Measurement Science and Technology

PAPER • OPEN ACCESS

Efficient energy-based orthogonal matching pursuit algorithm for multiple sound source localization with unknown source count

OPEN ACCESS
IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. **33** (2022) 045018 (10pp)
https://doi.org/10.1088/1361-6501/ac3d46

Efficient energy-based orthogonal matching pursuit algorithm for multiple sound source localization with unknown source count

Rongjiang Tang[1], Yingxiang Zuo[1], Weiya Liu[1,*], Weiguang Zheng[1,3], Liguo Tang[1] and Fei Xiao[2]

[1] School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, People's Republic of China
[2] Commercial Vehicle Technology Center, Dong Feng Liuzhou Automobile Co., Ltd, Liuzhou 545005, People's Republic of China
[3] School of Mechanical and Automotive Engineering, Guangxi University of Science and Technology, Liuzhou 545616, People's Republic of China

E-mail: weiya_liu813@126.com

Received 8 July 2021, revised 14 November 2021
Accepted for publication 25 November 2021
Published 31 January 2022

Abstract

In this paper, we propose a compressed sensing (CS) sound source localization algorithm based on signal energy to solve the problem of stopping the iteration condition of the orthogonal matching pursuit (OMP) reconstruction algorithm in CS. The orthogonal matching tracking algorithm needs to stop iteration according to the number of sound sources or the change of residual. Generally, the number of sound sources cannot be known in advance, and the residual often leads to unnecessary calculation. Because the sound source is sparsely distributed in space, and its energy is concentrated and higher than that of the environmental noise, the comparison of the signal energy at different positions in each iteration reconstruction signal is used to determine whether the new sound source is added in this iteration. At the same time, the block sparsity is introduced by using multiple frequency points to avoid the problem of different iteration times for different frequency points in the same frame caused by the uneven energy distribution in the signal frequency domain. Simulation and experimental results show that the proposed algorithm retains the advantages of the orthogonal matching tracking sound source localization algorithm, and can complete the iteration well. Under the premise of not knowing the number of sound sources, the maximum error between the number of iterations and the set number of sound sources is 0.31. The experimental results show that the proposed algorithm has good positioning accuracy and has certain anti-reverberation capability. Compared with other OMP algorithms, the proposed algorithm has better iterative ability and stability. This work is helpful in promoting the development of multiple sound source localization.

* Author to whom any correspondence should be addressed.

Keywords: compressed sensing, sound source localization, signal energy

(Some figures may appear in colour only in the online journal)

1. Introduction

Speech interaction technology has been rapidly developed and applied in many fields, such as in smart homes [1, 2], intelligent robots [3–6], and vehicle-mounted intelligent systems. Sound location technology based on microphone arrays plays an important role as its front-end processing. Traditional sound location technology is divided into three categories: (a) the conventional beam forming algorithm [7]; (b) the time difference of arrivals algorithm [8, 9]; and (c) the high-resolution spectrum estimation algorithm [10]. The traditional algorithm is relatively mature, but it still has the disadvantages of low resolution, vulnerability to environmental factors and the inability to locate coherent signals [11, 12].

In recent years, more and more scholars have begun to explore new measures for sound source localization. The measurement performance of the sound source positioning system not only depends on the hardware conditions of the microphone array [13], but is also greatly related to the software algorithms, including the four-element microphone array groups measure [14], machine learning, the steered response power-phase transform (SRP-PHAT) algorithm and multiple signal classification (Music) algorithm [15], the orthogonal matching pursuit (OMP) algorithm and so on. The new algorithm performs well in locating both multi-source and monophonic sources, and has applications in many fields [5, 6]. The OMP algorithm not only has better resolution, but can also complete sound source location with less information. Its contents can be divided into sparse signal representation, the design of the sensing matrix (or measurement matrix) and the reconstruction algorithm [16], since the sound source can be regarded as a point source, and the number of sound sources in space is very sparse. Based on compressed sensing (CS), by using the spatial scarcity, the localization can be transformed into the sparse recovery problem [17]. Angeliki *et al* [18] used the CS weighted l_1 norm minimization algorithm for the direction of arrival estimate. The influence of the coherence of the perceptual matrix and the signal-to-noise ratio on the performance of DOA estimation is analysed, and compared with the traditional algorithm, and the superiority of the compressive sensing algorithm is proved. For the mismatch of the CS basis caused by conventional grid-based measurement matrix construction, some scholars have proposed gridless compression sensing sound location [19, 20]. For the sparse reconstruction part of the signal, the greedy algorithm is widely used because of its lower computational cost and low complexity. Among them, the OMP algorithm is particularly prominent. Ning *et al* [21] pointed out that the OMP algorithm can still locate the sound source more accurately when the measurement matrix does not have the restricted isometry property. Then the author proposes the OMP-SVD algorithm which combines singular value decomposition (SVD) and the OMP algorithm to reduce the influence of noise. The OMP algorithm generally has two criteria for stopping: the number of known iteration times or the residual reaches a certain threshold. The residual threshold is difficult to determine in different environments for the source location [22]. Too few iterations will lead to incomplete reconstruction and the omission of sound sources. Too much noise causes non-zero values in most passive source locations, which will affect the positioning performances. The optimal number of iterations is the number of sound sources, but the actual number of sound sources is unknown.

To overcome the defects of the OMP algorithm in multiple sound source localization measurements, an energy-based OMP (EbOMP) is proposed, based on the energy of the reconstruction signal [23, 24]. When the signal energy corresponding to the newly-added non-zero position in the iteration is less than the set ratio of the signal energy corresponding to the reconstructed position, the OMP algorithm stops iteration. At the same time, because the OMP algorithm can only calculate the single frequency point at a time, as well as the energy distribution of each frequency point of the sound source being uneven, the block sparsity is introduced [25], all frequencies are put into the solution model and the sum of energy of all frequency points is calculated each iteration for comparison. The simulation and experimental results show that the algorithm solves the problem of the uncertainty of the number of iterations of the OMP algorithm, and at the same time solves the problem that only a single frequency point and the uneven energy distribution of the frequency points can be calculated at a time. The algorithm proposed in this paper can accurately locate the location of sound sources and estimate the number of sound sources.

2. Signal model and algorithm

2.1. Sparse model in sound source localization

We assume that the microphone array is a circular planar array of radius R, and that the array has N microphones. Because the microphone array plane is symmetric on both sides, only the space on one side of the array is considered. The positive direction space of the array is divided into M directions according to the angle, among which there are K sound sources from different directions. In addition, $N \ll M$, $K \ll M$, as shown in figure 1,

where θ_i is the elevation angle of the ith sound source, and φ_i is the azimuth angle of the ith sound source (elevation angle is from the x-y plane to the positive z axis, azimuth angle is from the positive x axis to the positive y axis.). Since only the space in front of the array plane is considered, $0^\circ \leqslant \theta_i \leqslant 90^\circ$, $0^\circ \leqslant \varphi_i \leqslant 360^\circ$.

In the anechoic environment, the centre of the array circle is taken as the reference position, and signals being received by the nth microphone could be described as:

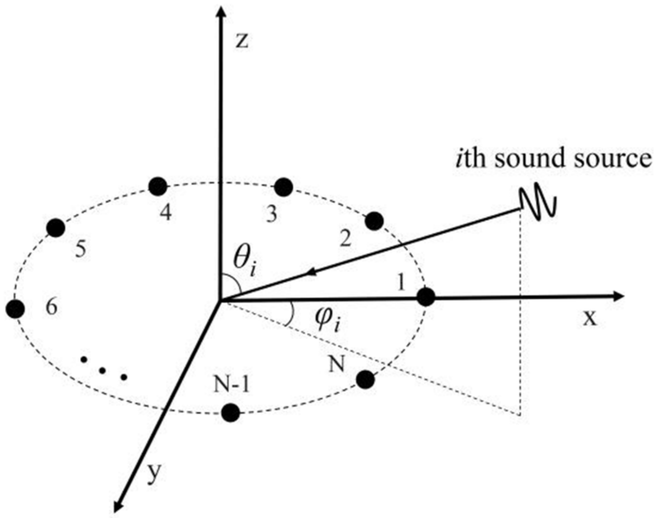


Figure 1. Diagram of microphone array receiving signal.

$$y_n(k) = \sum_{m=1}^{\mathrm{M}} s_m[k - \tau_{nm}] + v_n(k) \quad n = 1, 2, \cdots, N \tag{1}$$

where $s_m(k)$ ($m = 1, 2, \ldots$, M) is the signal being received at the reference position from the mth direction, $s_m(k)$ is non-zero only in the direction with the sound source, and it is assumed that the sound source signals in different directions are independent of each other; τ_{nm} is the time difference between the nth microphone and the reference position for the signal in the mth direction. $v_n(k)$ is the additive noise signal of the nth microphone, and it is assumed that it is unrelated to all source signals and observation noises of other microphones.

Transforming equation (1) to the frequency domain,

$$Y_n(\omega) = \sum_{m=1}^{\mathrm{M}} S_m(\omega) e^{-j\omega\tau_{nm}} + V_n(\omega) \quad n = 1, 2, \cdots, N \tag{2}$$

where $Y_n(\omega)$, $S_m(\omega)$ and $V_n(\omega)$ are the discrete Fourier transforms of $y_n(k)$, $s_m(k)$ and $v_n(k)$ respectively. The frequency domain signal model can be described as a vector.

$$\begin{bmatrix} Y_1(\omega) \\ Y_2(\omega) \\ \vdots \\ Y_{\mathrm{N}}(\omega) \end{bmatrix} = \begin{bmatrix} e^{-j\omega\tau_{11}} & e^{-j\omega\tau_{12}} & \cdots & e^{-j\omega\tau_{1\mathrm{M}}} \\ e^{-j\omega\tau_{21}} & e^{-j\omega\tau_{22}} & \cdots & e^{-j\omega\tau_{2\mathrm{M}}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-j\omega\tau_{\mathrm{N}1}} & e^{-j\omega\tau_{\mathrm{N}2}} & \cdots & e^{-j\omega\tau_{\mathrm{NM}}} \end{bmatrix} \begin{bmatrix} S_1(\omega) \\ S_2(\omega) \\ \vdots \\ S_{\mathrm{M}}(\omega) \end{bmatrix} + \begin{bmatrix} V_1(\omega) \\ V_2(\omega) \\ \vdots \\ V_{\mathrm{N}}(\omega) \end{bmatrix}. \tag{3}$$

This is simplified as

$$Y = AS + V \tag{4}$$

where $\boldsymbol{A}(N \times M)$ is called the measurement matrix. The number of sound sources $K \ll M$. Therefore, the signal vector $\boldsymbol{S}$ of the sound source is a sparse vector. Its non-zero value position is the space corresponding to the sound source direction.

2.2. OMP reconstruction algorithm

The essence of CS can be described by the problem of solving the sparse solution of the linear model $\boldsymbol{Y} = \boldsymbol{AS}$ through the measurement value $\boldsymbol{Y}$ and the measurement matrix $\boldsymbol{A}$, that is, the sparse sound source signal $\boldsymbol{S}$ is reconstructed. Due to $N \ll M$, there are multiple solutions to this problem. Usually, the l_0-norm is used to find the sparsest solution. In general, equation (4) can be expressed as the following constrained optimization problem

$$\min \|S\|_{l_0} \quad \text{s.t.} \ \|\boldsymbol{AS} - \boldsymbol{Y}\|_{l_2} \leqslant \varepsilon \tag{5}$$

where ε is determined by the upper limit of the given noise $\boldsymbol{V}$.

However, the solution of equation (5) is a non-deterministic polynomial hard (NP-hard) problem, which is usually difficult to calculate. Under certain conditions, the l_1-norm may be used instead of the l_0-norm [26]. By constraining the sparsity of $\boldsymbol{S}$, a better reconstruction effect of $\boldsymbol{S}$ can be achieved.

Signal reconstruction is the core problem of CS theory. At present, the OMP algorithm is widely used in sound source localization. The basic idea of this algorithm is to find out the atoms in the measurement matrix that can represent the measurement signal in a sparse linear manner by the iterative method.

The brief steps of the OMP algorithm are as follows.

Step 1. Determine measurement matrix $\boldsymbol{A}$, measurement value $\boldsymbol{Y}$ and iteration number K (number of sound sources).

Step 2. Initialize the residuals $\boldsymbol{r}_0 = \boldsymbol{Y}$, support set $\Lambda_0 = \oslash$, the number of iterations $t = 1$.

Step 3. The inner product of the residual and the measurement matrix was obtained, and the atom j_t with the largest result was selected, and the support set was updated as $\Lambda_t = \Lambda_{t\text{-}1} \cup j_t$.

$$j_t = \underset{j=1,2,\cdots,\mathrm{N}}{\operatorname{argmax}} \left| \langle \mathbf{r}_{t-1}, \boldsymbol{A}_j^{\mathrm{H}} \rangle \right| \tag{6}$$

where $\boldsymbol{A}^{\mathrm{H}}$ is the complex conjugate of $\boldsymbol{A}$.

Step 4. The current approximate solution, $\boldsymbol{S}_t$, is obtained by the least square method.

$$\boldsymbol{S}_t = \underset{z \in \mathbb{C}^{\mathrm{N}}}{\operatorname{argmin}} \left\{ \|\boldsymbol{Y} - \boldsymbol{A}_{\Lambda t} z\|_2 \right\}, z \in \Lambda_t. \tag{7}$$

Step 5. The approximate solution $\boldsymbol{Y}_t$ and the new residual $\boldsymbol{r}_t$ are calculated.

$$\boldsymbol{Y}_t = A\boldsymbol{S}_t \tag{8}$$

$$\mathbf{r}_t = Y - \boldsymbol{Y}_t. \tag{9}$$

Step 6. Determine whether the iteration number t reaches the set value K, if it does not reach the return step (3).

Step 7. Obtain an estimate of the original signal ***S, St***.

As can be seen from equation (3), the atoms in the measurement matrix that constitute the measured values can be gradually selected in step (3) of the algorithm iteration, and the optimal coefficient corresponding to the selected atoms can

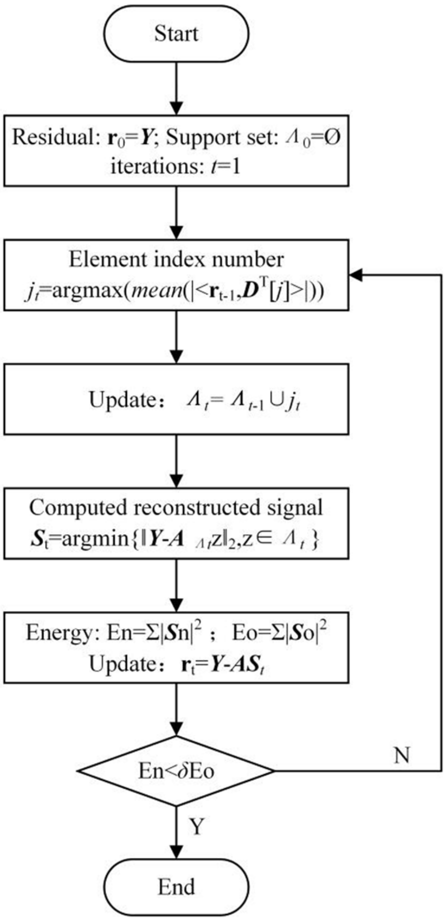


Figure 2. Algorithm flow steps.

be obtained by the least square method. In general, the number of iterations can be set to be consistent with the sparsity. In source location, the number of sound sources is the sparsity of the received signal. But in the real signal model, the number of sound sources K is often not available in advance. Another commonly used condition for the OMP algorithm to stop iteration is: stop when the change of residuals between two successive iterations is less than a certain threshold. However, it is difficult to determine the threshold value, and a value that is too small will lead to too many iterations, increase the non-zero value of the reconstruction coefficient vector, and increase unnecessary calculations. Too large a number of iterations may result in too few iterations to completely reconstruct all source positions. Only when the number of iterations is signal sparsity is the reconstruction result the optimal solution.

2.3. OMP algorithm based on signal energy

Based on the above analysis, signal energy is introduced as the stop iteration condition of the OMP algorithm, and the sound source sparsity is extended to the block sparsity considering the majority of frequency points. The algorithm flow steps in this paper are shown in figure 2.

According to equation (3), the non-zero elements of the reconstructed target sparse vector $\boldsymbol{S}$ are the received sound source signals in all directions. Because the sound source signal energy is more concentrated and general than the noise signal energy, and by the section 2.2 OMP algorithm steps, the number of iterations does not exceed that of the sound source, in the algorithm of step (4) the approximation of the solution is based on the sound source signal which has great energy, when the number of iterations is more than that of the sound source, new solutions for the noise caused by a non-zero value have less energy. Therefore, the change in reconstructed signal energy is taken as the condition for the algorithm to stop iteration.

In the part of the iteration judgment condition, the energy threshold δ is set. This determines the ratio of the newly-added non-zero position signal energy value to the reconstructed position signal energy in the iteration result. When the ratio is less than the set energy threshold, the iteration is stopped.

Because the general signal energy is not uniformly distributed in the frequency domain, the sparse signal model with a multi-frequency point construction block is selected. According to the literature, the signal model is extended to block sparsity, which can be expressed as the following formula.

$$\begin{bmatrix} Y_1(\omega_1) \\ \vdots \\ Y_1(\omega_F) \\ Y_2(\omega_1) \\ \vdots \\ Y_2(\omega_F) \\ \vdots \\ Y_N(\omega_F) \end{bmatrix} = \begin{bmatrix} D(1) \\ D(2) \\ \vdots \\ D(\mathrm{M}) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} S_1(\omega_1) \\ \vdots \\ S_1(\omega_F) \\ S_2(\omega_F) \\ \vdots \\ S_2(\omega_F) \\ \vdots \\ S_N(\omega_F) \end{bmatrix} + \begin{bmatrix} V_1(\omega_1) \\ \vdots \\ V_1(\omega_F) \\ V_2(\omega_F) \\ \vdots \\ V_2(\omega_F) \\ \vdots \\ V_N(\omega_F) \end{bmatrix} \tag{10}$$

where F is the number of extended frequencies. The measurement matrix is extended as $\boldsymbol{A} = [\boldsymbol{D}(1), \boldsymbol{D}(2), \ldots, \boldsymbol{D}(\mathrm{M})]$, where $\boldsymbol{D}(m)$ is the measurement matrix after frequency expansion of the measurement vector in the mth direction.

3. Simulation results and analysis

In order to verify the feasibility and advantages of the algorithm proposed in this paper, the parameters of the EbOMP algorithm in this paper are analysed in the computer simulation environment. They are compared to the SRP-PHAT

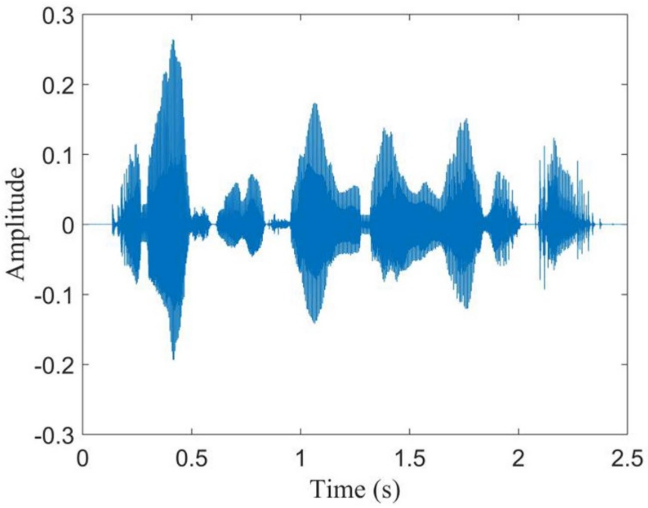


Figure 3. Sound signal.

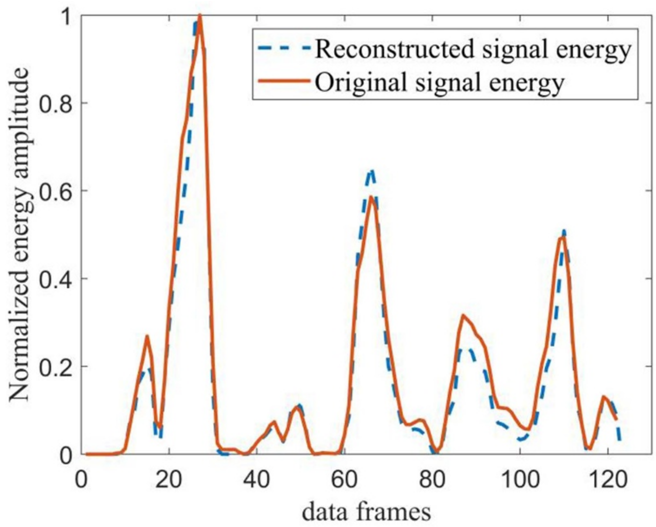


Figure 4. Signal energy correlation.

algorithm, the Music algorithm and the EbOMP algorithm in this paper. This paper focuses on the improvement of the iteration times of the OMP algorithm. The performance of the OMP algorithm under different sound source spacings and different frequencies has been fully studied by predecessors and will not be discussed in this paper. Therefore, the simulation and experiment focus on the effectiveness of the EbOMP algorithm [18, 27].

3.1. Simulation conditions

The array comprises a circular planar array of six microphones with a radius of 0.1 m, and the spatial azimuth angles and elevation angles in front of it are divided according to the spacing of 10°. The simulation environment is a free field, and the influence of noise on the signal is mainly considered. The voice signals were randomly selected from the TIMIT database, with a sampling rate of 16 kHz and a processing frame length of 512 points. The overlap rate between the two frames is 50%, each frame signal is added with a hamming window, and the discrete Fourier transform is taken. Since the energy of the speech signal is mainly distributed in the middle and low frequencies, 94 frequency points at equal intervals between 100 and 3 kHz are selected to construct the measurement matrix after the discrete Fourier transform.

3.2. Verification of reconstructed signal energy

In order to verify the accuracy of the reconstruction algorithm on the signal energy reconstruction, a speech signal was used for simulation verification, as shown in figure 3.

The simulation conditions in section 3.1 were used to calculate the energy of the original signal and the reconstructed signal. In order to see the energy details more clearly, the length of each frame was shortened to 512 points. The calculation results are shown in figure 4.

It can be seen from figure 4 that the reconstructed signal energy is approximately equal to the original signal energy. As the reconstruction algorithm does not use all frequency points of the signal to calculate, the reconstructed signal energy of some frames is less than the original signal.

3.3. Selection of energy threshold

The algorithm introduces an important parameter: energy threshold. According to the principle of the improved algorithm, this threshold is actually a representation of the relationship between the energy of the sound source signal and the energy of the noise signal, so it is necessary to choose a value that is applicable to most scenarios in daily life. A parameter also representing the relationship between signal energy and noise energy is the frequently-used signal-to-noise ratio. The signal-to-noise ratio can be calculated according to the signal energy ratio by the following formula:

$$\mathrm{SNR}\,(\mathrm{dB}) = 10 \times \log_{10} \frac{\sum |x(k)|^2}{\sum |v(k)|^2} \tag{11}$$

where $x(k)$ is the signal sequence and $v(k)$ is the noise sequence.

According to the international background noise measurement standard [28], the lowest adaptive environment of the algorithm is set as SNR = 5 dB, and then the energy threshold is selected as approximately 0.2. After the threshold is selected, it is verified whether the number of iterations of the EbOMP algorithm meets the expectation when the number of sound sources and the signal-to-noise ratio are different, as shown in figure 5.

It can be seen from the figure 5 that the number of iterations of the algorithm is approximately equal to the number of sound sources under different signal-to-noise ratios. However, it is also noted that when the number of sound sources

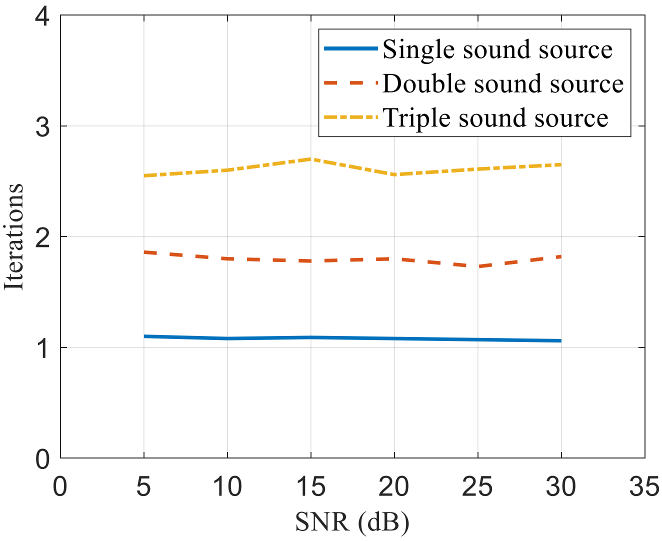


Figure 5. Number of iterations for different numbers of sound sources.

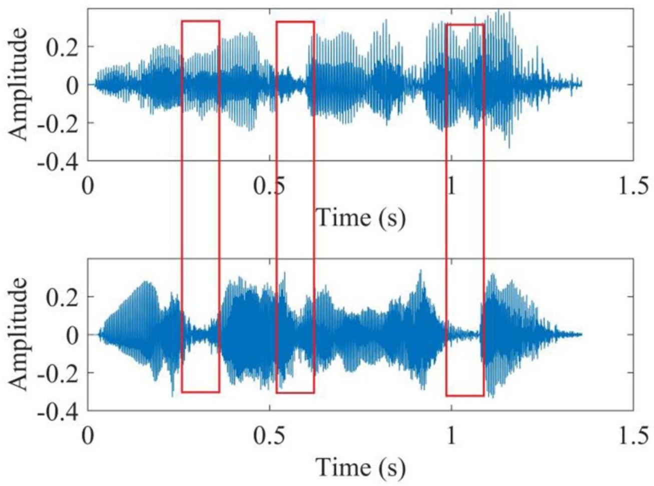


Figure 6. Audio signal contrast.

is greater than or equal to 2, regardless of the signal-to-noise ratio, the number of iterations obtained is slightly lower than the set number of sound sources. This is because the algorithm calculates the average number of iterations of all frame positioning results of audio, but there may be a gap with an amplitude of 0 or too small between the pronunciation of audio, audio and text, so that all sound sources cannot be detected during this period, resulting in a lower final result. As shown in the figure 6, the red box indicates where one is in the audio energy intensive period and the other is in the gap.

3.4. Comparison of location of different algorithms

Assume that there are two sound sources, its azimuth angle and elevation angle, respectively ($-60°$, $60°$), ($50°$, $30°$). Set the signal-to-noise ratio to 10 dB, and the three algorithms are simulated. Simulation environment details: the added noise is white noise, and the reverberation is not considered in the simulation, assuming a free field scenario.

The simulation results are as follows:

It can be seen from figures 7–9 that the EbOMP algorithm can accurately locate the positions of the two sound sources, and it has a narrow major lobe. The SRP-PHAT algorithm also locates two sound source locations, but the major lobe is wider and the calculation time of this algorithm is longer. Among the three algorithms, the EbOMP algorithm is not only narrow in the main lobe, but also does not need to know the number of sound sources in advance.

In order to further verify the performance of the EbOMP algorithm, the positioning accuracy of the three algorithms under different SNR conditions was counted. At the same time, the accuracy of the EbOMP algorithm with different numbers of sound sources is simulated. This experiment does not consider the positioning error, it only records whether the positioning result is correct. When the algorithm processes a frame signal and locates all the sound source positions, it is called accurate positioning. Therefore, positioning accuracy is defined as the ratio of locating accurate frames to all frames. The voice was selected from the TIMIT database, and a total of 1000 frames were tested. The length of a frame of signal is 512 sampling points, and the sampling frequency is 16kHz. White noise was added to clean speech, different SNR environments were simulated, and no reverberation was assumed.

As can be seen from figures 10 and 11, in the location of the two sound sources, the accuracy of the EbOMP algorithm is about 20% higher than that of the other two algorithms under different SNR simulations. It can be concluded that the algorithm can accomplish accurate positioning with less information.

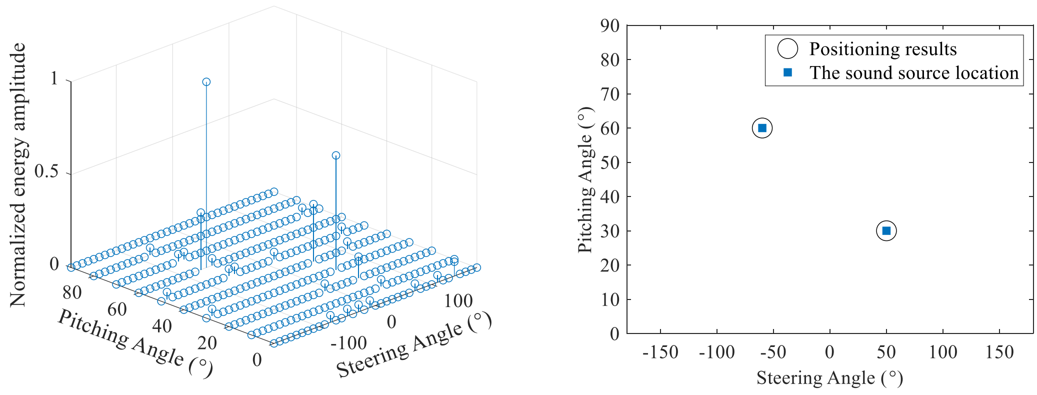
Through simulation and comparative analysis, it is found that the EbOMP algorithm proposed in this paper has a better main lobe performance and location success rate than the traditional algorithm. Moreover, compared with the OMP algorithm, it does not need to predict the number of sound sources and has good application performance.

4. Experimental results and analysis

In order to verify the performance of the algorithm proposed in this paper, tests were carried out in two scenarios; an empty square and an office. A seven-microphone circular microphone array with a radius of 0.1 m is adopted. The instruments used are shown in figure 12, and seven microphones are selected, as shown in figure 13.

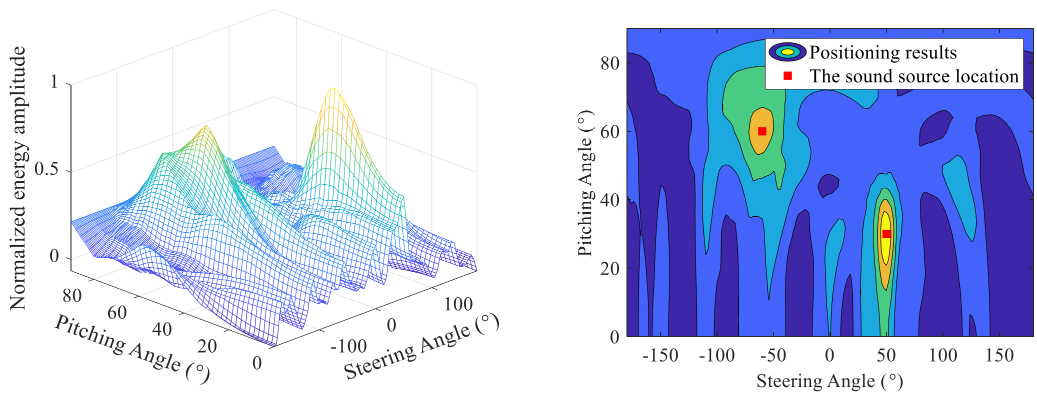
The most common scenario in life is a mono-source or dual-source scene, so the dual-source scene is selected to verify the EbOMP algorithm in practice. Two speakers are placed in the space, the distance between the sound source plane and the microphone array plane is set at 1 m and 2 m, respectively, as shown in figure 14.

The two speakers broadcast different voice signals with similar energy, and the positioning results are displayed in the picture taken.



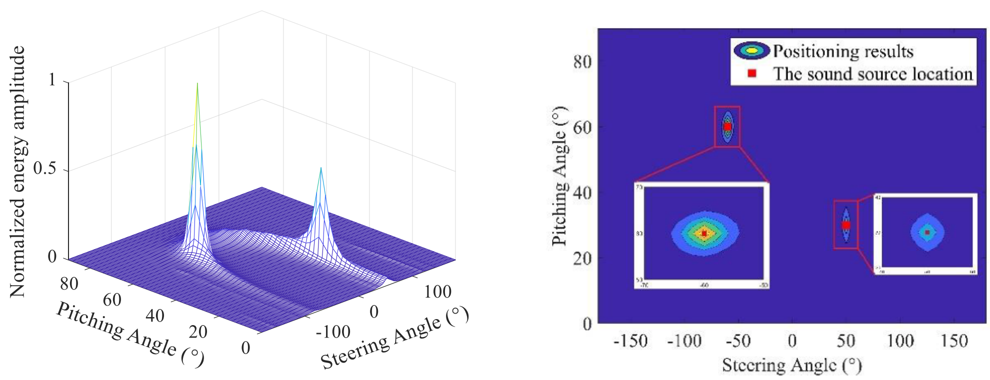
(a) Three-dimensional perspective (b) Vertical view

Figure 7. EbOMP algorithm localization results.



(a) Three-dimensional perspective (b) Vertical view

Figure 8. SRP-PHAT algorithm localization results.



(a) Three-dimensional perspective (b) Vertical view

Figure 9. MUSIC algorithm localization results.

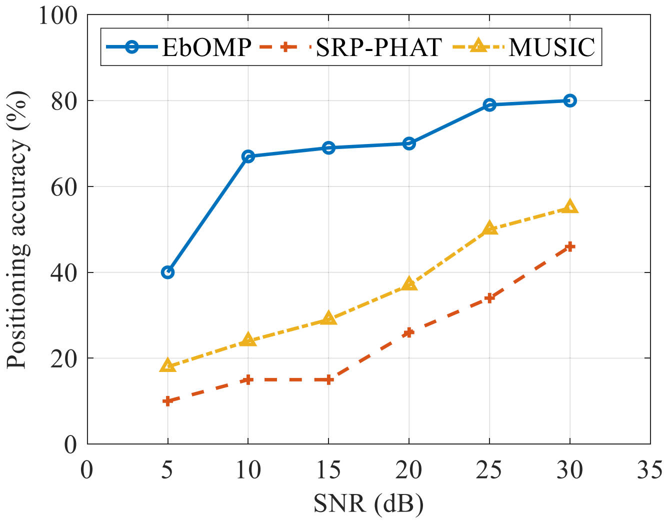


Figure 10. Positioning accuracy of the three algorithms in different SNRs.

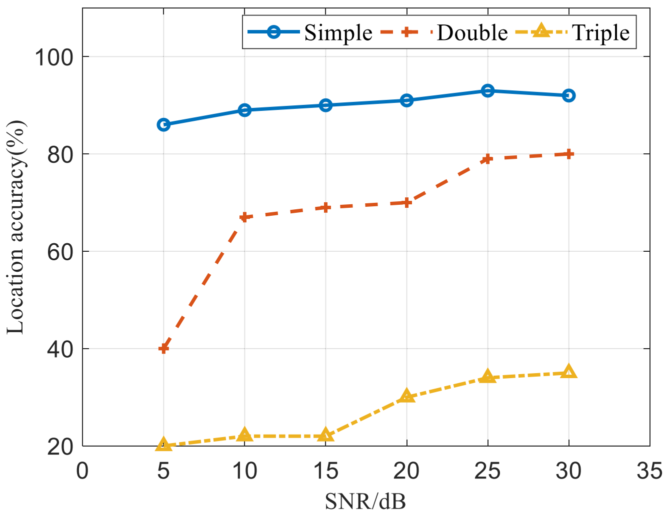


Figure 11. Location accuracy of different numbers of sound sources.

Figures 15 and 16 are the positioning results at different distances indoors and outdoors. It can be seen that the EbOMP algorithm proposed in this paper successfully locates the positions of two sound sources in different environments. It should be pointed out here that when modelling the received signal of the algorithm, the reverberation factor is not considered. However, in an indoor environment with reverberation, the improved algorithm can still locate the sound source, indicating that the algorithm has a certain resistance to reverberation. The ability to respond can be explained by the principle of the algorithm.

In each test, the average number of iterations of the EbOMP algorithm was recorded, and the results were as follows:

As can be seen from the table 1, the absolute value of iteration number deviation is 0.31 at the maximum and 0.07 at the minimum. In the simulation analysis in section 3.3, the number of iterations of the algorithm is generally less than the set number of sound sources. The actual statistical results show that the number of iterations is larger than that of the set sound source, because the actual sound source is not a very ideal point source in simulation. The sound box used in this experiment has a larger sounding 'surface' rather than a 'point', so when the sound pressure value of the sound source is large, the energy of the sound source signal will leak to the adjacent direction of the division, resulting in the increase of the number of iterations of some frames.

Figure 12. Microphone array system.

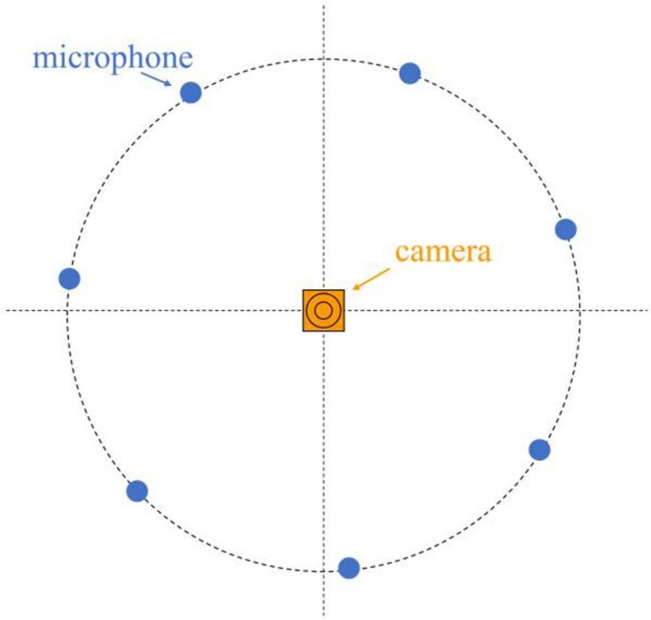


Figure 13. Schematic diagram of microphone array device.

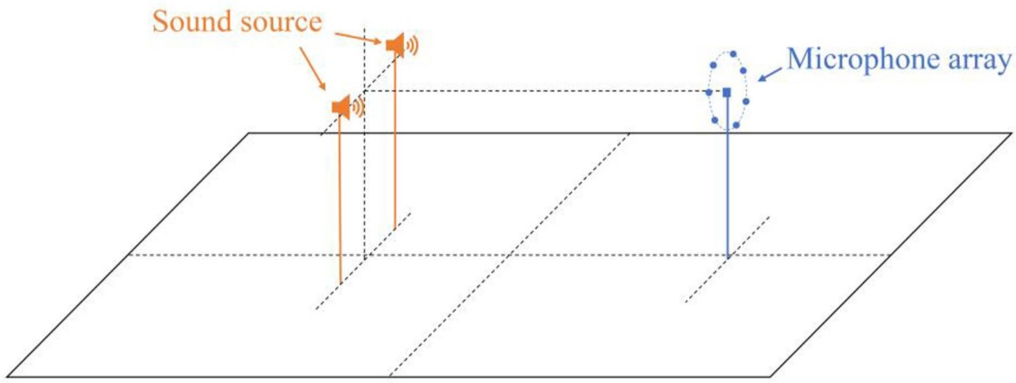


Figure 14. Schematic diagram of positioning experiment.

(a) 2m (b) 1m

Figure 15. Outdoor positioning results.

(a) 2m (b) 1m

Figure 16. Indoor positioning results.

Table 1. Iteration count of EbOMP algorithm in different scenarios.

Test environment	Outdoors		Indoors	
Distance	2 m	1 m	2 m	1 m
Number of iterations	1.93	2.17	1.69	2.24
Deviation	−0.07	0.17	−0.31	0.24

5. Conclusion

The main contribution of this paper is to propose an orthogonally-matched chase source localization algorithm based on reconstructed signal energy. The OMP algorithm needs to know the number of sound sources or bring unnecessary calculation amount in advance for the iteration stopping condition. After each iteration, the energy of reconstructed signals at different positions is compared to judge whether new sound sources are added. Simulation results show that the EbOMP algorithm not only has good directivity, but also does not need to know the number of sound sources in advance. In the experiment, the algorithm shows a certain ability to resist reverberation. Moreover, in different environments, the maximum absolute error between the number of iterations and the actual number of sound sources is 0.31, indicating that iteration can be stopped within a good range of iterations. Finally, it should be pointed out that due to the limitation of the algorithm principle, the algorithm can not distinguish multiple sound sources with large energy differences, which should be considered in future research.

Data availability statement

The datasets and codes for the simulations in this paper are available from the corresponding author if requested.

Conflict of interest

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

Funding statement

1. Innovation Project of Guet Graduate Education [2020YCXS009].
2. Liuzhou Science and Technology Project [2020PAA0604].
3. National Natural Science Foundation of China [52065013].
4. Natural Science Foundation of Guangxi, China [2018GXNSFAA281276].

ORCID iD

Yingxiang Zuo https://orcid.org/0000-0002-8579-5935

References

[1] Almusaylim Z A and Zaman N 2019 A review on smart home present state and challenges: linked to context-awareness internet of things (IoT) *Wirel. Netw.* **25** 3193–204
[2] Levy T and Ando S 2012 Applying direct algebraic sound source localization method for time-domain reflectometry of conference room *J. Acoust. Soc. Am.* **132** 1913
[3] Rascon C and Meza I 2017 Localization of sound sources in robotics: a review *Robot. Auton. Syst.* **96** 184–210
[4] Sangmoon L *et al* 2015 Three-dimensional sound source localization using inter-channel time difference trajectory *Int. J. Adv. Robot. Syst.* **12** 171
[5] Gala D, Lindsay N and Sun L 2019 Realtime active sound source localization for unmanned ground robots using a self-rotational Bi-microphone array *J. Intell. Robot. Syst.* **95** 935–54
[6] Gala D, Lindsay N and Sun L 2021 Multi-sound-source localization using machine learning for small autonomous unmanned vehicles with a self-rotating Bi-microphone array *J. Intell. Robot. Syst.* **103** 1–20
[7] Chiariotti P, Martarelli M and Castellini P 2019 Acoustic beamforming for noise source localization reviews, methodology and applications *Mech. Syst. Signal Process.* **120** 422–48
[8] Alameda-pineda X and Horaud R 2014 A geometric approach to sound source localization from time-delay estimates *IEEE Trans. Audio Speech Lang. Process.* **22** 1082–95
[9] Park C S, Jeon J H and Kim Y H 2014 Localization of a sound source in a noisy environment by hyperbolic curves in quefrency domain *J. Sound Vib.* **333** 5630–40
[10] Huang Q H, Zhang L and Fang Y 2018 Two-step spherical harmonics ESPRIT-type algorithms and performance analysis *IEEE Trans. Audio Speech Lang. Process.* **26** 1684–97
[11] Chu N, Picheral J, Mohammad-djafari A and Gac N 2014 A robust super-resolution approach with sparsity constraint in acoustic imaging *Appl. Acoust.* **76** 197–208
[12] Yong Z, Li-Yi Z, Jian-Feng H, Zhe B and Yi Y 2019 A new DOA estimation algorithm based on compressed sensing *Cluster Comput.* **22** 895–903
[13] Qian W J, He Q B, Ni Y, Peng Z-K, Gao R, Ren D-P and Qi Z-M 2019 Design of a three degrees-of-freedom biomimetic microphone array based on a coupled circuit *Meas. Sci. Technol.* **30** 165051
[14] Yang X, Xing H Y and Yan Y 2021 A data complementary method for sound source localization based on four-element microphone array groups *Meas. Sci. Technol.* **32** 095015
[15] Li L, Wang L F, Dong J, Fu J, Gao Y, Qiao Y and Zhu M 2021 SuperSoundcompass: a high-accuracy acoustic localization sensor using a small-aperture microphone array *Meas. Sci. Technol.* **32** 105106
[16] Donoho D L 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306
[17] Ke W, Zhang X, Yuan Y and Shao J 2017 Compressing sensing based source localization for controlled acoustic signals using distributed microphone arrays *Math. Probl. Eng.* **2017** 1–11
[18] Xenaki A, Gerstoft P and Mosegaard K 2014 Compressive beamforming *J. Acoust. Soc. Am.* **136** 260–71
[19] Xenaki A and Gerstoft P 2015 Grid-free compressive beamforming *J. Acoust. Soc. Am.* **137** 1923–35
[20] Yang Y, Chu Z, Ping G and Xu Z 2018 Resolution enhancement of two-dimensional grid-free compressive beamforming *J. Acoust. Soc. Am.* **143** 3860–72
[21] Ning F, Pan F, Zhang C, Liu Y, Li X and Wei J 2018 A highly efficient compressed sensing algorithm for acoustic imaging in low signal-to-noise ratio environments *Mech. Syst. Signal Process.* **112** 113–28
[22] Do T T, Gan L, Nguyen N and Tran T D 2008 Sparsity adaptive matching pursuit algorithm for practical compressed sensing *2008 42nd Asilomar Conf. on Signals, Systems and Computers* pp 581–7
[23] Deng F, Guan S, Yue X, Gu X, Chen J, Lv J and Li J 2017 Energy-based sound source localization with low power consumption in wireless sensor networks *IEEE Trans. Ind. Electron.* **64** 4894–902
[24] Dou Y, Wang H and Ma S 2016 An algorithm of sound source localization using range differences of arrival and energy ratios of arrival *2016 IEEE 11th Conf. on Industrial Electronics and Applications (ICIEA)* pp 1547–50
[25] Yin H, Li S and Fang L 2013 Block-sparse compressed sensing: non-convex model and iterative re-weighted algorithm *Inverse Probl. Sci. Eng.* **21** 141–54
[26] ISO 1996-2-2007 2007 Acoustics—description, measurement and assessment of environmental noise—part 2: determination of environmental noise levels 2nd edn (15 March 2007)
[27] Cotter S F, Rao B D, Engan K and Kreutz-Delgado K 2005 Sparse solutions to linear inverse problems with multiple measurement vectors *IEEE Trans. Signal Process.* **53** 2477–88
[28] Ning F L, Wei J G, Qiu L, Shi H and Li X 2016 Three-dimensional acoustic imaging with planar microphone arrays and compressive sensing *J. Sound Vib.* **380** 112–28