

Research Article

Compressing Sensing Based Source Localization for Controlled Acoustic Signals Using Distributed Microphone Arrays

Wei Ke,^{1,2,3} Xiunan Zhang,¹ Yanan Yuan,¹ and Jianhua Shao¹

¹School of Physics and Technology, Nanjing Normal University, Nanjing 210097, China

²Jiangsu Key Laboratory of Meteorological Observation and Information Processing,
Nanjing University of Information Science & Technology, Nanjing 210044, China

³Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application,
Nanjing 210023, China

Correspondence should be addressed to Wei Ke; wkykw@sina.com and Jianhua Shao; shaojianhua@njnu.edu.cn

Received 19 October 2016; Revised 9 January 2017; Accepted 19 February 2017; Published 8 August 2017

Academic Editor: Laurent Bako

Copyright © 2017 Wei Ke et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to enhance the accuracy of sound source localization in noisy and reverberant environments, this paper proposes an adaptive sound source localization method based on distributed microphone arrays. Since sound sources lie at a few points in the discrete spatial domain, our method can exploit this inherent sparsity to convert the localization problem into a sparse recovery problem based on the compressive sensing (CS) theory. In this method, a two-step discrete cosine transform- (DCT-) based feature extraction approach is utilized to cover both short-time and long-time properties of acoustic signals and reduce the dimensions of the sparse model. In addition, an online dictionary learning (DL) method is used to adjust the dictionary for matching the changes of audio signals, and then the sparse solution could better represent location estimations. Moreover, we propose an improved block-sparse reconstruction algorithm using approximate l_0 norm minimization to enhance reconstruction performance for sparse signals in low signal-noise ratio (SNR) conditions. The effectiveness of the proposed scheme is demonstrated by simulation results and experimental results where substantial improvement for localization performance can be obtained in the noisy and reverberant conditions.

1. Introduction

In the past decade, the microphone array based sound source localization technique has been developed rapidly and has been widely applied to video conference systems [1, 2], robot audition [3, 4], speech enhancement [5], and speech recognition [6]. Since distributed microphone arrays have larger array aperture than traditional microphone arrays [7, 8], it can provide better performance of sound source localization, and therefore it becomes a hot research area.

Existing sound source localization methods can be generally divided into three types: the first one is the beamforming method [9]; the second one is the high-resolution spectrum estimation method [10]; the third one is the time delay estimation method [11–14]. However, differently from positioning technology based on sonar and radar arrays, speech signals are wideband signals with property of short-term stationarity. Moreover, sound source localization is apt

to be affected by the reverberant and noisy environments [15, 16]. Hence, there is still a quite big room to improve the performance of sound source localization.

In recent years, with the rapid development of compressive sensing (CS) theory [17, 18], it has brought a revolutionary influence to the traditional sound source localization method. Since sound sources can be regarded as point sources and the number of the sound sources is quite limited in the positioning space, the sound source localization problem essentially implicates the spatial sparsity. According to this natural sparsity, Cevher and Baraniuk modeled the sound source localization problem as a sparse approximation problem [19] based on the sound propagation model and provided better positioning performance. In [20], the CS theory was utilized to estimate time difference of arrival (TDOA) and obtained higher estimation accuracy than the generalized cross-correlation (GCC) method. Unlike the above methods, [21] proposed a source localization algorithm

based on the two-level Fast Fourier Transform- (FFT-) based feature extraction method and spatial sparsity. The proposed feature extraction method leads to a sparse representation of audio signals. Further, Simard and Antoni exploited Green functions to establish the sparse localization model, which resulted in better sound source identification and localization performance than traditional beamforming methods [22]. In addition, [23] extended the sparse constraint to sound source imaging, which can not only locate sound source targets but also realize single target imaging. These above researches show that the CS theory has broad application prospects in sound source localization.

However, just like the traditional sound source localization methods, the CS-based sound source localization approaches are still confronted with the challenges from complex acoustic environments. The challenges mainly come from ambient noise and reverberation. On one hand, the sparse reconstruction performance is seriously affected by noises. At present, it is common to use l_1 norm instead of l_0 norm to enforce sparse constraint, since l_0 norm optimization is NP-hard to solve. However, under the l_1 norm constraint, the objective function will give large coefficients more constraints to ensure the convergence of the cost function. In fact, in the positioning applications large coefficients generally correspond to the targets, while small coefficients may correspond to the noises. Thus, the ordinary reconstruction algorithms such as the greedy algorithm [24] and the convex-optimization algorithm [25] may weaken the contribution of large coefficients in the reconstruction process, while small coefficients may not perform any constraint. Therefore, the reconstruction accuracy may be decreased significantly, and even in low SNR conditions, noise may be treated as targets wrongly. On the other hand, in indoor environments reverberation is ubiquitous. Moreover, the reflection and scattering effects of sound waves between walls or ceilings are hard to predict. When these happen, there will be some errors between the sparse model and real measurement, which also lead to localization performance degradation.

To improve the performance of sound source localization based on distributed microphone arrays in noisy and reverberant environments, a sound source localization method is proposed in this paper. This method exploits the inherent spatial sparsity to convert the localization problem into a sparse recovery problem based on the CS theory. Meanwhile, inspired by [21], this paper also uses sparse feature in transform domain to construct CS-based localization model. Since the DCT has the character of strong energy concentration and can keep in the real number area, we propose a two-step DCT-based feature extraction method to cover both short-time and long-time properties of the signal and reduce the dimension of the sparse model. In addition, we propose an improved block-sparse reconstruction algorithm based on approximate l_0 norm minimization to enhance reconstruction performance under low SNR conditions. The novel feature of this method is to use approximate l_0 norm to promote interblock sparsity and the optimization problem is solved by a sequential procedure in conjunction with a conjugate-gradient method for fast reconstruction. Moreover, a dictionary learning (DL) method is used to adjust

the dictionary for matching the changes of audio signals, and then the sparse solution could better represent location estimations.

The remainder of the paper is organized as follows. Section 2 describes the system model. In Section 3, we propose a novel two-step DCT-based feature extraction approach. The details of the proposed localization method are addressed in Section 4. Simulation and experimental results are given in Sections 5 and 6, respectively. Finally, Section 7 concludes the paper.

2. System Model

Let us consider a spatial distribution of M sensors, each accommodating $L \geq 2$ microphones. For the sake of notational simplicity and with no loss of generality, let us assume that each sensor has the same number of microphones. In this paper, we assume that all microphones begin to work at the same time and then keep working until each experiment ends. Moreover, the propagation medium of sound signals is isotropic. Let $\mathbf{p}_l^{(m)} = (x_l^{(m)}, y_l^{(m)})^T$, $l = 1, \dots, L$; $m = 1, \dots, M$, be the coordinates of each microphone which are generally known in the localization system, while the acoustic sources are located at $\mathbf{p}_k = (x_k, y_k)^T$, $k = 1, \dots, K$, which are unknown. Here, $(\cdot)^T$ means transposition. The whole localization area is uniformly divided into N grid points (generally $N \gg K$). Each microphone receives signals from sound sources and transmits them to the positioning center, where the localization algorithm is performed.

Since sound sources locate at one or a few grid points in the localization area, the positions of sound sources in the discrete spatial domain can be accurately represented as a sparse vector $\boldsymbol{\theta}$. The elements of vector $\boldsymbol{\theta}$ are equal to ones if sound sources locate at the corresponding grids in the localization area, while other elements of vector $\boldsymbol{\theta}$ are equal to zeros. In such case, the localization problem is converted to determine nonzero elements and their specific locations in the sparse vector $\boldsymbol{\theta}$ according to the received signals. Based on the CS theory, the sparse localization model can be represented as

$$\mathbf{y} = \Phi \boldsymbol{\theta}, \quad (1)$$

where $\mathbf{y} = [\tilde{\mathbf{s}}_1^T \ \tilde{\mathbf{s}}_2^T \ \dots \ \tilde{\mathbf{s}}_L^T \ \dots \ \tilde{\mathbf{s}}_{LM}^T]^T$ represents the entire feature vector, $\tilde{\mathbf{s}}_i$ ($i = 1, \dots, LM$) corresponds to the feature vector of received signals from i th microphone, and Φ is the dictionary matrix. The feature extraction method will be introduced in the next section. Since the dictionary Φ is another key factor for the sparse reconstruction, this paper firstly sets up an initial dictionary based on the sound signal propagation model in [19] and then exploits the DL technique to modify the initial dictionary for overcoming the model errors. According to [19], the l th received signal coming from the sound source located at grid n by the l th microphone of the m th sensor can be expressed as

$$\mathbf{r}_{ln}^{(m)} = \frac{1}{d_{ln}^{(m)}} \mathbf{q} \left(t - \frac{d_{ln}^{(m)}}{c} \right) + \mathbf{v}_{ln}^{(m)}, \quad (2)$$

$$l = 1, \dots, L; \quad m = 1, \dots, M, \quad n = 1, \dots, N,$$

where c is the speed of sound, $d_{ln}^{(m)}$ is the distance between the n th grid point and the l th microphone of the m th sensor, $\mathbf{q}(t)$ represents sound signals, and $\mathbf{v}_{ln}^{(m)}$ represents noises. Since the positions of grid points and microphones are already known, the distance $d_{ln}^{(m)}$ can be calculated in advance. Hence, under the noise-free conditions Φ can be calculated as

$$\Phi = \begin{bmatrix} \boldsymbol{\varphi}_1^{(1)} & \boldsymbol{\varphi}_2^{(1)} & \cdots & \boldsymbol{\varphi}_N^{(1)} \\ \boldsymbol{\varphi}_1^{(2)} & \boldsymbol{\varphi}_2^{(2)} & \cdots & \boldsymbol{\varphi}_N^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\varphi}_1^{(M)} & \boldsymbol{\varphi}_2^{(M)} & \cdots & \boldsymbol{\varphi}_N^{(M)} \end{bmatrix}, \quad (3)$$

where

$$\boldsymbol{\varphi}_n^{(m)} = \begin{bmatrix} T[\mathbf{r}_{1n}^{(m)}] \\ T[\mathbf{r}_{2n}^{(m)}] \\ \vdots \\ T[\mathbf{r}_{Ln}^{(m)}] \end{bmatrix} \quad (4)$$

represents the feature vector of the signal from the sound source located at the n th grid and received by the l th microphone of the m th sensor, in which $T[\cdot]$ represents the feature extraction operation. Since each sensor has L microphones, it should be noted that $\boldsymbol{\theta}$ is a block-sparse vector whose nonzero elements occur in clusters. Unlike the conventional CS framework, the sparsity patterns of neighboring coefficients are related to each other in the block-sparse model, and this block-sparsity has the potential to encourage structured-sparse solutions.

Assuming $K = 2$ sources as an example, we substitute the concrete forms of vectors \mathbf{y} and $\boldsymbol{\theta}$ and matrix Φ into (1), and then we can obtain

$$\begin{bmatrix} \tilde{\mathbf{s}}_1^T \\ \tilde{\mathbf{s}}_2^T \\ \vdots \\ \tilde{\mathbf{s}}_L^T \\ \vdots \\ \tilde{\mathbf{s}}_{LM}^T \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{\varphi}_1^{(1)} & \boldsymbol{\varphi}_2^{(1)} & \cdots & \boldsymbol{\varphi}_i^{(1)} & \cdots & \boldsymbol{\varphi}_j^{(1)} & \boldsymbol{\varphi}_N^{(1)} \\ \boldsymbol{\varphi}_1^{(2)} & \boldsymbol{\varphi}_2^{(2)} & \cdots & \boldsymbol{\varphi}_i^{(2)} & \cdots & \boldsymbol{\varphi}_j^{(2)} & \boldsymbol{\varphi}_N^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\varphi}_1^{(m)} & \boldsymbol{\varphi}_2^{(m)} & \cdots & \boldsymbol{\varphi}_i^{(m)} & \cdots & \boldsymbol{\varphi}_j^{(m)} & \boldsymbol{\varphi}_N^{(m)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\varphi}_1^{(M)} & \boldsymbol{\varphi}_2^{(M)} & \cdots & \boldsymbol{\varphi}_i^{(M)} & \cdots & \boldsymbol{\varphi}_j^{(M)} & \boldsymbol{\varphi}_N^{(M)} \end{bmatrix}}_{\Phi} \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{1} \\ \mathbf{0} \\ \vdots \\ \mathbf{1} \\ \mathbf{0} \end{bmatrix} = \boldsymbol{\theta} \quad (5)$$

$\begin{matrix} \text{ith column} & \text{jth column} \\ \text{ith block} & \text{jth block} \end{matrix}$

where $\mathbf{0}$ and $\mathbf{1}$ represent the $L \times 1$ all-zero and all-one column vectors, respectively. If there are no noises and reverberation, the specific locations of nonzero blocks in $\boldsymbol{\theta}$ point out the accurate positions of sound sources according to the corresponding grids. However, the reflection and scattering of sound signals and noises in indoor environments are unavoidable. Therefore, there are errors between the above sparse localization model and real measurements. This may ultimately lead to the degradation of localization performance. In order to show differences between the predefined

dictionary and the actual cases, the sparse model (1) is then modified as

$$\mathbf{y} = (\Phi + \Gamma)\boldsymbol{\theta} + \mathbf{v} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v}, \quad (6)$$

where \mathbf{v} denotes noises, $\mathbf{H} = (\Phi + \Gamma)$ denotes the actual dictionary, and Γ is the model error due to both reverberation and noise, which is time-varying and unknown in advance. To obtain the accurate position estimates of sound sources in real environments, a DL method is used to dynamically adjust the dictionary for matching the changes of actual signals. Meanwhile, an improved block-sparse reconstruction algorithm using approximate l_0 norm minimization is proposed to enhance reconstruction performance under the noisy conditions.

3. Feature Extraction

In the applications of sound source localization, many localization algorithms use long speech signals to calculate source locations [9, 21]. However, considering high dimensionality of the received signals, the computational complexity is high. Moreover, lots of interferences and noises are embedded in the long acoustic signals, which may affect the final localization performance. Hence, instead of using acoustic signals in time domain, in this paper a two-step DCT-based feature extraction method is proposed to overcome the above problems.

Considering the properties of audio signals, the long signals received by microphones are partitioned into successive frames using the small-size window function (Hamming window is used here). In the first step, the received signal vector \mathbf{r} is partitioned into J short frames; that is,

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_P \end{bmatrix}_{P \times 1} \xrightarrow{\text{Hamming Windowing}} \mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_J \end{bmatrix}_{J \times Q}, \quad (7)$$

where r_i ($i = 1, \dots, P$) is the i th element of the received signal vector \mathbf{r} and P is the length of entire input signal. \mathbf{Z} is a $J \times Q$ matrix, whose rows are the windowed frames of input signal. Q is the frame (window) length and J is the number of the frames.

Next, the type-II DCT is applied to the speech signal frames.

$$\mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_J \end{bmatrix}_{J \times Q} \xrightarrow{\text{DCT}} \tilde{\mathbf{Z}} = \begin{bmatrix} D(z_1) \\ D(z_2) \\ \vdots \\ D(z_J) \end{bmatrix}_{J \times Q} = \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \vdots \\ \tilde{z}_J \end{bmatrix}_{J \times Q}, \quad (8)$$

where $D(\cdot)$ represents the discrete cosine transform and \tilde{z}_j ($j = 1, \dots, J$) is the output of DCT. Then, the amplitudes of

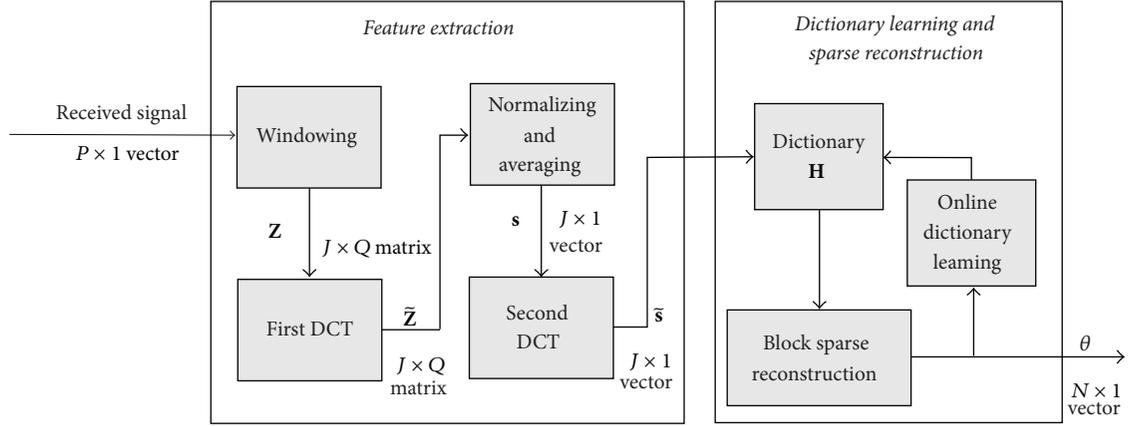


FIGURE 1: Block diagram of proposed algorithm.

the DCT coefficients in each frame are normalized by dividing them by their maximum value, and then all normalized frames are averaged; that is,

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \vdots \\ \tilde{z}_J \end{bmatrix}_{J \times Q} \xrightarrow{\text{Normalizing and Averaging}} \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_J \end{bmatrix}_{J \times 1}. \quad (9)$$

Although the acoustic signals are generally nonstationary over a relatively long-time interval, in this constructed short-time feature space the statistical properties of the resulting data become almost constant, so that the features can be considered as approximately stationary process. On the contrary, noise does not have the same short-time properties as acoustic signals and is usually assumed as obeying the certain distribution with zero mean, so noise may be mitigated after the averaging procedure. In the next step, the long-time properties of audio signals are considered. Apply the DCT to the vector \mathbf{s} again:

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_J \end{bmatrix}_{J \times 1} \xrightarrow{\text{DCT}} \tilde{\mathbf{s}} = D \left(\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_J \end{bmatrix} \right) = \begin{bmatrix} \tilde{s}_1 \\ \tilde{s}_2 \\ \vdots \\ \tilde{s}_J \end{bmatrix}_{J \times 1}. \quad (10)$$

It should be noted that $\tilde{\mathbf{s}}$ is still a $J \times 1$ vector, so the two-step DCT-based feature extraction method can reduce the dimension from the length of samples P to the length of frames J and thus decrease the computational load of sparse reconstruction. Meanwhile, since the DCT has the characteristic of strong energy concentration, big coefficients in the DCT domain will be relatively concentrated in certain locations of the vector $\tilde{\mathbf{s}}$, and other coefficients of the vector $\tilde{\mathbf{s}}$ are very small. In other words, the vector $\tilde{\mathbf{s}}$ is approximately sparse in the DCT domain, which is good for the following sparse reconstruction. Now, we can know that the transformation $T[\cdot]$ in (4) represents the operation from (7) to (10).

The proposed feature extraction process is summarized in Figure 1, and the localization algorithm will be proposed in the next section.

Remark 1. Although diverse variety of features have been used for the representation of acoustic signals, the important point for CS-based sound source localization is that the selected feature can accurately relate the spatial information of the sound source. Our feature extraction method originates from the sound signal propagation model in [19], which obviously includes the position information of sound sources. When these features are transformed to the transform domain, the spatial information of sound sources is not lost in the transform process, which is really important for sound source localization. In other words, the two-step DCT-based feature extraction method utilizes both short-time and long-time properties of acoustic signals to improve localization performance. The short-time properties are used to reduce noise based on the short-time stationary character of acoustic signals, while the long-time properties of acoustic signals are exploited to hold the spatial information of sound sources.

4. Localization Algorithm

Since the initial dictionary Φ is based on the signal propagation model, it cannot dynamically match the actual signals. Therefore, using initial dictionary Φ directly for sparse reconstruction will result in large localization errors. To overcome this problem, the DL method is used to dynamically adjust the dictionary for matching the changes of acoustic signals, and then the sparse solution could better represent location estimations.

According to the CS theory and DL technology [26], the above problem in (5) can be converted into the following problem:

$$\begin{aligned} \underset{\mathbf{H}, \Theta}{\operatorname{argmin}} \quad & \|\mathbf{Y} - \mathbf{H}\Theta\|_F^2 \\ \text{s.t.} \quad & \|\theta_i\|_0 \leq K, \\ & i = 1, \dots, S, \end{aligned} \quad (11)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_S]$ is the training dataset and $\Theta = [\theta_1, \dots, \theta_S]$ is the sparse matrix. $\|\cdot\|_F$ and $\|\cdot\|_0$ are the Frobenius norm and l_0 norm, respectively. K represents the sparsity, and S represents the number of samples. Since this problem is nonconvex, most DL methods generally use the alternating minimization method. In one step, a sparse recovery algorithm finds sparse representations of the training samples with a fixed dictionary. In the other step, the dictionary is updated to decrease the average approximation error while the sparse coefficients remain fixed. The alternating learning method is also used in this paper.

4.1. Improved Block-Sparse Reconstruction Algorithm Based on Approximate l_0 Norm Minimization. In this section, we consider how to obtain the unknown sparse solution whose nonzero elements occur in clusters, that is, block-sparse signal. While there are many approaches to perform the recovery, two famous recovery approaches are the block orthogonal matching pursuit (BOMP) method [27] and the mixed l_2/l_1 norm algorithm [28]. The BOMP method is proposed as an extension of the orthogonal matching pursuit (OMP) algorithm for block-sparse signals. Hence, BOMP belongs to the greedy-based algorithm and can run fast, but it does not always provide good estimation under the noisy condition. Unlike the greedy-based algorithm, the mixed l_2/l_1 norm algorithm modifies the l_1 norm cost function to exploit the convex-optimization method for enforcing block-sparsity constraint. Although the mixed l_2/l_1 norm method can achieve better recovery than BOMP in the presence of noise, it is very slow because of using second-order cone-programming (SOCP) to find the sparsest solution. Moreover, it should be emphasized that larger coefficients in a sparse vector are penalized more heavily than smaller coefficients in the l_1 norm penalty, unlike the more democratic penalization of the l_0 norm. The imbalance of the l_1 norm penalty will seriously influence the recovery accuracy, which may result in many false targets in the low SNR condition. To overcome the mismatch between l_0 norm minimization and l_1 norm minimization, in this section we propose an improved algorithm to use the approximate l_0 norm minimization as our cost function for block-sparse recovery. The improved block-sparse reconstruction algorithm based on approximate l_0 norm minimization can not only avoid the drawbacks of using l_1 norm to approximate l_0 norm but also overcome the NP-hard problem of l_0 norm constraints.

Since the sparse recovery method for each sparse vector θ_i is identical, we omit the subscript of θ for simplicity. Generally, a block-sparse signal can be stated as follow:

$$\theta = \begin{bmatrix} \theta_{11}, \dots, \theta_{1L}, \theta_{21}, \dots, \theta_{2L}, \dots, \theta_{N1}, \dots, \theta_{NL} \end{bmatrix}_{LN \times 1}^T, \quad (12)$$

where ξ_i , $i = 1, \dots, N$, is called the i th block of θ and L is the block size. It is clear that if $L = 1$, block-sparsity reduces

to conventional sparsity. To work with the block version of approximate l_0 norm minimization, let $\beta(x)$ be as follows:

$$\beta(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0. \end{cases} \quad (13)$$

Further, by denoting $\|\theta\|_{2,0} = \sum_{i=1}^N \beta(\|\xi_i\|_2)$, a vector θ is block K -sparse if $\|\theta\|_{2,0} \leq K$. Therefore, the CS-based optimization problem for localization will be given as

$$\operatorname{argmin}_{\theta} \|\theta\|_{2,0}, \quad (14)$$

$$\text{s.t.} \quad \|\mathbf{y} - \mathbf{H}\theta\|_2 \leq \varepsilon,$$

where ε is a user-defined parameter which reflects somehow the noise level. By defining $c_i = \|\xi_i\|_2$, we can rewrite $\|\theta\|_{2,0} = \sum_{i=1}^N \beta(c_i) = \|\mathbf{c}\|_0$, where $\mathbf{c} = [c_1, c_2, \dots, c_N]^T$. Then, the optimization problem is reformed as

$$\operatorname{argmin}_{\theta} \|\mathbf{c}\|_0, \quad (15)$$

$$\text{s.t.} \quad \|\mathbf{y} - \mathbf{H}\theta\|_2 \leq \varepsilon.$$

Note that (15) is a NP-hard problem to solve. Inspired by the idea of the smoothed l_0 norm algorithm in [29], we also choose a suitable continuous function to approximate l_0 norm and minimize it by means of a minimization algorithm for continuous functions. Differently from the Gaussian function in [29], we design the following combination function, which is expressed as

$$f_{\sigma}(x) = \lambda \left(1 - \frac{\sigma^2}{x^2 + \sigma^2} \right) + (1 - \lambda) \left(1 - e^{-x^2/\lambda\sigma^2} \right), \quad (16)$$

where $0 < \lambda < 1$. Obviously, $f_{\sigma}(x)$ is a continuous function and has the following properties:

$$\lim_{\sigma \rightarrow 0} f_{\sigma}(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0. \end{cases} \quad (17)$$

Therefore, this function can meet the conditions required for a smoothing function as [29] pointed out. Moreover, as shown in Figure 2, this combination function is steeper than the Gaussian function around $x = 0$, and thus it can be more accurately approximate to l_0 norm.

Let $F_{\sigma}(\mathbf{c}) = \sum_{i=1}^N f_{\sigma}(c_i)$; we have $\lim_{\sigma \rightarrow 0} F_{\sigma}(\mathbf{c}) = \|\mathbf{c}\|_0$. Therefore, when σ is small, we can obtain

$$\|\theta\|_{2,0} \approx F_{\sigma}(\theta) = \sum_{i=1}^N \left\{ \lambda \left(1 - \frac{\sigma^2}{\sum_{j=1}^L \theta_{ij}^2 + \sigma^2} \right) + (1 - \lambda) \left(1 - e^{(-\sum_{j=1}^L \theta_{ij}^2/\lambda\sigma^2)} \right) \right\}. \quad (18)$$

Then, the final optimization problem will be

$$\operatorname{argmin}_{\theta} F_{\sigma}(\theta), \quad (19)$$

$$\text{s.t.} \quad \|\mathbf{y} - \mathbf{H}\theta\|_2 \leq \varepsilon.$$

Input: Measurement vector \mathbf{y} and dictionary \mathbf{H} .
Initialization:
(1) set $\boldsymbol{\theta}^{(0)} = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{y}$; set μ and T_0 .
(2) choose a suitable decreasing sequence $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_{T_0}]$.
Iteration:
for $m = 1: T_0$
(i) let $\sigma = \sigma_m, \mu_m = \mu\sigma^2$;
(ii) solving $\boldsymbol{\theta}^{(m)}$ by FR algorithm:
(1) compute $\boldsymbol{\alpha}_m$ according to (22);
(2) compute $\boldsymbol{\theta}^{(m+1)}$ according to (21);
(3) project $\boldsymbol{\theta}^{(m+1)}$ back onto the feasible set:
 $\boldsymbol{\theta}^{(m+1)} \leftarrow \boldsymbol{\theta}^{(m+1)} - \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}(\mathbf{H}\boldsymbol{\theta}^{(m+1)} - \mathbf{y})$
end for
Output: $\boldsymbol{\theta} = \boldsymbol{\theta}^{(T_0)}$.

ALGORITHM 1: The main steps of the improved block sparse reconstruction algorithm based on approximate l_0 norm minimization (IBAL0).

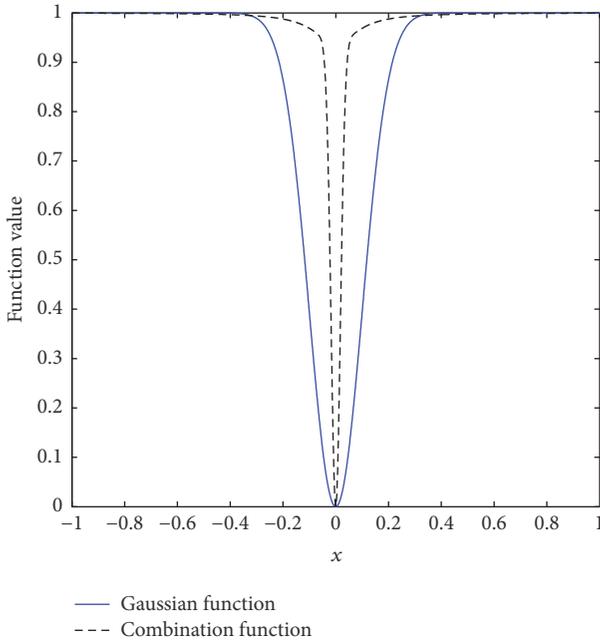


FIGURE 2: Comparisons of two smoothing functions ($\sigma = 0.1$; $\lambda = 1/16$).

Using the Lagrange multiplier method, (19) is converted into an unconstrained optimization problem:

$$\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = F_{\sigma}(\boldsymbol{\theta}) + \gamma \|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|_2^2, \quad (20)$$

where γ is the Lagrange multiplier. Now, we exploit the Fletcher-Reeves (FR) algorithm in [30] to solve the optimization problem in (20), instead of the steepest descent method used in [29]. The FR algorithm belongs to the class of conjugate-gradient methods which has the faster convergence speed than the steepest descent method and has the best numerical stability. Moreover, the FR algorithm does not need to calculate the second derivative of the cost

function (20) as the Newton method. According to the FR algorithm, in the m th iteration $\boldsymbol{\theta}^{(m+1)}$ is updated as

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + \mu_m \boldsymbol{\alpha}_m, \quad (21)$$

where μ_m is the step-size parameter and $\boldsymbol{\alpha}_m$ is the conjugate direction of $L(\boldsymbol{\theta})$ which can be calculated as

$$\boldsymbol{\alpha}_m = \begin{cases} -\mathbf{g}_0, & m = 0 \\ -\mathbf{g}_m + \eta_{m-1} \boldsymbol{\alpha}_{m-1}, & m \neq 0, \end{cases} \quad (22)$$

where \mathbf{g}_m is the gradient vector when $\boldsymbol{\theta} = \boldsymbol{\theta}^{(m)}$ and $\eta_{m-1} = \|\mathbf{g}_m\|_2^2 / \|\mathbf{g}_{m-1}\|_2^2$. For completeness, a full description of the algorithm is given in Algorithm 1.

It should be noted that when σ is small, the solution of (20) is easy to converge to the local optimal solution. In order to converge to the global optimal solution, we exploit the method in [29] to choose a suitable decreasing sequence; that is,

$$\sigma = [\sigma_0, \sigma_1, \dots, \sigma_{T_0}]. \quad (23)$$

Then, the result at $\sigma = \sigma_m$ ($m = 0, 1, \dots, T_0$) as the initial value is used to solve problem (20) at $\sigma = \sigma_{m+1}$, so that the improved approximate l_0 norm reconstruction algorithm converges gradually to the global optimal solution when $\sigma = \sigma_{T_0}$.

4.2. Online DL Method. At this stage the sparse vector is fixed, we update the dictionary \mathbf{H} by using the DL algorithm. Currently, there are some common DL algorithms, such as the K-SVD algorithm, the agent function optimization method, and the recursive least squares method [26]. However, these methods only can effectively handle training data off-line, so they are unable to meet the requirement of rapid positioning. To overcome this drawback, we tend to optimize the dictionary in the online manner. Fortunately, Mairal et al. [31] proposed a new online optimization algorithm based on stochastic approximations for DL, which can learn an updated dictionary to adapt to the dynamic training data changing over time.

Therefore, in order to be fit for the time-varying noise and reverberation, this paper chooses the online DL algorithm in [31]. The main idea of the online DL algorithm is to simply add an increment element to each column of the original dictionary for avoiding performing operations on large matrices, and then it can result in short execution time. According to [31], each column of the dictionary is updated as

$$\begin{aligned} \mathbf{u}_j &\leftarrow \mathbf{h}_j + \frac{(\mathbf{b}_j - \mathbf{H}\mathbf{a}_j)}{\mathbf{A}(j, j)}, \quad j = 1, 2, \dots, N \\ \mathbf{h}_j &\leftarrow \frac{1}{\max(\|\mathbf{u}_j\|_2, 1)} \mathbf{u}_j, \end{aligned} \quad (24)$$

where \mathbf{h}_j , \mathbf{b}_j , and \mathbf{a}_j are the j th column vector of \mathbf{H} , \mathbf{B}_n , and \mathbf{A}_n , respectively. \mathbf{A}_n and \mathbf{B}_n are intermediate variables of the online DL algorithm, whose definitions can be seen in [31] for detail. $\mathbf{A}(j, j)$ represents the element at the j th column and j th row of \mathbf{A}_n . Since the online DL algorithm can update the dictionary in a real-time manner, it can dynamically track the time-varying effects of sound source signals and effectively overcome the mismatch between the sparse model and actual measurements.

5. Simulation Results

To evaluate the localization accuracy and computational cost of the proposed method, simulation and experimental tests were conducted. In this section, we performed extensive simulations to evaluate the performance of the proposed method under different SNRs and different reverberation conditions and compared it with the previously reported CS-based sound source localization algorithms, namely, the convex-optimization based approach in [19] (called Algorithm 1) and the greedy-based two-level FFT method in [21] (called Algorithm 2).

In the simulations, we consider the sound sources positions in a two-dimensional (2D) plane for simplification, and thus the targets are coplanar with the sensors. The extension to three-dimensional (3D) space can also be done with the same steps described as follows. We localized the source using $M = 4$ identical sensors and each sensor had $L = 3$ microphones. We selected the sound source signals from NTT Chinese speech database. We assume that the source signals are known in advance, which is the same assumption as [21]. Although this assumption may be not fit for the time-varying sound sources such as speech signals, except for speaker localization and few other applications, it is usually not practical to require the target (e.g., people) to speak continuously for positioning. Therefore, in many sound source localization scenarios, synthetic acoustic signals are exploited to substitute the speech signals, and thus sound source signals can be assumed to be known in advance. The length of speech segment is $L = 65536$, which is divided into $J = 64$ frames. Thus, the length of each frame is 1024 and the sampling frequency F_s is 16 kHz. In the simulations, the image model in [32] was used to simulate a rectangular room with dimensions $5 \text{ m} \times 5 \text{ m} \times 3 \text{ m}$ (length \times width \times height), and then the received signals can be obtained by performing

convolution operation between the sound source signals and the room impulse response. Assume that the number of grids in the location area is $N = 20 \times 20$, which means yielding a 0.25 m resolution along both x - and y -axes. As for noise samples, we used white noise from Noisex92 database [33]. The localization center is controlled by a laptop with 3.1 GHz processor and 4 GB memory for real-time processing, and it handles the collection of data and runs the localization algorithms. Positioning performance is evaluated by average positioning error and root mean square error (RMSE). RMSE can indicate the statistical positioning results of different algorithms. However, it is the common case that large RMSE can still be gotten due to big estimate errors of few targets, although the positioning accuracy of most of targets is even very high. Therefore, we introduce the positioning failure rate (PFR) indicators for fairness. When the Euclidean distance between the true location of the target and the estimated location of the target is greater than 0.35 m, we consider the estimated result as a failed positioning. PFR is defined as the ratio of the total number of the positioning failures and the total number of experiments.

The position estimates at the five benchmark positions under the anechoic condition and under the reverberant condition are shown in Figures 3(a) and 3(b), respectively. The SNR is set as 10 dB and the reverberation time is 300 ms. In Figure 3(a), the positioning accuracies of three algorithms under the anechoic condition are very close, while the positioning results of three algorithms under the reverberant condition show a performance degradation more or less in Figure 3(b). Since reverberation seriously affects the accuracy of the sparse model, the estimated sound source positions of Algorithms 1 and 2 obviously deviate from the true positions. Differently from Algorithms 1 and 2, the proposed method makes use of the DL technique to dynamically adjust the dictionary for matching the sparse model, and thus the positioning errors of the proposed method do not increase significantly. The detailed statistical characters of different algorithms are summarized in Table 1. All the statistical results are the average of 100 repeated experiments, and in each experiment the position of the sound source is selected at random within the localization area.

As shown in Table 1, the positioning errors of three algorithms under reverberation-free conditions are all small. Although the average positioning errors of Algorithms 1 and 2 are both less than 0.1 m, the proposed algorithm has the lowest average positioning error. Moreover, the RMSE of the proposed algorithm is around 0.11 m and the PFR is also smaller than 3%. These two results are the smallest among three algorithms. The PFR of Algorithm 1 is very close to our proposed algorithm, which shows that the reconstruction method based on the convex-optimization can ensure that the positioning errors of most of sound sources are less than 0.35 m. However, the RMSE of Algorithm 1 is higher than our algorithm. This means that the positioning errors of the failed targets are very large. Since greedy-based OMP algorithm is easy to be affected by noises, Algorithm 2 has the lowest positioning accuracy.

Unlike the results under reverberation-free conditions, the average positioning errors of Algorithms 1 and 2 become

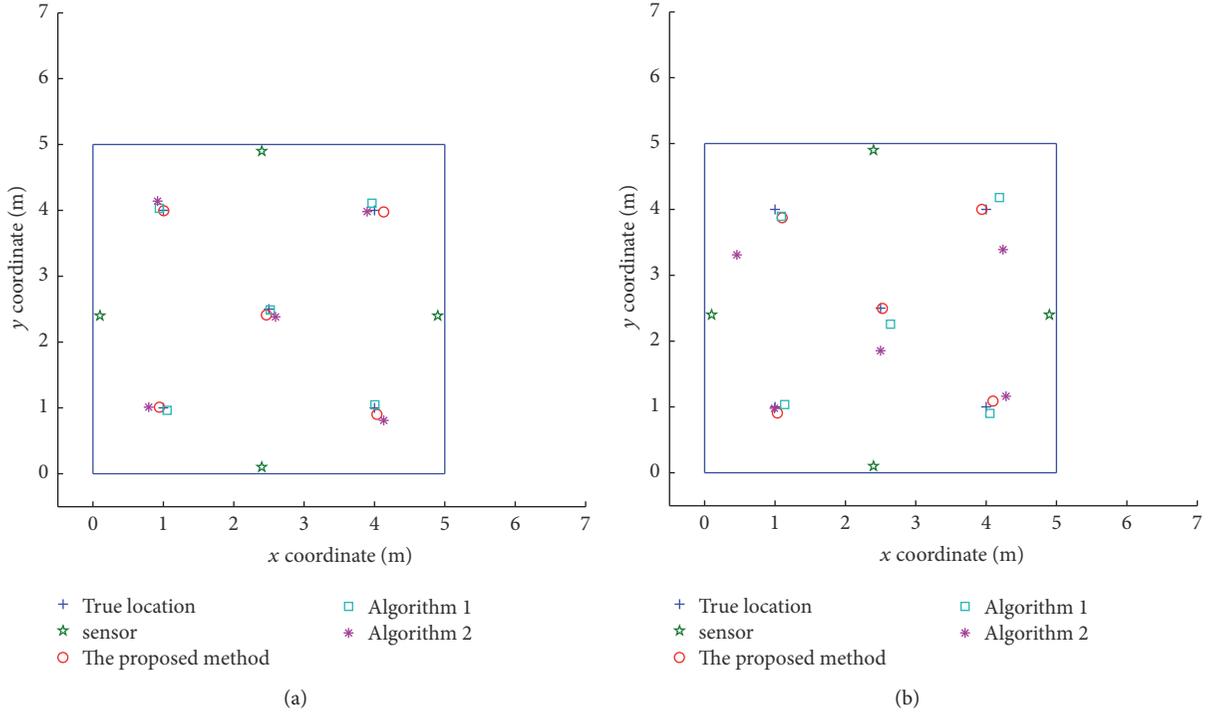


FIGURE 3: Localization results of three algorithms (a) under the anechoic condition and (b) under the reverberation condition.

TABLE 1: Comparisons of positioning performance under different conditions.

	Algorithm	Average	RMSE	PFR
Anechoic environments	Algorithm 1	0.063 m	0.166 m	3.79%
	Algorithm 2	0.077 m	0.187 m	4.95%
	Proposed method	0.051 m	0.113 m	2.82%
Reverberation environments	Algorithm 1	0.155 m	0.277 m	7.58%
	Algorithm 2	0.183 m	0.338 m	8.67%
	Proposed method	0.097 m	0.156 m	4.92%

evidently larger under reverberation environments, with the RMSE increasing 0.111 m and 0.137 m, respectively. This is because there are some errors between the initial dictionary obtained from the sound signal propagation model and real measurements under reverberation environments, thereby resulting in the performance degradation of sparse reconstruction. However, since the proposed algorithm exploits the online DL method to dynamically adjust the dictionary for matching the changes of audio signals, the positioning performance degradation is least in three algorithms. Moreover, the PFR of the proposed algorithm is still below 5%, which means that our algorithm can meet the requirement of effective positioning even under reverberation conditions.

In next part of simulations, the RMSEs of three algorithms under different SNRs are compared in Figure 4, where the reverberation time is also set as 300 ms. Since the signal feature obtained by two-step DCT processing is not sensitive to noises and the IBAL0 algorithm uses the continuous function to approximate l_0 norm constraint, the proposed algorithm can get results more robust against noises and

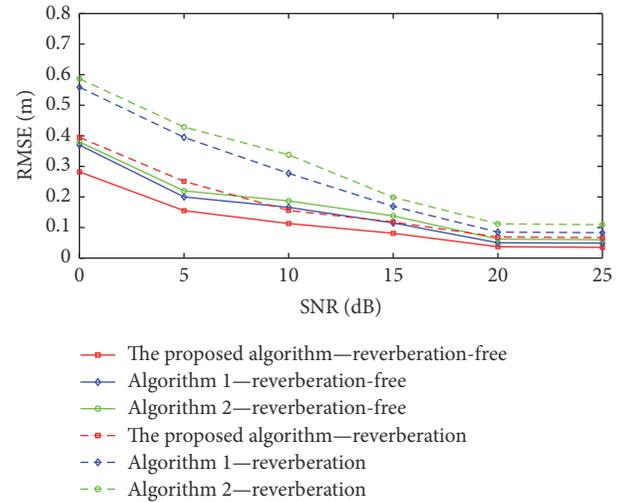


FIGURE 4: RMSEs of three algorithms under different SNR.

ensure the accuracy of the reconstruction. Therefore, the positioning performance of the proposed method is superior to the other two algorithms in Figure 4. The improvement of the proposed algorithm is more obvious at low SNR, where the positioning accuracy can be improved by about 0.1 m at 0 dB SNR under the anechoic condition and about 0.2 m at 0 dB SNR under the reverberation condition, respectively. These phenomena mean the robustness of the proposed method in the presence of noise.

Next, we compare the positioning performance of three algorithms under different reverberation conditions in Figure 5, when SNR is set to 10 dB. With the increase of the

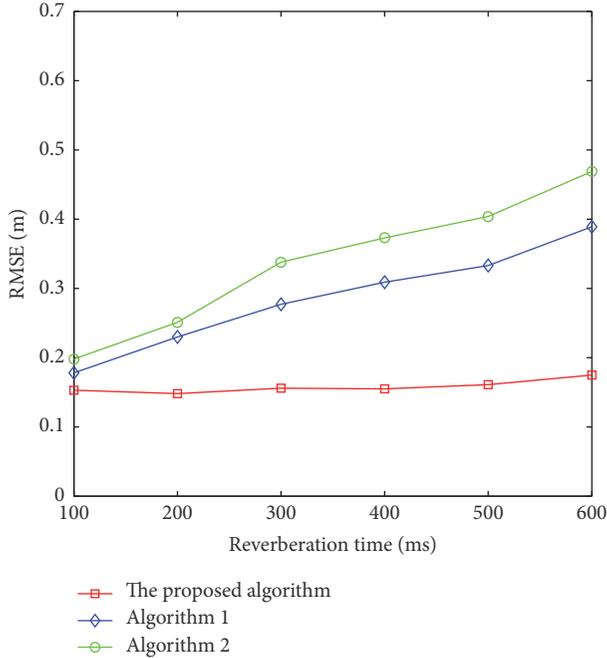


FIGURE 5: RMSEs with respect to the reverberation time.

reverberation time, the RMSEs of two compared algorithms increase quickly due to the high sensitivity to the reverberation. This is because the mismatches between the initial dictionary obtained from the sound signal propagation model and real measurements become more serious with longer reverberation time. On the contrary, since the online DL method can dynamically adjust the dictionary for matching the changes of acoustic signals under reverberation environments, the variations of RMSE in the proposed algorithm are still small.

Finally, the average running time is used as a measure to compare the computational complexity of three algorithms. In Figure 6, Algorithm 2 has the fastest running speed among three algorithms, while Algorithm 1 based on convex-optimization requires the longest running time. Due to adding the DL step in our algorithm, the running speed of the proposed method is also slower than Algorithm 2, although its execution time is far less than Algorithm 1. However, since real-valued DCT-based feature extraction can greatly reduce the dimension of the localization model and the online DL method can avoid performing operations on large matrices, the increment of execution time is not too much. In addition, we exploit the FR algorithm [30] to solve the optimization problem (20), instead of the steepest descent method in [29], which also can reduce execution time.

6. Experimental Results

Using the experimental setup of Figure 7, an experiment was also conducted to evaluate the performance of the proposed method, where the three-dimensional positioning space is a $6.33\text{ m} \times 5.15\text{ m} \times 2.65\text{ m}$ room. The signals of the four sensors were simultaneously sampled by an ESI MAYA44USB

TABLE 2: Comparisons of experimental results.

Algorithm	Average	RMSE	Running time
Algorithm 1	0.185 m	0.293 m	1.030 s
Algorithm 2	0.217 m	0.375 m	0.108 s
Proposed method	0.109 m	0.162 m	0.272 s

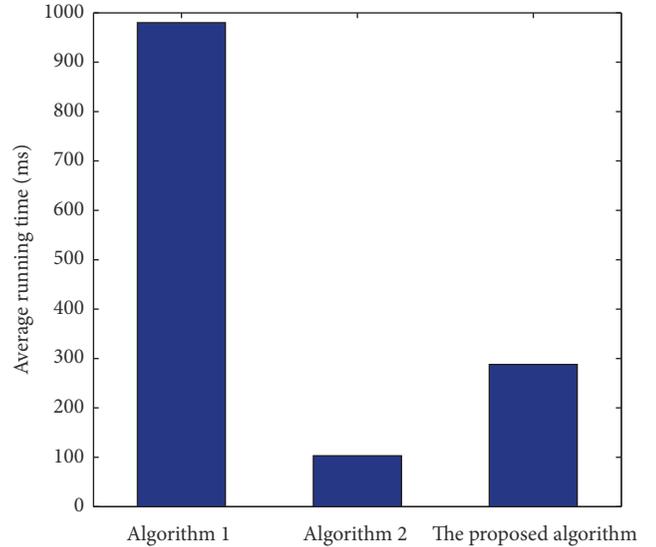


FIGURE 6: Running times of different algorithms.

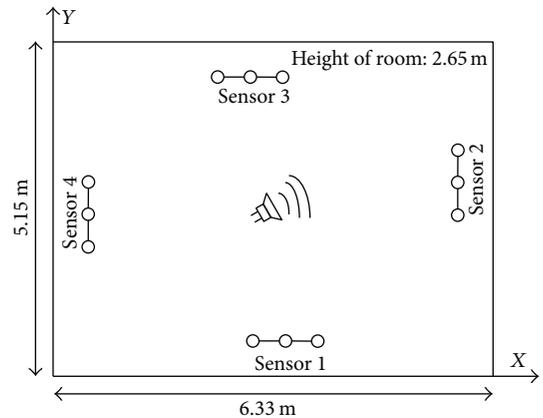


FIGURE 7: Experiment environment scheme.

external sound card via ASIO interface. Each sensor had three inexpensive omnidirectional microphones (\$2 each). The measured reverberation time was about 350 ms. The acoustic signal used in the experiment was a prerecorded male speech sample with duration of about 5 s, which was presented by a speaker with natural voice amplitude to be the sound source. The signal processing part was identical to that in the simulation. All the statistical results are the average of 25 repeated experiments, and in each experiment the position of the source is selected at random within the localization area.

The experimental results in terms of average positioning error, RMSE, and execution time are shown in Table 2. From

Table 2, we can find that the proposed method has the highest localization accuracy among the three algorithms, and the computational cost is moderate. These experimental results confirm the analysis of the simulation tests and also verify the feasibility of the proposed method under real applications.

7. Conclusion

In this paper, a novel sound source localization method based on distributed microphone arrays and CS theory is proposed to improve the localization performance under noisy and reverberant environments. In the proposed method, two-step DCT is used to extract speech signal features and reduce the dimension of the sparse model. The online DL method is further introduced to dynamically adjust the dictionary for matching the changes of audio signals and then overcome the problem of model mismatch. In addition, we propose an improved approximate l_0 norm minimization algorithm to enhance reconstruction performance for block-sparse signals in the low SNR conditions. The effectiveness of the proposed scheme for sound source localization has been demonstrated by simulation and experimental results where the proposed algorithm can get results more robust against noises and reverberations, and substantial improvement for localization performance is achieved. Future research will emphasize the influence of different microphone structures to positioning performance.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20133207120007), the Open Research Fund of Jiangsu Key Laboratory of Meteorological Observation and Information Processing (Grant no. KDXS1408), and Jiangsu Government Scholarship for Oversea Study.

References

- [1] W. S. Jiang, Z. B. Cai, M. F. Luo, and Z. L. Yu, "A simple microphone array for source direction and distance estimation," in *Proceedings of the 6th IEEE Conference on Industrial Electronics and Applications (ICIEA '11)*, pp. 1002–1005, IEEE, Beijing, China, June 2011.
- [2] C. Zhang, D. Florencio, D. E. Ba, and Z. Zhang, "Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 538–548, 2008.
- [3] 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, 2008.
- [4] A. Badali, J.-M. Valin, F. Michaud, and P. Aarabi, "Evaluating real-time audio localization algorithms for artificial audition in robotics," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '09)*, pp. 2033–2038, St. Louis, Mo, USA, October 2009.
- [5] X. Ma and K. Mu, "Study on 2-dimensional microphone circular array speech enhancement algorithm based on the DOA," in *Proceedings of the International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT '11)*, pp. 2505–2508, Harbin, China, August 2011.
- [6] E. Zwyszig, M. Lincoln, and S. Renals, "A digital microphone array for distant speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 5106–5109, Dallas, Tex, USA, March 2010.
- [7] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 439–443, 2013.
- [8] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, "A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 10, pp. 1563–1575, 2015.
- [9] Z. X. Lei, K. D. Yang, R. Duan, and P. Xiao, "Localization of low-frequency coherent sound sources with compressive beamforming-based passive synthetic aperture," *Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. EL255–EL260, 2015.
- [10] Z. X. Yao, J. H. Hu, and D. M. Yao, "A bearing estimation algorithm using an acoustic vector sensor array based on MUSIC," *Acta Acustica*, vol. 4, pp. 305–309, 2008.
- [11] C.-S. Park, J.-H. Jeon, and Y.-H. Kim, "Localization of a sound source in a noisy environment by hyperbolic curves in quefrency domain," *Journal of Sound and Vibration*, vol. 333, no. 21, pp. 5630–5640, 2014.
- [12] X. Alameda-Pineda and R. Horaud, "A geometric approach to sound source localization from time-delay estimates," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 6, pp. 1082–1095, 2014.
- [13] Y. Zhao, X. Chen, and B. Wang, "Real-time sound source localization using hybrid framework," *Applied Acoustics*, vol. 74, no. 12, pp. 1367–1373, 2013.
- [14] Y. Kan, P. F. Wang, F. S. Zha, M. T. Li, W. Gao, and B. Y. Song, "Passive acoustic source localization at a low sampling rate based on a five-element cross microphone array," *Sensors*, vol. 15, no. 6, pp. 13326–13347, 2015.
- [15] A. Pourmohammad and S. M. Ahadi, " N -dimensional N -microphone sound source localization," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, article 27, 19 pages, 2013.
- [16] N. Zhu and S. F. Wu, "Sound source localization in three-dimensional space in real time with redundancy checks," *Journal of Computational Acoustics*, vol. 20, no. 1, Article ID 1250007, 2012.
- [17] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling: a sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [18] M. Yin, K. Yu, and Z. Wang, "Compressive sensing based sampling and reconstruction for wireless sensor array network," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9641608, 11 pages, 2016.

- [19] V. Cevher and R. Baraniuk, "Compressive sensing for sensor calibration," in *Proceedings of the 5th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM '08)*, pp. 175–178, IEEE, Darmstadt, Germany, July 2008.
- [20] H. Jiang, B. Mathews, and P. Wilford, "Sound localization using compressive sensing," in *Proceedings of the 1st International Conference on Sensor Networks (SENSORNETS '12)*, pp. 159–166, Rome, Italy, February 2012.
- [21] M. B. Dehkordi, H. R. Abutalebi, and M. R. Taban, "Sound source localization using compressive sensing-based feature extraction and spatial sparsity," *Digital Signal Processing*, vol. 23, no. 4, pp. 1239–1246, 2013.
- [22] P. Simard and J. Antoni, "Acoustic source identification: experimenting the ℓ_1 minimization approach," *Applied Acoustics*, vol. 74, no. 7, pp. 974–986, 2013.
- [23] N. Chu, J. Picheral, A. Mohammad-Djafari, and N. Gac, "A robust super-resolution approach with sparsity constraint in acoustic imaging," *Applied Acoustics*, vol. 76, pp. 197–208, 2014.
- [24] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [25] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [26] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.
- [27] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [28] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.
- [29] H. Mohimani, Z. M. Babaie, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.
- [30] A. Antoniou and W. S. Lu, *Practical Optimization: Algorithms and Engineering Applications*, Springer, New York, NY, USA, 2006.
- [31] J. Mairal, F. Bach, and J. Ponce, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.
- [32] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [33] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, 1993.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

