Research Article

Acoustic Source Localization via Subspace Based Method Using Small Aperture MEMS Arrays

Xin Zhang,^{1,2} Enliang Song,^{1,2} JingChang Huang,^{1,2} Huawei Liu,^{1,2} YuePeng Wang,^{1,2} Baoqing Li,^{1,2} and Xiaobing Yuan^{1,2}

¹ Science and Technology on Micro-System Laboratory, Shanghai Institute of Microsystem and Information Technology, CAS, Shanghai 200050, China

² Graduate University of the Chinese Academy of Sciences, Beijing 100049, China

Correspondence should be addressed to Enliang Song; songenliang@mail.sim.ac.cn

Received 26 February 2014; Revised 1 June 2014; Accepted 8 June 2014; Published 16 July 2014

Academic Editor: Xinyong Dong

Copyright © 2014 Xin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Small aperture microphone arrays provide many advantages for portable devices and hearing aid equipment. In this paper, a subspace based localization method is proposed for acoustic source using small aperture arrays. The effects of array aperture on localization are analyzed by using array response (array manifold). Besides array aperture, the frequency of acoustic source and the variance of signal power are simulated to demonstrate how to optimize localization performance, which is carried out by introducing frequency error with the proposed method. The proposed method for 5 mm array aperture is validated by simulations and experiments with MEMS microphone arrays. Different types of acoustic sources can be localized with the highest precision of 6 degrees even in the presence of wind noise and other noises. Furthermore, the proposed method reduces the computational complexity compared with other methods.

1. Introduction

Microphone array based acoustic source localization is widely used in many application scenarios. However, a big array aperture may greatly limit its applications. In recent years, due to the development of microelectromechanical system (MEMS) technology, small aperture array shows its potential superiority for portable applications. Microphone arrays for hearing aid [1], videoconferencing [2], and robotic audition [3, 4] are often used in small offices or desktop applications and meanwhile high mobility equipment is required in vehicle localization [5] and gunshot localization [6] in the battlefield. Therefore, small aperture arrays have received high attention.

In general, acoustic source localization using microphone arrays is achieved by differences of the signals received from different microphones. The two aspects of the differences are the amplitude change and the phase difference.

The aspect of the amplitude change, concerning array apertures, is related to the solution of the spherical wave

model and the plane wave model of the signal [7, 8]. Under large and medium size apertures, the spherical wave model of acoustic source is applied to microphone arrays for high performance of direction of arrival (DOA) estimation [9, 10]. However, the plane wave model can be applied to small aperture arrays for the sake of simplification and not limiting the accuracy of DOA estimation.

The aspect of the phase difference forms the essence of localization using spatial information [11]. Methods for acoustic source localization can be generally divided into three categories based on computational complexity: time delay based methods [12–14], subspace based methods, and parametric methods [15–17]. The time difference of arrival (TDOA) in time delay based methods is obtained from the phase differences of microphones [18]; it is easy to compute the TDOA if delays are known to be less than the sampling period. In [19, 20], several adaptive filtering time delay based methods are proposed to compare the performance of speaker localization using small aperture arrays. However, when the array aperture is small, time delay based methods

tend to have a high sampling rate which aggravates the burden on the hardware system [21] and the accuracy of the localization estimates depends on the system's sampling rate. The proposed method in this paper is independent of the sampling rate as long as Nyquist-Shannon sampling law is satisfied. Parametric methods are featured by high computational cost [10] and thus not suitable for real time processing, while subspace based methods such as multiple signal classification (MUSIC) [22], root MUSIC [23], and ESPRIT algorithm [24] are computationally attractive with high accuracy. As for subspace based methods, researchers focus on the model of acoustic source or new focusing matrix for wideband signals. However, with respect to small aperture arrays, positive effects on subspace based method should be applied to improve the accuracy of DOA estimation and reduce the complexity of processing. Parameters concerning the localization performance should be studied for the aim of accuracy and generality of application scenarios.

In this paper, a low complexity subspace based method is proposed for small aperture microphone array by applying compact planar model. The array aperture, frequency, and variance of the source are discussed as parameters related to localization performance. Under a small aperture, the method reduces the computational complexity as well as achieving high accuracy in general scenarios.

The remainder of this paper is organized as follows. In Section 2, the proposed method was elaborated. In Section 3, localization performance and computational complexity are compared with other representative acoustic source localization methods. Performance of the proposed method is theoretically evaluated with simulations of parameters of the acoustic signals. Experiments with MEMS microphone array are given in Section 4 and conclusions are presented in Section 5.

2. Subspace Based Localization Method Using Small Aperture Arrays

We first establish here the notations used in this paper.

- (1) The text in bold denotes vectors.
- (2) The superscript H denotes the complex conjugate of the matrix.
- (3) f denotes the frequency domain, $\omega = 2\pi f$.
- (4) $S_m(f)$ is the received signal power as a function of frequency [25].
- (5) Let M be the number of microphones in the array.
- (6) Let N be the number of acoustic sources.
- (7) $B = f_{\max} - f_{\min}$ denotes the signal bandwidth.
- (8) d is the aperture of the array.

2.1. The MUSIC Algorithm. The MUSIC algorithm is a well-known subspace based method. Generally, the DOA estimation of the acoustic sources can be obtained through the use of the plane model under a small array aperture [8, 9].

The signal $\mathbf{X}$ received at each sensor is given by

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \mathbf{W}, \quad (1)$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = [\mathbf{a}(\theta_1) \ \mathbf{a}(\theta_2) \ \cdots \ \mathbf{a}(\theta_D)] \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_D \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_M \end{bmatrix}, \quad (2)$$

$$\mathbf{a}(\theta_i) = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{Mi} \end{bmatrix} = \begin{bmatrix} e^{-j\omega_0\tau_{1i}} \\ e^{-j\omega_0\tau_{2i}} \\ \vdots \\ e^{-j\omega_0\tau_{Mi}} \end{bmatrix}, \quad i = 1, 2, \dots, D, \quad (3)$$

where $\mathbf{F}$ is the signal source, $\mathbf{W}$ is the white Gaussian noise, and $\mathbf{a}(\theta_i)$ is the array manifold which describes the difference of the same signal received at different sensors with time delays caused by spatial diversity. The covariance matrix is constructed by

$$\mathbf{S} = \overline{\mathbf{X}\mathbf{X}^H} = \overline{\mathbf{A}\mathbf{F}\mathbf{F}^H}\mathbf{A}^H + \overline{\mathbf{W}\mathbf{W}^H}, \quad (4)$$

$$\mathbf{S} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H, \quad (5)$$

where $\mathbf{E}$ is the matrix of eigenvectors. The eigenvectors can be partitioned into two subspaces, the signal subspace and the noise subspace, so we have

$$\mathbf{S} = \mathbf{E}_s\mathbf{\Lambda}_s\mathbf{E}_s^H + \mathbf{E}_n\mathbf{\Lambda}_n\mathbf{E}_n^H, \quad (6)$$

where $\mathbf{E}_n$ is an $M \times N$ matrix whose columns are the N noise eigenvectors.

Array manifold (i.e., the columns of $\mathbf{A}$) and the noise eigenvectors $\mathbf{E}_n$ are orthogonal to each other [22]. The distance between the two subspaces is defined by the MUSIC spectrum as follows:

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)}, \quad (7)$$

where $P_{\text{MUSIC}}(\theta)$ exhibits peaks in the vicinity of the true DOA.

2.2. The Proposed Localization Method Using Small Aperture Arrays. The MUSIC algorithm is based on the fact that array manifold and the noise eigenvectors $\mathbf{E}_n$ are orthogonal to each other. Wideband MUSIC algorithms for acoustic source focus on the fact that the array manifold changes as the frequency varies and hence have either to calculate all the frequencies separately (incoherent signal method, ISM) or to find a focusing matrix and transform all the frequencies to a single one (coherent signal method (CSM)). However, the two methods will greatly increase the computational burden, and therefore, they are not suitable for certain portable real time applications, whereas the power supply is limited.

The array manifold changes as the frequency varies, while the decrease in the array aperture will make the change of array manifold smaller. In other words, the error caused by frequency dispersion declines as the array aperture becomes

FIGURE 1: Plane wave propagation in a small aperture.

smaller. The approximation error of array manifold is derived below with the array aperture and other parameters. In this paper, f is assumed to be fixed at f_c rather than changing from $f_{\min}$ to $f_{\max}$, which forms the main error source of our method. The array manifold changes as the frequency varies, while the decrease in the array aperture will make the change of array manifold smaller. In other words, the error caused by frequency dispersion declines as the array aperture becomes smaller. The approximation error of array manifold is derived below with the array aperture and other parameters.

According to (3), then

$$\begin{aligned} a_{ki} &= e^{-j\omega_0 \tau_{ki}} \\ &= e^{-j2\pi f \tau_{ki}}, \quad k = 1, 2, \dots, M. \end{aligned} \quad (8)$$

The reference point is the first sensor in Figure 1; then

$$\begin{aligned} \tau_{ki} &= \frac{r_k}{v_{\text{sound}} (\cos(\theta_i - \varphi_k))}, \\ \ln a_{ki} &= \frac{-j2\pi f r_k}{v_{\text{sound}} \cos(\theta_i - \varphi_k)}. \end{aligned} \quad (9)$$

By applying the error transfer function to the variables of f and θ_i [26, 27], then

$$\begin{aligned} \frac{\delta a_{ki}}{a_{ki}} &= \left(\frac{-j2\pi r_k}{v_{\text{sound}} \cos(\theta_i - \varphi_k)} \right) \delta f \\ &+ \left(\frac{j2\pi f r_k}{v_{\text{sound}} \sin(\theta_i - \varphi_k)} \delta \theta_i \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \left(\frac{\sigma a_{ki}}{a_{ki}} \right)^2 &= \left(\frac{-j2\pi r_k}{v_{\text{sound}} \cos(\theta_i - \varphi_i) \sqrt{\text{var}(f)}} \right)^2 \\ &+ \left(\frac{j2\pi f r_k}{v_{\text{sound}} \sin(\theta_i - \varphi_i) \sqrt{\text{var}(\theta_i)}} \right)^2. \end{aligned} \quad (11)$$

The variance of f is expressed as

$$\text{var}(f) = \int_{f_{\min}}^{f_{\max}} S_m(f) (f - f_0)^2 df, \quad (12)$$

where $S_m(f)$ is the received signal power as a function of frequency [25].

With respect to the definition of array aperture d , the maximum distance between any two sensors, hence

$$r_k \leq d. \quad (13)$$

Note that (13) allows the error of array manifold caused by our assumption in (10) to be related to array aperture d . Therefore when the array aperture is small enough, the change of array manifold caused by frequency change is negligible, allowing us to simplify the localization method. With respect to (11), as the basis of array manifold, the accuracy of DOA estimation is related to f and θ_i . As θ_i varies all over the plane, the array manifold also changes. Seeking the θ_i to find orthogonal array manifold vector to noise eigenvectors is related to the performance of our method. The quotient of the first part and the second part of (11) is defined as error ratio (ER). The first part denotes error of the assumption of fixed f and the second part represents the performance of the method. When ER is smaller, the proposed method will be more effective.

Hence,

$$\begin{aligned} \text{ER} &= \frac{(-j2\pi r/v_{\text{sound}} (\cos(\theta_i - \varphi_i)) \sqrt{\text{var}(f)})^2}{(j2\pi f r/v_{\text{sound}} (\sin(\theta_i - \varphi_i)) \sqrt{\text{var}(\theta_i)})^2} \\ &= \frac{\text{var}(f)}{\tan^2(\theta_i - \varphi_i) f \text{var}(\theta_i)}. \end{aligned} \quad (14)$$

It can be derived from (14) that, with relatively smaller $\text{var}(f)$ and larger f , ER will be decreased, which makes the estimation more accurate.

Consequently, the proposed method mainly focuses on two aspects: one is to simplify the array manifolds of different frequencies to a single frequency f and the other is to choose the higher useful bands of the signal according to prior knowledge of acoustic source and thus to improve the accuracy. For a given array of M microphones, the overall localization method is now presented in a step-by-step format.

Step 1. Collect L snapshots of data from the small aperture array of M sensors.

Step 2. Associate bandwidths of acoustic source and environment to limit the processed bandwidths to relatively higher bandwidths of $[f_{\min}, f_{\max}]$; calculate the Fourier transforms of the signals of different microphones to $\mathbf{X}$.

Step 3. Construct the M covariance matrices according to $\mathbf{X}$ using (4).

Step 4. Estimate f_c which minimizes the variance of the acoustic source spectrum as follows:

$$f_c = \arg \min_{f_c} \left(\int_{f_{\min}}^{f_{\max}} \frac{S_m(f) S_m^H(f)}{B} - \frac{S_m(f_c) S_m^H(f_c)}{B} \right) df. \quad (15)$$

In the case that $S_m(f)$ is not given, $f_c = 0.5(f_{\max} + f_{\min})$.

Step 5. Calculate $\mathbf{A}$ from (3) with f_c .

Step 6. Isolate the source locations as the maxima of the pseudospectrum $P_{\text{MUSIC}}(\theta)$ defined in (7).

3. Analysis and Simulation

3.1. Theory Analysis of Array Aperture and Other Factors. Traditionally, the acoustic sources such as vehicle, music, and human talking are treated as wideband sources. The method for acoustic localization includes ISM and CSM. Among them, the SSA, RSS, and TCT are representative methods. TCT have the best performance. The basic idea for TCT is focusing. However, this will greatly improve the computation. Instead of focusing, we choose a proper approximation for f_c without focusing on the cost of some performance of localization for computational complexity.

3.1.1. The Selection of the Focusing Frequency f_c . Concerning the simple uniform distribution,

$$s(f - f_0) = \begin{cases} \frac{1}{2B}, & |f - f_0| \leq B \\ 0, & |f - f_0| > B. \end{cases} \quad (16)$$

A reasonable approximation of f_c is $f_c = f_0$ [28, 29]. Concerning the approximation of other distributions,

$$\varepsilon = \sum_{j=1}^J \left\| \mathbf{A}(f_c) (\mathbf{R}_s(f_c) - \mathbf{R}_s(f_j)) \mathbf{A}^H(f_c) \right\|^2 \quad (17)$$

$$f_c = \min_{f_c} (\varepsilon).$$

ε is the error of approximation. When $\mathbf{R}_s(f_c) = \frac{1}{J} \sum_{j=1}^J \mathbf{R}_s(f_j)$, the error is minimum [30, 31]:

$$\begin{aligned} \mathbf{R}_s(f_c) &= S_m(f_c) S_m^H(f_c), \\ \mathbf{R}_s(f_c) &= \frac{1}{J} \sum_{j=1}^J \mathbf{R}_s(f_j) \\ &\stackrel{J=\infty}{=} \frac{1}{B} \int_{f_{\min}}^{f_{\max}} S_m(f) S_m^H(f) df. \end{aligned} \quad (18)$$

According to (17),

$$f_c = \arg \min_{f_c} \left(\int_{f_{\min}}^{f_{\max}} \frac{S_m(f) S_m^H(f)}{B} - \frac{S_m(f_c) S_m^H(f_c)}{B} \right) df. \quad (19)$$

When there is no prior knowledge about $S_m(f)$, $f_c = 0.5(f_{\max} + f_{\min})$ will be a compromising choice.

3.1.2. Discussion of Other Factors. The essence of the algorithm in this paper is that when the aperture of array is very small, the acoustic signal which is band-limited could be viewed as narrowband signal or approximate narrowband signal. However, there is no explicit definition of narrowband signal for array signal. The 3 expressions listed below show some of the characteristics of narrowband array signal which is widely accepted [32]:

- (i) $B/f_0 \ll 1$;
- (ii) $d/V_{\text{sound}} \ll 1/B$;
- (iii) if the bandwidth of a signal is such that the second eigenvalue of the signal's noise free covariance matrix is larger than the noise level in the signal plus noise covariance matrix, then that signal may not be described as narrowband [25].

The first expression is the basic understanding of narrow-band signal. The second expression describes the narrowband signal for array signal. The third definition is more complicated. The definition of narrowband signal is a function of bandwidth, AOA, SNR, array gain, and dimension of the array.

The conclusion drawn from the 3 definitions is consistent with Section 2.2 that, with smaller aperture (d), higher signal frequency (f), and smaller bandwidth ($\text{var}(f)$), the signal is narrower. In this case, the localization error caused by frequency error is smaller so that our method is more effective. To improve the performance for localization, we could use a band-pass filter to the signal and also make use of relatively higher band of the signal; after all, when the array aperture is very small, the error of acoustic source localization could be under control.

3.2. Simulations. Comparison with other methods is performed in this part. And the error of DOA caused by frequency error with aperture d , signal frequency f , and signal power variance $\text{var}(f)$ will also be simulated.

3.2.1. Comparison with Other Algorithms. Comparing our method with TCT(CSM) [30], RSS(CSM) [33], SSA(ISM) [34], and TDE method in RMSE and computational complexity.

Considering one source impinges on the array from 108.5 degrees. The aperture of the uniform circular array is 1/3 and 1/6 of the wavelengths of the center frequency. The bandwidth

FIGURE 2: RMSE of the DOA estimation versus SNR for the different methods. The aperture of the uniform circular array is (a) half and (b) 1/5 of the wavelengths of the center frequency.

TABLE 1: Time elapses for different methods.

TCT	RSS	SSA	TDE	Ours
0.0724 s	0.0614 s	0.1020 s	0.0022 s	0.0074 s

of the signal is 20% of the center frequency. RMSE is defined as

$$\theta_{\text{RMSE}} = \sqrt{E \left(\frac{1}{K} \sum_{k=1}^K |\theta_k - \hat{\theta}_k|^2 \right)} \quad (20)$$

based on 500 Monte Carlo experiments. In Figure 2, the RMSE of TCT and RSS methods are smaller. Our method has nearly the same performance as the SSA method and much smaller performance than the TDE method. When the SNR is higher than 10 dB, the RMSE of all the subspace based method are nearly the same. Comparing Figures 2(a) and 2(b), when the aperture of the array is smaller, the performances of the methods are closer.

The computational complexity is compared in Table 1. The snapshots of the signal are 1024, the sampling rate is 8192 Hz, the source is from 108.5 degrees, and the array is a uniform circular array with radius of 0.02 m. All of the methods are executed in the Matlab 2008a environment on a personal computer (dual core, 2.9 GHz-frequency processor, and 2 GB memory). The time elapses are shown in Table 1.

As shown in Table 1, the computation shows an order of magnitude less than the other TCT, RSS, and SSA. The proposed algorithm was also applied in ADSP21375 processor, with 75 MHz clock, the signal snapshots of 1024, and sampling rate of 8192 Hz. The DSP system could work out the DOA of the acoustic source in real time with DOA error less than

6 degrees. The computation process of every 1024 snapshots takes no longer than 0.05 s.

The performance of method is not worse than ISM and CSM under small aperture. However, the computational complexity is much lower than ISM and CSM.

3.2.2. The Aperture. As is discussed in Section 2.2, the frequency error will result in manifold error. However, with the decrease in array aperture, the error will be smaller. We focus on the frequency error that is defined as the difference between f used in the simulation method and f_c defined in Step 4.

The simulation signal is a wideband signal with bandwidth ranging from 200 to 1000 Hz. The real DOA is 108.5 degrees. The array is a uniform circular array. Table 2 shows the error frequency boundary which is defined as the frequency error when the difference between the estimated DOA and the real DOA is more than 10 degrees for the first time.

As is shown in Table 2, when the array aperture is less than 0.05 m, the error frequency boundary will be larger than 2000 Hz. This means if we confine the input signal to a certain bandwidth by analog or digital means, the DOA error caused by frequency error could be under control. In this case, the prior knowledge of the signal is not necessary, which could broaden the use of our method. Figure 3 shows the relationships between frequency error (x -axis), array aperture (y -axis), and DOA error (z -axis). At the same aperture, the DOA error is increased as the frequency error grows. At the same frequency error, the DOA error is reduced with the decrease of aperture size.

3.2.3. The Signal Frequency. We conclude in Section 2.2 that, with the increase of signal frequency, the DOA error caused

TABLE 2: Array aperture and error frequency boundary.

Array aperture (m)	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13
Error frequency boundary (Hz)	>2000	>2000	>2000	>2000	1584	1284	1058	884	764	634	542	464

FIGURE 3: Simulation of DOA error versus signal frequency and frequency error.

FIGURE 4: Simulation of DOA error versus signal frequency and frequency error.

by frequency error will be decreased. The simulation is done as follows. The array aperture is 0.02 m; the signal bandwidth is 400 Hz. The signal frequency is the center of the bandwidth. Figure 4 shows the relationships between frequency error (x -axis), signal frequency (y -axis), and DOA error (z -axis). At the same frequency error, the DOA error is decreased with the increase of signal frequency. The proposed algorithm is used to estimate DOA by choosing frequency error from 0 to 1000 Hz stepped by 2 Hz. The RMSE of DOA is shown in Figure 5. RMSE is reduced with the increase of signal frequency.

3.2.4. Variance of the Signal Power of Frequency. As presented in Section 2.2, as $\text{var}(f)$ is decreased, the error caused by frequency error is reduced. Two signals $s_1(t)$ and $s_2(t)$ are generated, whose powers are shown in Figure 6(a). The array is uniform circular array; the real DOA of the signals is 108.5 degrees. Obviously, the variance of signal 1 is larger than signal 2.

As shown in Figure 6(b), the estimated DOA of $s_2(t)$ is more accurate than that of $s_1(t)$ at the same frequency error, which supports our derivation.

3.3. Negative Influence of Small Aperture Array. Our method is based on the fact that small array aperture could decrease the manifold error caused by the frequency error. However, some negative influences of small aperture should be mentioned.

3.3.1. High Consistency in Amplifying Circuit. In a real microphone array system, the difference of the amplifying circuit cannot be avoided and will cause amplitude error and phase difference for different array elements, which will affect the array manifold.

Based on the Nyquist-Shannon sampling theorem, then

$$f_s \geq 2f_0 \quad (21)$$

and space sampling theorem is expressed as follows [35]:

$$d \leq 0.5\lambda, \quad (22)$$

where d is the minimum distance of any two array elements and λ is the shortest wavelength of the signal. Just as in time domain, higher space sampling rate means more information is kept, but it will also cost more in system design. $d > 0.5\lambda$ will cause an aliasing effect as described by Nyquist-Shannon sampling theorem. When the array size is not limited, 0.5λ is always chosen as the array aperture. This is because with the increase of the array aperture size, the tolerance of phase difference introduced by the amplifying circuit is increased.

The DOA errors in Figure 7 are caused by 1 degree of phase difference introduced by the amplifying circuit. The signal bandwidth is 400 Hz; the center of the band is 500 Hz. The array is a uniform circular array. Figure 7 shows that, with the increase of the array aperture size, the DOA error is reduced.

In other words, in order to achieve high accuracy, the small aperture array should be highly consistent in the amplifying circuit.

3.3.2. Lower Resolution. Another question caused by high space sampling rate (small array aperture) is the resolution, as for multiple signals, it is more difficult to separate closely spaced signals. When the array size is not limited, 0.5λ is always chosen as the array aperture to achieve higher resolution. However, the algorithm still works well for single signal.

Figure 8(a) shows the MUSIC spectra for two signals from 50 degrees and 60 degrees, by using a uniform linear array with 4 elements. λ is the wavelength of the signal. The distances between every two elements are 0.02λ , 0.1λ , and 0.5λ . Figure 8(b) shows the MUSIC spectra for one signal from 60 degrees, by using a uniform linear array with 4 elements.

Even though the resolution of small aperture is lower than that of large aperture, our method is still able to distinguish multiple signals. It works well for single signal.

FIGURE 5: RMSE versus signal frequency at different SNR.

FIGURE 6: (a) Signal power spectra of $s_1(t)$ (left) and $s_2(t)$ (right). (b) Estimated DOAs at different frequency errors for $s_1(t)$ and $s_2(t)$.

FIGURE 7: DOA errors under different array apertures.

FIGURE 8: (a) The MUSIC spectra for two signals of different aperture. (b) The MUSIC spectra for one signal of different aperture.

TABLE 3: Six arrays for validation of the algorithm.

Number	Array	Number of array elements	Aperture (m)
1	Uniform circular	4	0.04
2	Uniform circular	4	0.02
3	Uniform circular	4	0.005
4	Equilateral triangle	3	0.04
5	Equilateral triangle	3	0.02
6	Equilateral triangle	3	0.005

4. Experiment with MEMS Microphone Arrays

Several experiments are carried out by real microphone array in this section. MEMS microphone is chosen as the sensor to ensure the array aperture is small enough. The package dimensions of MEMS microphone below are $3.35 \times 2.5 \times 0.88$ mm.

4.1. Validation of the Method Using Different Arrays. After the array for localization was properly designed, six arrays were proposed to validate the algorithm with different shapes. Table 3 and Figure 9 list all of them.

A turntable was designed to test the changes of DOA. It was driven by a stepping motor controlled by a micro-controller with a constant rotating speed of 25.71 degrees/s. The acoustic source is a famous Chinese folk song Jasmine (Molihua) played by piano. The sampling rate of the array system is 8192 Hz, the band is limited to 200–2000 Hz, and f_c is 800 Hz. The experiment was conducted in a quiet room and the source was fixed to 0 degrees to the array at the distance of 2 m.

As shown in Figure 10, when the speed of the turntable is constant, the estimated DOA varies linearly with time. The slope means the velocity of the turntable. The intercept is the deviation from the real DOA when the turntable starts to revolve, and the deviation could be used to correct the estimated DOA as a whole. A relevant coefficient describes how the result obeys the linear law with 1 meaning 100%.

FIGURE 9: MEMS arrays in Table 3.

FIGURE 10: (a) The song Jasmine in frequency domain. (b) DOA estimated from array number 1.

Standard deviation means the DOA error. As shown in Table 4, in comparison with the results between arrays number 1, number 2, and number 3, and number 4, number 5, the error and the intercept are increased with the decrease in the array aperture. This is because of the inconsistency of the amplifying circuit. As mentioned in Section 3.3.1, the consistency of the amplifying circuit is especially important for small aperture array.

In comparison of arrays number 1, number 4 and number 2, number 5, a larger number of sensors result in better estimated DOA, because the former acquires more information. However, array number 6 could not be localized, because in order to minimize the aperture of an equilateral triangle, the sound hole was arranged as an equilateral triangle, but the MEMS microphone array was not equilateral. For the

position of the MEMS microphone that sense sound may not be the exact sound inlet, the shape of the array is not equilateral, which leads to the indeterminacy of the array manifold and the invalidity of the estimated DOA. Nevertheless, the result of number 3 shows that our method could work at very small aperture as 5 mm.

4.2. Validation of the Method Using Different Acoustic Sources.

An application of the proposed method to locate moving vehicles using a 0.04 m microphone array (array number 1) will be elaborated in this part. The experiment was conducted on a cement road after a light rain with wind levels from 0 to 2 during November 2012, on an island in Zhejiang Province, China. l is the vertical distance from the array to the car, and v is the velocity of the cars. The velocities of vehicles can be

TABLE 4: Results of linear fit of DOA for the arrays in Table 3.

	Number 1	Number 2	Number 3	Number 4	Number 5	Number 6
Intercept (degree)	1.17304	2.55845	5.44434	3.41981	7.40605	N/A ¹
Slope (degree/s)	26.05438	25.21226	24.65839	26.92926	26.29049	N/A
Relevant coefficient	0.99979	0.99935	0.99786	0.99924	0.99571	N/A
Standard deviation (degree)	2.106	3.74889	6.64365	4.13669	9.87195	N/A

¹N/A means not available because the DOA does not change with time linearly.

TABLE 5: Results of fit of DOA.

	a (degree)	V (m/s)	t_0 (s)	l (m)	Deviation of DOA (degree)	Relevant coefficient
E-bicycle ¹	89.28	10.23	13.36	3	5.82	0.99247
Tricycle	90.43	10.55	14.98	3	5.56	0.99598
Car	90.78	22.10	8.70	5	2.76	0.99958
Truck	91.71	13.50	16.52	5	3.61	0.99812

¹E-bicycle means electric bicycle.

FIGURE 11: Scheme of the localization experiment.

regarded as uniform. The car is running from left to right (Figure 11); then

$$\theta = \frac{\pi}{2} + \arctan\left(\frac{v(t_0 - t)}{l}\right), \quad t \in R. \quad (23)$$

Four typical types of vehicles are used for localization: an electric bicycle, a tricycle, a car, and a truck. Their signals are shown on (A) in Figure 12 and the signal power spectrum amplitudes are shown on (B) in sequence. A nonlinear fit of DOA is conducted with the fit function equation (23). The estimated DOAs of the four vehicles are depicted on (C) in Figure 12.

Considering the width of the road, as the driver drives on the right side, l is 3 m when the vehicle is driving from left to the right; l is 5 m when the vehicle is driving from the opposite side. The results of the fit of DOA for different vehicles are shown in Table 5.

V represents the relative speed of the acoustic source. The DOA estimation error is residual mean square, which is a standard for nonlinear fit of DOA. The relevant coefficient is a parameter that describes how the data fit (23) with 1 meaning

100%. The results of the fit of DOA for different vehicles are shown in Table 5.

According to our initial parameter analysis in Section 3.2, the performance of acoustic localization estimation is related to the signal frequency and SNR of the acoustic signal. The four types of vehicles consist of different combinations of signal frequencies and SNRs. The frequency spectra of the four vehicles are shown on (B) in Figures 12(a), 12(b), 12(c), and 12(d) and are related to the speed of the vehicles [36]. Sorting the vehicles in a descending order by frequency spectra, the order is as follows: car, truck, electric bicycle, and tricycle. Sorting the vehicles in a descending order by SNRs, the order is as follows: truck, tricycle, car, and electric bicycle. The frequency spectra of vehicles have a significant effect on the performance of the localization estimation. With the highest frequency spectrum and the third highest SNR, the car showed the best localization performance, followed closely by the truck, ranking as the second highest in frequency and highest in SNR. Taking into account the fact that the frequency spectra of electric bicycle and tricycle are nearly the same, despite the fact that they are different in SNRs, the performances are nearly the same.

In the application scenarios considered in this part, the proposed method has been applied to more than 200 experiments for over 20 types of vehicles. The maximum estimation error of DOA is no bigger than 6° . According to the results in this section and computational complexity comparing with other methods in Section 3.3, on one hand, the proposed method is excellent in accuracy of localization; on the other hand, the method is suitable for real time processing for portable applications.

5. Conclusion

In this paper, an acoustic source localization method using small aperture array was proposed. The method could be used in real time processing systems such as vehicle localization and videoconference due to low computation complexity. As the proposed algorithm is featured by small aperture, array response obeys a plane wave propagation model and the

FIGURE 12: Continued.

FIGURE 12: Acoustic signals (A), spectrum amplitudes (B), and fits of DOA (C) of (a) electric bicycle, (b) tricycle, (c) car, and (d) truck.

consistency of the amplifying circuit is very important, which is often ignored for the discussion of ordinary arrays. The accuracy of DOA estimation could be improved by taking advantage of the source's relatively higher band and confining the signal to a certain band. Simulations and experiment show that the proposed method can effectively localize different acoustic sources and has a lower computation complexity compared with the existing methods.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the associate editor and anonymous reviewers for their valuable comments and suggestions to improve this paper. This work has been supported by fund 9140C18020211ZK34.

References

- [1] A. Wang, K. Yao, R. E. Hudson et al., "High performance microphone array system for hearing aid applications," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)*, pp. 3197–3200, Atlanta, Ga, USA, May 1996.
- [2] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 19–22, 1997.
- [3] F. Keyrouz, W. Maier, and K. Diepold, "Robotic localization and separation of concurrent sound sources using self-splitting competitive learning," in *Proceedings of the IEEE Symposium on Computational Intelligence in Image and Signal Processing*, pp. 340–345, April 2007.
- [4] M. R. Bai and C. Lin, "Microphone array signal processing with application in three-dimensional spatial hearing," *Journal of the Acoustical Society of America*, vol. 117, no. 4 I, pp. 2112–2121, 2005.
- [5] X. Zhang, J. Huang, E. Song, H. Liu, B. Li, and X. Yuan, "Design of small MEMS microphone array systems for direction finding of outdoors moving vehicles," *Sensors*, vol. 14, no. 3, pp. 4384–4398, 2014.
- [6] A. K. Bandi, M. Rizkalla, and P. Salama, "A novel approach for the detection of gunshot events using sound source localization techniques," in *Proceedings of the IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS '12)*, pp. 494–497, IEEE, Boise, Idaho, USA, August 2012.
- [7] F. Asano, H. Asoh, and T. Matsui, "Sound source localization and separation in near field," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 83, no. 11, pp. 2286–2294, 2000.
- [8] J. C. Chen, K. Yao, and R. E. Hudson, "Acoustic source localization and beamforming: theory and practice," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 4, Article ID 926837, pp. 359–370, 2003.
- [9] A. Manikas, Y. I. Kamil, and M. Willerton, "Source localization using sparse large aperture arrays," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6617–6629, 2012.
- [10] T. Ju, Y. Xu, and Q. Peng, "Speech source localization in near field," in *Proceedings of the International Conference on Communications, Circuits and Systems*, vol. 2, pp. 769–772, Chengdu, China, June 2004.
- [11] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1 of *Springer Topics in Signal Processing*, Springer, Berlin, German, 2008.
- [12] A. H. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 527–533, 1981.
- [13] Z. H. Michalopoulou and R. Jain, "Particle filtering for arrival time tracking in space and source localization," *Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3041–3052, 2012.
- [14] N. Imam and J. Barhen, "Acoustic source localization via time difference of arrival estimation for distributed sensor networks using tera-scale optical core devices," *Journal of Sensors*, 2009.
- [15] J. Bohme, "Estimation of source parameters by maximum likelihood and nonlinear regression," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '84)*, vol. 9, pp. 271–274, March 1984.
- [16] J. L. Odom, J. L. Krolik, and J. S. Rogers, "Maximum-likelihood spatial spectrum estimation in dynamic environments with a short maneuverable array," *Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 311–322, 2013.
- [17] P. Stoica and A. Nehorai, "MUSIC, maxium likelihood and Cramer- Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 17, no. 5, pp. 2296–2299, 1989.
- [18] N. L. Owsley and G. R. Swope, "Time delay estimation in a sensor array," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 519–523, 1981.
- [19] I. McCowan and D. Moore, *Small Microphone Array: Algorithms and Hardware*, IDIAP Communication, 2003.
- [20] P. S. Chang and A. N. Willson Jr., "Performance of 3D speaker localization using a small array of microphones," in *Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 328–332, November 1997.
- [21] Y. T. Chan, R. V. Hattin, and J. B. Plant, "The least squares estimation of time delay and its use in signal detection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 3, pp. 217–222, 1978.
- [22] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [23] B. D. Rao and K. V. S. Hari, "Performance analysis of Root-Music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.
- [24] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, pp. 296–312, 1990.
- [25] M. Zatman, "How narrow is narrowband," *IEE Proceedings: Radar, Sonar and Navigation*, vol. 145, no. 2, pp. 85–91, 1998.
- [26] A. L. Shestakov, "Dynamic error correction method," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 1, pp. 250–255, 1996.
- [27] T. Hilaire, "On the transfer function error of state-space filters in fixed-point context," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 56, no. 12, pp. 936–940, 2009.
- [28] Y. U. Lee, "Distributed source modeling and direction-of-arrival estimation techniques," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 960–969, 1997.

- [29] T. P. Jantti, "The influence of extended sources on the theoretical performance of the MUSIC and ESPRIT methods: narrowband sources," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Volumes 1–5 (ICASSP '92)*, pp. B429–B432, IEEE, San Francisco, Calif, USA, March 1992.
- [30] S. Valaee and P. Kabal, "Wideband array processing using a two-sided correlation transformation," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 160–172, 1995.
- [31] "The methods and algorithm of spatial spectrum estimation".
- [32] H. Yu, J. Liu, Z. Huang, and Y. Zhou, "Modeling method and simulation of wideband signals received by an array antenna," *Electronic Warfare*, vol. 109, no. 4, pp. 7–12, 2006.
- [33] H. Hung and M. Kaveh, "Focussing matrices for coherent signal-subspace processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 8, pp. 1272–1281, 1988.
- [34] G. Su and M. Morf, "The signal subspace approach for multiple wide-band emitter location," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 6, pp. 1502–1522, 1983.
- [35] M. A. Doron and E. Doron, "Wavefield modeling and array processing .I. Spatial sampling," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2549–2559, 1994.
- [36] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave patterns," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 30–47, 2009.