Robust Localization of Single Sound Source Based on Phase Difference Regression

Zhaoqiong Huang, Ge Zhan, Dongwen Ying, and Yonghong Yan

Key Laboratory of Speech Acoustics and Content Understanding, Chinese Academy of Sciences
huangzhaoqiong@hccl.ioa.ac.com

Abstract

Phase difference regression (PDR) was widely utilized to estimate Direction-of-Arrival (DOA) for linear arrays because of its high time resolution and high computational efficiency. However, conventional regression methods were seldom reported to estimate DOA using planar arrays. This paper proposes a regression method to derive DOA from all phase differences on all frequencies for a planar array. The DOA is represented as the function of the array topology and phase differences between all microphones. Moreover, the proposed method considers another two problems that were often ignored by most regression methods. One is the problem about the period of phase difference in the regression cost function. The other is the signal enhancement that can effectively suppress the acoustic interference. We conducted some experiments in simulated environment to evaluate the proposed method using a 9-element circular array. The experimental results confirmed its superiority in both computational efficiency and robustness.

Index Terms: Planar array, phase difference regression, signal enhancement, sound source localization.

1. Introduction

Single speech source localization is of great significance to speech signal processing such as speaker tracking [1]. Single source localization generally requires high computational efficiency and acoustic robustness in reality. The direction of arrival (DOA) of sound source can be straightforwardly represented as the inverse triangular function of the time delays for linear arrays [2]. GCC-PHAT [3] was often utilized to estimate the DOA of speech source because of its high computational efficiency [4] [5]. The drawback of GCC-PHAT is the low time resolution. Even though the time resolution can be improved by increasing the sampling frequency [4], the computational efficiency will be deteriorated. The bin-wise time delay regression methods can estimate DOA with high time resolution [6], [7], [8]. For linear arrays, the time delay is denoted as the slope of the scatter figure that is plotted by the phase difference and angular frequency. However, regression methods are likely to suffer from the acoustic interference. Their acoustic robustness was conventionally addressed by weighting bin-wise delays based on SNR in [6], [8]. Both the acoustic robustness and computational efficiency were considered in conventional regression methods.

However, there are still three points to be improved. First, conventional regression methods do not work on planar arrays since the DOA can no longer be represented as the inverse triangular function of delays. Second, the period of phase difference was often ignored in the regression cost function. The cost function was generally defined as the square error between the straightforward phase difference and the DOA-derived phase difference. In theory, the phase difference error (PDE) should be limited in the range of [-π,π] since PDE is a periodical variable. This point was seldom mentioned in most regression-based methods. It should be noticed that, even if the spatial aliasing does not occur, the range of PDE should be limited. Last, the acoustic robustness is not well guaranteed. The SNR-based weight can mitigate the effect of additive noise to some extent. But SNR can not reflect that, to what extent, the source signal is deteriorated by reverberation. Moreover, the weighting method ignores the low-SNR speech components, which may be still helpful to DOA estimation.

This paper proposes a regression-based method for a planar array, where the cost function is taken as the weighted square error between the straightforward phase difference and the DOA-derived phase difference over all microphone pairs. By solving the first-order derivative of the cost function with respect to zero, DOA is represented as a linear function of all bin-wise delays and the array topology. Both the spatial aliasing and the range of PDE is considered in the proposed method. Moreover, the signal enhancement [9], [10] and weighting factor are introduced to mitigate the effect of the acoustic interference.

2. Phase difference regression

Let's consider a planar array consisting of K omni-directional microphones, and a single speech sound impinges on the array in a far-field scenario. It is assumed that the size of the array aperture is small relative to the distance from the source to the array. Therefore, the attenuation factors on all microphones are assumed to be equivalent. Denoting the desired source signal by $s(t)$, the sampled signal $y_k(t)$ received by the kth microphone is described as

$$y_k(t) = s(t - \tau_k) + n_k(t), \tag{1}$$

where t denotes the time, τ_k denotes the propagation time from the source to the kth microphone, and $n_k(t)$ denotes the acoustic interference, which comprises the additive noise and reverberation.

The short-term Fourier transform (STFT) of $y_k(t)$ is given by

$$Y_k(\omega_f) = e^{-j\omega_f \tau_k} S(\omega_f) + N_k(\omega_f), \tag{2}$$

where f denotes the frequency index, $0 \leq \omega_f \leq 2\pi$ denotes the digital frequency, and $j = \sqrt{-1}$ denotes the imaginary unit. The received signal vector is denoted as

$$\mathbf{y}(\omega_f) = \mathbf{a}(\omega_f) S(\omega_f) + \mathbf{n}(\omega_f), \tag{3}$$

where

$$\mathbf{y}(\omega_f) = [Y_1(\omega_f), \cdots, Y_K(\omega_f)]^T,$$
$$\mathbf{a}(\omega_f) = [e^{-j\omega_f \tau_1}, \cdots, e^{-j\omega_f \tau_K}]^T,$$
$$\mathbf{n}(\omega_f) = [N_1(\omega_f), \cdots, N_K(\omega_f)]^T,$$

where $(.)^T$ denotes the transpose. $\mathbf{a}(\omega_f)S(\omega_f)$ expresses the directional component. The signal enhancement is conducted on $\mathbf{y}(\omega_f)$ to suppress acoustic interference, the details of which are given in the next section. The enhanced signal is represented as

$$\mathbf{u}(\omega_f) = [u_1(\omega_f), \cdots, u_K(\omega_f)]^T.$$

There are in total $M = K(K-1)/2$ microphone pairs. For a given time-frequency (TF) bin, the mth pairwise phase difference between the pth and qth microphone can be expressed as

$$\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) = \angle u_p(\omega_f) - \angle u_q(\omega_f) + 2\pi h_{m,f}, \tag{4}$$

where $\angle(.)$ denotes the phase operation and the integer $h_{m,f}$ denotes the number of aliasing periods. $h_{m,f}$ may have several candidates for widely spaced microphones, which leads to several candidates for each phase difference. The potential phase difference is given by a set:

$$B_{m,f} = \Big\{\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big)\Big| -\frac{\omega_f d_m}{c} \leq \widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) \leq \frac{\omega_f d_m}{c}\Big\},$$

where c denotes the sound speed and d_m denotes the distance between the mth microphone pair. The cardinality of the set $B_{m,f}$ is determined by the integer $h_{m,f}$. If $|B_{m,f}| > 1$, spatial aliasing occurs.

In this paper, DOA is represented by a unit direction vector, which can be derived from the elevation and azimuth of the source. For a given unit direction vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3]^T$, phase difference can also be described by

$$\widehat{\psi}_{m,f}(\boldsymbol{\gamma}) = \omega_f d_m \mathbf{g}_m^T \boldsymbol{\gamma}/c, \tag{5}$$

where the unit vector $\mathbf{g}_m = [g_{m,1}, g_{m,2}, 0]^T$ denotes the direction of the mth microphone pair. The array topology is expressed by a set of vectors $[\mathbf{g}_1^T, \mathbf{g}_2^T, \cdots, \mathbf{g}_M^T]$ and their third dimension being set to zero indicates that all microphones lie in a plane.

Without acoustic interference, $\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big)$ is infinitely close to $\widehat{\psi}_{m,f}(\boldsymbol{\gamma})$. Under adverse environments, however, there exists an error between $\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big)$ and $\widehat{\psi}_{m,f}(\boldsymbol{\gamma})$. The error is defined as the weighted sum of PDEs' square, given by

$$\varepsilon(\boldsymbol{\gamma}) = \sum_{m=1}^{M}\sum_{f=1}^{F} w_{m,f}\Big[\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) - \widehat{\psi}_{m,f}(\boldsymbol{\gamma}) + 2\pi l_{m,f}\Big]^2 \tag{6}$$

where F denotes half STFT length and $w_{m,f}$ denotes a coefficient weighting the (m, f)-th phase difference. Because PDE is a variable with a period of 2π, the integer $l_{m,f}$ is used to limit the PDE in the range of $[-\pi, \pi]$, which is given by

$$l_{m,f} = \arg_{l}\big\{-\pi < \widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) - \omega_f d_m \mathbf{g}_m^T \boldsymbol{\gamma}/c + 2\pi l < \pi\big\}. \tag{7}$$

It should be noticed that $l_{m,f}$ is quite different from $h_{m,f}$. The latter describes the spatial aliasing, which is given by the microphone space. On contrast, the former limits the range of PDE, which should be taken into account even if the spatial aliasing does not occur. Limiting the range of PDE was seldom reported in conventional regression-based methods.

The unit direction vector is estimated by minimizing the error given by

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} &= \min_{\boldsymbol{\gamma}} \varepsilon(\boldsymbol{\gamma}), \\ \text{subjected to : } & \boldsymbol{\gamma}^T \boldsymbol{\gamma} = 1. \end{aligned} \tag{8}$$

The optimal estimator in sense of (8) is constructed by using the Kuhn-Tucker necessary condition for constrained minimization. The gradient Lagrangian equation is given by

$$Z(\boldsymbol{\gamma}, \mu) = \varepsilon(\boldsymbol{\gamma}) + \mu(\boldsymbol{\gamma}^T\boldsymbol{\gamma} - 1), \tag{9}$$

where μ is the Lagrangian multiplier. Eq. (9) can be confirmed to be a concave function with only one minimum. From $\nabla\boldsymbol{\gamma} Z(\boldsymbol{\gamma}, \mu) = 0$, the closed-form solution to the unit directional vector is given by

$$\begin{aligned} \begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} &= \Big[\sum_{m=1}^{M}\sum_{f=1}^{F} w_{m,f}\omega_f^2 d_m^2 \mathbf{g}_m^{'} \mathbf{g}_m^{'T}/c\Big]^{-1} \\ &\quad \times \Big[\sum_{m=1}^{M}\sum_{f=1}^{F} w_{m,f}(\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) + 2\pi l_{m,f})\omega_f d_m \mathbf{g}_m^{'}\Big], \\ \widehat{\gamma}_3 &= \sqrt{1 - \widehat{\gamma}_1^2 - \widehat{\gamma}_2^2}, \end{aligned} \tag{10}$$

where $\mathbf{g}_m' = [g_{m,1}, g_{m,2}]^T$. Each phase difference is weighted by the PDE

$$\delta_{m,f} = \widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) - \widehat{\psi}_{m,f}(\widehat{\boldsymbol{\gamma}}) + 2\pi l_{m,f}. \tag{11}$$

Suppose that PDE conforms a zero-mean Gaussian distribution with variance

$$\sigma^2 = \sum_{m=1}^{M}\sum_{f=1}^{F} \delta_{m,f}^2 \Big/ MF. \tag{12}$$

The normalized weight of each phase difference is derived from the likelihood as

$$w_{m,f} = \frac{\exp(-\delta_{m,f}^2/\sigma^2)}{\sum_{m=1}^{M}\sum_{f=1}^{F}\exp(-\delta_{m,f}^2/\sigma^2)}. \tag{13}$$

Phase difference outliers usually associate with larger errors and smaller weights, which leads to the decline of the importance of outliers in determining the unit direction vector.

3. Signal enhancement

The performance of signal enhancement depends on the estimate of the spatial correlation matrix. Let's consider a correlation matrix on a TF bin, which is given by

$$\mathbf{R}_\ell(\omega_f) = E[\mathbf{y}_\ell(\omega_f)\mathbf{y}_\ell^H(\omega_f)] \tag{14}$$

where ℓ denotes the frame index, $E(.)$ denotes expectation over time, and $(.)^H$ denotes the conjugate transpose. By substituting (3) into (14), the correlation matrix is written as

$$\begin{aligned} \mathbf{R}_\ell(\omega_f) &= \mathbf{R}_\ell^{(s)}(\omega_f) + \mathbf{R}_\ell^{(n)}(\omega_f), \\ &= \frac{1}{2\rho+1}\sum_{\ell'=\ell-\rho}^{\ell+\rho} |S_{\ell'}(\omega_f)|^2 \times \mathbf{a}_{\ell'}(\omega_f)\mathbf{a}_{\ell'}^H(\omega_f) + \mathbf{R}_\ell^{(n)}(\omega_f), \end{aligned} \tag{15}$$

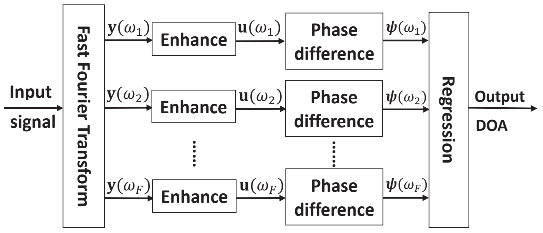

Figure 1: Block diagram of the proposed method.

where $\mathbf{R}_\ell^{(s)}(\omega_f)$ denotes the correlation matrix of the source signal, $\mathbf{R}_\ell^{(n)}(\omega_f)$ denotes the correlation matrix of the noise signal, and $2\rho+1$ denotes the number of used frames.

Let's consider three types of time-frequency bins. The first type concerns the bins that are dominated by the speech source signal, where the speech signal power is much larger than the noise power, namely $\mathbf{R}_\ell(\omega_f) \approx \mathbf{R}_\ell^{(s)}(\omega_f)$. Since $\mathbf{R}_\ell(\omega_f)$ is a rank-1 correlation matrix whose principal component is given as

$$\mathbf{u}(\omega_f) \approx e^{-j\omega_f \xi_f} \mathbf{a}(\omega_f)/\|\mathbf{a}(\omega_f)\|, \tag{16}$$

where ξ_f is an arbitrary real constant that is introduced by the complex eigenvalue decomposition. The second type concerns the bins with the noise power roughly equivalent to the speech source power. Since the correlation matrix of the source signal is a rank-1 matrix, the normalized eigenvalue of the principal component is close to 1. Most of the source energy concentrates at the principal component. On contrast, the rank of the noise correlation matrix is often much greater than 1, which indicates that the noise energy is much more uniformly distributed in the signal space. For a special case that the noise is spatially white, the principal component of $\mathbf{R}_\ell(\omega_f)$ is equivalent to that of $\mathbf{R}_\ell^{(s)}(\omega_f)$. For general noises, Eq. (16) still holds truth if the noise signal is much more spatially white than the speech signal. The third type concerns the speech-absence bins, wherein the enhancement has no negative effects on phase spectrum although it can not enhance the source signal. For the first two types of speech-presence bins, the proposed method can enhance the speech source signal, where the principal component is taken as the steering vector and the majority of the noise and reverberation remains in the subspace spanned by other eigenvectors.

4. Implementation

The block diagram of the proposed method is shown in Fig.1, where $\boldsymbol{\psi}(\omega_f) = [\widehat{\psi}_{1,f}, \cdots, \widehat{\psi}_{M,f}]^T$. Phase spectrogram is firstly enhanced by eigenanalysis, and then, phase difference of each bin is calculated. Finally, the unit direction vector is estimated by phase difference regression. The azimuth and elevation are obtained from the unit direction vector.

The spatial aliasing should be noticed when calculating the phase difference $\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big)$. According to (4), there may be several candidates for a phase difference. The optimal phase difference is selected from those candidates at each iteration. In the first iteration, the initial phase differences of the mth microphone pair are determined by using a histogram of the time delay set, denoted as $\big\{B_{m,1}/\omega_1, \cdots, B_{m,F}/\omega_F\big\}$. This approach is based on the consideration that most speech frequency components are located at low frequencies and the spatial aliasing will not occur at those frequencies. The spatial aliasing on high frequencies can be unwrapped by using the low-frequency components in the histogram. On all frequencies, the initial phase differences of the mth pair are given by

$$\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) = \arg \min_{\widehat{\psi} \in B_{m,f}} \big|\widehat{\psi} - \omega_f \varphi_m\big|, f \in \{1, \cdots, F\}, \tag{17}$$

where φ_m is the time delay with the maximal occurrence in the histogram. In other iterations, phase difference is calculated by

$$\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big) = \arg \min_{\widehat{\psi} \in B_{m,f}} \big|\widehat{\psi} - \omega_f d_m \mathbf{g}_m^T \widehat{\boldsymbol{\gamma}}/c\big|, f \in \{1, \cdots, F\}, \tag{18}$$

where $\widehat{\boldsymbol{\gamma}}$ is derived in the last iteration.

After the spatial de-aliasing is completed, $l_{m,f}$ is checked to minimize the PDE which is described by (7) in each iteration. An initial unit direction vector is firstly obtained by equally treating all phase differences, and then this vector is used to calculate new weights. The new weights are used to estimate the new unit direction vector. This iteration continues until the unit vector converges. The procedure is summarized in Algorithm.1, where ϵ is a constant greater than but close to zero.

Algorithm 1 : DOA estimation algorithm

```
Calculate spatial correlation matrix using (14), and obtain principal components by eigenvalue decomposition.
Initialize $\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big)$ using (17) and calculate $l_{m,f}$ using (7).
Initialize the unit direction vector $\widehat{\boldsymbol{\gamma}}$ by taking all weights as $1/(MF)$ using (10).
repeat.
    Let $\boldsymbol{\zeta} = \widehat{\boldsymbol{\gamma}}$, and calculate the new weights using (11), (12), and (13).
    Re-calculate $\widehat{\psi}_{m,f}\big(\mathbf{u}(\omega_f)\big)$ using (18) and re-calculate $l_{m,f}$ using(7).
    Re-calculate $\widehat{\boldsymbol{\gamma}}$ with the new weights using (10).
until $(1 - \boldsymbol{\zeta}^T \widehat{\boldsymbol{\gamma}} < \epsilon)$.
```

5. Evaluation

The proposed algorithm was tested using a 9-element circular array. One microphone was placed at the circular center and the other were uniformly distributed at the circumference. The circular radius was 0.08 m. Since the planar array is horizontally placed, the array is incapable of providing precise elevation discrimination. Therefore, the accuracy of the arrival azimuth, namely $\arctan(\gamma_2/\gamma_1)$, was used to calculate error rate, i.e., the percentage of the incorrectly estimated DOA frames, whose azimuth error was greater than a given threshold, to all frames. SRP-PHAT [11] was used as the competing algorithm. Given that the proposed algorithm utilized a 7-frame sliding window to calculate the spatial correlation matrix. For the sake of fairness, SRP-PHAT made use of 7-frame data to calculate the steered response power. SRP-PHAT performed hypothesis test at 1-degree intervals in azimuth and elevation. Both algorithms employed 32-ms windows without frame overlap. The relationship between the error rate and the error threshold of both algorithms was illustrated by a curve. PDR standing for "Phase Difference Regression" denotes the proposed algorithm. All experiments were conducted in a far-field scenario. The source signal is a continuous speech with 8000 Hz sampling rate.

In order to control the reverberation time, the room with size 7.4 × 3.4 × 2.6 meters was simulated by using an image model [12]. The continuous speech taken from the TIMIT [13] database was used as the source signal. The reverberation time,

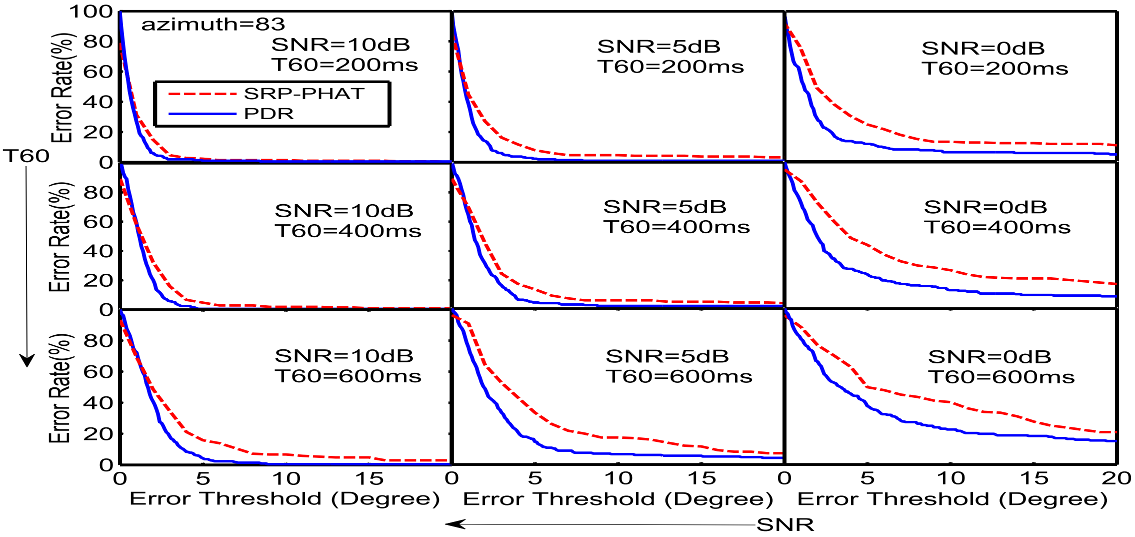

Figure 2: Error rate versus error threshold under various simulated environments.

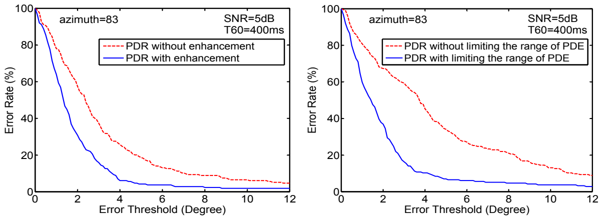

Figure 3: Error rate curves: (a) PDR performance with/without enhancement; (b) PDR performance with/without limiting the range of PDE.

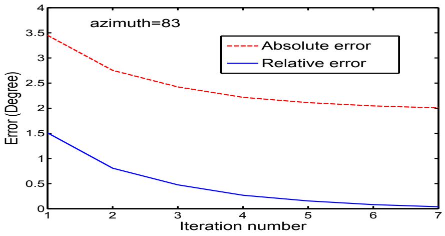

Figure 4: Convergence error versus iteration number.

T60, was respectively set as 200, 400, and 600ms. The noise, which was recorded on the side of a road with heavy-traffic by using our array, was artificially added to the simulated signal at SNR of 0, 5, 10 dB. The experimental results in Fig.2. confirm that the proposed algorithm consistently achieves better performance under all test conditions.

In addition, two cases with different enhancement methods are designed to investigate the importance of enhancement.

•**Case**"**a**": The signal are not enhanced. For the sake of fairness, the phase differences are calculated from the smoothed Fourier coefficients that are averaged over seven successive frames, given by

$$\widetilde{\mathbf{y}}(\omega_f) = \mathbf{a}(\omega_f)\widetilde{S}(\omega_f) + \widetilde{\mathbf{n}}(\omega_f), \tag{19}$$

•**Case**"**b**": The phase differences are calculated with the enhancement based on the temporal correlation, where the spatial correlation matrix is averaged over seven successive bins.

The error rate of azimuths versus error threshold are plotted in Fig.3(a). A comparison between cases "a" and "b" indicates that the eigenanalysis-based enhancement substantially improves the robustness of regression-based DOA estimation. Furthermore, The PDR performance with/without limiting the range of PDE was compared in Fig.3(b). The experiment result indicates that limiting the range of PDE is of great significance to improve the performance.

Finally, the relative convergence error and absolute convergence error averaged over all data are denoted in Fig.4. Relative error illustrates the error deviating from the final estimate, and thereby the error should be close to zero in the last iteration. The absolute error plots the error deviating from the real target. These curves indicate that the proposed algorithm converges toward the real target. The proposed algorithm converges after 4 iterations on average. The computational load of both algorithms is compared on a desktop computer. The experiments show that PDR runs three times faster than SRP-PHAT does according to the CPU time.

6. Conclusions

This paper presents a closed-form method to estimate DOA of a single speech source using a planar array. Spatial aliasing and the range of PDE are taken into consideration in the cost function, which could improve the accuracy of the DOA estimation. In order to mitigate the acoustic interference, the eigenanalysis-based method is introduced to enhance the phase spectrum. So the proposed algorithm has advantages over SRP-PHAT in both computational efficiency and acoustic robustness. The proposed method can be extended to multiple source localization, an issue that will be addressed in our future work.

7. Acknowledgment

This work was supported by the National Program on Key Basic Research Project (2013CB329302), the National Natural Science Foundation of China (Nos. 61271426, 11461141004, 91120001), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Nos. XDA06030100, XDA06030500), and by the CAS Priority Deployment Project (KGZD-EW-103-2).

8. References

[1] H. Krim and M. Viberg, “Two decades of array signal processing research: The parametric approach,” *IEEE Signal Process. Mag.*, vol. 13, pp. 67C94, 1996.

[2] J. Chen, J. Benesty, and Y. Huang, “Time Delay Estimation in Room Acoustic Environments: An Overview,” *EURASIP J. on App. Signal Process*, pp. 1C19, 2006.

[3] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, 1976.

[4] A. Pourmohammad and S. M. Ahadi, “Real time high accuracy 3-D PHAT-based sound source localization using simple 4-microphone arrangement,” *IEEE Sys-tems Journal*, vol. 6, no. 3, pp. 455–468, 2012.

[5] J. Stachurski, L. Netsch, and R. Cole. “Sound source localization for video surveillance camera,” in *IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 93–98, Aug 2013.

[6] Y. Chan, R. Hattin, and J. Plant, “The least squares estimation of time delay and its use in signal detection,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 3, pp. 217–222, Jun. 1978.

[7] M. Brandstein and H. Silverman, “A robust method for speech signal time-delay estimation in reverberant rooms,” in *Proc. IEEE Int. Conf.Acoust., Speech, Signal Process.*, Apr. 1997, vol. 1, pp. 375–378.

[8] W. Zhang and B. D. Rao, “A two microphone-based approach for source localization of multiple speech sources,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 8, pp.1913–1928 2010.

[9] Ying, D. and Yan, Y., “Robust and fast localization of single speech source using a planar array,” *IEEE Signal Process. lett.*, 20(9):909C912, 2013.

[10] Ying, D., Zhou, R., Li, J., Pan, J. and Yan, Y., “Direction-of-Arrival Estimation of Multiple Speakers Using a Planar Array,” in *INTERSPEECH*, pp. 2223–2227, 2014.

[11] J.H. DiBiase, “A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays,” Ph. D. dissertation, Brown Univ., Providence, RI, USA, May 2000.

[12] J. Allen and D. Berkley, “Image method for efficiency simulating small-room acoustics,” *J. Acoust. Amer.*, vol. 65, pp. 943–950, Apr.1979.

[13] J. S. Garofolo, “Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database,” in *Nat. Inst. Stand. Technol. (NIST)*, Gaithersburg, MD, USA, Dec. 1988.