Hindawi
Journal of Sensors
Volume 2019, Article ID 4639850, 13 pages
https://doi.org/10.1155/2019/4639850

Research Article

A Sound Source Localization Device Based on Rectangular Pyramid Structure for Mobile Robot

Guoliang Chen and Yang Xu

School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, China

Correspondence should be addressed to Guoliang Chen; glchen@whut.edu.cn

Received 5 April 2019; Revised 7 June 2019; Accepted 24 July 2019; Published 27 August 2019

Academic Editor: Calogero M. Oddo

A sound source localization device based on a multimicrophone array with the rectangular pyramid structure is proposed for mobile robot in some indoor applications. Firstly, a time delay estimation method based on the cross-power spectral phase algorithm and a fast search strategy of peak value based on the geometric distribution of microphones are developed to estimate the sound propagation delay differences between two microphones. Moreover, a rejection strategy is presented to evaluate the correctness of the delay difference values. And then, the device's geometric equations based on the time-space mapping relationship are established to calculate the position of the sound source. For fast solving the equations, the multimicrophone array space is divided into several subspaces to narrow the solution range, and Newton iteration algorithm is introduced to solve the equations, while its solution is evaluated by an evaluation mechanism based on coordinate thresholds. Finally, some experiments are carried out to verify the performance of the device, of which the results show that the device can achieve sound source localization with a high accuracy.

1. Introduction

As an important branch of the robot family, mobile robots for home services are definitely the trend that must lead a new life style of human being [1]. Compared to industrial robots, the kind of robots is required to offer safer, more flexible, and more intelligent services to human being because they serve humans directly. So, home service robots are paid higher requirements of intelligence. Robot audition stimulated by human hearing capabilities which is an important part of robot intelligence has attracted great attention of many scholars in recent years. Numerous works have been proposed by a growing community, with contributions ranging from sound source localization (SSL) and separations in realistic reverberant conditions to speech or speaker recognition [2].

In contrast to vision, robot audition has its unique advantages in perception. Robot's perception of sound is nearly omnidirectional and independent of the lighting conditions. Similarly, robots are able to detect sound signals even in the presence of obstructions [3, 4]. Robot audition can not only locate sound sources, but also recognize their origins and interpret their contents, which is a prominent capacity for human-robot harmonious interaction. Therefore, robot audition has broad application prospects [5]. For example, home service robots are expected to receive voice commands from elderly people with limited exercise and provide corresponding services for them. On the other hand, robot audition can help a robot to recognize its environment and then achieve some special tasks, such as detecting and tracking an abnormal sound to prevent the occurrence of dangerous situation in time [6, 7].

SSL technology allows a robot determine the direction and position of a sound source using only sound data, which is essential in the overall scheme of robot audition, and has an important impact on other robot audition modules[8]. In order to achieve a higher localization accuracy, SSL systems usually need many microphones with a complex array, which increases the computational complexity greatly. So, there is a trade-off between the localization accuracy and the complexity of microphone array. A SSL system for home service robots has several unexpected constraints. It is mounted on a robot platform, which limits its size and the number of microphones. Moreover, in order to meet the requirements of real-time human-robot interaction and target tracking, the

location algorithm adopted by SSL system should have a low computational complexity and high computational accuracy [9]. Therefore, the trade-off is particularly important for the mobile robots' SSL system.

Taking the localization of real-time and high accuracy and a relatively simple structure into account, this paper designs a new microphone array with a regular quadrature pyramid structure for mobile robots in indoor environment. The proposed microphone array provides a nonplanar reference point with the vertex of the regular quadric pyramid, which can help the array to obtain sound source signals and improve the localization accuracy. According to the microphone array, this paper develops its estimation model of SSL and computing method and proposes a double screening mechanism to improve the reliability of position results.

The rest of the paper is as follows. Section 2 introduces some of the relevant works in the field of robot audition. Section 3 describes the principle of time delay estimation. Section 4 presents the process of the localization algorithm based on iterative optimization. Section 5 shows the experimental prototype and receives the final experimental results and then analyzes the experimental errors. The last section obtains the conclusion with plans for the future studies.

2. Related Work

Irie [10] applies SSL technology to mobile robots for the first time. Since then, a number of methods have been proposed to enrich SSL technology. At present, there are three kinds of well-known methods for SSL based on microphone array, including (1) directional technology based on high-resolution spectral estimation (HRSA); (2) controllable beamforming technology based on the biggest output power (BS); (3) technology based on time difference of arrival (TDOA) [11].

All above SSL methods are always based on a certain number of microphones. Pavlidi et al. [12] present a uniform circular microphone array to overcome the ambiguities of linear arrays. Cho et al. [13] develop a SSL system for mobile robots that uses a square microphone array of 0.17×0.17 m^2 with four sensors attached on the shoulder of a plaster cast of the small size home robot. Ren et al. [14] use a triangular pyramid microphone array with four omnidirectional microphones for multiple sparse source localization. Its lateral faces are isosceles right triangles. Each omnidirectional microphone is placed at the vertex of the triangular pyramid, and the sensor located at origin point is taken as the reference sensor. Valind et al. [15] construct an 8-microphone array with a cuboid structure to locate the axial angle and elevation angle of a sound source. Huang et al. [16] use four microphones to form a three-dimensional microphone array to locate axial and elevation angles of sound sources.

The above-mentioned microphone arrays have a regular structure, such as linear, triangular, polygonal, circle, and polyhedral arrays, which have the ability to locate sound sources in two-dimensional and three-dimensional space, respectively. The number of microphones and their topology in SSL system mainly depend on the SSL method adopted. Generally, the number of microphones is required by TDOA, HRSA, and BS increases successively.

Among these SSL localization methods, TDOA method is more suitable for robot auditions, in which the azimuth and horizontal distance of a sound source should be determined in real-time. So, the localization methods considered here are all based on TDOA technique. The basic idea of SSL based on TDOA technique is using the observed time difference between signals of a sound source arriving two microphones to construct a hyperboloid in space, on which the location of the sound source can be constrained to lie [17]. According to the idea, TDOA-based localization technique can be divided into two steps, including (1) the step of time delay estimation (TDE) to compute the time delays between the sound source and each sensor of a microphone array; (2) the step of localization to compute the direction and position of the sound source based on the geometric model of the microphone array.

The accuracy of TDE estimation is related to the performance of a SSL system. Some research results show, as the source is moved further away from the robot, an important error growth in distance estimation, as opposed to azimuth and elevation estimation [6]. The generalized cross-correlation (GCC) method weighted by the phase transform (PHAT) is by far the most used in TDE technique for single direction-of-arrival estimation in robot audition because of its robustness and easy implementation. The paper just gives some related research works based on GCC-PHAT method.

Perez-Lorenzo et al. [18] evaluate some GCC methods in typical conditions of real scenarios with the presence of external, stationary, and in most cases, correlated noise sources. Chen et al. [19] analyze the performance of several weighting functions in TDOA algorithm based on GCC and indicate that PHAT weighting is the best choice for SSL using GCC method for its small fluctuations, sharp peak, and strong antijamming ability. Padoisa et al. [20] develop an improved GCC method to detect the source positions. The method introduces two criteria to improve the noise source map, which are based on the geometric properties of the microphone array, the scan zone whereas, and the energy of the spatial likelihood function, respectively. Valind et al. [15] develop an estimate method based on GCC-PHAT to compute the time delay by suppressing the impact of narrowband noise by adjusting the weight according to the different signal noise ratios (SNR). In view of the problems of diffraction and ambiguity of SSL technique based on GCC-PHAT method used on a binaural robot platform, Kim et al. [21] develop an improved SSL method to overcome the diffraction problem by incorporating a new time delay factor into GCC-PHAT method under the assumption of a spherical robot head and the ambiguity problem by utilizing a amplification effect of the pinnae for localization over the entire azimuth. Zhang et al. [22] propose a comprehensive PHAT method and a maximum likelihood method based on GCC to reduce the influence of reverberation on the signal effectively, which obtains more accurate time difference estimations.

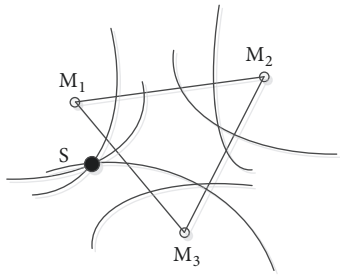

Figure 1: SSL of hyperbolic curve.

3. Method of Sound Source Localization Based on Rectangular Pyramid

3.1. Principle of Sound Source Localization. As shown in Figure 1, S is a sound source, and M_1, M_2, and M_3 are three microphones. τ_{21} is the time difference of S propagating to two microphones M_1 and M_2. From the nature of the hyperbolic curve, the absolute value of any distance from the two points on the hyperbolic curve is constant, that is, the length of the real axis. Thus, the sound source will be located on the branch line of the hyperbolic curve near M_1, which takes the positions of the two microphones as the focal points, and the distance difference of the sound source is the length of the real axis. Multiple hyperbolic curves can be obtained when multiple microphones are used, and the intersection point of multiple hyperbolic curves is the position of the sound source [23].

3.2. Time Delay Estimation

3.2.1. Review of GCC. GCC method based on the basic cross-correlation method is often used to estimate the time difference. The main functions of GCC are to suppress noise, prevent the occurrence of multipeak, and highlight the effective peak.

3.3. Time Delay Estimation

3.3.1. Review of Generalized Cross Correlation. If $x_1(t)$ and $x_2(t)$ are the signal of a sound source $s(t)$ received by M_1 and M_2, respectively, then the implementation process of GCC can be shown as Figure 2. First, the sound signal is transformed from the time domain to the frequency domain by fast Fourier transform (FFT), and the desired signal is highlighted by the weighting function in the frequency domain. Second, GCC function $R_{12}^{g}(\tau)$ is obtained by inverse fast Fourier transform (IFFT), of which the abscissa of the peak value corresponds to the time difference τ_{21}. Finally, τ_{21} is obtained by detecting the peak of the cross-correlation function.

GCC function is defined as

$$R_{12}^{g}(\tau) = \int_{-\infty}^{+\infty} \varphi(\omega) G_{12}(\omega) e^{j\omega\tau} d\omega \tag{1}$$

where $\varphi(\omega)$ is a weighting function and $G_{12}(\omega)$ is the cross-correlation power spectral function of the signals $x_1(t)$ and $x_2(t)$.

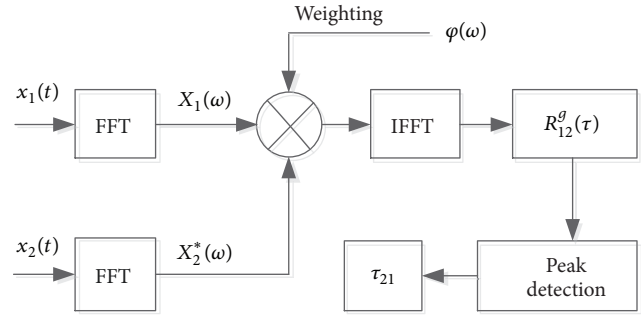

Figure 2: Implementation process of time delay estimation based on GCC.

3.3.2. Cross-Power Spectral Phase Algorithm. When there is reverberation in the environment, sound signals received by the two microphones can be expressed as

$$\begin{aligned} x_1(t) &= \alpha_1 s(t-\tau_1) + n_{z1}(t) + n_{h1}(t) \\ x_2(t) &= \alpha_2 s(t-\tau_2) + n_{z2}(t) + n_{h2}(t) \end{aligned} \tag{2}$$

where $n_{z1}(t)$ and $n_{z2}(t)$ are uncorrelated noise received by the two microphones, $n_{h1}(t)$ and $n_{h2}(t)$ represent the reverberation signals received by the two microphones, τ_1 and τ_2 are the delay times, and α_1 and α_2 are the decay factors.

In (2), $n_{z1}(t)$, $n_{z2}(t)$, and $s(t)$ are independent of each other. Hence, the cross-correlation power spectrum function of $x_1(t)$ and $x_2(t)$ can be calculated as follows:

$$\begin{aligned} G_{12}(\omega) &= X_1(\omega) X_2^*(\omega) \\ &= \alpha_1\alpha_2 S(\omega) S^*(\omega) e^{-j\omega(\tau_1-\tau_2)} \\ &\quad + \alpha_1 S(\omega) e^{-j\omega\tau_1} N_{h2}^*(\omega) \\ &\quad + \alpha_2 S^*(\omega) e^{-j\omega\tau_2} N_{h1}(\omega) + N_{h1}(\omega) N_{h2}^*(\omega) \end{aligned} \tag{3}$$

When the reverberation intensity is weak, the values of the last three items in formula (3) are very small and can be neglected. So, $G_{12}(\omega)$ indicates that $x_1(t)$ and $x_2(t)$ are only affected by noises. When the reverberation intensity is strong, the signals received by the microphone contain not only the direct sound signal, but also various reflection signals of the sound. So, in this case, the performance of TDE algorithm will be decreased because the values of the last three items cannot be ignored. To improve the performance of TDE algorithm, in the paper, the weighting function of the cross-power spectrum phase is modified as

$$\varphi(\omega) = \frac{1}{|G_{12}(\omega)|} \tag{4}$$

Formula (4) can be modified to be

$$\varphi_g(\omega) = \frac{1}{\gamma |G_{12}(\omega)|} \tag{5}$$

$$\gamma = \frac{1}{1+\sigma} \tag{6}$$

where γ is the weighting coefficient and σ is the ratio of the reverberation energy to the direct sound energy.

3.3.3. Estimation Method of γ. Reverberation, as a physical quantity in acoustic space, has two characteristic parameters of reverberation time and direct to reverberation ratio (DRR).

Reverberation time is a very important parameter to describe the degree of the sound decay in the room. Specifically, it refers to the time that the energy attenuates 60dB after the sound is stopped in the diffusion field and the residual sound energy is refracted for many times. Reverberation time increases with the increase of the volume of the room and decreases with the increase of the sound absorption coefficient of the room. If α is the sound absorption coefficient, A is the surface area of the room, V is the volume of the room, and c is the speed of sound. Reverberation time T can be computed as follows:

$$T = \frac{24\ln(10)}{c}\frac{V}{\alpha A} = \frac{0.163V}{\alpha A} \tag{7}$$

DRR refers to the power spectral ratio of a direct sound signal to a reverberant signal. Its mathematical expression is as follows:

$$DRR = 10\log_{10}\left(\frac{\sum_{t=0}^{t_d} h^2(t)}{\sum_{t=t_d-1}^{\infty} h^2(t)}\right) \tag{8}$$

$$h(t) = \begin{cases} h_d(t) & 0 \le t < t_d \\ h_r(t) & t_d \le t < \infty \\ 0 & else \end{cases} \tag{9}$$

where $h(t)$ is the impulse response of the sound channel, $h_d(t)$ is the direct sound impulse response, and $h_r(t)$ is the reverberation impulse response.

The value of DRR depends on the distance between the sound source and the microphone and reverberation time. Further, DRR can be expressed as

$$DRR = 10\log_{10}\left(\frac{QR}{16\pi D^2}\right) \tag{10}$$

$$R = \frac{A\alpha}{1-\alpha} \tag{11}$$

where Q is the directivity factor, R is the room constant, and D is the distance between the sound source and the microphone.

The value of σ can be obtained by the model of room reverberation. According to the definition of DRR, (8) can be transformed into

$$DRR = 10\log_{10}\left(\frac{1}{\sigma}\right) \tag{12}$$

Substituting (10) into (12), σ can be computed as follows:

$$\sigma = \frac{16\pi D^2}{QR} \tag{13}$$

Further, according to (7) and (11), σ is

$$\sigma = \frac{16\pi D^2 (AT - 0.163V)}{0.163QAV} \tag{14}$$

So, γ is

$$\gamma = \frac{1}{1+\sigma} = \frac{0.163QAV}{16\pi D^2 (AT - 0.163V) + 0.163QAV} \tag{15}$$

Some simulation experiments based on MATLAB are carried to verify the performance of the improved algorithm of PHAT weight under reverberation time of 100ms, 200ms, and 300ms, respectively. The experiments estimate the time delay between two microphone signals. One signal is the sampled signal of a section of sound signal with the sampling frequency of 20kHz and the sampling digit of 16bit, which is considered to contain no noise. Another signal is the sampled signal with a delay of 800 sampling cycles. According to the sampling frequency, the time delay between the two signals is 40ms. Gaussian white noise is added to the two signals with a SNR of 5dB.

Figure 3 are the experimental results. It can be seen from these results that, under the conditions of same SNR and reverberation time, the delay estimation performance of the improved algorithm of PHAT weight is better than that of the unimproved PHAT weight. During reverberation time at 300ms, the time delay peak of the unimproved algorithm is difficult to extract because it is covered by other peaks, but the improved algorithm can still accurately obtain the time delay peak.

The multiplier of the sound increases due to the reflection of the surface of the object when the sound source signals are in different positions in the room. The improved algorithm of PHAT weight takes full account of the influencing factors of reverberation, which makes the time delay estimation algorithm based on the cross-power spectrum phase not only retain the effective suppression of noise, but also play a good role in eliminating reverberation.

However, the improved weight algorithm needs to know some parameters of the room and the rough position of the sound source in advance, which limits its universality and requires further improvements.

3.3.4. Fast Search Strategy of the Peak. It will be a way to improve the real-time performance of TDE algorithm through reducing the searching interval of the peak of cross-correlation function. So, the paper proposes a fast search strategy of the peak based on the geometric model of the microphone array. There are two geometric relationships between a sound source and two microphones on a plane. As shown in Figure 4, S_1, S_2, and S_3 represent the sound source S at three different locations, and M_1 and M_2 are two microphones. One is the linear relationship such as the straight line $S_2M_1M_2$ or $S_3M_2M_1$, and the other is the triangular relationship like the triangle $S_1M_1M_2$.

According to Figure 4, M_1 and M_2 are fixed. l is the distance between M_1 and M_2, and the difference of the distances between each microphone and S is d. As we know, the difference of any two sides is smaller than the third side in a triangle. Thus, $d$will be $-l$ to l only if S is collinear with M_1 and M_2, such as the sound source located at S_2 or S_3. It is easy to know that the sampling range of the peak is$-l \times f/c$ to $l \times f/c$, where f is the sampling frequency. As shown in Figure 5, the cross-correlation function of the two

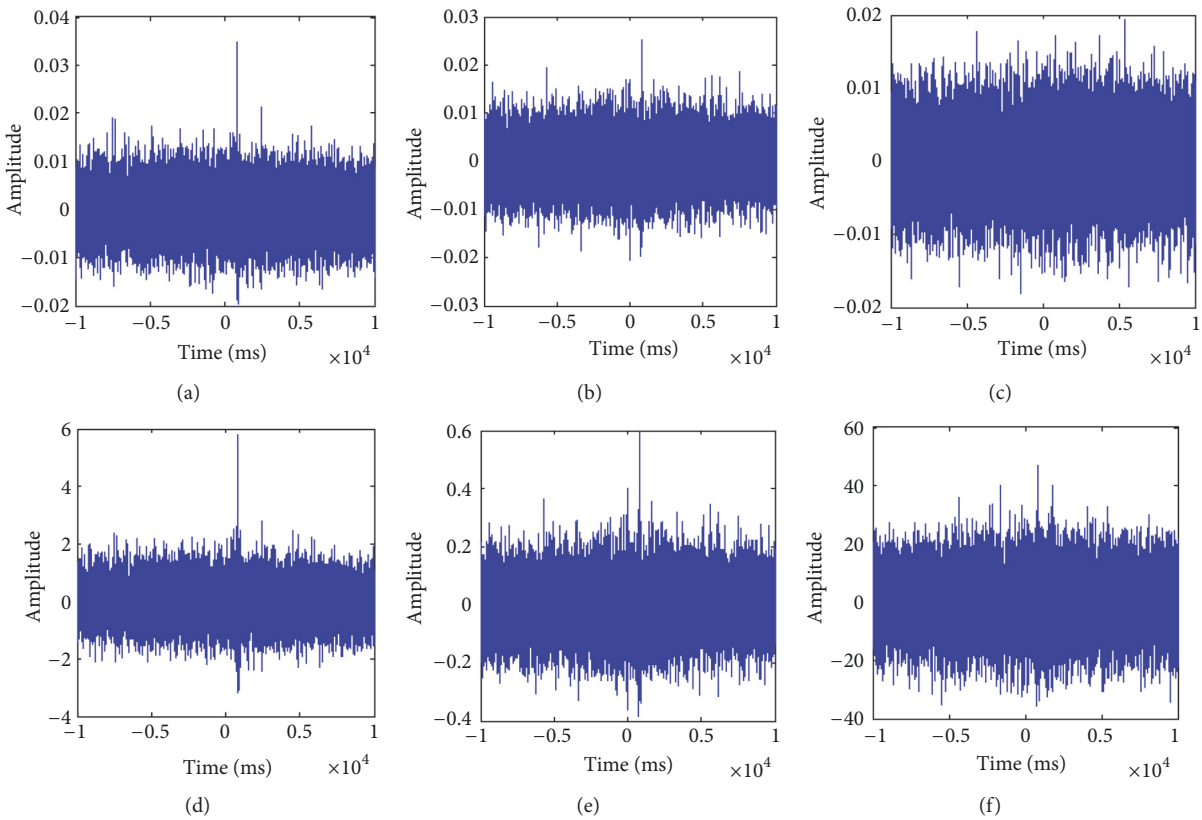

FIGURE 3: The results of simulation experiment of the improved algorithm PHAT weight under different reverberation time. (a), (b), and (c) are the results of the unimproved algorithm PHAT weight under reverberation time of 100ms, 200ms, and 300ms, respectively. (d), (e), and (f) are the results of the improved algorithm PHAT weight under reverberation time of 100ms, 200ms, and 300ms, respectively.

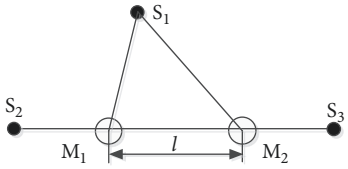

FIGURE 4: Geometric relationships between three sound sources and two microphones.

signals is obtained by using TDE algorithm, and the peak of the function corresponds to the delay value between the two signals. Therefore, it is possible to reduce the estimating time of TDE by setting the search boundary (dotted by line 1 and dotted by line 2).

3.3.5. Screening Strategy of Time Difference Values. The error of delay difference will lead to an erroneous positioning result. On the other hand, it is also easy to produce a directional error of positioning if the error of delay difference existed near the area's boundary. Therefore, the paper proposes a screening strategy of delay difference based on [24]. The implementation processes of the strategy are as follows:

(a) Ideally, the delay differences between the microphones i, j, and k have the following relationship:

$$\tau_{ij} = \tau_{ik} + \tau_{kj} \tag{16}$$

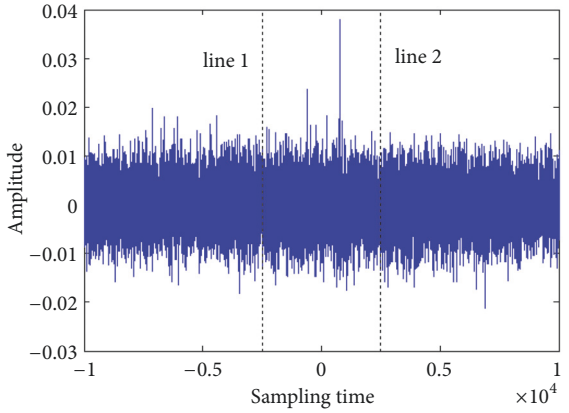

FIGURE 5: The search interval of the peak.

However, (16) is usually not valid due to the error of time delay estimation.

(b) Under the condition of the error of time delay estimation, (16) can be rewritten as the following inequality:

$$\left|\tau_{ij} - \left(\tau_{ik} + \tau_{kj}\right)\right| \leq \varepsilon \tag{17}$$

where ε is reasonable threshold.

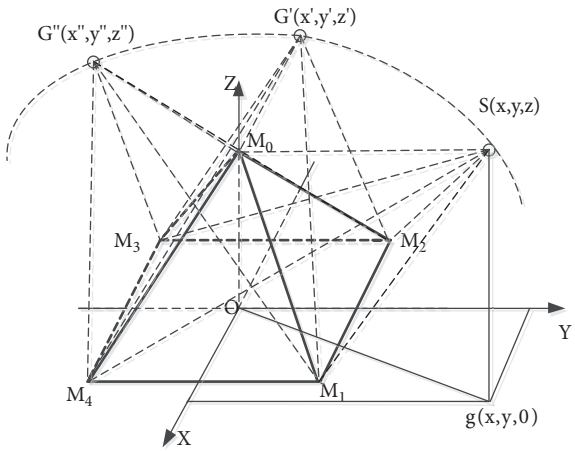

Figure 6: The geometry of the microphone array based on the structure of rectangular pyramid.

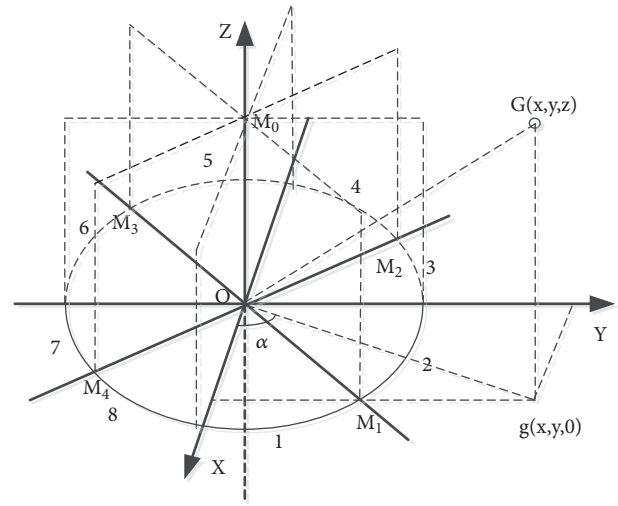

Figure 7: Schematic diagram of location interval partition.

(c) When the inequality is not satisfied, it indicates that the error of this time delay estimation is large and should be discarded. Conversely, this time delay estimation is accepted.

3.4. Sound Source Localization of Rectangular Pyramid Based on TDOA. A three-dimensional microphone array based on the structure of rectangular pyramid is designed to locate a sound source. As shown in Figure 6, five microphones M_0-M_4 are installed on five vertices of rectangular pyramid, of which coordinates are (x_i, y_i, z_i) $(i = 0, 1, \ldots, 4)$. In addition, the coordinates of the sound source S are (x, y, z). The distance from S to the position of each microphone is d_i.

The distance difference between the distance from the sound source to the other microphone and the distance from the sound source to M_0 is

$$d_{j0} = d_j - d_0 = v_{sound}\tau_{j0} \quad j = 1, 2, 3, 4 \tag{18}$$

where τ_{j0} is the delay difference between the sound source to M_j and M_0.

According to the spatial coordinate relation shown in Figure 6, the mathematical model of sound source localization is given as

$$\begin{aligned} d_j &= \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \\ d_0 &= \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \\ d_{j0} &= d_j - d_0 \quad (j = 1, 2, 3, 4) \end{aligned} \tag{19}$$

4. The Localization Algorithm Based on Iterative Optimization

The coordinates of the sound source can be obtained by using Newton iterative algorithm to solve (19). According to (19), four sets of equations and their corresponding solutions can be obtained. The paper takes the mean of the four solutions as the final results, which is the location of the sound source.

4.1. Partition Iterative Process. In Newton iteration, s set of appropriate initial coordinates are helpful to improve the convergence rate and get an accuracy solution. So, the paper proposes a strategy of partition iterative to reduce the number of iterations and ensure that each initial coordinates range has a unique optimization point. According to the structural characteristics of rectangular pyramid, the microphone array can get redundant delay differences. For example, the coordinates of the sound source should meet $\{(x, y, z) \mid x \geq 0, y \geq 0 \ \& \ x \geq y\}$, and its azimuth angle must also be in the range of $[0^\circ, 45^\circ]$ when $\tau_{21} \geq 0$, $\tau_{41} \geq 0$ and $\tau_{21} \geq \tau_{41}$. Therefore, it is possible that the solution interval can be divided by making full use of these redundant delay differences. As shown in Figure 7, the localization space is equally divided into eight areas according to the rule. The coordinates of the sound source in one of the areas can be locked initially when all delay differences between different microphones in the microphone array are determined, which can narrow the range of the solution to accelerate the computation speed.

According to the principle of Newton iteration algorithm and the localization model of the microphone array, (19) can be transformed into

$$\begin{aligned} f_j(x, y, z) = &\sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} \\ &- \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \\ &- d_{j0} \quad (j = 1, 2, 3) \end{aligned} \tag{20}$$

Jacobian matrix of $f_j(x, y, z)$ is

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{bmatrix} \tag{21}$$

So, the coordinates of the sound source can be calculated as

$$\begin{bmatrix} x^{(k+1)} \\ y^{(k+1)} \\ z^{(k+1)} \end{bmatrix} = \begin{bmatrix} x^{(k)} \\ y^{(k)} \\ z^{(k)} \end{bmatrix} - J^{-1} \begin{bmatrix} f_1\left(x^{(k)}, y^{(k)}, z^{(k)}\right) \\ f_2\left(x^{(k)}, y^{(k)}, z^{(k)}\right) \\ f_3\left(x^{(k)}, y^{(k)}, z^{(k)}\right) \end{bmatrix} \quad (22)$$

4.2. Screening Strategy of Localization Results. The screening of the delay difference provides a certain guarantee for the accurate space division and solution of the iterative operation. Therefore, when the iterative operation is running in the correct interval, the four localization results will be more accurate, and the difference between them will be smaller. However, when the sound source is very close to the boundary line, a small error of delay difference will also cause a wrong solution interval, which leads to a localization result with a large error. The paper proposes a screening strategy based on geometric model, which sets three coordinate thresholds δ_x,δ_y, and δ_z in the X, Y, and Z directions, respectively, and compares the four localization results (x_i, y_i, z_i) $(i = 1, 2, 3, 4)$ as follows:

$$\begin{aligned} \left|x_i - x_j\right| &\le \delta_x \\ \left|y_i - y_j\right| &\le \delta_y \\ \left|z_i - z_j\right| &\le \delta_z \end{aligned} \quad (23)$$

where $i, j = 1, 2, 3, 4$ and $i \neq j$.

The localization results will be considered correct when they meet the above three inequalities; otherwise they are marked as a localization failure (in fact, it is just for this paper to describe and does not mean that the localization result is not obtained).

The average value of the four localization results obtained by the double screening mechanism is used as the final localization result(x_F, y_F, z_F), and we can solve distance R_F, the azimuth angle α, and the elevation angle β from the sound source to the origin.

$$R_F = \sqrt{{x_F}^2 + {y_F}^2 + {z_F}^2} \quad (24)$$

$$\alpha = \arctan\frac{y_F}{x_F} \quad (25)$$

$$\beta = \arctan\frac{\sqrt{{x_F}^2 + {y_F}^2}}{z} \quad (26)$$

5. Experimental Evaluation

5.1. Experimental Prototype. Figure 8 shows a prototype of the microphone array with the length of 25cm and the height of 12.5cm, which is suitable for small and medium mobile robots. Microphone mounting holes are designed at the five vertices of rectangular pyramid to provide a more robust connection between the microphones and the array. The prototype is installed on a mobile robot platform by screws and the communication connection between them is established.

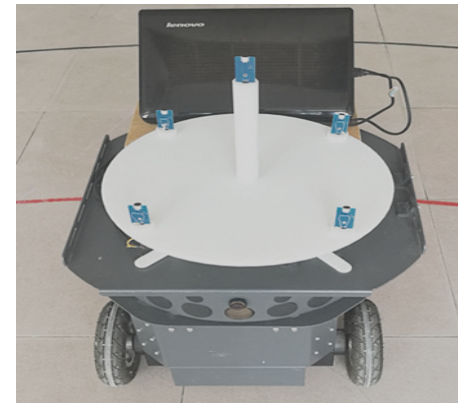

Figure 8: The prototype of the rectangular pyramid microphone array mounted on a mobile robot.

Table 1: The parameters of microphone.

Index	Parameter
dimensions	9.7×6.7mm
sensitivity	-48-66dB
frequency range	20-16kHz
S/N ratio	greater than 58dB
operation voltage range	1.5-10V
directivity	omnidirectional

Table 2: The parameters of data acquisition card.

Index	Parameter
the type of AD	9.7×6.7mm
accuracy	16 bits
input range	-10~+10V
numbers of voltage channels	8 single-ended
range of sampling rate	1-100 KHz

The selected microphone model is the DGO-6050CD-P, and its specific parameters are in Table 1.

The model of data acquisition card is USB_ HRF4626, which is a high-speed and high-precision synchronous data acquisition card based on USB bus, and its specific parameters are shown in Table 2.

Two terminals of the acquisition card are connected with the microphone array through 5 data wires and the computer through USB, respectively. The sound signals collected by the acquisition card are transmitted to the host computer.

5.2. Experimental Setup. ALL experiments are performed in a laboratory setting with a reverberation time of approximate 200ms shown in Figure 9(a). Six concentric circles with radius from one to six meters are set up in the lab, and six test points are designed at intervals of 60° on each circle. A sketch

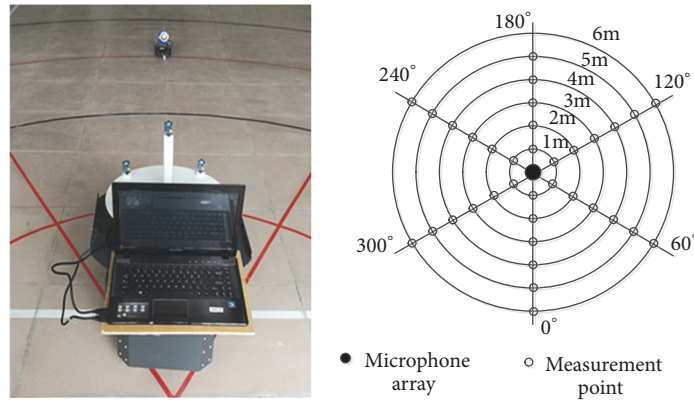

Figure 9: The experimental environment and condition, (a) the experimental environment in a laboratory; (b) a view sketch of the sound source used in the experiment.

Table 3: The success rate of delay estimation based on the screening strategy of delay difference.

Threshold Condition	*Success Rate*					
	1m	2m	3m	4m	5m	6m
ε = 0.1ms,	65	68	67	72	70	66
ε = 0.2ms,	76	78	75	76	77	76
ε = 0.3ms,	85	86	87	87	86	85
ε = 0.4ms,	91	89	91	90	88	91
ε = 0.5ms,	92	92	91	93	91	92

with labeled test positions is shown in Figure 9(b). There are no desks, chairs, and other types of equipment except the experimental system in the lab. The effects of doors and windows are ignored. There are some kinds of noise in the lab, such as electrical noise of the mobile robot, the noise made by the computer fan, etc. This noise distribution is unimodal and appears roughly Gaussian according to presampling analysis.

The test signals are sampled by a single channel at 20kHz. The sound source is generated by playing prerecorded clapping sound, which is easy to acquire and analyze because its energy is concentrated, and its silent segment is clearly distinguished from the sound segment. In the experiment, the laboratory is considered as a quiet environment, and all environmental noises are ignored. White noise is superimposed on the sound source to get the required SNR, so the actual SNR must be less than the set value.

Three kinds of experiments are designed as follows:

(1) Experiment 1 is carried out to verify the performance of the proposed delay estimation method. In this experiment, one of the six setting points at each distance shown in Figure 9(b) is randomly selected, and SNR of the signal source is set at 45dB. Each pair of four microphones on base of the prototype and the vertex microphone is selected to form a testing microphone set. So, there are six testing sets. Each point is tested 100 times, and the delay difference between each pair of microphones is evaluated by (17). Finally, the success rate of TDE is evaluated.

(2) Experiment 2 is used to test the localization performance of the prototype. All 36 testing points shown in Figure 9(b) are located. Five available positioning results are gotten at each testing point, which are applied to evaluate the performance of the proposed microphone array. In this experiment, the signal source has a fixed SNR of 45dB.

(3) Experiment 3 is designed to analyze the influence of different noise conditions on the localization accuracy of the prototype. Just one point with the azimuth angle $\alpha = 180^\circ$ at each distance shown in Figure 9(b) is selected for this experiment, and SNR of the signal source is set at 10dB, 20db, 30dB, and 40dB, respectively. At each SNR condition, each test point is positioned 10 times, and then the average of 10 localization results is computed.

5.3. Results Analysis

5.3.1. The Results of Experiment 1. In the experiment 1, the success rate of the delay estimation is defined as

$$R_{success} = \frac{N_{pass}}{N_{total}} \times 100\% \tag{27}$$

where N_{pass} is the success times of the results of TDE passed by (17) and N_{total} is the total times of test and equal to 100.

Different time thresholds ε from 0.1ms to 0.5 ms are used to evaluate the delay estimation results, respectively. The success rates of the delay estimation at each testing point are shown in Table 3.

Table 3 shows that the success rates of the delay difference estimation at all distances are above 85% while ε is taken to be 0.3ms. Obviously, the success rate will be higher while ε

Table 4: The evaluation results based on the screening strategy of coordinate thresholds.

Threshold Condition	*Times of Localization Failure*					
	1m	2m	3m	4m	5m	6m
$\delta_x = \delta_y = \delta_z = 0.1$m	10	13	9	13	15	21
$\delta_x = \delta_y = \delta_z = 0.2$m	7	8	8	9	9	15
$\delta_x = \delta_y = \delta_z = 0.3$m	3	4	5	4	6	6

taken a larger value. However, under the condition of the rectangular pyramid microphone array with a definite sizes, a large ε means that a large error of TDE is allowed, which will increase the localization error. Therefore, in experiment 2 and experiment 3, ε is taken to be 0.3ms.

In fact, we only need the accurate delay differences of three microphones to calculate the exact position of the sound source. So, the estimate results of delay difference are considered acceptable only if three of the six estimate results meet (17). Based on these, the success rate will increase to over 95% at $\varepsilon = 0.3$ms.

5.3.2. The Results of Experiment 2. In experiment 2, the localization results of each testing point calculated according to Newton iterative are evaluated by formula (23). In the paper, three coordinate thresholds δ_x, δ_y, and δ_z are set to be equal and are taken to be 0.1m, 02m, and 0.3m, respectively, to evaluate the localization results successively. There are a few cases of localization failure because the localization results do not meet formula (23). The times of localization failure at difference distance are shown in Table 4.

If the localization results do not meet formula (23) when the coordinate thresholds at 0.2m, the system will repeat the experiment and start a new localization process. As five valid localization results are needed at each point, the test and evaluation times at each distance should be at least 180 times. So, even at the coordinate thresholds of 0.1m, the rate of localization failure is no more than 10% except the result at the distance of 6m. Considering the localization accuracy and real-time, all coordinate thresholds are set to be 0.2m. The subsequent experiments are based on this.

The positioning error curves of distance and azimuth angle, as shown in Figure 10, can be obtained by using the root mean square error as the judgment standard of positioning accuracy.

Figures 10(a) and 10(b) show that, at the same azimuth angle, the positioning errors of distance and azimuth angle have different change curves with the increase of the testing distance between the sound source and the microphone array. In detail, the positioning errors of distance will increase with the increase of testing distance, and its maximum and minimum are about 0.25m and 0.05m, respectively. By contrast, the positioning errors of azimuth angle do not change significantly with the increase of testing distance and are within 1.5°.

Figures 10(c) and 10(d) show that the positioning errors of distance and azimuth angle vary little as the azimuth angle changes from 0 to 300°. In detail, the variation range of the positioning errors of distance is kept within 0.08m, and the variation range of the positioning errors of azimuth angle is maintained within 0.6°. From the analysis of the array structure, due to the symmetry of the rectangular pyramid, it is more balanced to receive the sound source signal from each angle. So the positioning errors have no significant variations with the change of the azimuth angle.

The relative error can be calculated as

$$\delta = \frac{|V_e - V_a|}{V_a} \times 100\% \tag{28}$$

where δ is the relative error, V_e is the experimental value, and V_a is the actual value.

Figure 11 shows that the relative positioning errors of distance and azimuth angle. In detail, the maximum relative positioning error of distance is about 9% at 1m, while at the other testing distances, the relative errors quickly reduced to less than 5%. The relative errors of azimuth angle at each testing azimuth angle are less than 1.5%.

According the results shown by Figures 10 and 11, both the absolute error and the relative error of the position results are within the acceptable range. So, the positioning accuracy of the prototype can meet the needs of mobile robot locating and tracking a sound target.

Currently, there are many kinds of microphone arrays for SSL, such as linear arrays, planar arrays, and three-dimensional arrays. From the perspective of positioning accuracy, the positioning accuracy of the three-dimensional array is obviously higher than that of other arrays. In [24], a tetrahedron-based sound source localization system is designed and carried out experimental tests at different distances and azimuth angle. Table 5 shows the comparisons of the positioning results between our array and the array in [24]. As can be seen from Table 5, compared with the tetrahedral array, the rectangular pyramid array designed in the paper has improved both the positioning accuracy.

5.3.3. The results of Experiment 3. Figure 12 is the result of localization experiment 3. Figures 12(a) and 12(b) are the positioning error of distance and the positioning error of azimuth angle, respectively, which show that both errors will increase as SNR decreases. Both kinds of error at 30dB are not much different from those of at 40dB and have an increase significantly when SNR is reduced to 20dB. When SNR at 10dB, the maximum distance error is about 0.4m at the testing distance of 6m, the maximum relative distance error is about 12.5% at the testing distance of 1m, and while the testing distances greater than 3m, the relative errors quickly reduced to less than 10%. The paper also carries out the localization experiment at SNR of 5dB, and its results are not acceptable because the correct delay estimation cannot be obtained. In

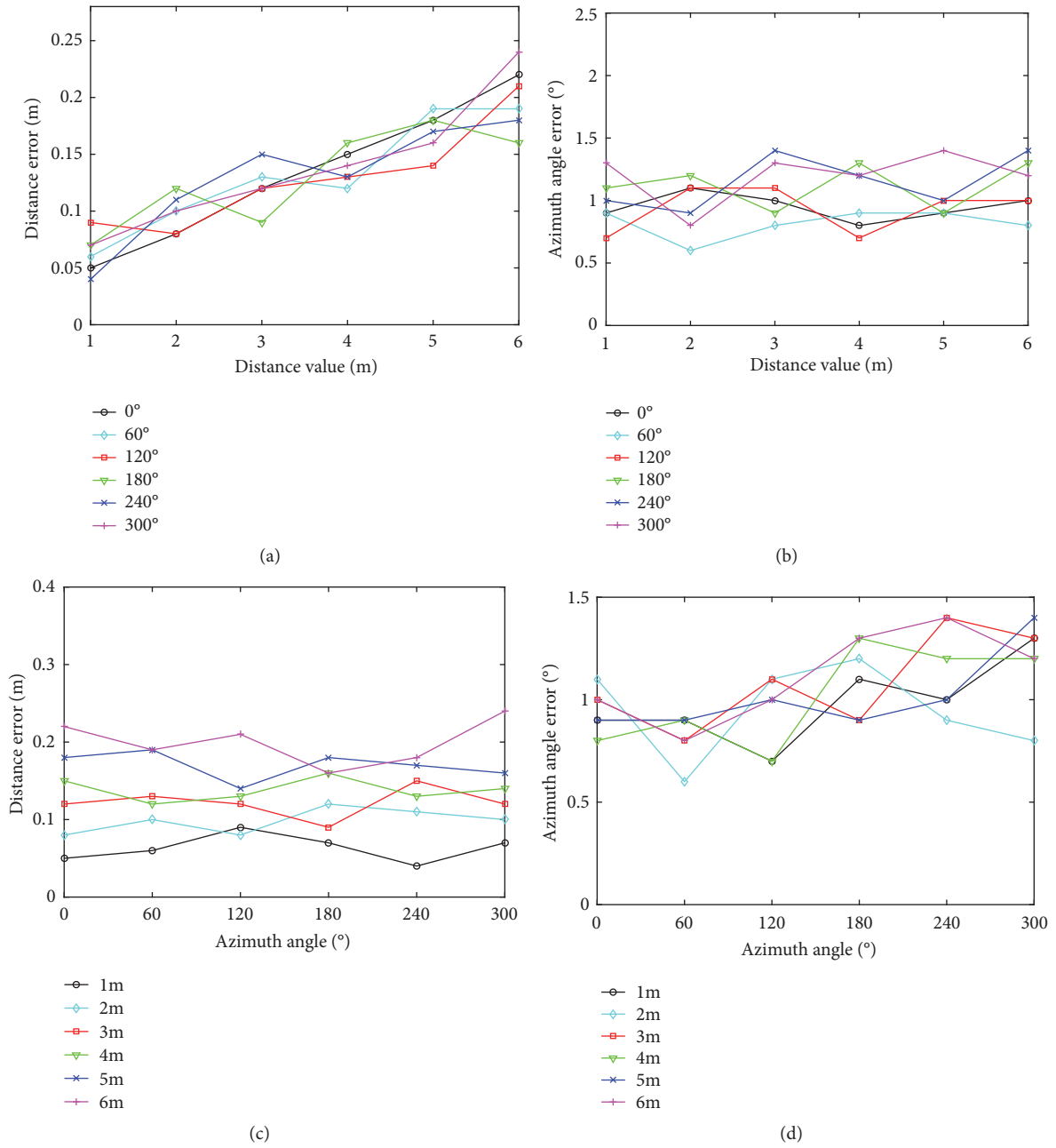

Figure 10: The results of localization experiment for all test points at SNR of 45dB, (a) the relation curves between the positioning error of distance and the testing distance; (b) the relation curves between the positioning error of azimuth angle and the testing distance; (c) the relation curves between the positioning error of distance and the testing azimuth angle; (d) the relation curves between the positioning error of azimuth angle and the testing azimuth angle.

Table 5: Comparison of positioning accuracy between two microphone arrays.

Array model	Distance positioning error	Azimuth positioning error
tetrahedral	≤ 0.40m	≤ 2.0°
Rectangular pyramid	≤ 0.24m	≤ 1.5°

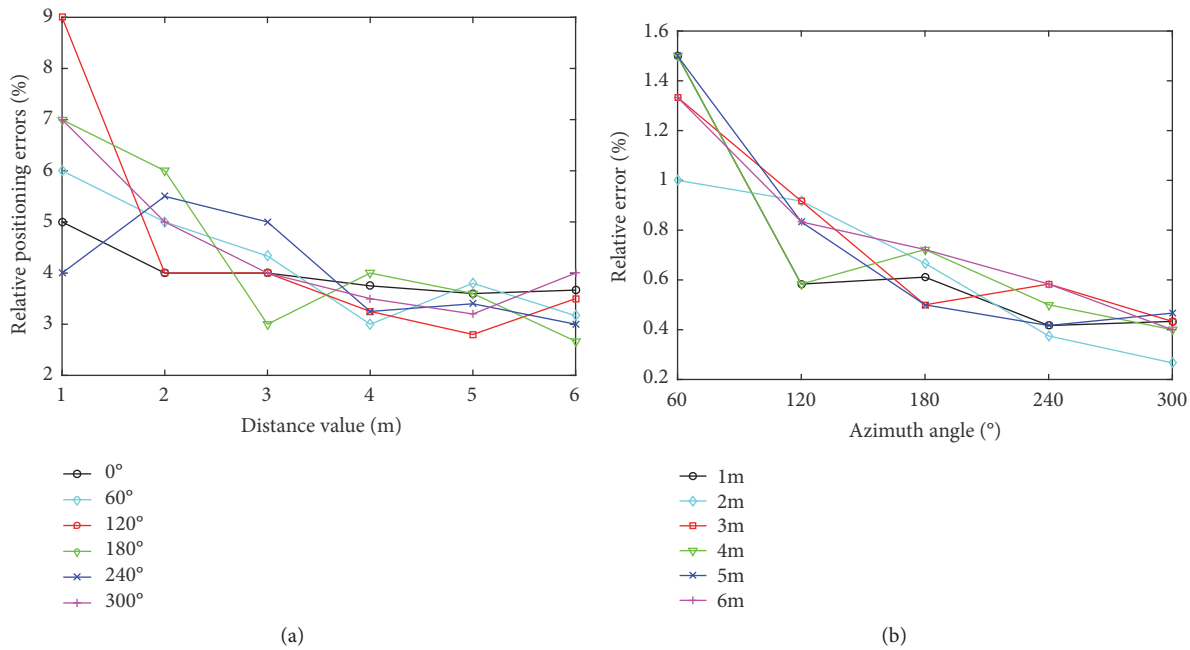

FIGURE 11: The relative positioning errors of localization experiment for all test points at SNR of 45dB, (a) the relative positioning errors of distance; (b) the relative positioning errors of azimuth angle.

general, under the conditions of SNR above or even slightly below 10dB, the prototype has certain anti-interference ability and a good positioning performance.

Considering that indoor environment is relatively quiet and the mobile robot mainly locates a purposeful active sound source with a strong sound signal, SNR is considered to be generally greater than 10dB. The distance error no more than 0.4m and the angle error of no more than 2° have little impact on mobile robot tracking because robot is able to approach the target by constantly modifying its positioning results. Therefore, the prototype can be used for mobile robot positioning the acoustic target in indoor environment.

Paper [25] develops a SSL system with a 60 cm circular array of eight omnidirectional microphones. In a noisy conference room of an average SNR of 7 dB, for a stationary source at 1.5m distance, the system has the angular error of less than one degree and the RMS error of distance of 10%. But, the paper did not give any other positioning location results at other distances. Compared with the system, at the testing distance of 1.5m, our prototype's positioning distance accuracy is similar to that of the system, and its positioning angle accuracy is slightly lower than that of this system. But our prototype has the advantages of smaller size and fewer microphones, which makes it be more suitable for mobile robots.

6. Conclusions

A SSL system for indoor mobile robots is proposed, including (1) a microphone array with the structure of rectangular pyramid, (2) an improved algorithm of TDE based on GCC-PHAT, and (3) a localization model and its solution method based on Newton iterative algorithm with a partition strategy of delay difference. The SSL system needs about 450ms positioning time for a complete positioning process including sampling, estimating, localizing, and evaluating (in fact, five times valid localization are done in the positioning time of 450ms, and the positioning time can be reduced to about 100ms in the robot's practical tracking application if we only do one localization), and has positioning errors less than 10% in most conditions and certain anti-interference ability, which can meet the requirements of mobile robot locating a sound source in indoor environment with SNR greater than 10dB and reverberation time less than 300ms.

The proposed method has some novelties compared with the existing methods, including (1) a new structure of microphone array is developed; (2) an improved TDE method to improve the ability to suppress reverberation; (3) Newton iterative algorithm based on geometric partition to reduce the computing time; (4) dual evaluation mechanism based on the delay difference threshold and the coordinate thresholds to improve the reliability of positioning results.

The next research is to combine the device with a mobile robot to realize the autonomous positioning and tracking a sound source target, which includes (1) optimizing the SSL algorithm by removing redundant computation during two evaluation processes for reducing positioning time and (2) researching the active sensing strategy of robot in tracking for achieving more efficient sound acquisition and robust positioning.

Data Availability

The research library related to the dissertation will be established in GitHub, where you can access the folders

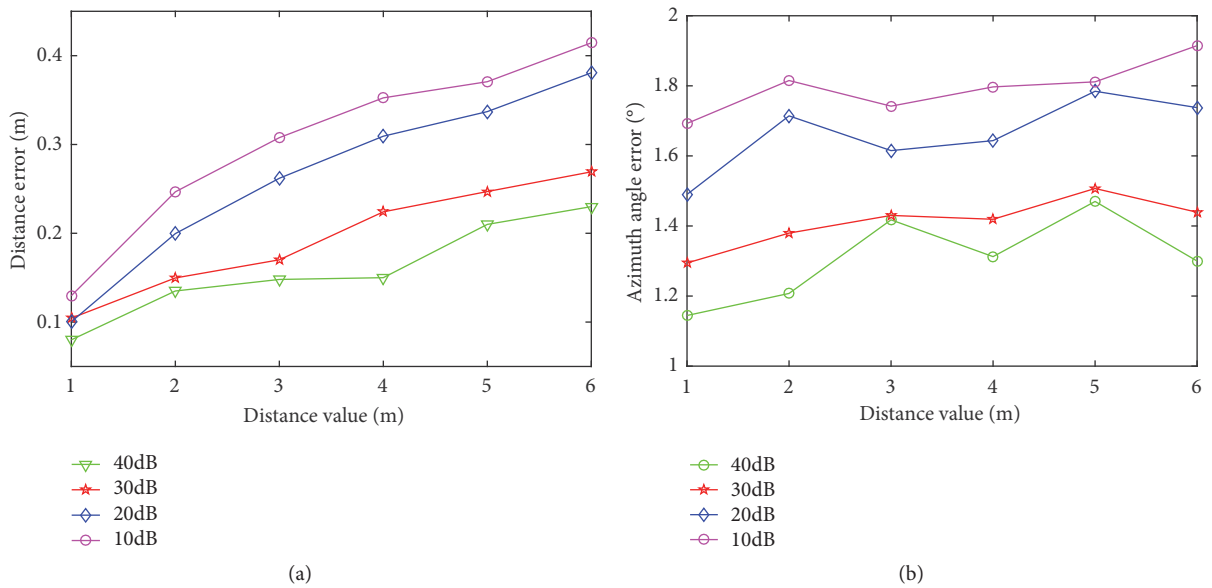

Figure 12: The results of comparison experiment of prototype with different SNR, (a) the curve between the positioning error of distance and the testing distance; (b) the curve between the positioning error of azimuth angle and the testing distance.

and find experimental data and lists: GitHub: https://github.com/glchenwhut.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Guoliang Chen conceived the idea, designed the experiments, and wrote the paper; Yang Xu helped with the algorithm and analyzing the experimental data.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61672396.

References

[1] F. J. Ren and X. Sun, "Present situation and development of intelligent robots," *Science and Technology Review*, vol. 1, pp. 32–38, 2015.

[2] S. Argentieri, P. Danès, and P. Souères, "A survey on sound source localization in robotics: from binaural to array processing methods," *Computer Speech & Language*, vol. 1, pp. 87–112, 2015.

[3] A. Magassouba, N. Bertin, and F. Chaumette, "Aural Servo: Sensor-Based Control From Robot Audition," *IEEE Transactions on Robotics*, vol. 4, pp. 572–585, 2018.

[4] P. Aarabi and S. Zaky, "Robust sound localization using multi-source audiovisual information fusion," *Information Fusion*, vol. 5, pp. 209–223, 2006.

[5] A. Aly, S. Griffiths, and F. Stramandinoli, "Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics," *Cognitive Systems Research*, vol. 43, pp. 313–323, 2017.

[6] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie, "A model-based sound localization system and its application to robot navigation," *Robotics and Autonomous Systems*, vol. 27, no. 4, pp. 199–209, 1999.

[7] K.-C. Kwak and S.-S. Kim, "Sound source localization with the aid of excitation source information in home robot environments," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 852–856, 2015.

[8] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robotics and Autonomous Systems*, vol. 96, pp. 184–210, 2017.

[9] X. F. Li and H. Liu, "A survey of sound source localization for robot audition," *CAAI Transactions on Intelligence Systems*, vol. 1, pp. 9–20, 2012.

[10] R. E. Irie, *Robust Sound Localization: An Application of An Auditory Perception System for a Humanoid Robot*, Department of Electrical Engineering and Computer Science, MIT, Cambridge, USA, 1995.

[11] X. Li and H. Liu, "Sound Source Localization for HRI Using FOC-Based Time Difference Feature and Spatial Grid Matching," *IEEE Transactions on Cybernetics*, vol. 43, no. 4, pp. 1199–1212, 2013.

[12] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 10, pp. 2193–2206, 2013.

[13] Y. Cho, D. Yook, S. Chang, and H. Kim, "Sound source localization for robot auditory systems," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 1663–1668, 2009.

[14] M. Ren and Y. X. Zou, "A Novel Multiple Sparse Source Localization Using Triangular Pyramid Microphone Array," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 83–86, 2012.

[15] J. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and System*, pp. 1228–1233, Las Vegas, NV, USA, 2003.

[16] J. Huang, K. Kume, A. Saji, M. Nishihashi, T. Watanabe, and W. L. Martens, "Robotic spatial sound localization and its 3-D sound human interface," in *Proceedings of the First International Symposium on Cyber Worlds*, pp. 191–197, Tokyo, Japan, 2002.

[17] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on Advances in Signal Processing*, vol. 1, pp. 1–19, 2006.

[18] J. Perez-Lorenzo, R. Viciana-Abad, P. Reche-Lopez, F. Rivas, and J. Escolano, "Evaluation of generalized cross-correlation methods for direction of arrival estimation using two microphones in real environments," *Applied Acoustics*, vol. 73, no. 8, pp. 698–712, 2012.

[19] L. Chen, Y. C. Liu, F. C. Kong, and N. He, "Acoustic source localization based on generalized cross-correlation time-delay estimation," *Procedia Engineering*, vol. 15, pp. 4912–4919, 2011.

[20] T. Padois, F. Sgard, O. Doutres, and A. Berry, "Acoustic source localization using a polyhedral microphone array and an improved generalized cross-correlation technique," *Journal of Sound and Vibration*, vol. 386, pp. 82–99, 2017.

[21] U. Kim, K. Nakadai, and H. G. Okuno, "Improved sound source localization in horizontal plane for binaural robot audition," *Applied Intelligence*, vol. 42, no. 1, pp. 63–74, 2014.

[22] L. Y. Zhang, X. G. Zhang, and C. Liu, "The improvement of time delay estimation in the microphone array sound localization system," *Journal of Nanjing University*, vol. 1, pp. 25–30, 2015 (Chinese).

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

[24] H. Sun, W. C. Zhong, and H. Y. Liu, "Research on tetrahedral microphone array sound source localization model," *J. of Computer Simulation*, vol. 2, pp. 378–382, 2015.

[25] J. Valin, F. Michaud, and J. Rouat, "Robust 3D localization and tracking of sound sources using beamforming and particle filtering," in *Proceedings of the 2006 IEEE International Conference on Acoustics Speed and Signal Processing*, pp. IV821–IV825, Toulouse, France, 2006.