

Article

Sound Source Localization Based on Multi-Channel Cross-Correlation Weighted Beamforming

Mengran Liu, Junhao Hu, Qiang Zeng, Zeming Jian * and Lei Nie

Hubei Key Laboratory of Modern Manufacturing Quantity Engineering, School of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China; mrlu@hbut.edu.cn (M.L.); hujunhao199802@163.com (J.H.); zengqiang199907@163.com (Q.Z.); leinie@hbut.edu.cn (L.N.)

* Correspondence: jianzemingx@163.com

Abstract: Beamforming and its applications in steered-response power (SRP) technology, such as steered-response power delay and sum (SRP-DAS) and steered-response power phase transform (SRP-PHAT), are widely used in sound source localization. However, their resolution and accuracy still need improvement. A novel beamforming method combining SRP and multi-channel cross-correlation coefficient (MCCC), SRP-MCCC, is proposed in this paper to improve the accuracy of direction of arrival (DOA). Directional weight (DW) is obtained by calculating the MCCC. Based on DW, suppressed the non-incoming wave direction and gained the incoming wave direction to improve the beamforming capabilities. Then, sound source localizations based on the dual linear array under different conditions were simulated. Compared with SRP-PHAT, SRP-MCCC has the advantages of high positioning accuracy, strong spatial directivity and robustness under the different signal–noise ratios (SNRs). When the SNR is -10 dB, the average positioning error of the single-frequency sound source at different coordinates decreases by 5.69%, and that of the mixed frequency sound sources at the same coordinate decreases by 5.77%. Finally, the experimental verification was carried out. The results show that the average error of SRP-MCCC has been reduced by 8.14% and the positioning accuracy has been significantly improved, which is consistent with the simulation results. This research provides a new idea for further engineering applications of sound source localization based on beamforming.

Keywords: sound source localization; multi-channel cross-correlation coefficient; microphone array; beamforming



Citation: Liu, M.; Hu, J.; Zeng, Q.; Jian, Z.; Nie, L. Sound Source Localization Based on Multi-Channel Cross-Correlation Weighted Beamforming. *Micromachines* **2022**, *13*, 1010. <https://doi.org/10.3390/mi13071010>

Academic Editors: Libo Gao and Zhuoqing Yang

Received: 20 May 2022

Accepted: 23 June 2022

Published: 26 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sound source localization techniques have a wide range of application prospects in civil and military systems, such as intelligent medical systems, security monitoring and sonar detection [1–5]. Existing sound source localization techniques can be divided into the subspace, time delay estimation and beamforming. The subspace approach uses the orthogonality between the signal and noise subspaces to determine the waveform direction, including multiple signal classification (MUSIC) and estimating signal parameters via rotational invariance techniques (ESPRIT) [6,7]. Their performance is heavily dependent on the covariance matrix estimation, which is influenced by the signal-to-noise ratio (SNR). Hu et al. proposed an improved MUSIC algorithm to calculate the spatial spectrum and achieve azimuth estimation [8]. Herzog et al. developed a novel EB-ESPRIT method to estimate the incoming wave direction of a sound source [9]. The direction estimation accuracy of studies [8] and [9] is higher than that of MUSIC and ESPRIT, respectively. However, the accuracy is still significantly reduced at a lower SNR. The time delay estimation method achieves the source location based on the arrival time difference [10]. In [11], sound source localization in an indoor environment is completed by using two dual-element arrays to estimate time delay. However, the time delay estimation method is susceptible to noise.

When the number of array elements is increased, there is redundant information in the signals. The multi-channel cross-correlation coefficient (MCCC) method in [12] can improve the robustness of delay estimation through multi-element array signals. Beamforming is to obtain the direction of the sound source by summing weighted array signals, and is classified into frequency domain beamforming and time domain beamforming [13,14]. In frequency domain beamforming, many methods have been proposed to improve its performance [15,16]. Reference [15] improved the robustness by adding norm constraints and spatial smoothing techniques. Reference [16] proposed nested arrays to improve beamforming performance. Time domain beamforming is compared with frequency domain beamforming in [17]. The two beamforming methods have similar performance. Time domain beamforming is a natural broadband method which is suitable for single-frequency and multi-frequency signals, and it does not require repeated multiple frequency processing. In time domain beamforming, the steered-response power (SRP) is commonly used. Steered-response power delay and sum (SRP-DAS) is used for direction estimation based on microphone arrays. The steered-response power phase transform (SRP-PHAT) algorithm is an optimization of the SRP-DAS, which is easy to implement and has stronger robustness than SRP-DAS by whitening the signals [18,19]. The SRP-PHAT algorithm as time domain beamforming has been widely used in target tracking and distributed localization [20–22]. Salvati et al. reported the SRP-PHAT modification algorithm, which can speed up the operation [23]. However, the directivity based on SRP-PHAT is not outstanding in azimuth estimation, and the localization accuracy still needs further improvement.

Therefore, combining the advantages of MCCC and SRP, a new beamforming method (SRP-MCCC) is proposed in this paper. In this method, the wave direction weight (DW) is calculated by the MCCC, which adjusts the SRP value to enhance the directivity of microphone arrays and improve spatial resolution. In this paper, Section 2 describes the positioning principle of the proposed method. Then, the sound source localization simulation is reviewed in Section 3. Section 4 verifies the performance of the proposed method under experimental conditions. Finally, conclusions are given in Section 5.

2. Positioning Principle

Figure 1 shows the calculating flow of the position. Suppose the numbers of arrays and each array elements are M and N , respectively. For each array, the MCCC is evaluated by the N -elements' signals. The DW is constructed by using redundant information of MCCC. After obtaining the DW, the weighted beamforming is performed, and finally, the relative direction $(\theta_I, \theta_{II}, \dots, \theta_M)$ is found. When $M \geq 2$, the source position can be calculated from $(\theta_I, \theta_{II}, \dots, \theta_M)$.

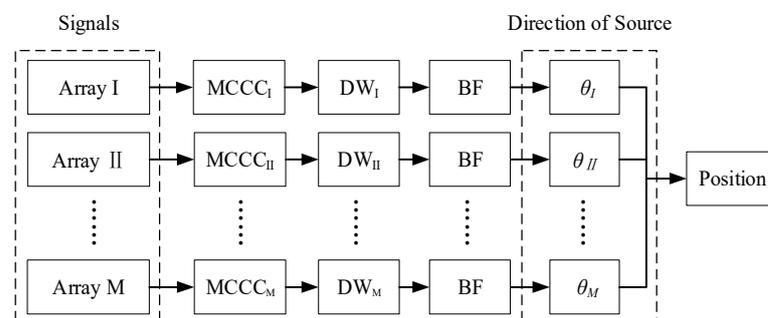


Figure 1. Positioning flowchart.

2.1. Signal Model

The far-field signal propagation and the N -element linear microphone array model are shown in Figure 2. θ_0 represents the direction from the far-field source to the array, and d_i ($i = 1, 2, \dots, N - 1$) denotes the spacing between the two array elements in the array.

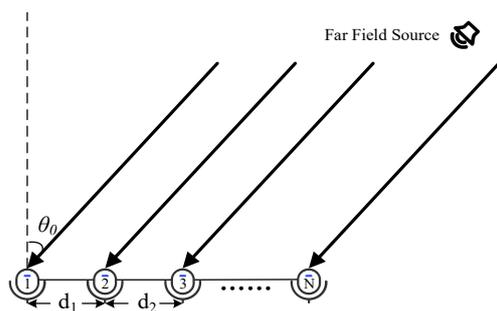


Figure 2. Signal and array model.

Assume that the source signal is $s(n)$, which is a time series. The signals received by the N -element linear microphone array can be expressed as $X(n)$.

$$X(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \\ \dots \\ x_{N-1}(n) \\ x_N(n) \end{bmatrix} = \begin{bmatrix} \alpha_1 s(n - t_1) + v_1(n) \\ \alpha_2 s(n - t_2) + v_2(n) \\ \dots \\ \alpha_{N-1} s(n - t_{N-1}) + v_{N-1}(n) \\ \alpha_N s(n - t_N) + v_N(n) \end{bmatrix} \tag{1}$$

where $x_i(n) (i = 1, 2, \dots, N)$ represents the signal received by the i -th microphone, α_i is the attenuation coefficient of the signal received by the i -th microphone, t_i is the propagation time from the sound source localization to the i -th microphone and $v_i(n)$ is the noise signal received by the i -th microphone.

Taking the first microphone as the reference, the aligned signal $y(n, \tau)$ can be written as:

$$y(n, \tau) = [\alpha'_1 x_1(n - \tau_1) \quad \alpha'_2 x_2(n - \tau_2) \quad \dots \quad \alpha'_{N-1} x_{N-1}(n - \tau_{N-1}) \quad \alpha'_N x_N(n - \tau_N)]^T \\ = [\alpha'_1 s'_1(n - \tau_1) \quad \alpha'_2 s'_2(n - \tau_2) \quad \dots \quad \alpha'_{N-1} s'_{N-1}(n - \tau_{N-1}) \quad \alpha'_N s'_N(n - \tau_N)]^T \\ + [v_1(n - \tau_1) \quad v_2(n - \tau_2) \quad \dots \quad v_{N-1}(n - \tau_{N-1}) \quad v_N(n - \tau_N)]^T \tag{2}$$

where $s'_i(n - \tau_i)$ represents the i -th signal after alignment. $\tau_i (i = 1, 2, \dots, n)$ is the time delay from the i -th microphone to the first microphone, and its value is related to the position of the sound source and the microphone array structure, where $\tau_1 = 0$. α'_i is the relative attenuation coefficient, and it is calculated by

$$\alpha'_i = \frac{\alpha_i}{\alpha_1} \quad (i = 1, 2, \dots, N) \tag{3}$$

For a linear array, the following relationship exists between τ_i and d_n (the distance between the n -th and $(n + 1)$ -th microphone):

$$\tau_i = \begin{cases} 0 & (i = 1) \\ \sum_{n=1}^{i-1} d_n & \\ \frac{n-1}{d_1} \tau_2 & (i \geq 2) \end{cases} \tag{4}$$

2.2. Direction Weight

The Pearson coefficient of the normalized signals is used to replace the value of the correlation function [24,25]. The correlation coefficient matrix of different time delays is $\rho(\tau)$:

$$\begin{aligned} \rho(\tau) &= P y(n, \tau) y(n, \tau)^T P \\ &= \begin{bmatrix} \rho_{11}(\tau_1, \tau_1) & \rho_{12}(\tau_1, \tau_2) & \dots & \rho_{1N}(\tau_1, \tau_N) \\ \rho_{21}(\tau_2, \tau_1) & \rho_{22}(\tau_2, \tau_2) & \dots & \rho_{2N}(\tau_2, \tau_N) \\ \dots & \dots & \dots & \dots \\ \rho_{N1}(\tau_N, \tau_1) & \rho_{N2}(\tau_N, \tau_2) & \dots & \rho_{NN}(\tau_N, \tau_N) \end{bmatrix} \end{aligned} \tag{5}$$

where

$$P = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \dots & \\ & & & \frac{1}{\sigma_N} \end{bmatrix} \tag{6}$$

$\rho_{ij}(\tau)$ is the correlation coefficient between the i -th and the j -th alignment signal, which can be expressed as:

$$\begin{aligned} \rho_{ij}(\tau_i, \tau_j) &= \frac{x_i(n-\tau_i) \cdot x_j(n-\tau_j)}{\sigma_i \sigma_j} \\ &= \frac{\alpha'_i s'_i(n-\tau_i) \cdot \alpha'_j s'_j(n-\tau_j)}{\sigma_i \sigma_j} + \frac{\alpha'_i s'_i(n-\tau_i) \cdot v_j(n-\tau_j)}{\sigma_i \sigma_j} \\ &\quad + \frac{\alpha'_j s'_j(n-\tau_j) \cdot v_i(n-\tau_i)}{\sigma_i \sigma_j} + \frac{v'_i(n-\tau_i) \cdot v'_j(n-\tau_j)}{\sigma_i \sigma_j} \\ &= r_{ij} + r_{iv_j} + r_{jv_i} + r_{v_i v_j} \end{aligned} \tag{7}$$

where $\sigma_i = \sum x_i^2$ is the energy of the i -th signal and \cdot indicates the inner product. Observing Equation (7), it can be deduced that r_{ij} is the correlation coefficient between the i -th signal and the j -th signal. r_{iv_j} represents the correlation coefficient between the i -th signal and the j -th noise, and r_{jv_i} is the correlation coefficient between the j -th signal and the i -th noise. $r_{v_i v_j}$ is expressed as the correlation coefficient between the i -th noise and the j -th noise. It can be seen that when $s'_i(n - \tau_i) = s'_j(n - \tau_j)$, r_{ij} has the maximum value. The $h_{DW}(\tau)$ is constructed by using redundant information of the MCCC. $h_{DW}(\tau)$ is calculated from the correlation coefficient matrix, which is directly related to the signal itself, and it can be represented as:

$$h_{DW}(\tau) = \prod_{i=1}^N \prod_{j=1}^i \rho_{ij}(\tau_i, \tau_j) \tag{8}$$

2.3. Weighted Beamforming

The weighted beamforming process of a single array after obtaining the $h_{DW}(\tau)$ is shown in Figure 3. First, the compensation time delay τ'_i ($i = 1, 2 \dots \dots N$) is performed on the signals $x_i(t)$ received by the array i_A ($i_A = 1, 2 \dots \dots M$) to obtain the compensation signals $x_i(t - \tau'_i)$. Then, the weighted signals $h_{DW}(\tau)x_i(t - \tau'_i)$ are calculated by multiplying $x_i(t - \tau'_i)$ and $h_{DW}(\tau)$. Subsequently, the spatial integration of $h_{DW}(\tau)x_i(t - \tau'_i)$ at the same instant is performed, and the time integration of all signals at different times is operated. Finally, $D_{i_A}(\theta)$ (beam output of array i_A) is calculated. Since $h_{DW}(\tau)$ contains spatial information, it can enhance the incoming wave direction and suppress the non-incoming wave direction, which can improve the directional ability of $D_{i_A}(\theta)$.

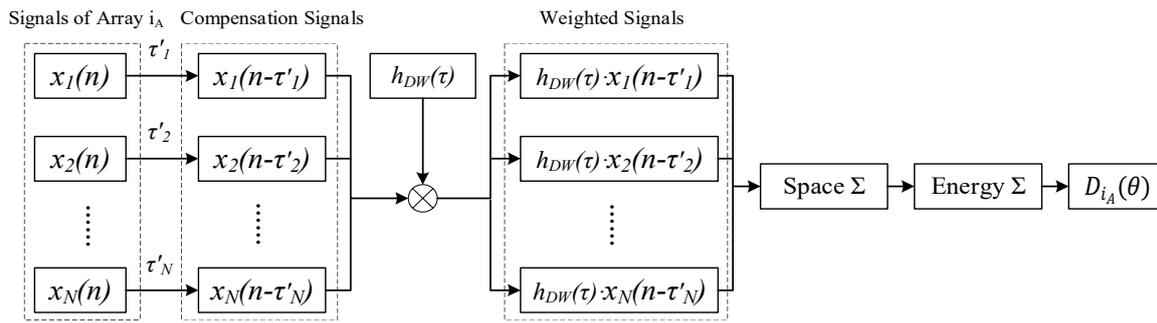


Figure 3. Weighted beamforming.

$D_{i_A}(\theta)$ can be expressed as:

$$D_{i_A}(\theta) = \sum |h^T y(n, \tau)|^2 \tag{9}$$

where $h^T = h_{DW}(\tau)[1 \ 1 \ \dots \ 1]$, $\tau = \frac{d_1 \sin \theta}{c}$ and c is the speed of sound in the environment. When $\theta = \theta_0$, $D_{i_A}(\theta)$ has a maximum value; at this time, θ_0 is the direction of the sound source.

The frequency domain expression can be written as:

$$D_{i_A}(\theta) = \frac{1}{2\pi} \prod_{j=1}^N \prod_{i=1}^j \int_{-\pi}^{\pi} \rho_{ij}(\tau_i, \tau_j) X_i(e^{j\omega}) X_j^*(e^{j\omega}) e^{j\omega(\tau_i - \tau_j)} d\omega \tag{10}$$

where $X_i(e^{j\omega})$ and $X_j(e^{j\omega})$ represent the Fourier transforms of the signals $x_i(t)$ and $x_j(t)$, respectively.

As the number of arrays is M , the coordinates (x, y) of the sound source can be obtained by the following equation:

$$D_{all} = \sum_{i_A=1}^M D_{i_A}(f_{i_A}(x, y)) \tag{11}$$

$$(x, y) = \underset{(x, y)}{\operatorname{argmax}} D_{all}$$

where $f_{i_A}(x, y) = \theta_{i_A}$ represents the direction of the source coordinate (x, y) relative to the normal direction of the array i_A . $f_{i_A}(x, y)$ represents a mapping from coordinate space to direction space. When D_{all} is maximized, the corresponding (x_0, y_0) is the maximum probability of the sound source.

3. Simulation Analysis

We used two sets of equally spaced linear arrays to simulate the sound source localization. The coordinates of the i -th microphone of the array I are $((i - 1) \times d, 0)$ ($i = 1, 2, 3 \dots 16$), and the coordinates of the i -th microphone of the array II are $((i - 1) \times d, 10)$. d is the distance between two adjacent microphones and $d = 0.043$ m. D is the distance between the array I and array II angle and $D = 10$ m. The sound source and array locations are shown in Figure 4. θ_1 and θ_2 are sound source directions relative to array I and array II, respectively. Assuming that the normal array is the beginning edge and the incoming wave direction is the end edge, θ_1 and θ_2 are positive counterclockwise and are negative clockwise. Moreover, θ_1 and $\theta_2 \in [-90^\circ, 90^\circ)$. The center of the array is used as a reference point, and the sound source coordinates (x, y) can be calculated by the following:

$$\frac{y}{x - 7.5d} = \tan(\theta_1 + 90^\circ)$$

$$\frac{y - D}{x - 7.5d} = \tan(\theta_2 - 90^\circ) \tag{12}$$

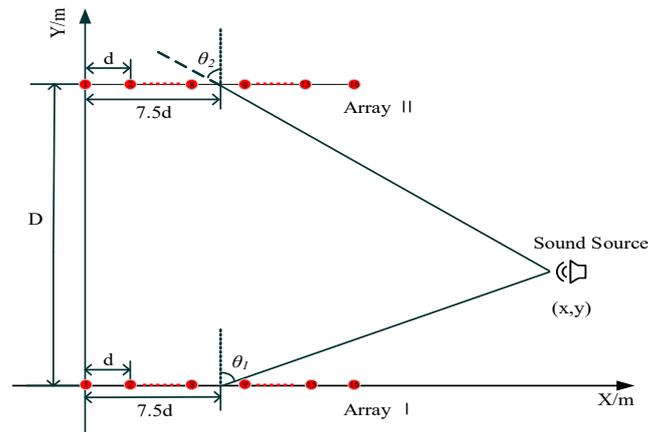


Figure 4. Sound source and microphone array simulation location.

Figure 5 shows the direction estimation results of the SRP-PHAT and the SRP-MCCC when the source signal is a 600 Hz sine signal with source coordinates (14, 4) and the SNR is -5 dB. The brightness of a point represents the probability that the point is the location of the sound source. Figure 6 shows the location estimation results of the SRP-PHAT and the SRP-MCCC when the SNR is -5 dB.

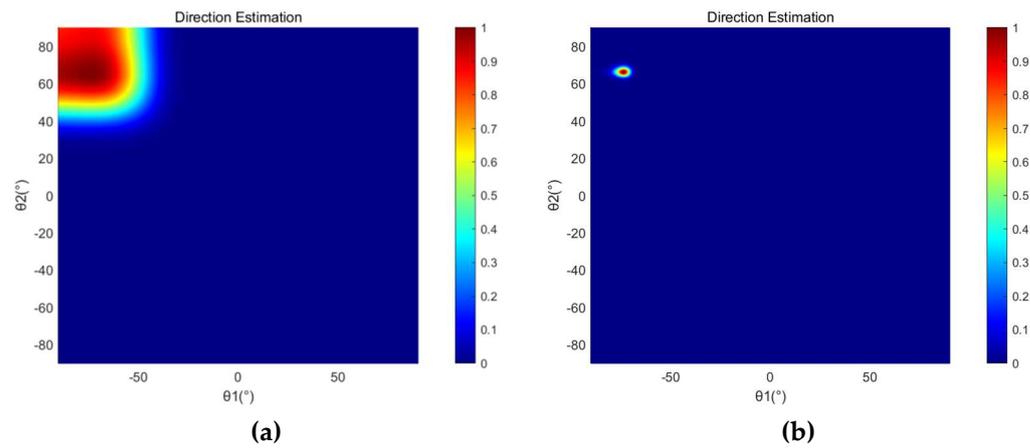


Figure 5. The direction estimation simulation results of (a) SRP-PHAT and (b) SRP-MCCC when the SNR is -5 dB.

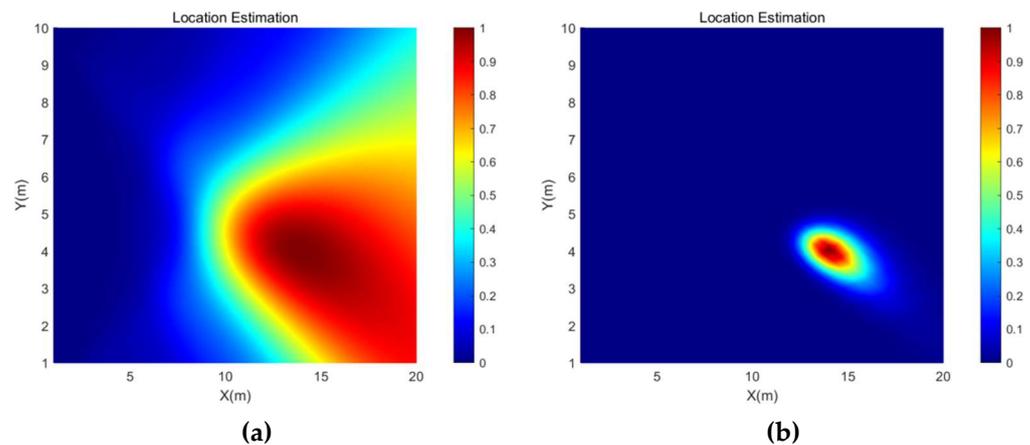


Figure 6. The location estimation simulation results of (a) SRP-PHAT and (b) SRP-MCCC when the SNR is -5 dB.

From Figures 5 and 6, we can observe that the SRP-MCCC has a superior localization convergence capability and can improve the localization accuracy by aggregating the localization results into smaller areas and improving spatial pointing. To further explore the advantages of this method, simulation analyses of SRP-PHAT and the SRP-MCCC were conducted under different SNR conditions, and the results are shown in Table 1.

Table 1. Simulation results for different SNR conditions (600 Hz, (14, 4)).

SNR/dB	SRP-PHAT			SRP-MCCC		
	Position/m	Directional Deviation/°	Distance Error/%	Position/m	Directional Deviation/°	Distance Error/%
10	(14.25, 3.94)	(−0.36, 0.12)	1.77	(13.96, 4.00)	(0.04, −0.08)	0.27
0	(14.27, 3.86)	(−0.46, −0.28)	2.09	(14.19, 3.96)	(−0.26, −0.03)	1.33
−5	(15.24, 3.84)	(−1.76, 1.12)	8.59	(13.78, 3.90)	(−0.11, −0.83)	1.66
−10	(12.59, 4.10)	(1.59, −2.03)	9.71	(14.31, 4.13)	(0.24, 0.87)	2.31
−15	(15.61, 4.06)	(−1.33, 2.29)	11.06	(13.48, 4.10)	(1.01, −0.46)	3.64
−20	(16.14, 4.07)	(−1.76, 3.02)	14.71	(14.57, 4.26)	(0.49, 1.57)	4.30

As shown in Table 1, when the SNR is high, the localization errors of both methods are comparatively insignificant. As the SNR decreases, the localization error of the SRP-PHAT increases significantly, while that of the SRP-MCCC remains at a low level. When the SNR is reduced from 10 dB to −20 dB, the position error of the SRP-PHAT increases from 1.77% to 14.71%, and the error of the SRP-MCCC only increases from 0.27% to 4.30%. Simulation experiments show that the SRP-MCCC is feasible. The localization effect is excellent, and the robustness is outstanding.

To further explore the method proposed in this paper, the single-frequency signal at different source positions and the mixed-frequency signals at the same coordinate were simulated when SNR = −10 dB. The results are shown in Tables 2 and 3, respectively. We can infer that the average error of the single-frequency signal at different coordinates decreased by 5.69%, and that of the mixed-frequency signals decreased by 5.77% at the same coordinate.

Table 2. Simulation results for different source locations (SNR = −10 dB, 600 Hz).

Source Location/m	SRP-PHAT			SRP-MCCC		
	Position/m	Directional Deviation/°	Distance Error/%	Position/m	Directional Deviation/°	Distance Error/%
(12, 7)	(10.96, 6.62)	(1.28, −3.53)	8.81	(12.22, 6.79)	(−0.96, −0.84)	2.19
(15, 6)	(16.26, 6.18)	(1.41, 1.11)	7.88	(15.52, 6.03)	(−0.46, 0.51)	3.22
(17, 4.5)	(18.27, 4.04)	(−2.34, −0.23)	7.51	(17.47, 4.60)	(−0.02, 0.85)	2.73
(19, 6.5)	(20.99, 6.84)	(0.78, 1.87)	10.05	(19.81, 6.92)	(0.46, 1.58)	4.05
(21, 5)	(21.21, 4.15)	(1.08, −0.63)	8.83	(21.52, 4.90)	(−0.62, 0.50)	2.45

Table 3. Simulation results for different frequencies (SNR = −10 dB, (15, 5)).

Frequency/Hz	SRP-PHAT			SRP-MCCC		
	Position/m	Directional Deviation/°	Distance Error/%	Position/m	Directional Deviation/°	Distance Error/%
600	(13.59, 5.09)	(2.21, −1.52)	8.93	(14.61, 5.09)	(0.81, −0.16)	2.53
600, 900	(16.17, 5.21)	(−0.62, 2.00)	7.52	(15.41, 5.13)	(−0.03, 0.92)	2.72
600, 900, 1500	(13.68, 5.08)	(2.03, −1.42)	8.36	(14.73, 5.23)	(1.15, 0.50)	2.24

4. Experiment

To verify the feasibility of the SRP-MCCC, a field localization experiment was conducted, as shown in Figure 7. The array structure and the sound source frequency are

consistent with the simulation during the experiment, as shown in Figure 4. This paper presents an experimental comparison between the SRP-PHAT and the SRP-MCCC.



Figure 7. (a) Experimental scene. (b) Experimental equipment.

The direction estimation and position estimation results of SRP-PHAT and SRP-MCCC when the sound source coordinate is (10, 5) are shown in Figures 8 and 9, respectively. From Figures 8 and 9, it can be found that the spot size of the SRP-MCCC, in both angle and position, is smaller than SRP-PHAT. Therefore, the SRP-MCCC has a stronger angular resolution and can control the beam width in a smaller range, which can lead to more accurate direction and position estimation.

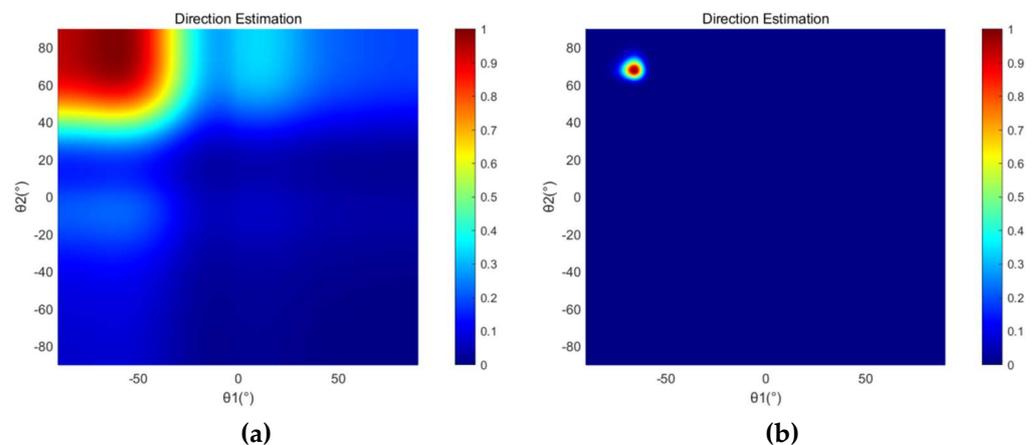


Figure 8. The direction estimation experimental results of (a) SRP-PHAT and (b) SRP-MCCC when the sound source coordinate is (10, 5).

Subsequently, localization experiments were conducted for the sound source coordinates of (11, 2) and (12, 7.5), and the experimental results are shown in Table 4. Compared with the SRP-PHAT, the SRP-MCCC has higher accuracy in both direction and position estimation, and the average error is reduced by 8.14%. Therefore, SRP-MCCC has a better localization effect, which is in line with the expected results and verifies the feasibility and superiority of the SRP-MCCC.

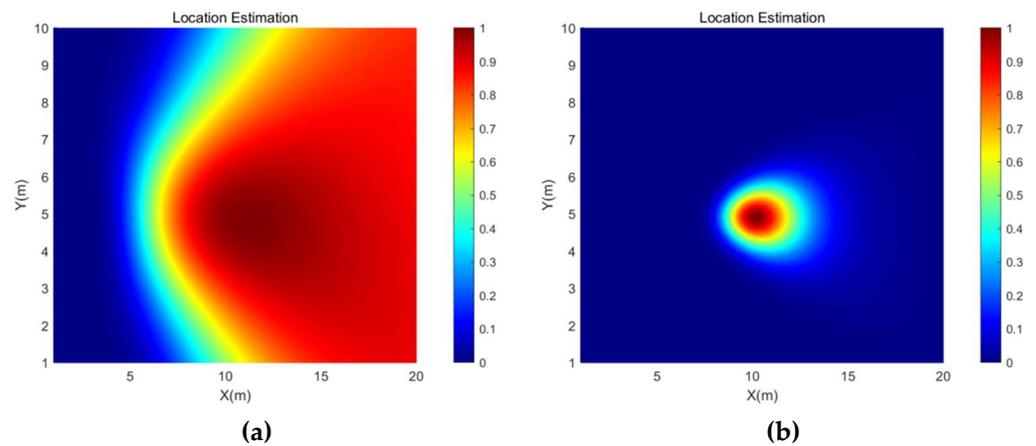


Figure 9. The location estimation experimental results of (a) SRP-PHAT and (b) SRP-MCCC when the sound source coordinate is (10, 5).

Table 4. Localization results of SRP-PHAT and MCCC-SRP at different coordinates.

Source Location/m	$(\theta_1, \theta_2)/^\circ$	SRP-PHAT			SRP-MCCC		
		Direction/ $^\circ$	Position/m	Error/%	Direction/ $^\circ$	Position/m	Error/%
(10, 5)	(−64.16, 64.16)	(−65.60, 66.08)	(11.14, 4.80)	10.35	(−64.99, 64.18)	(10.20, 4.91)	1.96
(11, 2)	(−79.98, 54.76)	(−81.94, 55.73)	(11.83, 1.72)	7.83	(−79.11, 54.75)	(10.81, 2.13)	2.06
(12, 7.5)	(−58.67, 78.53)	(−61.77, 79.12)	(13.39, 7.36)	12.62	(−59.52, 78.26)	(12.24, 7.39)	2.36

5. Conclusions

A new high-precision beamforming algorithm (SRP-MCCC) is proposed to improve the positioning accuracy by combining SPR and MCCC. A detailed theoretical analysis of the method is presented here, and the simulations and experiments verify its feasibility. The results show that the method has the advantages of strong robustness and high localization accuracy. Furthermore, the SRP-MCCC has better spatial resolution and localization capability than the SRP-PHAT. Both the simulations and experiments verify the effectiveness of the method. These results provide a new idea for the weighted beamforming algorithm, which is essential for researching high-precision sound source localization in complex environments.

Author Contributions: Conceptualization, M.L. and Z.J.; methodology, M.L. and J.H.; validation, M.L., Z.J. and J.H.; formal analysis, Z.J.; investigation, J.H. and Q.Z.; data curation, J.H. and Q.Z.; writing—original draft preparation, J.H. and Q.Z.; writing—review and editing J.H. and Q.Z.; visualization, J.H. and Q.Z.; supervision, Z.J. and L.N.; project administration; M.L., Z.J. and L.N.; funding acquisition, L.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China [Grant Nos. 51805154 and 51975191], Green Industry Technology Leading Project of Hubei University of Technology [XJ2021004901], and the Scientific Research Foundation of Hubei University of Technology [GCRC 2020010].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aich, A.; Palanisamy, P. On-grid DOA estimation method using orthogonal matching pursuit. In Proceedings of the International Conference on Signal Processing and Communications (ICSPC) IEEE, Coimbatore, India, 28–29 July 2017; pp. 483–487. [\[CrossRef\]](#)
2. Argentieri, S.; Danès, P.; Souères, P. A survey on sound source localization in robotics: From binaural to array processing methods. *Comput. Speech Lang.* **2015**, *34*, 87–112. [\[CrossRef\]](#)
3. Pradhan, D.; Bera, R. Direction of Arrival Estimation via ESPRIT Algorithm for Smart Antenna System. *Int. J. Comput. Appl.* **2015**, *118*, 5–7. [\[CrossRef\]](#)
4. Manamperi, W.; Abhayapala, T.D.; Zhang, J.; Samarasinghe, P.N. Drone audition: Sound source localization using on-board microphones. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 508–519. [\[CrossRef\]](#)
5. Kong, D.F.N.; Shen, C.; Tian, C.; Zhang, K. A New Low-Cost Acoustic Beamforming Architecture for Real-Time Marine Sensing: Evaluation and Design. *J. Mar. Sci. Eng.* **2021**, *9*, 868. [\[CrossRef\]](#)
6. Schmidt, R.O. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [\[CrossRef\]](#)
7. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [\[CrossRef\]](#)
8. Hu, Y.; Abhayapala, T.D.; Samarasinghe, P.N. Multiple source direction of arrival estimations using relative sound pressure based music. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *29*, 253–264. [\[CrossRef\]](#)
9. Herzog, A.; Habets, E.A.P. Eigenbeam-ESPRIT for DOA-Vector Estimation. *IEEE Signal Process. Lett.* **2019**, *26*, 572–576. [\[CrossRef\]](#)
10. Dvorkind, T.G.; Gannot, S. Time difference of arrival estimation of speech source in a noisy and reverberant environment. *Signal Process.* **2005**, *85*, 177–204. [\[CrossRef\]](#)
11. Chung, M.-A.; Chou, H.-C.; Lin, C.-W. Sound Localization Based on Acoustic Source Using Multiple Microphone Array in an Indoor Environment. *Electronics* **2022**, *11*, 890. [\[CrossRef\]](#)
12. He, H.; Wu, L.; Lu, J.; Qiu, X.; Chen, J. Time Difference of Arrival Estimation Exploiting Multichannel Spatio-Temporal Prediction. *IEEE Trans. Audio Speech, Lang. Process.* **2012**, *21*, 463–475. [\[CrossRef\]](#)
13. Chiariotti, P.; Martarelli, M.; Castellini, P. Acoustic beamforming for noise source localization—Reviews, methodology and applications. *Mech. Syst. Signal Process.* **2018**, *120*, 422–448. [\[CrossRef\]](#)
14. Yang, Y.; Zhigang, C. A Review of High-performance Beamforming Methods for Acoustic Source Identification. *Chin. J. Mech. Eng.* **2021**, *57*, 166–183. [\[CrossRef\]](#)
15. Zhao, Y.; Jensen, J.R.; Jensen, T.L.; Chen, J.; Christensen, M.G. Experimental study of robust acoustic beamforming for speech acquisition in reverberant and noisy environments. *Appl. Acoust.* **2020**, *170*, 107531. [\[CrossRef\]](#)
16. Yang, J.; Liao, G.; Li, J. Robust adaptive beamforming in nested array. *Signal Process.* **2015**, *114*, 143–149. [\[CrossRef\]](#)
17. Padois, T.; Fischer, J.; Doolan, C.; Doutres, O. Acoustic imaging with conventional frequency domain beamforming and generalized cross correlation: A comparison study. *Appl. Acoust.* **2021**, *177*, 107914. [\[CrossRef\]](#)
18. Cobos, M.; Marti, A.; Lopez, J.J. A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization With Scalable Spatial Sampling. *IEEE Signal Process. Lett.* **2010**, *18*, 71–74. [\[CrossRef\]](#)
19. Boora, R.; Dhull, S.K. Performance Evaluation of Iterative SRP-PHAT Techniques for Acoustic Source Localization. In Proceedings of the First International Conference on Computational Electronics for Wireless Communication, Singapore, 3 January 2022; Springer: Berlin/Heidelberg, Germany; pp. 403–418. [\[CrossRef\]](#)
20. Wu, K.; Khong, A.W.H. Acoustic source tracking in reverberant environment using regional steered response power measurement. In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, Taiwan, 29 October–1 November 2013; pp. 1–6. [\[CrossRef\]](#)
21. Huang, J.; Jin, N.; Wang, L.; Chen, X.; Sheng, X.; Yang, S.; Zhao, X.; Sun, L.; Zhu, M. Acoustic Source Localization with Distributed Smartphone Arrays. In Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015; pp. 1–6. [\[CrossRef\]](#)
22. Diaz-Guerra, D.; Miguel, A.; Beltran, J.R. Robust Sound Source Tracking Using SRP-PHAT and 3D Convolutional Neural Networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *29*, 300–311. [\[CrossRef\]](#)
23. Salvati, D.; Drioli, C.; Foresti, G.L. Sensitivity-based region selection in the steered response power algorithm. *Signal Process.* **2018**, *153*, 1–10. [\[CrossRef\]](#)
24. Dunn, O.J.; Clark, V.A. *Applied Statistics: Analysis of Variance and Regression*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 1986. [\[CrossRef\]](#)
25. Rezayee, A.; Gazor, S. An adaptive KLT approach for speech enhancement. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 87–95. [\[CrossRef\]](#)