INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Speaker Direction-of-Arrival Estimation Based On Frequency-Independent Beampattern

Feng Guo[1,3]*, Yuhang Cao*[2]*,Zheng Liu*[3]*, Jiaen Liang*[2]*, Baoqing Li*[1]*, Xiaobing Yuan*[1]

[1]Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences
[2]Beijing Unisound Information Technology Co. Ltd.
[3]HuaWei Technologies Co. Ltd.

gfeng7@mail.ustc.edu.cn,caoyuhang@unisound.com,liu.zheng@huawei.com,libq@mail.sim.ac.cn

Abstract

The differential microphone array (DMA) becomes more and more popular recently. In this paper, we derive the relationship between the direction-of-arrival (DoA) and DMA's frequency-independent beampatterns. The derivation demonstrates that the DoA can be yielded by solving a trigonometric polynomial. Taking the dipoles as a special case of this relationship, we propose three methods to estimate the DoA based on the dipoles. However, we find these methods are vulnerable to the axial directions under the reverberation environment. Fortunately, they can complement each other owing to their robustness to different angles. Hence, to increase the robustness to the reverberation, we proposed another new approach by combining the advantages of these three dipole-based methods for the speaker DoA estimation. Both simulations and experiments show that the proposed method not only outperforms the traditional methods for small aperture array but also is much more computationally efficient with avoiding the spatial spectrum search.

Index Terms: DoA estimation, Frequency-independent beampattern, differential microphone array, dipoles

1. Introduction

Speaker DoA estimation is of great importance in signal processing. Many methods have been proposed to address this problem in the past decades [1, 2]. The two-stage time-difference of arrival methods are typical ones for this task [3] with its computational efficiency. To further increase its robustness, the steered response power using the phase transform (SRP-PHAT) was proposed [4]. Nevertheless, the small aperture of microphone array will significantly degrade its performance. The steered minimum variance (STMV) was presented in [5] and it demonstrated a good performance. However, it brings a heavy computational cost.

Owing to the MEMS microphone, the microphone arrays with a small size become more and more popular [6]. In our previous work [7, 8], a small aperture microphone array is used for the DoA estimation. The MUSIC (Multiple signal classification) and its improved version SMSC-MUSIC have been employed for the DoA estimation in the wildfield. However, the performance of these methods will suffer much loss under the reverberation environment. In addition, the traditional methods usually need to search the entire spatial spectrum to find the DoA, which limits the result to the grid values as well as costs much processing resource [9]. Although the Root-MUSIC can also avoid the spectrum search by solving a polynomial, it is usually only applicable to the uniform linear array (ULA) [2].

In recent years, the DMA has gained a wide attention [10, 11]. The merits such as the small size and frequency-independent beampatterns make the DMA more and more attractive, especially for consumer electronics. The DMA is usually used as a beamfomer for signal enhancement [6]. The famous beampatterns of DMA include dipoles and cardioids. Many researches have been focused on their designs [12–14].

Nevertheless, DMA may also be used for DoA estimation. Z.L. Yu in [15] took two back-to-back cardioids to estimate the DoA. In their method, the signal was beamformed by the two back-to-back cardioids. Afterwards, two adaptive filters were used to filter the beamformed signals. The DoA was achieved by the optimal parameters of the adaptive filters. Nonetheless, this method still just uses the DMA for the signal enhancement and the DoA is determined by the optimal filter parameters which are vulnerable to the noise.

Attributing to a small size, DMA has the frequency-independent beampatterns. Our derivation demonstrates that the DoA could be directly acquired by the frequency-independent beampatterns. First, we estimate the signal subspace by using the eigenvalue decomposition (EVD) of the noisy signal covariance matrix. Then, a DoA candidate in the first quadrant would be obtained by solving a trigonometric polynomial problem. Due to the beampattern's symmetry property, we can also acquire the other candidates in other quadrants respectively. Finally, the traditional beamforming (BF) method is adopted to select the DoA from the candidates [16].

Since the dipoles have a very simple expression in the frequency-independent beampattern, three methods based on dipoles are presented. Nevertheless, the distortionless response in the reference angle is not well satisfied under the reverberation. These methods are vulnerable to these reference directions. Fortunately, their reference directions lie on the different axes. Based on this fact, we propose another new robust method by combining them. Both simulations and experiments show that the proposed method outperforms the traditional methods not only for the small aperture microphone array but also in terms of the computational efficiency.

2. DoA and Beampattern

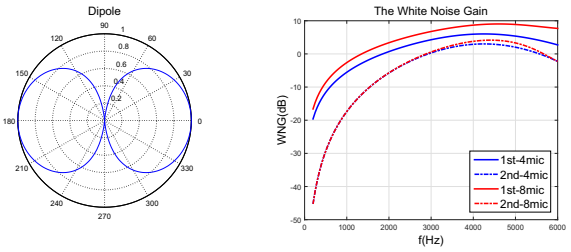
The beampattern indicates the spatial directivity of microphone array. Thus its ability of receiving signals varies with the DoA and we may derive the DoA with a known beampattern.

2.1. Signal model

Assume a far field source signal $\mathbf{S}$ with a frequency of ω_0 propagates in an anechoic acoustic environment and impinges on an M-element uniform circular array (UCA) [10]. Thus,

$$\mathbf{X} = \mathbf{d}(\theta) * \mathbf{S} + \mathbf{N}_0 \tag{1}$$

http://dx.doi.org/10.21437/Interspeech.2017-229



(a) **(b)**

Fig. 1. ***(a)*** *The dipole;* ***(b)*** *The WNG comparison.*

where $\mathbf{X}$ denotes the received signal, $\mathbf{N}_0$ is zero-mean white noise, and $\mathbf{d}(\theta) = [exp(-j\omega_0\tau_1(\theta)), \cdots, exp(-j\omega_0\tau_i(\theta)), \cdots, exp(-j\omega_0\tau_M(\theta))]^T$ is the steering vector, where $\tau_i(\theta)$ is the signal time delay between the ith microphone and the array center. Denote the weight of the beamformer by $\mathbf{h}$. The beampattern can be expressed as

$$\mathbf{B}(\theta) = \mathbf{h}^H * \mathbf{d}(\theta) \tag{2}$$

Moreover, with a small enough interelement spacing, Eq.(2) can be expressed as the frequency-independent beampattern formed by using the MacLaurin's series decomposition [11, 14]. For an Nth-order DMA, the frequency-independent beampattern is

$$\mathbf{B}(\theta) = \sum_{n=0}^{N} a_{N,n} cos^N(\theta) \tag{3}$$

where $\sum_{n=0}^{N} a_{N,n} = 1$. Specially, if $a_{N,N} = 1, a_{N,n} = 0, n < N$, Eq.(3) becomes an Nth-order dipole,

$$\mathbf{B}(\theta) = cos^N(\theta) \tag{4}$$

Taking a UCA with a radius of 2 centimeters as an example, we design the DMA using the minimum-norm filter method [10, 14]. Fig. 1(a) is the beampattern of a 2nd-order dipole. Fig. 1(b) is the comparison on the white noise gain (WNG) among the 1st-order dipole with 4 microphones, 2nd-order dipole with 4 microphones, 1st-order dipole with 8 microphones, and 2nd-order dipole with 8 microphones. It shows that the WNG could be improved by increasing the number of microphones.

The errors between the designed beampattern and the frequency-independent one are shown in Fig. 2(a). It indicates that the design errors are very small and the lower the frequency is, the smaller the differences are. This result follows the principle that a lower frequency makes the MacLaurin's series approximation more accurate. However, as shown in Fig. 1(b), the low frequency would suffer a bad WNG, the tradeoff between the design errors and the WNG should be taken into consideration if we expect a good performance of the designed DMA. Furthermore, we will derive another representation of the beampattern by the signal subspace in the following.

2.2. Beampattern and signal subspace

Denote the beamformed signal by $\mathbf{Y}$, then

$$\mathbf{Y} = \mathbf{h}^H \mathbf{X} \tag{5}$$

Thus, $E_{\mathbf{Y}} = \frac{1}{L}\mathbf{Y}\mathbf{Y}^H$ is the output power of the beamformed signal. It follows that from Eq.(1).

$$\begin{aligned} E_{\mathbf{Y}} &= \frac{1}{L}(\mathbf{h}^H\mathbf{d}(\theta) * \mathbf{S} + \mathbf{h}^H\mathbf{N}_0)(\mathbf{h}^H\mathbf{d}(\theta) * \mathbf{S} + \mathbf{h}^H\mathbf{N}_0)^H \\ &= E_{\mathbf{S}} * (\mathbf{h}^H\mathbf{d}(\theta)\mathbf{d}(\theta)^H\mathbf{h}) + \mathbf{h}^H\mathbf{R}_{\mathbf{N}_0}\mathbf{h} \end{aligned} \tag{6}$$

where $E_{\mathbf{S}} = \frac{1}{L}\mathbf{S}\mathbf{S}^H$ is the signal power, $\mathbf{R}_{\mathbf{N}_0} = \frac{1}{L}\mathbf{N_0}\mathbf{N_0}^H$, is the noise correlation matrix. For the white noise, denote the noise power of single channel by $E_{\mathbf{N}_0}$, we have

$$E_{\mathbf{Y}} = E_{\mathbf{S}} * (\mathbf{h}^H\mathbf{d}(\theta)\mathbf{d}(\theta)^H\mathbf{h}) + E_{\mathbf{N_0}}\mathbf{h}^H\mathbf{h} \tag{7}$$

With the distortionless constraint in the reference direction $0°$, $\mathbf{h}^H\mathbf{d}(0°) = 1$. Denote the WNG by G_{WN},

$$G_{WN} = \frac{1}{\mathbf{h}^H\mathbf{h}} \tag{8}$$

According to Eq.(2), Eq.(6) can be changed as follows:

$$E_{\mathbf{Y}} = E_{\mathbf{S}} * (\mathbf{B}(\theta))^2 + \mathbf{h}^H\mathbf{R}_{\mathbf{N}_0}\mathbf{h} \tag{9}$$

Moreover, we rearrange Eq.(6) as:

$$\begin{aligned} E_{\mathbf{Y}} &= \mathbf{h}^H * \mathbf{R_{XX}} * \mathbf{h} \\ &= \mathbf{h}^H * \mathbf{R_{SS}} * \mathbf{h} + \mathbf{h}^H * \mathbf{R_{N_0}} * \mathbf{h} \end{aligned} \tag{10}$$

where

$$\begin{aligned} \mathbf{R_{XX}} &= \frac{1}{L}\mathbf{X}\mathbf{X}^H = \mathbf{R_{SS}} + \mathbf{R_{N_0}} \\ \mathbf{R_{SS}} &= \frac{1}{L}(\mathbf{d}(\theta) * \mathbf{S})(\mathbf{d}(\theta) * \mathbf{S})^H \end{aligned} \tag{11}$$

Thus, from Eq.(9) and Eq.(10) we can obtain that

$$E_{\mathbf{S}} * (\mathbf{B}(\theta))^2 = \mathbf{h}^H * \mathbf{R_{SS}} * \mathbf{h} \tag{12}$$

Furthermore, $\mathbf{R_{XX}}$ can be expressed as Eq.(13) by EVD.

$$\mathbf{R_{XX}} = [\mathbf{U_S}\ \mathbf{U_N}]\mathbf{\Sigma}[\mathbf{U_S}\ \mathbf{U_N}]^H \tag{13}$$

where $\mathbf{\Sigma}$ is the diagonal matrix with the eigenvalues on its diagonal line. $\mathbf{U_S}$ represents the eigenvectors corresponding to the large eigenvalues and is called the signal subspace while $\mathbf{U_N}$ represents the eigenvectors corresponding to the small eigenvalues and is the noise subspace [9, 17]. Denote the eigenvalues by $\lambda_i, i \in [1, M]$ where λ_i are in a descending order. Define the diagonal matrix $\mathbf{\Sigma_S}$ with the diagonal line $[\lambda_1, ..., \lambda_P, \mathbf{0}_{M-P}]$ where P is the number of signal sources. Therefore, we have

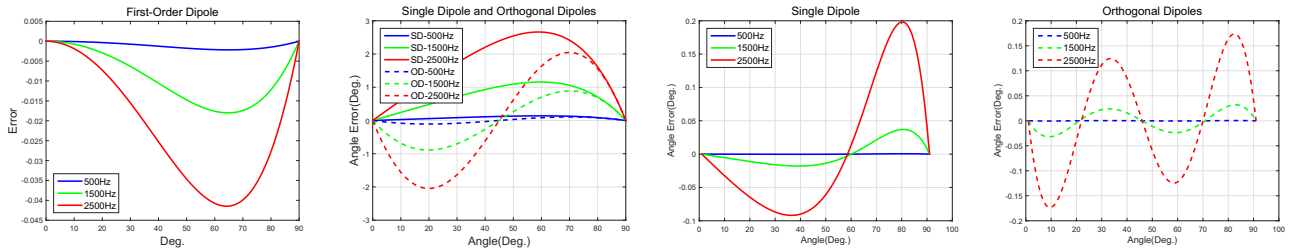
$$\mathbf{R_{SS}} = \mathbf{U_S}\mathbf{\Sigma_S}\mathbf{U_S}^H \tag{14}$$

Specially, the speech distributes sparsely in the time-frequency domain which indicates at most one speech source is prominent while the others are negligible at the same time [18]. Hence, $P = 1$ and $\mathbf{U_S}$ becomes $\mathbf{U_S^1}$ that is the eigenvector corresponding to the largest eigenvalue. In addition, we can derive that $\lambda_1 = M * E_{\mathbf{S}}$ [17]. Therefore,

$$\begin{aligned} \mathbf{R_{SS}} &= \mathbf{U_S^1}\lambda_1\mathbf{U_S^1}^H \\ &= M * E_{\mathbf{S}} * \mathbf{U_S^1} * \mathbf{U_S^1}^H \end{aligned} \tag{15}$$

It follows from Eq.(12) and Eq.(15) that

$$|\mathbf{B}(\theta)| = \sqrt{M * \mathbf{h}^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}} \tag{16}$$



(a) **(b)** **(c)** **(d)**

Fig. 2. *The errors. (**a**) is the beampattern errors between the designed one and the frequency-independent one. (**b**) is the angle errors without revision;(**c**) and (**d**) are the angle errors after revision from SD and OD respectively.*

3. Speaker DoA estimation

3.1. DoA and DMA

With a small enough interelement spacing, Eq.(16) can be written as follows by Eq.(3):

$$|\sum_{n=0}^{N} a_{N,n} cos^N(\theta)| = \sqrt{M * \mathbf{h}^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}} \quad (17)$$

Hence, we can obtain the DoA candidates by solving the trigonometric polynomial depicted in Eq.(17). As dipoles have a simple expression which is easy to acquire the solution, we use dipoles for the estimation. Specially, denote the candidate in the first quadrant by θ_1. For the Nth-order dipole,

$$\begin{aligned} \cos(\theta_1) &= \sqrt[2N]{M * \mathbf{h}^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}} \\ \hat{\theta}_1 &= \arccos(\sqrt[2N]{M * \mathbf{h}^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}}) \end{aligned} \quad (18)$$

Furthermore, using the dipole orthogonal to the first one,

$$\begin{aligned} |cos(\theta_1 + \frac{\pi}{2})| &= |sin(\theta_1)| \\ &= \sqrt[2N]{M * \mathbf{h}_O^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}_O} \\ \hat{\theta}_1 &= \arcsin(\sqrt[2N]{M * \mathbf{h}_O^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}_O}) \end{aligned} \quad (19)$$

where $\mathbf{h}_O$ is the weight of the orthogonal dipole and just a permutation of $\mathbf{h}$. Hence, the candidate can also be gotten by

$$\hat{\theta}_1 = \arctan(\sqrt[2N]{\frac{\mathbf{h}_O^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}_O}{\mathbf{h}^H * \mathbf{U_S^1} * \mathbf{U_S^1}^H * \mathbf{h}}}) \quad (20)$$

Therefore, three dipole-based DoA estimation methods are presented in Eq.(18), Eq.(19), and Eq.(20). These methods indicate that we can obtain the DoA with a known beampattern.

3.2. Estimation error reduction

However, difference usually exists between the designed beampattern and the frequency-independent one. Define $x_c = \cos(\theta_1), x_s = \sin(\theta_1), x_t = \tan(\theta_1)$ to represent the trigonometric values of ideal one. The errors between the designed and the ideal one can thus be denoted as Δx_c and Δx_s. On the other hand, denote the angle error with respect to the beampattern error by $\Delta\theta_1$. For the single dipole (SD) case,

$$\Delta\theta_1 = \arccos(x_c + \Delta x_c) - \arccos(x_c) \quad (21)$$

As $\frac{d\theta_1}{dx_c} = -\frac{1}{\sqrt{1-x_c^2}}$, we have

$$\begin{aligned} \frac{\Delta x_c}{x_s} \le |\Delta\theta_1| \le \frac{\Delta x_c}{\sqrt{1-(x_c+\Delta x_c)^2}} \quad if\ \Delta x_c \ge 0 \\ \frac{|\Delta x_c|}{\sqrt{1-(x_c+\Delta x_c)^2}} < |\Delta\theta_1| < \frac{|\Delta x_c|}{x_s} \quad if\ \Delta x_c < 0 \end{aligned} \quad (22)$$

For the orthogonal dipole (OD) pair case,

$$\begin{aligned} \Delta\theta_1 &= \arctan(\frac{x_s+\Delta x_s}{x_c+\Delta x_c}) - \arctan(\frac{x_s}{x_c}) \\ &= \arctan(\frac{x_s}{x_c}(1 + \frac{\Delta x_s x_c - \Delta x_c x_s}{x_c x_s + \Delta x_c x_s})) - \arctan(\frac{x_s}{x_c}) \end{aligned} \quad (23)$$

As $\frac{d\theta_1}{dx_t} = \frac{1}{1+x_t^2}$, we have

$$\begin{aligned} \frac{\Delta x_t}{1+(x_t+\Delta x_t)^2} \le |\Delta\theta_1| \le \frac{\Delta x_t}{1+x_t^2} \quad if\ \Delta x_t \ge 0 \\ \frac{|\Delta x_t|}{1+x_t^2} < |\Delta\theta_1| < \frac{|\Delta x_t|}{1+(x_t+\Delta x_t)^2} \quad if\ \Delta x_t < 0 \end{aligned} \quad (24)$$

where $\Delta x_t = x_t \frac{\Delta x_s - \Delta x_c x_t}{x_s + \Delta x_c x_t}$. Moreover, since $\arctan(\frac{x_c+\Delta x_c}{x_s+\Delta x_s}) + \arctan(\frac{x_s+\Delta x_s}{x_c+\Delta x_c}) = 90^\circ$, its angle errors are odd symmetrical to 45°. The errors caused by the difference between designed beampattern and the frequency-independent one are depicted in Fig. 2(b). It validates a low frequency contributes to the angle error reduction. Besides, the angle errors of OD around 45° are largely reduced in comparison with SD.

Eq.(22) and Eq.(24) indicate that although the estimation errors vary with the angles, the errors of two closed angles are similar. Based on this property, after getting $\widehat{\theta}_1$ by Eq.(17), we calculate its corresponding error $\Delta\widehat{\theta}_1$. For instance, if $\widehat{\theta}_1$ is taken by Eq.(18), $\Delta\widehat{\theta}_1$ will be computed by the corresponding equation Eq.(21). Afterwards, we estimate the real angle θ_1 by using Eq.(25). The estimation errors after the revision are shown in Fig. 2(c) and Fig. 2(d). It shows that the error is largely reduced after the revision so that it becomes negligible.

$$\theta_1 = \widehat{\theta}_1 - \Delta\widehat{\theta}_1 \quad (25)$$

3.3. Speaker DoA estimation based on DMA

Our previous work [7, 8] demonstrates that for a small aperture array, the covariance matrix of the received signal $\mathbf{R_{XX}}$ can be

estimated by Eq.(11), where $\mathbf{X}$ is the Fourier transforms of the acoustic signals. Furthermore, as the distortionless response in the reference angle is not well satisfied due to the reverberation, the DoA candidates attained by Eq.(18) or Eq.(19) (SD case) will suffer much loss from either 0° or 90° respectively, leading to an overestimate or underestimate in those directions. These results also make the DoA candidate attained by Eq.(20) (OD case) vulnerable to both the directions. Nevertheless, the OD is still much robust to the 45° owing to the symmetry.

On the other side, the DoA candidates got by the three equation Eq.(18), Eq.(19), and Eq.(20) are robust to different directions and thus can compliment each other. Besides, as dipoles are symmetrical to both the x-axis and y-axis, the three DoA candidates in the three other quadrants are $180 \pm \theta_1$ and $360 - \theta_1$. Therefore, to increase the robustness, we combine the two SD methods with the OD method and propose the SOD method to estimate the speaker DoA. The steps of the SOD are:

Speaker DoA estimation based on DMA

Preparation

1: Divide the speech into frames and choose the voice activity frames (VAF) by VAD methods [19].
2: Compute the matrix $\mathbf{R_{XX}}$ by Eq.(11) for the VAF.

DoA candidates acquisition

1: Obtain the signal subspace $\mathbf{U_S^1}$ by EVD.
2: Design an Nth-order orthogonal dipole pair and get $\widehat{\theta}_1$ by Eq.(20).
3: If $\widehat{\theta}_1 \geq 45^\circ$, obtain the $\widehat{\theta}_{1s}$ by Eq.(18) else by Eq. (19).

DoA selection

1: Revise both $\widehat{\theta}_1$ and $\widehat{\theta}_{1s}$ by Eq.(25) to get θ_1 and θ_{1s}.
2: Take the other candidates $180 \pm \theta_1$, $360 - \theta_1$, $180 \pm \theta_{1s}$, and $360 - \theta_{1s}$.
3: Use the BF to select the DoA from the candidates [16].

Although the computational costs of both the SOD and the MUSIC are $O(M^3)$ due to the EVD [9], the EVD would not take much processing resource with a small M and the SOD is much more computationally efficient than the traditional methods with removing the entire spatial spectrum search.

4. Simulations and Experiments

Both simulations and real experiments are conducted to validate the proposed SOD approach. A 4-element UCA with a radius of 2cm is used. The image method presented in [20] is used to produce the simulated speeches. The reverberation time $T60$ is 200 milliseconds. The speech segments from the TIMIT [21] are used as the source signal. The audio was resampled to 8kHz. Gaussian white noise was added. The RMSE is used to evaluate the performance.

The proposed method is compared with the SRP-PHAT [4], STMV [5], and the TCT [22]. The MUSIC method used in [7] is also used to do a comparison. The frequency used to design the DMA is 1500Hz which would get a good tradeoff. The frame size is 256. We average the RMSEs of the DoAs from 0° to 90° with an interval of 1°. Results are shown in Fig. 3.

Fig. 3(a) indicates that the SOD achieves a much better performance than the traditional methods. Both the SRP-PHAT and TCT are very sensitive to the white noise. Both the MUSIC and the STMV are comparable to the SOD only at a much higher SNR. In Fig. 3(b), the SD1 method just uses Eq.(18) to take the DoA candidates while the SD2 adopts the Eq.(19). Fig. 3(b) demonstrates that the performance of OD, SD1, and SD2 suffer much loss from the axial directions. However, their

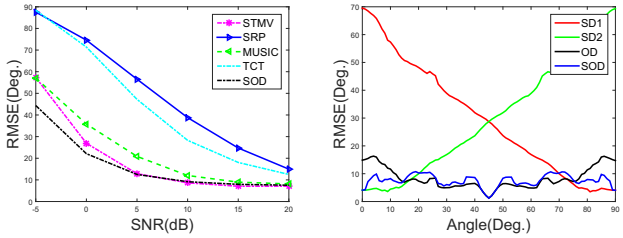


Fig. 3. *The performance comparision.* ***(a)*** *The averaged RMSEs;* ***(b)*** *The RMSEs with respect to the angle.*

combination is helpful to increase the performance due to their robustness to different angles. Therefore, the SOD performs the best with the help of the complimentary property.

Table 1. *Performance comparision.*

	STMV	**SRP-PHT**	**TCT**	**MUSIC**	**SOD**
45°	13.59°	19.09°	18.03°	18.68°	**13.22°**
90°	14.36°	13.69°	24.49°	13.84°	**9.02°**

The real experiment was conducted by a 4-element UCA with the radius of 3.5cm. The reverberation time was about 400 milliseconds measured by DIRAC. A pair of speakers were in the room located in 45° and 90° respectively. At most one speaker was speaking at the same time. The noise level is less than 45dBA. Five pairs were collected. The sampling rate was 16kHz, the audio duration is 25 minutes. The RMSEs are shown in Table 1. It is clear that the SOD achieves the best results.

Table 2. *Time Elapses.*

	STMV	**SRP-PHT**	**TCT**	**MUSIC**	**SOD**
Time(s)	846.89	34.21	2.52	1.9	**0.35**

Their computational complexities are compared in Table 2. These methods were conducted 1000 times and all the evaluations were performed using the Matlab $2014b$ on a computer (quad-core, 3.4 GHz CPU, and 8 GB memory). Results from Table 2 indicates that the proposed method is much more computationally efficient than the traditional methods.

5. Conclusion

In this paper, we investigated the relationship between the DoA and the frequency-independent beampatterns. The derived formulations indicate that we can obtain the DoA candidates by solving a trigonometric polynomial. Subsequently, we proposed three DoA estimation methods based on dipoles as the special cases. Furthermore, to increase the robustness, we combine them and propose the SOD method for the speaker DoA estimation. Both simulations and experiments show that the SOD outperforms the traditional methods for small aperture array and is much more computationally efficient. In the future work, multiple speech sources at the same time will be addressed and more experiments will also be conducted.

6. References

[1] M. R. Azimi-Sadjadi, N. Roseveare, and A. Pezeshki, "Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 4, pp. 1585–1599, Oct. 2008.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research - the parametric approach," *IEEE Signal Proc. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[3] E. V. Charles Blandin, Alexey Ozerov, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, no. 92, pp. 1950–1960, 2012.

[4] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction(src) on a large-aperture microphone array," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP*, vol. 1, Conference Proceedings, pp. I–121–I–124.

[5] J. Krolik and D. Swingler, "Multiple broad-band source location using steered covariance matrices," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 37, no. 10, pp. 1481–1494, 1989.

[6] A. Palla, L. Fanucci, R. Sannino, and M. Settin, "Wearable speech enhancement system based on MEMS microphone array for disabled people," in *Design, Technology of Integrated Systems in Nanoscale Era (DTIS), 2015 10th International Conference on*, Apr. 2015, Conference Proceedings, pp. 1–5.

[7] X. Zhang, J. C. Huang, E. L. Song, H. W. Liu, B. Q. Li, and X. B. Yuan, "Design of small mems microphone array systems for direction finding of outdoors moving vehicles," *Sensors*, vol. 14, no. 3, pp. 4384–4398, 2014.

[8] G. Feng, L. Huawei, H. Jingchang, Z. Xin, Z. Xingshui, L. Baoqing, and Y. Xiaobing, "Design of a direction-of-arrival estimation method used for an automatic bearing tracking system," *Sensors*, vol. 16, no. 7, p. 1145, Jul. 2016.

[9] S. U. Pillai and B. H. Kwon, "Performance analysis of MUSIC-type high-resolution estimators for direction finding in correlated and coherent scenes," *IEEE Trans. Acoustics Speech and Signal Processing*, vol. 37, no. 8, pp. 1176–1189, Aug 1989.

[10] J. C. Jacob Benesty and I. Cohen, *Design of Circular Differential Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2015.

[11] L. Zhao, J. Benesty, and J. Chen, "Design of robust differential microphone arrays," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1455–1466, 2014.

[12] J. Benesty, M. Souden, and Y. Huang, "A perspective on differential microphone arrays in the context of noise reduction," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 699–704, 2012.

[13] E. D. Sena, H. Hacihabiboglu, and Z. Cvetkovic, "On the design and implementation of higher order differential microphones," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 162–174, 2012.

[14] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2012.

[15] Y. Zhuliang and S. Rahardja, "DOA estimation using two closely spaced microphones," in *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, vol. 2, 2002, Conference Proceedings, pp. II–193–II–196 vol.2.

[16] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[17] F. Guo, J. Huang, X. Zhang, Y. Cheng, H. Liu, and B. Li, "A two-stage detection method for moving targets in the wild based on microphone array," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5795–5803, Oct 2015.

[18] Z. Huang, G. Zhan, D. Ying, and Y. Yan, "Robust multiple speech source localization using time delay histogram," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3191–3195.

[19] G. Aneeja and B. Yegnanarayana, "Single frequency filtering approach for discriminating speech and nonspeech," *IEEE-ACM Trans. on Audio Speech and Language Processing*, vol. 23, no. 4, pp. 705–717, 2015.

[20] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal Acoustic Society of America*, vol. 65, no. 4, p. 943, Apr. 1979.

[21] J. Garofolo, "Timit acoustic-phonetic continuous speech corpus LDC93S1," *Web Download. Philadelphia: Linguistic Data Consortium*, 1993.

[22] V. S. and K. P., "Wideband array processing using a two-sided correlation transformation," *IEEE Trans. On SP*, vol. 43, no. 1, pp. 160–172, Jan. 1995.