is one challenging research topic and has vast promising applications. One efficient solution is transforming the problem of estimating DOAs of multiple sources simultaneously to estimate DOA of one source at each time-frequency (TF) point by using the sparsity attribute of speech signal. The TPMA-RIPD algorithm proposed in [1] is one multiple sources DOA estimation algorithm following this route, which has shown high DOA estimation accuracy and robustness to additive white noise from the experimental results. In this paper, the robustness of the TPMA-RIPD algorithm is analyzed. The TPMA-RIPD algorithm under the real setup experiments has been evaluated, which further validates the good accuracy and robustness of the proposed algorithm.**

Keywords-Direction of arrival; performance analysis; real scenario; time-frequency sparsity; TPMA-RIPD algorithm

I. INTRODUCTION

Direction of arrival (DOA) estimation is an important fundamental technique in the array signal processing field [2]. There are various arrays that can be adopted for DOA estimation. Large arrays provide higher DOA estimation accuracy but are much costly due to the hardware of array and computational complexity. Comparing with the large arrays, small arrays achieve better performance in size, weight, cost and power consumption. Therefore, small arrays are more popular in portable and mobile devices. DOA estimation for speech signal using small microphone arrays, which is the focus of this paper, has various applications, such as service robot, teleconference system, and recorder pen. Considering that such applications usually have to deal with situations where the active sources outnumber the sensors [2], we mainly study the DOA estimation algorithms which adapt to these situations.

One sort of those algorithms utilizes the sparsity attribute of speech signal transforming the problem of estimating DOAs of multiple sources simultaneously to estimate the DOA of one source at each time-frequency (TF) point, and realizes DOA estimation in the situations where the active sources outnumber the sensors. This kind of algorithm calculates Inter-sensor Phase Difference (IPD) or Inter-sensor Time Difference (ITD) at each TF point, and obtains DOA estimation of multiple sources by clustering IPDs or ITDs calculated in TF domain. The classical algorithms in this sort estimate the DOAs of multiple sources just using two microphones [3-5]. From experimental results, we notice that the DOA estimation accuracy decreases in end-fire of array, and the DOA estimation range is limited in half plane. Besides this, the angular resolution is enslaved to the cosine relationship between IPD or ITD and DOA. To surmount these limitations, the novel TPMA-RIPD algorithm has been proposed [1]. The TPMA-RIPD algorithm provides higher angular resolution by using tangent relationship between RIPD and DOA. Meanwhile, it improves the DOA estimation accuracy in all directions, and realizes 3D space DOA estimation. Encouraged by the good performance of the TPMA-RIPD algorithm shown in [1], in this paper, we address the performance analysis of the TPMA-RIPD algorithm in noisy environment. One experiment setup in unconstrained office environment is further carried out to validate the robustness of the TPMA-RIPD algorithm. The rest of this paper is organized as follows. Section II presents TPMA-RIPD algorithm. Section III analyzes the performance of TPMA-RIPD algorithm under additive white noise. Section IV provides the simulation results of theoretical analysis and experimental results in real scenario. Finally, we conclude our work in Section V.

II. TPMA-RIPD ALGORITHM

In the TPMA-RIPD algorithm proposed in [1], we consider a regular triangular pyramid microphone array, which is shown in Fig. 1.

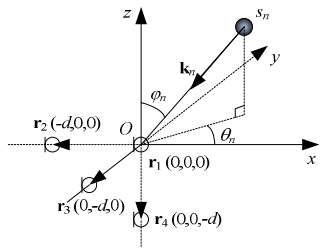

Figure 1. Spatial localization of nth speech source and the regular triangular pyramid microphone array (n is the source index, n = 1,2,…,N).

This work was partially supported by the Chinese National 863 Program (2007AA11Z224), Shenzhen Science & Technology Program (JC201105170727A), and the special foundation of president of PKUSZ (2010009).

978-1-4673-2193-8/12/$31.00 ©2012 IEEE

Four omnidirectional microphones are placed at four vertices. The lateral faces of triangular pyramid are isosceles right triangles. The sensor located at origin point O is taken as the reference sensor. $\mathbf{r}_m$ denotes position vector of mth microphone, which can be denoted as $\mathbf{r}_1 = (0, 0, 0)^T$, $\mathbf{r}_2 = (-d, 0, 0)^T$, $\mathbf{r}_3 = (0, -d, 0)^T$, and $\mathbf{r}_4 = (0, 0, -d)^T$, respectively. Wave vector of nth speech source S_n is denoted as $\mathbf{k}_n$, which is determined by

$$\mathbf{k}_n = \begin{bmatrix} k_{n1} \\ k_{n2} \\ k_{n3} \end{bmatrix} = -\frac{\Omega_n}{c} \begin{bmatrix} \sin\varphi_n \cos\theta_n \\ \sin\varphi_n \sin\theta_n \\ \cos\varphi_n \end{bmatrix} \tag{1}$$

where $\theta_n\left(\theta_n \in [0°, 360°)\right)$ and $\varphi_n\left(\varphi_n \in (0°, 180°)\right)$ denote azimuth and zenith of arrival direction of nth source (S_n), respectively, Ω_n denotes frequency of nth source, and c denotes sound velocity.

Assuming a far field source scenario, free field and noiseless condition, the Short-Time Fourier Transform (STFT) of speech signal captured by mth microphone at TF point (ω, t) is given as

$$X_m(\omega, t) = \sum_{n=1}^{N} S_n(\omega, t) e^{-j\mathbf{k}_n^T \mathbf{r}_m} \tag{2}$$

where t and ω are the frame index and the digital angular frequency respectively. $S_n(\omega,t)$ represents the STFT of the signal captured by reference sensor when only the nth source impinges on the array. Considering the sparsity attribute of speech signal, the following assumption is acceptable. At each TF point, at most one source is dominating and the contributions from the other sources are negligible [4-6]. If the TF point (ω, t) is dominated by S_n, (2) can be rewritten as

$$X_m(\omega, t) = S_n(\omega, t) e^{-j\mathbf{k}_n^T \mathbf{r}_m} \tag{3}$$

The IPD between mth microphone (m = 2,3,4) and reference microphone is given by

$$\phi_{m1}(\omega, t) = \angle\left[\frac{X_m(\omega, t)}{X_1(\omega, t)}\right] = -\mathbf{k}_n^T\left(\mathbf{r}_m - \mathbf{r}_1\right) \tag{4}$$

The Ratio of Inter-sensor Phase Difference (RIPD) is defined as

$$R_{IPD}(\omega, t) \triangleq \phi_{31}(\omega, t) / \phi_{21}(\omega, t) \tag{5}$$

Substitute (1) and (4) in (5), we get

$$R_{IPD}(\omega, t) = \frac{\mathbf{k}_n^T(\mathbf{r}_3 - \mathbf{r}_1)}{\mathbf{k}_n^T(\mathbf{r}_2 - \mathbf{r}_1)} = \frac{k_{n2} d}{k_{n1} d} = \frac{\sin\varphi_n \sin\theta_n}{\sin\varphi_n \cos\theta_n} = \tan\theta_n \tag{6}$$

From (6), the azimuth θ_n can be calculated by

$$\theta_n = \arctan^{360°} R_{IPD}(\omega, t) \tag{7}$$

The Ratio of Combined Squared Inter-sensor Phase Difference (RCSIPD) is defined as

$$R_{CSIPD}(\omega, t) \triangleq -\frac{\sqrt{\phi_{21}^2(\omega, t) + \phi_{31}^2(\omega, t)}}{\phi_{41}(\omega, t)} \tag{8}$$

Substitute (1) and (4) in (8), we get

$$R_{CSIPD}(\omega, t) = -\frac{\sqrt{(k_{n1} d)^2 + (k_{n2} d)^2}}{k_{n3} d} = \frac{\sin\varphi_n}{\cos\varphi_n} = \tan\varphi_n \tag{9}$$

Similarly, the zenith φ_n can be calculated by

$$\varphi_n = \arctan^{180°} R_{CSIPD}(\omega, t) \tag{10}$$

So far, we get the DOA estimation of speech source S_n at the TF point (ω, t) under an ideal condition. The detailed TPMA-RIPD algorithm can be found in [1].

III. Performance Analysis of TPMA-RIPD Algorithm

In this section, the performance of TPMA-RIPD algorithm under white noise is analyzed.

According to Section II, only the TF points with high local SNR are used to estimate the DOAs of sources. Therefore, we will look at the performance of the TPMA-RIPD at these TF points. Without lost of generality, the impact of the noise on azimuth estimation is discussed. Let's look at the partial derivative of $R_{IPD}(\omega, t)$ defined in (5), which is given as

$$\frac{\partial R_{IPD}(\omega, t)}{\partial \phi_{31}(\omega, t)} = \frac{1}{\phi_{21}(\omega, t)}, \quad \frac{\partial R_{IPD}(\omega, t)}{\partial \phi_{21}(\omega, t)} = -\frac{\phi_{31}(\omega, t)}{\phi_{21}^2(\omega, t)} \tag{11}$$

From (11), it is obvious that the partial derivative of $R_{IPD}(\omega, t)$ will go to infinite when $\phi_{21}(\omega, t)$ goes close to zero, which means that $R_{IPD}(\omega, t)$ will be very sensitive to the change of $\phi_{21}(\omega, t)$ and $\phi_{31}(\omega, t)$ when $\phi_{21}(\omega, t)$ is very small. Specifically, this relationship indicates that the small change of $\phi_{21}(\omega, t)$ or $\phi_{31}(\omega, t)$ due to the interference noise when $\phi_{21}(\omega, t)$ is very small will cause large variation of $R_{IPD}(\omega, t)$. In this case, from (7), we get to know that the estimation of θ_n will be biased.

Let's look at the occurring condition of small $\phi_{21}(\omega, t)$ (close to zero). From (1) and (4), we have the following relation:

$$\left.\begin{aligned} \theta &\to 90°, 270°, \text{ for } \forall \varphi, \omega \\ \varphi &\to 0°, 180°, \text{ for } \forall \theta, \omega \\ \omega &\to 0, \text{ for } \forall \theta, \varphi \end{aligned}\right\} \Rightarrow \phi_{21}(\omega, t) \to 0 \tag{12}$$

Keeping above information in mind, we come to investigate the impact of the additive noise on $\phi_{21}(\omega, t)$ and $\phi_{31}(\omega, t)$ when the sources located at the angles as shown in (12) or the frequency is considerably low. Assuming additive white noise at each microphone is independent to each other. The signal captured by mth (m = 1, 2, 3, 4) microphone is given by

$$Y_m(\omega, t) = S_n(\omega, t) e^{-j\mathbf{k}_n^T \mathbf{r}_m} + \sigma_m(\omega) e^{j\eta_m(\omega)} \tag{13}$$

where $Y_m(\omega, t)$ denotes the STFT of the signal with additive noise captured by the mth microphone, $\sigma_m(\omega)$ and $\eta_m(\omega)$ denote the magnitude response and phase response of white noise respectively, and $\eta_m(\omega)$ is within range of $[-\pi, \pi]$. Denoting the phase angle of $S_n(\omega, t)$ as ξ_n, for explanation simplification, we define

$$\alpha_m \triangleq \frac{\sigma_m(\omega)}{S_n(\omega,t)} = \frac{\sigma_m(\omega)}{|S_n(\omega,t)|} e^{-j\xi_n(\omega)} = |\alpha_m| e^{-j\xi_n(\omega)} \quad (14)$$

Considering there exists noise interference and the relationship $-\mathbf{k}_n^T\mathbf{r}_1 = 0$ (where $\mathbf{r}_1$=(0,0,0)^T), the ratio between $Y_m(\omega,t)$ and $Y_1(\omega,t)$ can be calculated as

$$\begin{aligned}\frac{Y_m(\omega,t)}{Y_1(\omega,t)} &= \frac{S_n(\omega,t)e^{-j\mathbf{k}_n^T\mathbf{r}_m} + \sigma_m(\omega)e^{j\eta_m(\omega)}}{S_n(\omega,t)e^{-j\mathbf{k}_n^T\mathbf{r}_1} + \sigma_1(\omega)e^{j\eta_1(\omega)}} \\ &= \frac{e^{-j\mathbf{k}_n^T\mathbf{r}_m} + \alpha_m e^{j\eta_m(\omega)}}{1+\alpha_1 e^{j\eta_1(\omega)}} \quad (m=2,3,4)\end{aligned} \quad (15)$$

In the TF points with high local SNR, we have $\sigma_m(\omega) \ll |S_n(\omega,t)|$, i.e., $|\alpha_m| \ll 1$ (m=1,2,3,4). Taking the Taylor expansion of $(1+\alpha_1 e^{j\eta_1(\omega)})^{-1}$ in (15), and ignoring the quadratic and higher order terms, we get

$$(1+\alpha_1 e^{j\eta_1(\omega)})^{-1} \approx 1-\alpha_1 e^{j\eta_1(\omega)} \quad (16)$$

Substituting (16) in (15) and ignoring the highest order term for $|\alpha_m| \ll 1$, we reach the following

$$\frac{Y_m(\omega,t)}{Y_1(\omega,t)} \approx e^{-j\mathbf{k}_n^T\mathbf{r}_m} + \alpha_m e^{j\eta_m(\omega)} - \alpha_1 e^{-j\mathbf{k}_n^T\mathbf{r}_m} e^{j\eta_1(\omega)} \quad (17)$$

The IPD $\phi_{m1}(\omega,t)$ with noise interference is given by

$$\phi_{m1}(\omega,t) = \angle\left[\frac{Y_m(\omega,t)}{Y_1(\omega,t)}\right] = \arctan\frac{-\sin\left(\mathbf{k}_n^T\mathbf{r}_m\right) + N_{\mathrm{Im}}}{\cos\left(\mathbf{k}_n^T\mathbf{r}_m\right) + N_{\mathrm{Re}}} \quad (18)$$

where $N_{\mathrm{Im}} = |\alpha_m|\sin[\eta_m(\omega)-\xi_n] + |\alpha_1|\sin[\mathbf{k}_n^T\mathbf{r}_m - \eta_1(\omega)+\xi_n]$, and $N_{\mathrm{Re}} = |\alpha_m|\cos[\eta_m(\omega)-\xi_n] - |\alpha_1|\cos[\mathbf{k}_n^T\mathbf{r}_m - \eta_1(\omega)+\xi_n]$. Note that, (18) leads to (4) when $N_{\mathrm{Im}} = N_{\mathrm{Re}} = 0$.

According to the derivations above, we discuss the influence of noise on $\phi_{21}(\omega,t)$ and $\phi_{31}(\omega,t)$ under the situations in (12) as follows. Firstly, let's consider the influence of noise on $\phi_{21}(\omega,t)$. From (1) and (4), we can see that $\mathbf{k}_n^T\mathbf{r}_2 \to 0$ when each of the three situations in (12) presents. Then we have $\sin(\mathbf{k}_n^T\mathbf{r}_m) \to 0$ and there is a high likelihood that N_{Im} is about or even larger than the signal component $\sin(\mathbf{k}_n^T\mathbf{r}_m)$ in the numerator of (18). Therefore, under the situations described in (12), the noise will have big impact on the estimation of $\phi_{21}(\omega,t)$, and cause large variation of $R_{IPD}(\omega,t)$ when the situations in (12) presents as discussed before. The influence of noise on $\phi_{31}(\omega,t)$ is similar to that on $\phi_{21}(\omega,t)$, and noise will cause large variation of $R_{IPD}(\omega,t)$ for the change of $\phi_{31}(\omega,t)$ when the 2nd and 3rd situations in (12) presents. In conclusion, when $\theta \to 90°, 270°$, $\varphi \to 0°, 180°$ or $\omega \to 0$, the estimation of $R_{IPD}(\omega,t)$ is very sensitive to the additive noise.

To reduce the adverse impact of noise on $R_{IPD}(\omega,t)$ when $\theta \to 90°, 270°$. In [1], the value of $R_{IPD}(\omega,t)$ is replaced by θ_n when the situation of $\theta \to 90°, 270°$ meets. From (7), we have

$$\frac{d\theta_n}{dR_{IPD}(\omega,t)} = \frac{1}{1+R_{IPD}^2(\omega,t)} \quad (19)$$

From (6), we can see that if $\theta \to 90°, 270°$, $R_{IPD}(\omega,t) = \tan\theta_n \to \pm\infty$. According to (19), the quantity of $[1+R_{IPD}^2(\omega,t)]^{-1}$ will go to zero. That means the big variation of $R_{IPD}(\omega,t)$ due to the additive noise can be scaled down in θ domain.

Moreover, from (12), it is noted that $R_{IPD}(\omega,t)$ is also very sensitive to noise when $\omega \to 0$. To tackle this problem, it is a practical way to predefine a frequency threshold ω_L (>>0), and only the TF points with the frequency higher than ω_L are selected. Furthermore, from (12), noise will also cause large interference on $R_{IPD}(\omega,t)$ when $\varphi \to 0°, 180°$, it is a limitation for the TPMA-RIPD algorithm.

The similar approach can be applied on the performance analysis of zenith in the additive noise. We present the conclusion directly due to the page limitation in this paper: when $\varphi \to 90°$ or $\omega \to 0$, $R_{CSIPD}(\omega,t)$ is sensitive to additive noise and $R_{CSIPD}(\omega,t)$ is replaced by φ_n when $\varphi \to 90°$.

IV. EXPERIMENTAL RESULTS

In this section, the simulation results of theoretical analysis are provided first to verify the conclusion obtained in Section 3. Then, the experimental results are shown to demonstrate the robustness of the TPMA-RIPD algorithm [1] in real scenario.

A. Simulation Results of Theoretical Analysis

In this simulation study, one simulation is given to evaluate the robustness of TPMA-RIPD algorithm under different azimuths. The similar simulation can be carried to different zeniths. The array setup is shown in Fig. 1 where d = 8cm is used. The distance between speech source and origin point O is set to be 2m. One speech source with 32kHz sample rate is considered. Sound velocity is set to be 345m/s. The simulation data is generated using the Image Method [7] where room size is 10×10×4 m^3 and the reflection coefficients are 0. The 1024-point STFT is used to calculate the frequency spectra. Hamming window with 30 ms duration and 20 ms overlap is used to segment signals. Additive white noise at each microphone is independent and SNR of data captured by each microphone is set to be 5dB. One high local SNR TF point (750Hz, 64Frame) is selected. φ is set to be 90º. R_{IPD} at the selected TF point is calculated using (18) and (5). Each experiment is repeated with 100 times, and only retain R_{IPD} estimation results in the high local SNR TF points (SNR>14dB).

The root mean squared (RMS) error of estimated results is used as the performance measurement. Fig. 2 shows RMS error

of estimated R_{IPD} versus azimuth, which aims at illustrating the performance of the R_{IPD} by (5) under the additive noise. From Fig. 2, we can see that big variations appear when $\theta \in (70°,110°)$ or $\theta \in (250°,290°)$, which corresponds to the analysis in Section 3. Therefore, the simulation results suggest the angle range of R_{IPD} where R_{IPD} is very sensitive to the additive noise. As the result, we replace the estimated R_{IPD} by θ when $\theta \in (70°,110°)$ or $\theta \in (250°,290°)$, and the estimation result in the whole range of azimuth ($\theta \in [0°,360°)$) is denoted as R_{IPD}_θ. The RMS error of the R_{IPD}_θ are shown in Fig. 3. Comparing to Fig. 2, it's clear to see that the resulting R_{IPD}_θ is much robust to the additive noise in the whole range of azimuth.

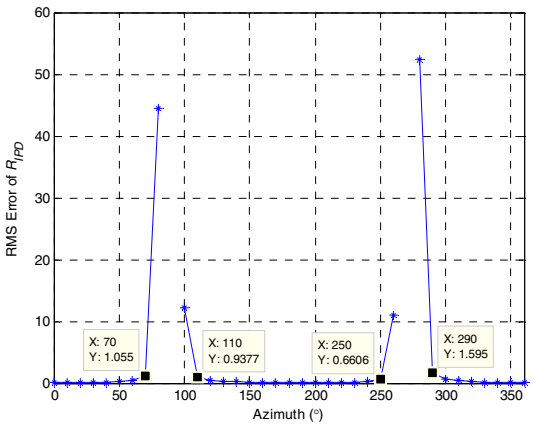

Figure 2. Bias and RMS error of RIPD vs. azimuth.

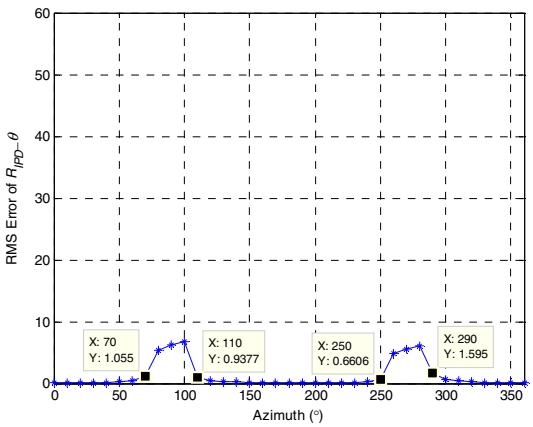

Figure 3. Bias and RMS error of $RIPD\_\theta$ vs. azimuth.

B. Experimental Results in Real Scenario

1) Experimental setup.

In this experiment study, two experiments are given to evaluate the robustness of TPMA-RIPD algorithm in real scenario. Experimental setup is shown in Fig. 4. The room size is 8.5×3×5 m^3. There are various interferences in the environment, including additive noise and reverberation. The SNR is about 5dB. The distance between speaker and microphone array is 2m. The microphone spacing is 8cm. The analog signals captured by the microphone array are converted to digital signal by NI PXI signal-capture card, and then transferred to the computer. The captured signal is processed on the LabVIEW platform. The sample rate of speech signals is 16kHz. The 512-point STFT is used to calculate frequency spectra. Hamming window with 30 ms duration and 20 ms overlap is adopted to segment signals. Sound velocity is set to be 345m/s.

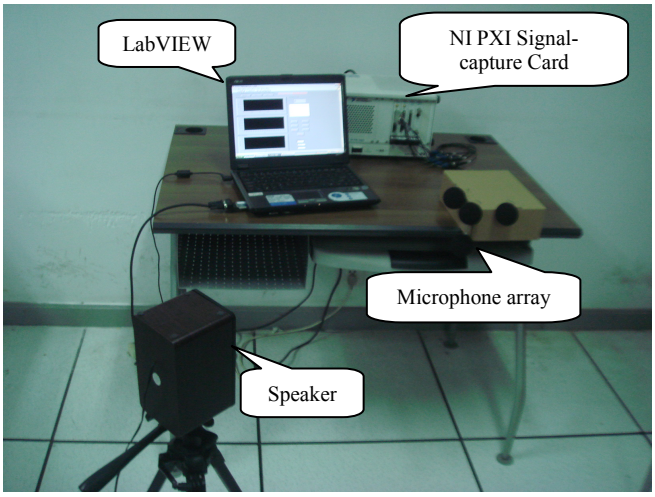

Figure 4. Experimental setup in real scenario.

2) Robustness of TPMA-RIPD algorithm in real scenario.

The robustness of the TPMA-RIPD algorithm is tested in two conditions: a) only one active speaker is in the environment, b) two active speakers are in the environment. In two experiments, only azimuths of speakers are varied, and the zeniths are fixed at 90º. Each experiment is repeated 3 times.

Table. 1 shows the experimental results when there is only one active speaker in the environment. The real azimuth of speaker is shown in the first row of Table I, and the experimental results are shown in the following rows. From Table I, we can see that all of the biases of DOA estimation results for the three azimuths are smaller than 3º. The TPMA-RIPD algorithm is robust to noise and gives consistent estimating accuracy for all DOAs in this real scenario.

TABLE I. EXPERIMENTAL RESULTS OF ONE ACTIVE SOURCE

Real Azimuth	45º	90º	135º
Trial 1-Azimuth	43.88º	87.50º	136.12º
Trial 2-Azimuth	43.88º	92.50º	136.12º
Trial 3-Azimuth	43.88º	87.50º	136.12º

Confined by the experimental conditions, we setup two active speakers in this experiment to evaluate the performance of estimating the DOAs of multiple speakers in the real scenario. The azimuths of two active speakers are shown in the first row of Table II, and the estimated azimuths are shown in the following rows accordingly. From Table II, we can see that the absolute biases of the estimation results for the three trials are smaller than 4º when there are two active sources. It can be

concluded that the TPMA-RIPD algorithm performs quite well when multiple sources present in terms of high DOA estimation accuracy and the robustness to noises.

TABLE II. EXPERIMENTAL RESULTS OF TWO ACTIVE SOURCES

Real Azimuths	45º 90º	60º 105º	80º 125º
Trial 1-Azimuths	43.88º 92.50º	56.50º 102.50º	77.50º 123.50º
Trial 2-Azimuths	43.88º 87.50 º	56.50º 107.50º	77.50º 128.97º
Trial 3-Azimuths	43.88º 87.50 º	56.50º 102.50º	77.50º 128.97º

V. CONCLUSIONS

In this paper, we analyze the performance of TPMA-RIPD algorithm in noisy environment. Additive white noise is considered and its influence is carefully analyzed. The simulation results well support the theoretical analysis and demonstrate that the TPMA-RIPD algorithm is robust to additive white noise. Furthermore, experiments in the real environment show that the TPMA-RIPD algorithm is robust to the additive noise and reverberation in the office room in both single active source and two active sources scenarios.

REFERENCES

[1] M. Ren and Y. Zou, "A Novel Multiple Sparse Source Localization Using Triangular Pyramid Microphone Array," *IEEE Signal Processing Letters,* vol. 19, pp. 83-86, 2012.

[2] S. Araki, H. Sawada, R. Mukai, S. Makino,"DOA Estimation for Multiple Sparse Sources with Arbitrarily Arranged Multiple Sensors," *Journal of Signal Processing Systems,* vol. 63, pp. 265-275, 2011.

[3] C. Liu, B. C. Wheeler, W. D. O'Brien Jr., R. C. Bilger, C. R. Lansing, and A. S. Feng, "Localization of multiple sound sources with two microphones," *The Journal of the Acoustical Society of America,* vol. 108, pp. 1888-1905, 2000.

[4] W. Zhang and B. D. Rao, "A Two Microphone-Based Approach for Source Localization of Multiple Speech Sources," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 18, pp. 1913-1928, 2010.

[5] M. I. Mandel, R. J. Weiss, and D. Ellis, "Model-Based Expectation-Maximization Source Separation and Localization," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 18, pp. 382-394, 2010.

[6] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing,* vol. 52, pp. 1830-1847, 2004.

[7] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America,* vol. 65, pp. 943-950, 1979.