sensors

MDPI

Article

Acoustic Source Tracking Based on Probabilistic Data Association and Distributed Cubature Kalman Filtering in Acoustic Sensor Networks

Yang Chen, Yideng Cao and Rui Wang *

School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213164, China
* Correspondence: wangrui@cczu.edu.cn

Abstract: A probabilistic data association-based distributed cubature Kalman filter (PDA-DCKF) method is proposed in this paper, whose performance on tracking single moving sound sources in the distributed acoustic sensor network was verified. In this method, the PDA algorithm is first used to sift the observations from neighboring nodes. Then, the sifted observations are fused to update the state vectors in the CKF. Since nodes in a sensor network have different reliabilities, the final tracking result integrates the estimations from the local nodes, which are weighted with the parameters depending on the mean square error of the estimation and the energy of the received signal. The experimental results illustrated that the proposed PDA-DCKF method is superior to the other DCKF methods in tracking sound sources even under severe noise and reverberant conditions.

Keywords: acoustic source tracking; distributed acoustic sensor networks; distributed cubature Kalman filter; probabilistic data association

Citation: Yang, C.; Cao, Y.; Wang, R. Acoustic Source Tracking Based on Probabilistic Data Association and Distributed Cubature Kalman Filtering in Acoustic Sensor Networks. *Sensors* **2022**, *22*, 7160. https://doi.org/10.3390/s22197160

Academic Editors: Wei Yi and Xiansheng Guo

Received: 15 August 2022
Accepted: 20 September 2022
Published: 21 September 2022

1. Introduction

The problem of acoustic source localization and tracking has always been one of the research hotspots in the field of speech processing. It has been widely used in many aspects, such as audio and video conferencing systems, human-computer interaction and speech enhancement, etc. [1–4]. Traditional acoustic localization and tracking methods usually require the microphone array to have a regular geometric structure, and generally use a centralized data processing method [5]. With the continuous advancement of technology, some traditional microphone arrays gradually show some deficiencies. The distributed microphone network has attracted more and more research work because it has no strict restrictions on the arrangement of microphones, and is a network composed of multiple nodes arbitrarily distributed in space, usually each node contains a set of microphones [6–10].

So far, there have been many studies on acoustic source localization using distributed microphone networks [11]. But they only locate the acoustic source based on the current observations of multiple microphones, which can locate the acoustic source when the background noise and reverberation are small. In noisy and reverberant environments, spurious observations may even mask observations from real acoustic sources, degrading localization performance. To avoid this problem, a Bayesian filter [12] combined the current observation with a series of past observations for current position estimation, which is more effective for dealing with the adverse effects of noise or reverberation. Theoretically, Bayesian filters describe the tracking problem with a state-space model that includes a dynamic model that describes the motion of the target and an observation model that describes the relationship between the observations and the state of the acoustic source. When the state space model is linear and Gaussian, Kalman filter can replace Bayesian

filter. However, in acoustic source tracking scenarios, the observation function is usually nonlinear, and some conditions and properties applicable to linear systems no longer hold, and the performance of the Kalman filter may be severely degraded.

Extended Kalman filter (EKF) was first proposed in [13]and is the simplest and most widely used nonlinear filtering method. EKF only intercepts the first-order term in the linearization process to approximate the system function, but its error is relatively large. In order to solve this problem, the literature [14] proposed iterative extended Kalman filter (IEKF), which can improve the accuracy of EKF through several iterations. However, when the nonlinearity of the system model is very strong, whether it is EKF or IEKF, their effect is not good, and there are disadvantages such as poor stability and easy divergence. The particle filter (PF) is a Monte Carlo implementation of a Bayesian filter that approximates the state by a series of weighted particles extracted from the proposal function [15]. PF can handle nonlinear and non-Gaussian situations well, and many PF-based sound source tracking methods have been developed. Vermaak and Blake [16] first introduced PF to sound source tracking. The particle filter method breaks through the limitations of linearity and Gaussian, but its computational load is too large. In practical applications, particle filtering will only be considered when the approximate filter fails. The Sigama point Kalman filter method [17] was similar to the idea of particle filter. It does not use the method of approximating nonlinear systems, but directly uses the real system and observation model by selecting a set of effective deterministic sampling sets, namely Sigama points, which can achieve second-order accuracy. According to the different selection methods of Sigama points, it can be divided into UKF, CKF, QKF and so on. In general, several Gaussian filtering methods introduced above are centralized, that is, the data of all nodes is collected and transmitted to the central processing unit to perform the task of acoustic source tracking. This method is generally unreliable, as any failure of the central processor renders the entire network untraceable.

In order to solve the unreliable problem of centralized methods, many distributed methods have been developed for sound source tracking. No central processor is needed in the distributed method, and all nodes realize the estimation of the global state only by exchanging data with their neighbors. In reference [18], a distributed extended Kalman particle filter (DEKPF) for speaker tracking was developed, which combined the current TDOA observations into EKF to propose particle filter. In reference [19], a distributed particle filter (DPF) was proposed, which applied the improved iterative covariance intersection (MICI) algorithm and interactive multiple model (IMM) to speaker tracking in distributed microphone networks. In reference [20], a distributed iterative EKF was proposed to estimate the time-varying speaker position in the microphone array. In reference [21], a Distributed Unscented Kalman Filter (DUKF) is proposed to overcome the nonlinearity of the measurement model in speaker tracking. The time difference of arrival (TDOA) was used as the observation and then the distributed IMM-UKF was used to track the location of the sound source.

In the actual environment, the existence of noise or reverberation usually produces unreliable observations with false peaks, which may lead to serious performance degradation. Usually, the current observations contained in these methods are only extracted from the largest peak value of a certain observation function. In some bad cases, the peak value related to the real acoustic source may be masked by the stray acoustic source. Therefore, it is more reasonable to extract multiple observations from the observation function, rather than one observation, and then incorporate it into the above tracking scheme. Probabilistic Data Association (PDA) [22] is an effective method to combine multiple observations into Kalman filter state update, which has been proved to be suitable for target tracking in clutter environment. In reference [23], an improved distributed unscented Kalman particle filter (DUKPF) was proposed to track a single moving acoustic source using a distributed microphone network in noise and reverberation environments. This method proposed to extract multiple observations from the observation function of each node and combined them into the status update of UKF through probabilistic data

association (PDA) technology, so as to generate PDA-UKF, and then brought in particle filter. In reference [24], a microphone array network distributed multi speaker tracking method based on tasteless particle filter and data association was proposed. The available observations were associated with each speaker at each node using data association technology to track the speaker. Reference [25] proposed a volume information filter based on joint probabilistic data association (JPDA) for multi acoustic source tracking based on distributed acoustic vector sensor (AVS) array, in which JPDA was used to deal with the correlation between observations and targets. Issues related to multi-source tracking are beyond the scope of this article. However, most of particle filter-based methods require excessive computational costs, which limits them in real-time applications. Besides, in existing speaker tracking methods, the PDA algorithm is applied to sift the observations without considering the information from neighboring nodes.

Probabilistic data association with cubature Kalman filtering are combined in this paper, and they are applied to the problem of single-acoustic source tracking in noisy and reverberant environments with distributed acoustic sensor networks. The contributions of this paper are as follows:

- Combining the cubature Kalman filter (CKF) with PDA, the probabilistic data association-cubature Kalman filter (PDA-CKF) was developed. In PDA-CKF, multiple possible observations were merged into the state update of CKF by the PDA technique.
- In this paper, PDA-CKF was applied to the distributed acoustic sensor network, and the probabilistic data association-distributed cubature Kalman filter (PDA-DCKF) was developed by combining the observation information of each node's neighbor nodes in the network.
- Considering the reliability of the local state, it was proposed to combine the mean square error (MSE) of the position estimation of each node and the received signal energy to adjust the weighting coefficient of distributed acoustic sensor data fusion. In this way, the local state of high-quality nodes is enhanced, and each node can achieve global consistency and good speaker tracking performance.

The structure of this paper is as follows. Section 2 presents the problem formulation, background knowledge, and some prior knowledge of acoustic source tracking. Section 3 first introduces the single-node PDA-CKF and then details the distributed PDA-DCKF. Section 4 presents the experimental results and discussion. Section 5 summarizes some conclusions.

2. Background Knowledge

2.1. Problem Formulation

Consider a distributed sensor network with N nodes deployed as shown in Figure 1. The positions of the nodes can be obtained in advance by calibration [26]. Each node in the DMA consists of two microphones at distance L. All nodes of the network are modeled as vertices of the graph $\mathcal{G}_1 = (\varepsilon, \upsilon)$, where $\upsilon = \{1, 2, ..., N\}$ is the vertex set, $\varepsilon \subset \{(p,q) \mid p,q \in \upsilon\}$ is the edge set, and $(p,q) \in \varepsilon$ represents the network's communication constraints, i.e., node p can send information to node q, and vice versa. Let $\mathcal{N}_{p,k} = \{q \in \upsilon \mid (p,q) \in \varepsilon\} \bigcup \{p\}$ denote the set of neighbors of node p at time k, where a node is a neighbor of itself certainly.

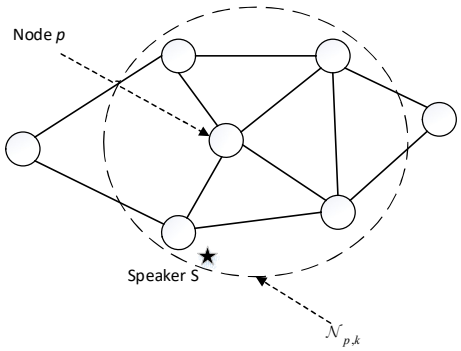


Figure 1. Diagram of speaker tracking in the distributed acoustic sensor network.

2.2. Signal Model and TDOA Estimation

In acoustic sensor networks, the discrete-time signal acquired by the $l-$ th microphone ($l=1,2$) of node p can be modeled as [23]

$$y_{p,l}(t) = h_{p,l}(t) * y(t) + e_{p,l}(t), \forall p \in \upsilon \tag{1}$$

where t is the discrete-time index, $h_{p,l}(t)$ is the room impulse response (RIR) between the microphone and the acoustic source, $*$ denotes the convolution operator, $y(t)$ is the source signal, and $e_{p,l}(t)$ is the additive noise.

Traditionally, the generalized cross-correlation function (GCC) [27] is used for TDOA estimation. Assuming that $Y_1(k)$ and $Y_2(k)$ are the acoustic signal received by a microphone pair at time k and $Y_l(f) = FFT\{Y_l(k)\}, l=1,2$ is the frequency domain representation of the corresponding acoustic signal in a time frame, the generalized cross-correlation function of the acoustic signal received by the microphone pair is

$$R_{12}(\tau) = \int_{-\infty}^{+\infty} \frac{Y_1(f)Y_2^*(f)}{|Y_1(f)Y_2^*(f)|} e^{j2\pi f\tau} df \tag{2}$$

where $Y_1(f)$ and $Y_2(f)$ represent the frequency-domain microphone signals at the node, and $*$ represents the complex conjugation operation. Therefore, the delay estimation is [27]

$$\hat{\tau} = \int_{-\infty}^{+\infty} \underset{\tau\in[-\tau\max,\tau\max]}{\arg\max} \; R_{12}(\tau) \tag{3}$$

$\tau\max$ is the largest time delay estimation.

However, in the real indoor environment, reverberation and noise will bring false maxima of $R_{12}(\tau)$ and obtain invalid TDOA estimation. In order to solve this problem, the local largest of the first Q largest peaks of $R_{12}(\tau)$ are taken as the candidate measurement value of multiple TDOA of node p at time k. In this paper, multiple TDOA observations were extracted through a two-step selection process, taking node p as an example [23].

(1) Select Q delays according to the peak amplitude of the GCC, i.e.,

$$z_{p,k} = [\tilde{\tau}_{p,k}^{(1)}, \tilde{\tau}_{p,k}^{(2)}, ..., \tilde{\tau}_{p,k}^{(Q)}] \tag{4}$$

where $\tilde{\tau}_{p,k}^{(i)}$ is the delay of node p related to the $i-$th largest peak of $R_{12}(\tau)$ at time k.

(2) Further, select $m_{p,k}$ observations from (4) as local observations, and the selection rules are shown in Section 3.

$$\breve{\boldsymbol{z}}_{p,k}=[\tilde{\tau}_{p,k}^{(1)},\tilde{\tau}_{p,k}^{(2)},...,\tilde{\tau}_{p,k}^{(m_{p,k})}] \tag{5}$$

where each delay $\hat{\tau}_{p,k}^{(j)}, j=1,2,...,m_{p,k}$ is deemed as a TDOA candidate.

2.3. Dynamic Model of Acoustic Source

Without loss of generality, the two-dimensional tracking is considered herein, since the height of a moving acoustic source would usually not change significantly. Speakers move in a room with a distributed acoustic sensor network, and Langevin model [24] can accurately and simply describe the time-varying position of speakers. At time k, the state of the speaker is defined as $\boldsymbol{x}_k=[x_k,y_k,\dot{x}_k,\dot{y}_k]^T$, where $(x_k,y_k)^T$ and $(\dot{x}_k,\dot{y}_k)^T$ represent the position and moving speed of the speaker, respectively. In this model, the speaker's motion in the Cartesian coordinate system is considered to be independent and modeled as [23]

$$\boldsymbol{x}_k=\begin{bmatrix}\boldsymbol{I}_2 & a\Delta T\otimes\boldsymbol{I}_2\\ \boldsymbol{0} & a\otimes\boldsymbol{I}_2\end{bmatrix}\boldsymbol{x}_{k-1}+\begin{bmatrix}b\Delta T\otimes\boldsymbol{I}_2 & \boldsymbol{0}\\ \boldsymbol{0} & b\otimes\boldsymbol{I}_2\end{bmatrix}\boldsymbol{u}_{k-1} \tag{6}$$

where $a=e^{-\beta\Delta T}$, and $b=\bar{\upsilon}\sqrt{1-a^2}$; β and $\bar{\upsilon}$ are the rate constant and the steady velocity parameter, respectively. $\boldsymbol{I}_s$ denotes the s-order identical matrix, $\otimes$ stands for the Kronecker product, ΔT is the sampling period for position estimation, and $\boldsymbol{u}_{k-1}$ is the zero-mean white Gaussian noise with identity covariance matrix, which describes the uncertainty of the acoustic source motion.

2.4. Bayesian Framework for Speaker Tracking

Bayesian filtering is the basis of Kalman filtering. This section briefly reviews the basic principles of the Bayesian filtering algorithm.

Assuming that the state variable at time k is $\boldsymbol{x}_k\in\mathbb{R}^p$ and its observation value is $\boldsymbol{y}_k\in\mathbb{R}^q$, where $\mathbb{R}^n$ represents the n-dimensional real vector space, the state equation and observation equation are expressed as [21]:

$$\boldsymbol{x}_{k+1}=f_k(\boldsymbol{x}_k)+\boldsymbol{\Gamma}_k w_k \tag{7}$$

$$\boldsymbol{y}_k=h_k(\boldsymbol{x}_k)+v_k \tag{8}$$

where $f_k(\cdot)$ is the nonlinear state transfer function, $h_k(\cdot)$ is the nonlinear observation function, $\boldsymbol{\Gamma}_k$ is the noise transfer matrix, w_k is the process noise, and v_k is the observation noise, which meets [21]

$$E\left\{\begin{bmatrix}w_k\\ v_k\end{bmatrix}\begin{bmatrix}w_l\\ v_l\end{bmatrix}^T\right\}=\begin{bmatrix}\boldsymbol{Q}_k\delta_{k,l} & 0\\ 0 & \boldsymbol{R}_k\delta_{k,l}\end{bmatrix} \tag{9}$$

where the superscript T represents the transpose of the matrix, $E\{\cdot\}$ represents the expected operator, and $\delta_{k,l}$ represents the Kronecker delta function. $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$

are the covariance matrices of noise w_k and v_k, respectively, and it is assumed that they are both positive definite.

The Bayesian filtering problem is to infer the estimated value of the state variable $\boldsymbol{x}_k$ at time k given the observation information $\boldsymbol{y}_{1:k} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_k\}$ at time k, i.e., to estimate the posterior probability density $p(\boldsymbol{x}_k \mid \boldsymbol{y}_{1:k})$. Assuming that the initial probability density function $p(x_0)$ of the state variable is known as prior knowledge, the posterior probability density $p(\boldsymbol{x}_k \mid \boldsymbol{y}_{1:k})$ can be obtained recursively by the following equations [20]:

$$p(\boldsymbol{x}_k \mid \boldsymbol{y}_{1:k-1}) = \int p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1} \mid \boldsymbol{y}_{1:k-1}) d\boldsymbol{x}_{k-1} \tag{10}$$

$$p(\boldsymbol{x}_k \mid \boldsymbol{y}_{1:k}) = \frac{p(\boldsymbol{y}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \boldsymbol{y}_{1:k-1})}{p(\boldsymbol{y}_k \mid \boldsymbol{y}_{1:k-1})} \tag{11}$$

In Equations (10) and (11), the state transition probability density function $p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1})$ is defined by the state equation; the observation likelihood probability density function $p(\boldsymbol{y}_k \mid \boldsymbol{x}_k)$ is defined by the observation equation.

3. Improved Distributed Cubature Kalman Filter

In the CKF, the observation corresponding to the largest peak of the observation function is used for the state update. This approach works well under moderate acoustic environments, while its performance degrades in severe noise and reverberation conditions because the spurious peaks from noise or reverberation may cover up the peaks from real acoustic sources. To alleviate this problem, multiple observations are selected from the multiple local maxima of the observation function. A general framework for state updates that integrates multiple possible observations is provided by the probabilistic data association (PDA). Inspired by this idea, the probabilistic data association-cubature Kalman filter (PDA-CKF) was derived in this paper. Next, PDA-CKF was used for acoustic source tracking in distributed acoustic sensor networks, and an improved PDA-DCKF algorithm was developed. The observations of multiple nodes in the neighborhood are filtered by PDA and then merged into the state update of CKF to integrate the information of multiple nodes to realize distributed tracking.

Before introducing PDA-CKF, the preliminary knowledge of cubature Kalman—cubature point set $\{\boldsymbol{\xi}_i, \omega_i\}$ [28]—should be introduce first.

The standard Gaussian weighted integral is calculated using the spherical-radial cubature rule, i.e., [28]

$$E[\boldsymbol{x} \mid \boldsymbol{z}] = \int_R f(\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; 0, \boldsymbol{P}) d\boldsymbol{x} \approx \sum_{i=1}^{2n} \omega_i f(\boldsymbol{\xi}_i) \tag{12}$$

In Equation (12), $f(\cdot)$ is the nonlinear state transfer function or observation function, n is the dimension of the state variable, $\mathcal{N}(\boldsymbol{x}; 0, \boldsymbol{P})$ is a Gaussian distribution function with a mean of zero and a variance of $\boldsymbol{P}$, and $\boldsymbol{\xi}_i$ is the cubature points.

$$\boldsymbol{\xi}_i = \sqrt{n}\left[\boldsymbol{I}\right]_i, i = 1, 2, ..., 2n \tag{13}$$

$$\omega_i = \frac{1}{2n} \tag{14}$$

$[\boldsymbol{I}]_i$ represents the point set of n (n-dimensional state) dimensional space, i.e.,

$$[\boldsymbol{I}]_i = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix} \right\} \tag{15}$$

3.1. PDA-CKF Algorithm

(a) Initialization

When $k = 0$, assuming $\boldsymbol{x}_0 \sim \mathcal{N}(\overline{\boldsymbol{x}}_0, \boldsymbol{P}_0)$, the initial value of the process noise and observation noise matrix are set to $\boldsymbol{Q}_0$ and $\boldsymbol{R}_0$, respectively. Then, the optimal initialization of the filter is

$$\begin{aligned} \hat{\boldsymbol{x}}_{p,0|0} &= \overline{\boldsymbol{x}}_0 \\ \hat{\boldsymbol{P}}_{p,0|0} &= \boldsymbol{P}_0 \end{aligned} \tag{16}$$

(b) State Prediction

For each node p, the state estimate and covariance matrix $\hat{\boldsymbol{x}}_{p,k-1|k-1}, \hat{\boldsymbol{P}}_{p,k-1|k-1}$ at time $k-1$ are given, and the positive definite noise matrix $\boldsymbol{Q}_{p,k-1}, \boldsymbol{R}_{p,k-1}$ are given. Using Equations (13) and (14), the state predicted cubature points $\boldsymbol{\chi}^i_{p,k-1|k-1}$ is calculated as:

$$\boldsymbol{S}_{p,k-1|k-1} = \sqrt{\hat{\boldsymbol{P}}_{p,k-1|k-1}} \tag{17}$$

$$\boldsymbol{\chi}^i_{p,k-1|k-1} = \hat{\boldsymbol{x}}_{p,k-1|k-1} + \boldsymbol{S}_{p,k-1|k-1}\boldsymbol{\xi}_i, i = 1, 2, \ldots, 2n \tag{18}$$

According to the state transition model, the cubature points are propagated nonlinearly, i.e.,

$$\boldsymbol{\chi}^i_{p,k|k-1} = f(\boldsymbol{\chi}^i_{p,k-1|k-1}), i = 1, 2, \ldots, 2n, p = 1, 2, \ldots, N \tag{19}$$

where n represents the dimension of the state variable, and N represents the number of nodes in the distributed acoustic sensor network. At this time, the state prediction $\hat{\boldsymbol{x}}_{p,k|k-1}$ and its error matrix $\hat{\boldsymbol{P}}_{p,k|k-1}$ are calculated as:

$$\hat{\boldsymbol{x}}_{p,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n} \boldsymbol{\chi}^i_{p,k|k-1}, i = 1, 2, \ldots, 2n, p = 1, 2, \ldots, N \tag{20}$$

$$\hat{\boldsymbol{P}}_{p,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n} (\boldsymbol{\chi}^i_{p,k|k-1} - \hat{\boldsymbol{x}}_{p,k|k-1})(\boldsymbol{\chi}^i_{p,k|k-1} - \hat{\boldsymbol{x}}_{p,k|k-1})^T + \boldsymbol{Q}_{p,k}, p = 1, 2, \ldots, N \tag{21}$$

(c) Status Update

From the estimated $\hat{\boldsymbol{x}}_{p,k|k-1}$ and variance $\hat{\boldsymbol{P}}_{p,k|k-1}$ at time k, the state update cubature points $\boldsymbol{\chi}^i_{p,k|k-1}$ is calculated as:

$$\boldsymbol{S}_{p,k|k-1} = \sqrt{\hat{\boldsymbol{P}}_{p,k|k-1}} \tag{22}$$

$$\boldsymbol{\chi}^{i}_{p,k|k-1} = \hat{\boldsymbol{x}}_{p,k|k-1} + \boldsymbol{S}_{p,k|k-1}\boldsymbol{\xi}_i \tag{23}$$

$\boldsymbol{\chi}^{i}_{p,k|k-1}$ is propagated through the observation equation,

$$\hat{\boldsymbol{Z}}^{i}_{p,k|k-1} = h(\boldsymbol{\chi}^{i}_{p,k|k-1}), i = 1,2,...,2n, p = 1,2,...,N \tag{24}$$

Further, the observation prediction $\hat{\boldsymbol{z}}_{p,k|k-1}$ and the observation prediction error variance $\boldsymbol{P}^{zz}_{p,k|k-1}$ are, respectively, obtained by

$$\hat{\boldsymbol{z}}_{p,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n}\hat{\boldsymbol{Z}}^{i}_{p,k|k-1} \tag{25}$$

$$\boldsymbol{P}^{zz}_{p,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n}(\hat{\boldsymbol{Z}}^{i}_{p,k|k-1} - \hat{\boldsymbol{z}}_{p,k|k-1})(\hat{\boldsymbol{Z}}^{i}_{p,k|k-1} - \hat{\boldsymbol{z}}_{p,k|k-1})^T + \boldsymbol{R}_{p,k} \tag{26}$$

Then, according to the probabilistic data association, the verification area of node p can be constructed by [29]:

$$\{\boldsymbol{z}_{p,k} : (\boldsymbol{z}_{p,k} - \hat{\boldsymbol{z}}_{p,k|k-1})^T (\boldsymbol{P}^{zz}_{p,k|k-1})^{-1} (\boldsymbol{z}_{p,k} - \hat{\boldsymbol{z}}_{p,k|k-1}) \le \gamma\} \tag{27}$$

where γ is the gate threshold. Suppose $m_{p,k}(m_{p,k} \ge 0)$ observations fall into the validated region (27) at time k. Define validate observations $\breve{\boldsymbol{z}}_{p,k}$, i.e.,

$$\breve{\boldsymbol{z}}_{p,k} = \boldsymbol{z}^{(j)}_{p,k}, j = 1,2,...,m_{p,k} \tag{28}$$

Actually, only one of the above observations is related to the real source; the others are due to noise or reverberation, or none of them are related to the real source. Correspondingly, for $m_{p,k}$ validated observations, there maybe be $m_{p,k}+1$ possible hypothesis, i.e.,

$$\begin{cases} \mathrm{H}_{p,0} = \{\text{All observations are independent of real sound sources}\}, & j = 0 \\ \mathrm{H}_{p,j} = \{\boldsymbol{z}^{(j)}_{p,k} \text{ is associated with the true source}\}, & j = 1,2,...,m_{p,k} \end{cases} \tag{29}$$

According to Equation (29), the equation for calculating $\hat{\boldsymbol{x}}^{(j)}_{p,k|k}, j = 0,1,...,m_{p,k}$ is as follows:

$$\hat{\boldsymbol{x}}_{p,k|k} = \sum_{j=0}^{m_{p,k}} \mathrm{E}\{\boldsymbol{x}_{p,k} \mid \mathrm{H}_{p,j}, \boldsymbol{z}_{p,1:k}\} p(\mathrm{H}_{p,j} \mid \boldsymbol{z}_{p,1:k}) = \sum_{j=0}^{m_{p,k}} \beta^{(j)}_{p,k} \hat{\boldsymbol{x}}^{(j)}_{p,k|k} \tag{30}$$

where $\beta^{(j)}_{p,k} \triangleq p(\mathrm{H}_{p,j} \mid \boldsymbol{z}_{p,1:k})$ is the prior probability of event $\mathrm{H}_{p,j}$, $0 \le \beta^{(j)}_{p,k} \le 1$, and $\sum_{j=0}^{m_{p,k}} \beta^{(j)}_{p,k} = 1$, $\hat{\boldsymbol{x}}^{(j)}_{p,k|k} \triangleq \mathrm{E}\{\boldsymbol{x}_{p,k} \mid \mathrm{H}_{p,j}, \boldsymbol{z}_{p,1:k}\}$ is the updated estimate conditioned on the event $\mathrm{H}_{p,j}, j = 0,1,...,m_{p,k}$, and

$$\hat{\boldsymbol{x}}^{(0)}_{p,k|k} = \hat{\boldsymbol{x}}_{p,k|k-1} \tag{31}$$

$$\hat{\boldsymbol{x}}^{(j)}_{p,k|k} = \hat{\boldsymbol{x}}_{p,k|k-1} + \boldsymbol{K}_{p,k}\boldsymbol{v}^{(j)}_{p,k}, j = 1,2,...,m_{p,k} \tag{32}$$

where $\boldsymbol{v}^{(j)}_{p,k} = \boldsymbol{z}^{(j)}_{p,k} - \hat{\boldsymbol{z}}_{p,k|k-1}$ is the innovation related to the observation $\boldsymbol{z}^{(j)}_{p,k}$, $\boldsymbol{K}_{p,k}$ is the Kalman gain of node p, and

$$\boldsymbol{P}_{p,k|k-1}^{xz} = \frac{1}{2n}\sum_{i=1}^{2n}(\boldsymbol{\chi}_{p,k|k-1}^{i} - \hat{\boldsymbol{x}}_{p,k|k-1})(\hat{\boldsymbol{Z}}_{p,k|k-1}^{i} - \hat{\boldsymbol{z}}_{p,k|k-1})^{T} \tag{33}$$

$$\boldsymbol{K}_{p,k} = \boldsymbol{P}_{p,k|k-1}^{xz}(\boldsymbol{P}_{p,k|k-1}^{zz})^{-1} \tag{34}$$

where $\boldsymbol{P}_{p,k|k-1}^{xz}$ is the cross covariance between the state and observation $\boldsymbol{z}_{p,k}$ of node p.

Given the innovation $\boldsymbol{v}_{p,k}^{(j)}$ and its covariance $\boldsymbol{P}_{p,k|k-1}^{zz}$, the probability $\beta_{p,k}^{(j)}$ is generally computed as [30]

$$\begin{aligned}
\beta_{p,k}^{(j)} &= \begin{cases} \dfrac{e_{p,j}}{b_p + \sum_{i=1}^{m_{p,k}} e_{p,i}}, j = 1,2,...,m_{p,k} \\ \dfrac{b_p}{b_p + \sum_{i=1}^{m_{p,k}} e_{p,i}}, j = 0 \end{cases} \\
e_{p,j} &= e^{-\frac{1}{2}(v_{p,k}^{(j)})^{T}(\boldsymbol{P}_{p,k|k-1}^{zz})^{-1}v_{p,k}^{(j)}} \\
b_p &= \lambda \,|\, 2\pi \boldsymbol{P}_{p,k|k-1}^{zz} \,|^{\frac{1}{2}} \frac{1 - P_{p,D}P_G}{P_{p,D}}
\end{aligned} \tag{35}$$

where λ is the spatial probability, $P_{p,D}$ is the probability that the acoustic source is detected by sensor p, and P_G is the gate probability.

Finally, the state estimate value $\hat{\boldsymbol{x}}_{p,k|k}$ and error covariance $\hat{\boldsymbol{P}}_{p,k|k}$ can be obtained by

$$\hat{\boldsymbol{x}}_{p,k|k} = \hat{\boldsymbol{x}}_{p,k|k-1} + \boldsymbol{K}_{p,k}\boldsymbol{v}_{p,k} \tag{36}$$

$$\hat{\boldsymbol{P}}_{p,k|k} = \beta_{p,k}^{(0)}\hat{\boldsymbol{P}}_{p,k|k-1} + (1-\beta_{p,k}^{(0)})\dot{\boldsymbol{P}}_{p,k|k} + \ddot{\boldsymbol{P}}_{p,k|k} \tag{37}$$

where $\boldsymbol{v}_{p,k} = \sum_{j=1}^{m_{p,k}} \beta_{p,k}^{(j)}\boldsymbol{v}_{p,k}^{(j)}$ is the probability weighted innovation, and the covariances $\dot{\boldsymbol{P}}_{p,k|k}$ and $\ddot{\boldsymbol{P}}_{p,k|k}$ are respectively given by [29,30]

$$\dot{\boldsymbol{P}}_{p,k|k} = \hat{\boldsymbol{p}}_{p,k|k-1} - \boldsymbol{K}_{p,k}\boldsymbol{P}_{p,k|k-1}^{zz}\boldsymbol{K}_{p,k}^{T} \tag{38}$$

$$\ddot{\boldsymbol{P}}_{p,k|k} = \boldsymbol{K}_{p,k}\{\sum_{j=1}^{m_{p,k}} \beta_{p,k}^{(j)}(\boldsymbol{v}_{p,k}^{(j)} - \boldsymbol{v}_{p,k})(\boldsymbol{v}_{p,k}^{(j)} - \boldsymbol{v}_{p,k})^{T}\}\boldsymbol{K}_{p,k}^{T} \tag{39}$$

To summarize, the pseudo-code of the PDA-CKF method of using the observations from a single node is depicted in Algorithm 1.

Algorithm 1: PDA-CKF Algorithm

Initialization: $\hat{\boldsymbol{x}}_{p,0|0} = \bar{\boldsymbol{x}}_0, \hat{\boldsymbol{P}}_{p,0|0} = \boldsymbol{P}_0$

Input: $\hat{\boldsymbol{x}}_{p,k-1|k-1}, \hat{\boldsymbol{P}}_{p,k-1|k-1}, \boldsymbol{z}_{p,k}$

Output: $\hat{\boldsymbol{x}}_{p,k|k}, \hat{\boldsymbol{P}}_{p,k|k}$

Iteration: for $k = 1,2,...$

1: Prediction step:

2: Compute the state predicted cubature points $\boldsymbol{\chi}^i_{p,k-1|k-1}$ at time $k-1$ with (18).

3: Compute the predicted estimate $\hat{\boldsymbol{x}}_{p,k|k-1}$ and covariance $\hat{\boldsymbol{P}}_{p,k|k-1}$ with (20) and (21), respectively.

4: Update step:

5: Compute the state update cubature points $\boldsymbol{\chi}^i_{p,k|k-1}$ with (23).

6: Compute the predicted observations $\hat{\boldsymbol{z}}_{p,k|k-1}$ with (25).

7: Compute the innovation covariance $\boldsymbol{P}^{zz}_{p,k|k-1}$ with (26).

8: Select the validated observations $\breve{\boldsymbol{z}}_{p,k}$ according to (28).

9: Compute the cross-covariance $\boldsymbol{P}^{xz}_{p,k|k-1}$ with (33).

10: Compute the Kalman gain $\boldsymbol{K}_{p,k}$ with (34).

11: Compute the association probability $\beta^{(j)}_{p,k}$ with (35), $j = 1,2,...,m_k$

12: Compute the covariances $\dot{\boldsymbol{P}}_{p,k|k}$ and $\ddot{\boldsymbol{P}}_{p,k|k}$ with (38) and (39), respectively.

13: Compute the updated estimate $\hat{\boldsymbol{x}}_{p,k|k}$ and covariance $\hat{\boldsymbol{P}}_{p,k|k}$ with (36) and (37), respectively.

The PDA-CKF algorithm makes full use of the observation information of the node itself, which improves the tracking accuracy. However, this algorithm will fail when a node is damaged or the environmental noise and reverberation are severe. Therefore, this paper generalized PDA-CKF to a distributed version that can be used in distributed sensor networks. The improved method was named the probabilistic data association-based distributed cubature Kalman filter (PDA-DCKF). The specific process is shown in Section 3.2.

3.2. PDA-DCKF Algorithm

3.2.1. PDA-DCKF

The neighborhood information of nodes are fused in PDA-DCKF to form local node networks. Then, the local state estimations and error covariances for the local node networks are calculated separately. Finally, the local results are fused to obtain the global state estimation.

On the basis of the above steps, the following is defined:

$$\begin{aligned} &\hat{\boldsymbol{Z}}^i_{Np,k|k-1} = [\hat{\boldsymbol{Z}}^i_{p,k|k-1}; \hat{\boldsymbol{Z}}^i_{q,k|k-1}]_{num(\mathcal{N}_{p,k})\times 2n}, i = 1,2,...,2n, p = 1,2,...,N, \\ &q \in \mathcal{N}_{p,k} = \{q \in \upsilon \,|\, (p,q) \in \varepsilon\} \bigcup \{p\} \end{aligned} \tag{40}$$

where q represents the neighborhood nodes adjacent to node p, $\upsilon = \{1,2,...,N\}$ is the vertex set, $\varepsilon \subset \{(p,q) \,|\, p,q \in \upsilon\}$ is the edge set of the distributed acoustic sensor network, $num(\mathcal{N}_{p,k})$ indicates the number of nodes in the neighborhood of node p.

$\mathcal{N}_{p,k} = \{q \in \upsilon \mid (p,q) \in \varepsilon\} \bigcup \{p\}$ denotes the set of neighbors of node p at time k, where a node is a neighbor of itself certainly.

Further, the resulting observations are fused into a matrix. Then, the observed prediction and prediction error variance are, respectively, given by

$$\hat{\boldsymbol{z}}_{Np,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n} \hat{\boldsymbol{Z}}^{i}_{Np,k|k-1} \tag{41}$$

$$\boldsymbol{P}^{zz}_{Np,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n} (\hat{\boldsymbol{Z}}^{i}_{Np,k|k-1} - \hat{\boldsymbol{z}}_{Np,k|k-1})(\hat{\boldsymbol{Z}}^{i}_{Np,k|k-1} - \hat{\boldsymbol{z}}_{Np,k|k-1})^{T} + [\boldsymbol{R}_{P,k}; \boldsymbol{R}_{q,k}]_{num(\mathcal{N}_{p,k})\times num(\mathcal{N}_{p,k})} \tag{42}$$

For single node p, $\boldsymbol{v}^{(j)}_{p,k} = \boldsymbol{z}^{(j)}_{p,k} - \hat{\boldsymbol{z}}_{p,k|k-1}$ is the innovation vector related to observation $\boldsymbol{z}^{(j)}_{p,k}$, and $\boldsymbol{K}_{p,k}$ is the Kalman gain of node p. As far as multiple nodes are concerned, the information of node p and surrounding nodes q is fused to obtain

$$\boldsymbol{P}^{xz}_{Np,k|k-1} = \frac{1}{2n}\sum_{i=1}^{2n} (\boldsymbol{\chi}^{i}_{p,k|k-1} - \hat{\boldsymbol{x}}_{p,k|k-1})(\hat{\boldsymbol{Z}}^{i}_{Np,k|k-1} - \hat{\boldsymbol{z}}_{Np,k|k-1})^{T} \tag{43}$$

$$\boldsymbol{K}_{Np,k} = \boldsymbol{P}^{xz}_{Np,k|k-1}(\boldsymbol{P}^{zz}_{Np,k|k-1})^{-1} \tag{44}$$

where $\boldsymbol{P}^{xz}_{Np,k|k-1}$ is the cross covariance between the state and the observed value of node p after fusing the information of neighboring nodes, and $\boldsymbol{K}_{Np,k}$ is the Kalman gain of node p at time k after the fusion.

the probability weighted innovation vector of local nodes is defined as

$$\boldsymbol{v}_{NP,k} = [\boldsymbol{v}_{p,k}; \boldsymbol{v}_{q,k}]_{num(\mathcal{N}_{p,k})\times 1}, p = 1,2,...,N, q \in \mathcal{N}_{p,k} = \{q \in \upsilon \mid (p,q) \in \varepsilon\} \bigcup \{p\} \tag{45}$$

The following is defined as

$$\beta^{(0)}_{Np,k} = (\beta^{(0)}_{p,k} + \sum_{q=1}^{num(\mathcal{N}_{p,k})-1} \beta^{(0)}_{q,k}) / num(\mathcal{N}_{p,k}) \tag{46}$$

where $\{\sum_{j=1}^{m_{p,k}} \beta^{(j)}_{p,k}(\boldsymbol{v}^{(j)}_{p,k} - \boldsymbol{v}_{p,k})(\boldsymbol{v}^{(j)}_{p,k} - \boldsymbol{v}_{p,k})^{T}\}$ is defined in the covariance $\ddot{\boldsymbol{P}}_{p,k|k}$ of node p as $\boldsymbol{w}_{p}$; when the information of node p and surrounding nodes is fused, the expression of $\boldsymbol{w}_{p}$ is computed as

$$\begin{aligned} &\boldsymbol{w}_{Np} = \mathrm{diag}(\boldsymbol{w}_{p}, \boldsymbol{w}_{q})_{num(\mathcal{N}_{p,k})\times num(\mathcal{N}_{p,k})}, p = 1,2,...,N, \\ &q \in \mathcal{N}_{p,k} = \{q \in \upsilon \mid (p,q) \in \varepsilon\} \bigcup \{p\} \end{aligned} \tag{47}$$

where $\boldsymbol{w}_{q} = \{\sum_{j=1}^{m_{q,k}} \beta^{(j)}_{q,k}(\boldsymbol{v}^{(j)}_{q,k} - \boldsymbol{v}_{q,k})(\boldsymbol{v}^{(j)}_{q,k} - \boldsymbol{v}_{q,k})^{T}\}$.

Finally, the state estimate $\hat{\boldsymbol{x}}_{Np,k|k}$ and the error covariance $\hat{\boldsymbol{P}}_{Np,k|k}$ for node p are expressed as

$$\hat{\boldsymbol{x}}_{Np,k|k} = \hat{\boldsymbol{x}}_{p,k|k-1} + \boldsymbol{K}_{Np,k}\boldsymbol{v}_{Np,k} \tag{48}$$

$$\hat{\boldsymbol{P}}_{Np,k|k} = \beta_{Np,k}^{(0)} \hat{\boldsymbol{P}}_{p,k|k-1} + (1-\beta_{Np,k}^{(0)}) \dot{\boldsymbol{P}}_{Np,k|k} + \ddot{\boldsymbol{P}}_{Np,k|k} \tag{49}$$

$$\dot{\boldsymbol{P}}_{Np,k|k} = \hat{\boldsymbol{p}}_{p,k|k-1} - \boldsymbol{K}_{Np,k} \boldsymbol{P}_{Np,k|k-1}^{zz} \boldsymbol{K}_{Np,k}^{T} \tag{50}$$

$$\ddot{\boldsymbol{P}}_{Np,k|k} = \boldsymbol{K}_{Np,k} \boldsymbol{w}_{Np} \boldsymbol{K}_{Np,k}^{T} \tag{51}$$

3.2.2. Fusion Strategy

After calculating the estimation of each local node in the distributed acoustic sensor network, these data need to be fused to obtain a global estimate. Since nodes in a sensor network have different reliabilities, the final tracking result integrates the estimations from the local nodes, which are weighted with the parameters depending on the mean square error of the estimation and the energy of the received signal.

(a) Energy

The energy of the signal received by each node in the acoustic sensor network is calculated [31], and the equation is described as:

$$E_p = \lim_{T\to\infty} \int_{-T}^{T} |x_p(t)|^2 \, dt \tag{52}$$

where $x_p(t)$ represents the sound signal received by node p. In practice, analog signal $x(t)$ is converted into digital signal $x(n)$, and $x(n)$ needs to be framed and windowed. Then, the framed signal is donated by $x(n)\cdot\omega(n)$. In this paper, the Hamming window was selected for the window function $\omega(n)$. Further, the energy of each frame can be obtained by

$$E_{p,n} = \sum_{m=-\infty}^{\infty} [x_p(m)\omega(n-m)]^2 = \sum_{m=-\infty}^{\infty} {x_p}^2(m) h(n-m) = {x_p}^2(n) * h(n) \tag{53}$$

Where $h(n) = \omega^2(n)$, and $E_{p,n}$ represents the short-term energy of node p when the window function starts at the $n-th$ point of the signal. The short-term energy can be regarded as the output of the square of the speech signal passing through a linear filter, and the unit impulse response of the linear filter is $h(n)$.

(b) MSE

In Equation (48), the local estimate $\hat{\boldsymbol{x}}_{Np,k|k}$ of node $p(p=1,2,...,N)$ is calculated, and $\hat{\boldsymbol{r}}_{p,k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \hat{\boldsymbol{x}}_{Np,k|k}$ is expressed as the estimated acoustic source position of node p at time k. The following is defined:

$$\hat{\boldsymbol{r}}_{N,k} = \frac{1}{N} \sum_{p=1}^{N} \hat{\boldsymbol{r}}_{p,k} \tag{54}$$

where $\hat{\boldsymbol{r}}_{N,k}$ represents the global position estimation result weighted with the average consensus coefficients and calculates the MSE between the position obtained by each local node and $\hat{\boldsymbol{r}}_{N,k}$, defined as

$$M_p = \| \hat{\boldsymbol{r}}_{p,k} - \hat{\boldsymbol{r}}_{N,k} \|^2 \tag{55}$$

After calculating the energy E_p and the mean square error M_p of node p at time k, the following is defined:

$$C_p = \frac{E_p}{M_p} \tag{56}$$

$$\eta_p = \frac{C_p}{\sum_{p=1}^{N} C_p} \tag{57}$$

where η_p represents the weight of node p during global fusion. A global consistency analysis is performed on the results obtained by each node according to $\eta_p, p = 1, 2, ..., N$:

$$\hat{\boldsymbol{x}}_{k|k} = \sum_{p=1}^{N} \eta_p \hat{\boldsymbol{x}}_{Np,k|k} \tag{58}$$

$$\hat{\boldsymbol{P}}_{k|k} = \sum_{p=1}^{N} \eta_p \hat{\boldsymbol{P}}_{Np,k|k} \tag{59}$$

To summarize, the PDA-DCKF is depicted in Algorithm 2.

Algorithm 2: PDA-DCKF Algorithm

Initialization: $\hat{\boldsymbol{x}}_{p,0|0} = \bar{\boldsymbol{x}}_0, \hat{\boldsymbol{P}}_{p,0|0} = \boldsymbol{P}_0$

Input: $\hat{\boldsymbol{x}}_{k-1|k-1}, \hat{\boldsymbol{P}}_{k-1|k-1}, \boldsymbol{z}_k$

Output: $\hat{\boldsymbol{x}}_{k|k}, \hat{\boldsymbol{P}}_{k|k}$

Iteration: for $k = 1, 2, ...$

For any node $p(p = 1, 2, ..., N)$ in sensor network

1: Prediction step:

2: Compute the state predicted cubature points $\boldsymbol{\chi}^i_{p,k-1|k-1}$ at time $k-1$ with (18).

3: Compute the predicted estimate $\hat{\boldsymbol{x}}_{p,k|k-1}$ and covariance $\hat{\boldsymbol{P}}_{p,k|k-1}$ with (20) and (21), respectively.

4: Update step:

5: Compute the state update cubature points $\boldsymbol{\chi}^i_{p,k|k-1}$ with (23).

6: Compute the observed values of predicted local nodes $\hat{\boldsymbol{z}}_{Np,k|k-1}$ with (41).

7: Compute the innovation covariance of predicted local nodes $\boldsymbol{P}^{zz}_{Np,k|k-1}$ with (42).

8: Select the validated observations $\bar{\boldsymbol{z}}_{p,k}$ according to (28), $p = 1, 2, ..., N$.

9: Compute the cross-covariance of predicted local nodes $\boldsymbol{P}^{xz}_{Np,k|k-1}$ with (43).

10: Compute the Kalman gain $\boldsymbol{K}_{Np,k}$ with (44).

11: Compute the probability weighted innovation vector of local nodes $\boldsymbol{v}_{Np,k}$ with (45)

12: Compute the association probability $\beta^{(j)}_{p,k}$ with (35), $j = 1, 2, ..., m_{p,k}$.

13: Compute the association probability $\beta^{(0)}_{Np,k}$ with (46).

14: Compute $\boldsymbol{vl}_{Np}$ with (47).

15: Compute the updated estimate $\hat{\boldsymbol{x}}_{Np,k|k}$ and covariance $\hat{\boldsymbol{P}}_{Np,k|k}$ of local nodes with (48) and (49), respectively.

16: Compute the weight η_p of node $p(p=1,2,...N)$ with (57).

17: Compute the updated estimate $\hat{\boldsymbol{x}}_{k|k}$ and covariance $\hat{\boldsymbol{P}}_{k|k}$ with (58) and (59), respectively.

The advantages of probabilistic data association and distributed acoustic sensor networks are combined in the PDA-DCKF proposed in this paper. In this method, the PDA algorithm is used to sift the observations from neighboring nodes. Then, the sifted observations are fused to update the state vectors in the CKF. This method not only makes the observation value obtained by each node more accurate, but also makes full use of the information of neighborhood nodes.

Meanwhile, a weighted fusion method based on local node-received signal energy and position estimation mean square error was proposed. This dynamic weighted consistency fusion considers the reliability of the local state of the nodes and provides a good global estimation performance.

4. Experiments and Results Discussion

To verify the performance of the proposed speaker tracking method, the evaluations are performed in a simulated room environment. Under the same conditions, the comparative experiments between PDA-DCKF and current methods are carried out, including centralized method (CCKF), DUKF, DCKF, iteration based DCKF [20] (DICKF) and DEKF. The results obtained by all methods are the average of 100 Monte Carlo runs.

The root mean square error (RMSE) is used here to evaluate the tracking performance. $\boldsymbol{r}_k$ is expressed as the ground truth value of time k, and $\hat{\boldsymbol{r}}_{N,k}$ represents the global consistency position calculated by the acoustic sensor network at this time. The RMSE is defined as [32]

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} \| \boldsymbol{r}_k - \hat{\boldsymbol{r}}_{N,k} \|^2} \tag{60}$$

where K denotes the number of frames. Generally, the smaller the RMSE, the better the tracking result.

4.1. Simulation Setups

The simulation environment was a typical room of size 6 m × 6 m × 3 m, with an acoustic sensor network of 12 nodes ($N=12$). Each node contained a pair of microphones 0.5 m apart. The communication diagram of the distributed acoustic sensor network is shown in Figure 2, where the communication radius is 2.5 m, and each circle represents a node. The simulated trajectory 1 was a line from (0.5,0.8) to (2.5, 2.8), and trajectory 2 was an arc from (1, 2) to (4.86, 2.1), as shown in Figure 3. In different experiments, the speech sampled at the frequency of $F_s = 16\mathrm{KHz}$ was used as the acoustic source signal; the speech was a female recording, and the waveform and spectrum of the signal are shown in Figure 4a. The sound speed was $c = 342\mathrm{m/s}$. The microphone signals were simulated with the Image method [33]. Specifically, different RIRS are generated by virtual sound source method to reflect different reverberation time. These RIRSs were convolved with the speech signal and then added to the Gaussian white noise with a determined mean and covariance to produce a received microphone signal with a mixture of reverberation and noise. The different covariance of Gaussian noise determines the different value of the signal-to-noise ratio (SNR), which reflects different environmental noise conditions. The microphone signal was divided into different signal frames along the sound source track, where the frame length of speech signal was $N_f = 512$ and each signal frame was used for state estimation. Taking node 1 as an example, Figure 4b shows

the waveform and spectrum of the speech signal received by the first microphone of node 1. For the observation TDOA, a total of eight time delays were chosen according to the magnitude of the GCC peak. From these delays, further TDOA observations were selected, where the relevant parameters were set as $\lambda = 10$, $\gamma = 4$, $P_G = 0.93$, and $P_D = 0.95$. The standard deviation of TDOA measurement error was $\sigma = 50\mu s$. In the acoustic dynamical model, the parameters were $\beta = 10 s^{-1}$ and $\bar{\upsilon} = 1 ms^{-1}$. In the average consistency calculation of the global state estimation and its error covariance, the Metropolis weight was used, the number of consistency iterations [34] was $N_{con} = 10$, and the number of iterations in the iterative CKF was 3.

This paper conducted four experiments to evaluate the tracking performance of PDA-DCKF. In Experiment 1, trajectory 1 was used as the acoustic source trajectory. The initial prior $p(x_0)$ of the acoustic source position was set as a Gaussian distribution with mean $x_0 = [0.5, 0.8, 0.02, 0.02]^T$ and covariance $P_0 = diag([0.05, 0.05, 0.0025, 0.0025])$. In experiment 2, the sound source signal and track were the same as experiment 1. Using simple average fusion rules, the influence of fusion rules on PDA-DCKF tracking performance was discussed. Experiment 3 discussed the robustness of the algorithm. The acoustic source and trajectory were the same as the previous two experiments. In Experiment 4, trajectory 2 was used as the acoustic source track to check the tracking results of the acoustic source when the track was nonlinear.

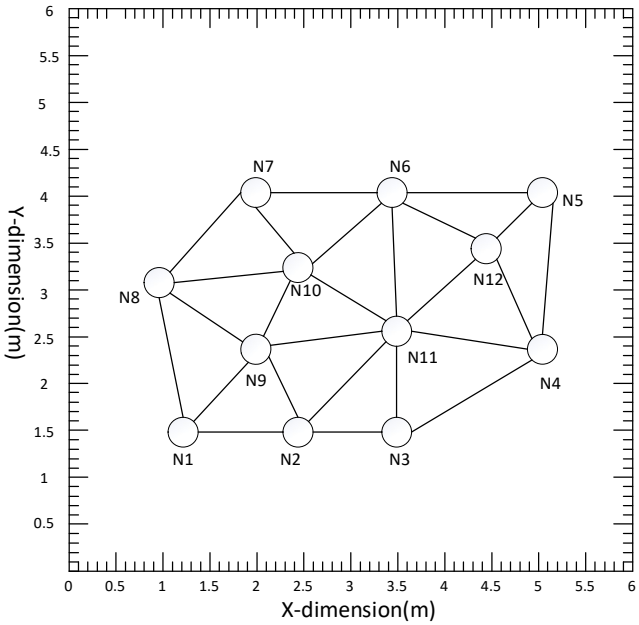


Figure 2. Diagram of a distributed acoustic sensor network with 12 nodes; circles represent nodes in the network. A pair of microphones was placed on each node, and the lines between the nodes indicate that the nodes can communicate with each other.

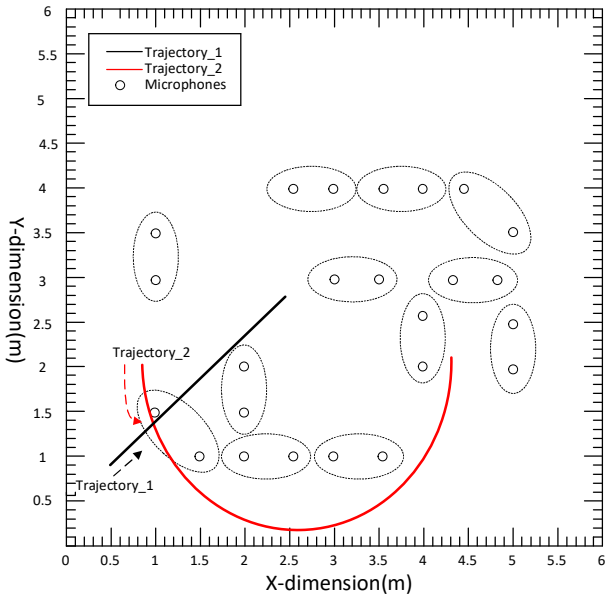
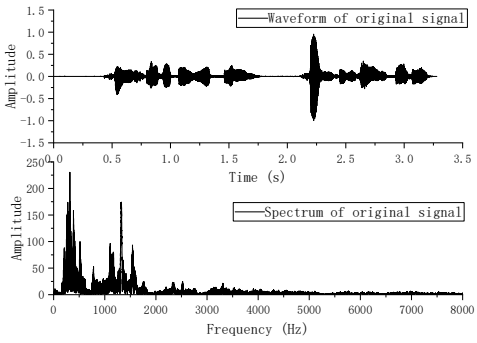
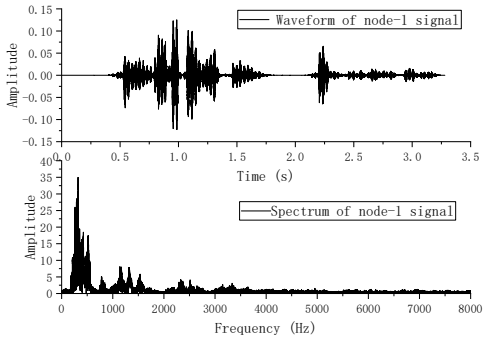


Figure 3. Microphone deployments and acoustic source trajectories: the black line denotes trajectory 1, the black dashed arrow denotes the motion direction of trajectory 1, the red semicircle denotes trajectory 2, the red dashed arrow denotes the motion direction of trajectory 2.



(**a**)



(**b**)

Figure 4. (**a**) The waveform and spectrum of the original speech signal, and (**b**) the waveform and spectrum of node 1 speech signal.

4.2. Simulation Results

4.2.1. Experiment 1

In this experiment, the tracking performance was evaluated under different ambient and reverberant conditions. First, the impact of environmental noise on tracking performance was investigated. Figure 5 depicts the RMSE results as a function of SNR for a reverberation time of $T_{60} = 200\text{ms}$. In Figure 5, it is observed that the RMSE of all methods decreases with the increase of SNR, which means that the tracking accuracy increases with the increase of SNR. This is because when the SNR becomes larger, the microphone signal is less affected by ambient noise, resulting in better tracking performance. In addition, under the same SNR, PDA-DCKF performs better than traditional distributed Kalman filtering, such as extended Kalman filtering, unscented Kalman filtering, and cubature Kalman filtering. Since only one time-delayed observation of the GCC largest peak is used in traditional methods, peaks associated with real sources may be masked by spurious peaks caused by noise or reverberation, resulting in erroneous state estimates. In contrast, multiple time-difference observations of multiple largest peaks of GCC are employed in PDA-DCKF, resulting in ideal tracking performance. At the same time, compared with DICKF in this experiment, the results show that the effect of PDA-DCKF is better than that of DICKF. Because DICKF is aimed at the DCKF method, and DCKF has problems such as slow response speed and low tracking accuracy. However, the tracking performance and convergence speed of the algorithm can be improved through several local iterations in DICKF. However, still only one time-delay observation of the GCC largest peak is used in DICKF, which also causes it to be inaccurate, but as can be seen from Figure 5, as the SNR increases, the gap between DICKF and PDA-DCKF becomes smaller because the observations are more reliable when the SNR becomes larger. In addition, Figure 5 shows that PDA-DCKF is not as good as CCKF because the observation information of all nodes is used in CCKF, but PDA-DCKF achieved an effect very close to the CCKF effect, and its computational cost and the burden of the network is less than that of CCKF.

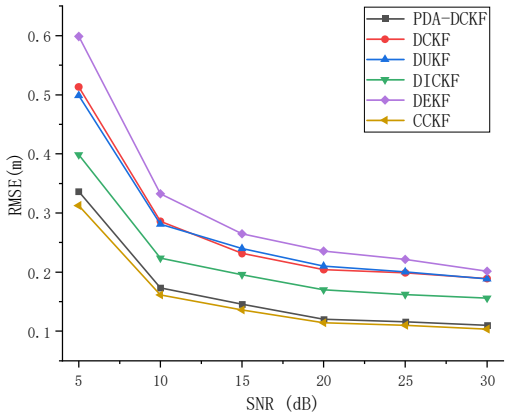


Figure 5. RMSE versus SNR for different tracking algorithms with $T_{60} = 200\text{ms}$.

The effect of reverberation on tracking performance was also studied in this paper. Figure 6 depicts the RMSE results as a function of T_{60} with SNR = 20 dB. From the results, we can observe that the RMSEs of all the methods increased as T_{60} became larger, which signifies the degradation of the tracking accuracies. This may be because the microphone signal is more affected by reverberation as T_{60} becomes larger, the time difference observations extracted from only the largest peak or multiple largest peaks are not reliable,

and the tracking performance of these methods deteriorates. In addition, it can be found from Figure 6 that the tracking performance of PDA-DCKF is better than DEKF, DUKF, DCKF, and DICKF. In fact, in traditional methods, the time-difference observations included in the scheme are only extracted from the largest peak of the GCC, while the peaks associated with the true hypocenter may be masked by false peaks caused by reverberation. In contrast, PDA-DCKF incorporates TDOA observations of multiple largest peaks of GCC into the scheme, which can alleviate the adverse effects of reverberation to a certain extent. Furthermore, the effect is not as good as CCKF showed in Figure 6, but it also achieves a very close effect.

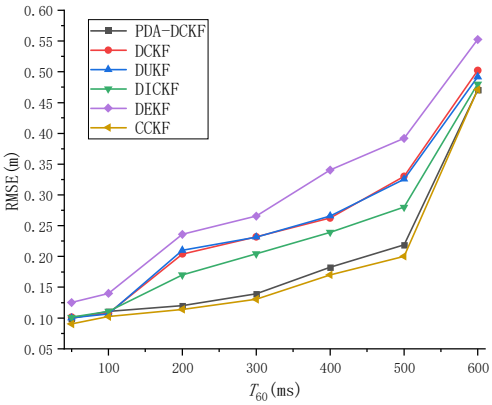


Figure 6. RMSE versus T_{60} for different tracking algorithms with SNR = 20 Db.

4.2.2. Experiment 2

The effect of the fusion strategy proposed in this paper on the results is discussed in Experiment 2. When PDA-DCKF adopts a simple average fusion rule, it is called PDA-DCKF-avg. In this section, different SNR and different reverberations are used to test the effectiveness of the fusion strategy. The experimental results are shown in Figures 7 and 8.

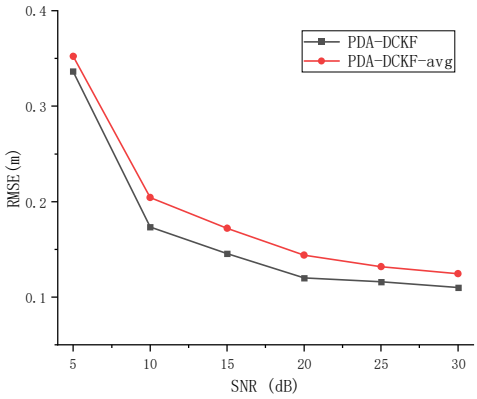


Figure 7. RMSE versus SNR for different fusion rules with $T_{60} = 200\text{ms}$.

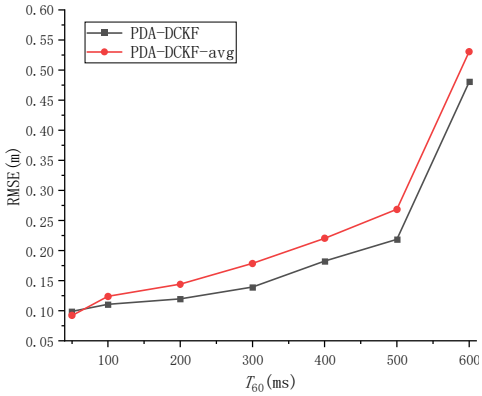
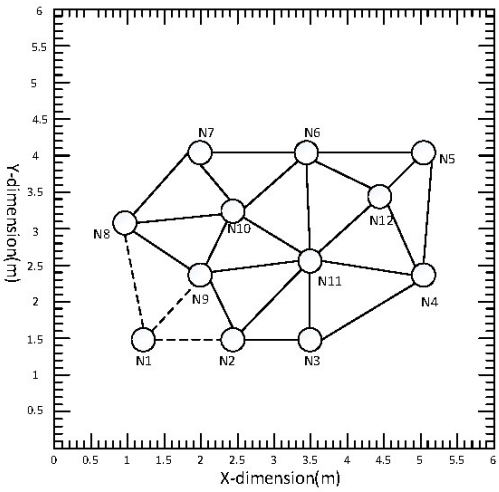


Figure 8. RMSE versus T_{60} for different fusion rules with SNR = 20 dB.

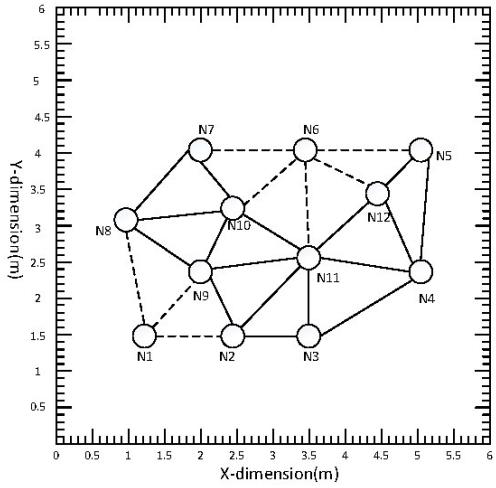
As depicted in Figure 7, with the increase of the SNR, the RMSEs for the PDA-DCKF methods with both these two fusion strategies decrease, but the proposed one is more effective. Figure 8 also shows that, with the increase of the reverberation time, the error also increases. In addition, only under 50 ms reverberation, the error of the average fusion strategy is smaller than that proposed in this paper, and the fusion strategy proposed in this paper was better than the average fusion effect under 100–600 ms. Comparing Figures 5–8, it can be found that, even if the average fusion strategy is used, the PDA-DCKF in this paper is still smaller than the error obtained by the above comparison test, which further proves the effectiveness of the method in this paper.

4.2.3. Experiment 3

In practical applications, a network may be damaged by nodes, and when a node in a network is damaged, whether the network can still work normally will test the robustness of the system. In this subsection, the node damage in the distributed acoustic sensor network is simulated, and the tracking results of the acoustic source after the damage are compared with those before the damage. When node 1 in the network is damaged, it is called graph $\mathcal{G}2$, as shown in Figure 9a. When node 1 and node 6 in the acoustic sensor network are damaged, it is called graph $\mathcal{G}3$, as shown in Figure 9b. The experimental results are shown in Tables 1 and 2.



(**a**) Graph $\mathcal{G}2$



(**b**) Graph $\mathcal{G}3$

Figure 9. (**a**) Indicates that node 1 is broken, (**b**) indicates that node 1 and node 6 are broken. The solid line indicates that the nodes can communicate with each other, and the dotted line indicates that they cannot communicate with each other.

Table 1. RMSE versus SNR under different graphs with $T_{60} = 200\text{ms}$.

SNR(dB)	$\mathcal{G}1$	$\mathcal{G}2$	$\mathcal{G}3$
5	0.3363	0.3521	0.3637
10	0.1735	0.1803	0.2057
15	0.1457	0.1511	0.1786
20	0.1201	0.1284	0.1543
25	0.1161	0.1203	0.1457
30	0.1101	0.1169	0.1376

Table 2. RMSE versus T_{60} under different graphs with SNR = 20 dB.

T_{60}(ms)	$\mathcal{G}1$	$\mathcal{G}2$	$\mathcal{G}3$
50	0.0992	0.1013	0.1164
100	0.1108	0.1195	0.1351
200	0.1201	0.1284	0.1543
300	0.1393	0.1501	0.1754
400	0.1824	0.1903	0.2158
500	0.2187	0.2305	0.2439
600	0.4703	0.4897	0.5062

It can be seen from Tables 1 and 2 that the acoustic source can still be tracked in the case of node damage. Although the accuracy has decreased, the amplitude of the drop is not large and the acoustic source can still be tracked accurately. This can prove that the method proposed in this paper has good robustness under this network.

4.2.4. Experiment 4

In order to further verify the effectiveness of the algorithm in this paper, the semicircle of trajectory 2 was used as the acoustic source trajectory, and comparative experiments were carried out under different SNR and reverberation. The experimental data are shown

in Tables 3 and 4. Figure 10 shows the tracking results with SNR = 15 dB and $T_{60} = 400\text{ms}$.

Table 3. RMSE versus SNR with $T_{60} = 200\text{ms}$.

SNR(dB)	RMSE
5	0.3751
10	0.1869
15	0.1423
20	0.1299
25	0.1214
30	0.1167

Table 4. RMSE versus T_{60} with SNR = 20 dB.

T_{60}(ms)	RMSE
50	0.1174
100	0.1251
200	0.1299
300	0.1322
400	0.1635
500	0.2216
600	0.5027

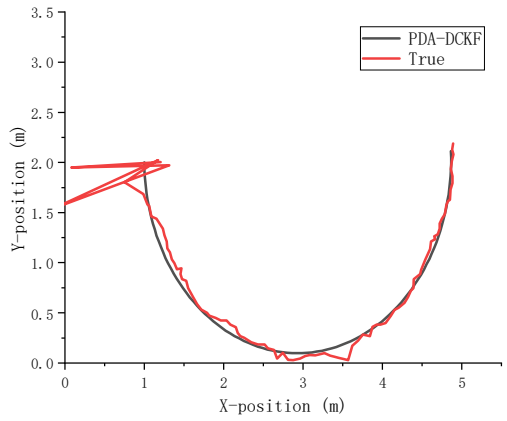


Figure 10. The tracking result of the semicircle trajectory when the SNR = 15 dB and $T_{60} = 400\text{ms}$.

From the above Tables 3 and 4 and Figure 10, it can be seen that the algorithm in this paper can still accurately track the sound source in the face of such a strong nonlinear trajectory.

5. Conclusions

An improved PDA-DCKF method was proposed in this paper, which proved to be able to solve the problem of tracking a single mobile acoustic source with distributed acoustic sensor networks in the noise and reverberation environment. First, in order to reduce the adverse effects of noise and reverberation, the prediction value of observation is obtained by using the prediction state and the observation model of distributed nodes.

Secondly, the actual observations are screened according to the predicted value. Multiple TDOA observations are extracted at each node and incorporated into the status update of CKF through PDA to generate PDA-CKF. PDA-CKF was applied to distributed acoustic sensor networks, and PDA-DCKF was further developed. In PDA-DCKF, the PDA algorithm is first used to sift the observations from neighboring nodes. Then, the sifted observations are fused to update the state vectors in the CKF. Each node runs PDA-DCKF for local state estimation and TDOA observation. Then, a new fusion strategy is proposed using energy and MSE to merge all single local estimates in a distributed manner for global state estimation. In order to apply the improved PDA-DCKF to the acoustic source tracking problem, the Langevin model was used to model the acoustic source dynamics, and a method to extract the time difference observation was proposed. Finally, a distributed acoustic source tracking framework was obtained. In order to evaluate the effectiveness of PDA-DCKF in acoustic source tracking, comparative experiments were carried out with existing methods (DCKF, DUKF, DEKF, and DICKF) under different ambient noise and reverberation conditions. The results show that the PDA-DCKF has better tracking performance than DCKF, DUKF, DEKF, and DICKF under most noise and reverberation conditions. In addition, the PDA-DCKF achieved the same tracking performance as the centralized CKF. Furthermore, it can even track the acoustic source stably in the case of node damage.

Author Contributions: Methodology, R.W.; Software, Y.C. (Yideng Cao); Writing—review & editing, Y.C. (Yang Chen). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Changzhou Science and Technology Funds] grant number [CJ20220100]. And The APC was funded by [Changzhou University].

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

References

1. Wang, L.; Reiss, J.D.; Cavallaro, A. Over-Determined Source Separation and Localization Using Distributed Microphones. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 1573–1588. https://doi.org/10.1109/taslp.2016.2573048.
2. Li, X.; Chen, J.; Qi, W.; Zhou, R. A distributed sound source surveillance system using autonomous vehicle network. In Proceedings of the 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 31 May–2 June 2018; pp. 42–46. https://doi.org/10.1109/iciea.2018.8397686.
3. Kapralos, B.; Jenkin, M.R.M.; Milios, E. Audiovisual localization of multiple speakers in a video teleconferencing setting. *Int. J. Imaging Syst. Technol.* **2003**, *13*, 95–105.
4. Green, T.; Hilkhuysen, G.; Huckvale, M.; Rosen, S.; Brookes, M.; Moore, A.; Naylor, P.; Lightburn, L.; Xue, W. Speech recognition with a hearing-aid processing scheme combining beam-forming with mask-informed speech enhancement. *Trends Hear.* **2022**, *26*, 23312165211068629.
5. Gerstoft, P.; Hu, Y.; Bianco, M.J.; Patil, C.; Alegre, A.; Freund, Y.; Grondin, F. Audio scene monitoring using redundant adhoc microphone array networks. *IEEE Internet Things J.* **2021**, *9*, 4259–4268.
6. Laufer-Goldshtein, B.; Talmon, R.; Gannot, S. A hybrid approach for speaker tracking based on TDOA and data-driven models. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 725–735.
7. Ruiz, S.; Van Waterschoot, T.; Moonen, M. Distributed combined acoustic echo cancellation and noise reduction in wireless acoustic sensor and actuator networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 534–547.
8. Dang, X.; Zhu, H. A feature-based data association method for multiple acoustic source localization in a distributed microphone array. *J. Acoust. Soc. Am.* **2021**, *149*, 612–628.
9. Ziegler, J.; Schröder, L.; Koch, A.; Schilling, A. A Neural Beamforming Front-end for Distributed Microphone Arrays. In *Audio Engineering Society Convention 151*; Audio Engineering Society: New York, NY, USA, 2021.
10. Guo, X.; Yuan, M.; Zheng, C.; Li, X. Distributed node-specific block-diagonal LCMV beamforming in wireless acoustic sensor net-works. *Signal Process.* **2021**, *185*, 108085.
11. Faraji, M.M.; Shouraki, S.B.; Iranmehr, E.; Linares-Barranco, B. Sound Source Localization in Wide-Range Outdoor Environment Using Distributed Sensor Network. *IEEE Sens. J.* **2019**, *20*, 2234–2246.
12. Yang, B.; Yan, G.; Wang, P.; Chan, C.-Y.; Song, X.; Chen, Y. A Novel Graph-Based Trajectory Predictor with Pseudo-Oracle. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–15. https://doi.org/10.1109/tnnls.2021.3084143.
13. Wishner, R.P.; Tabaczynski, J.A.; Athans, M. A comparison of three non-linear filters. *Automatica* **1969**, *5*, 487–496.

14. Nicoletti, O. MDS-IEKF: A Delayed-State Invariant Extended Kalman Filter for Monocular Visual-Inertial Navigation. Ph.D. Thesis, McGill University, Montreal, QC, Canada, 2020.
15. Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Processing* **2002**, *50*, 174–188.
16. Vermaak, J.; Blake, A. Nonlinear filtering for speaker tracking in noisy and reverberant environments. In Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No. 01CH37221), Salt Lake City, UT, USA, 7–11 May 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 5, pp. 3021–3024.
17. Sung, K.; Song, H.J.; Kwon, I.H. A Local Unscented Transform Kalman Filter for Nonlinear Systems. *Mon. Weather. Rev.* **2020**, *148*, 3243–3266.
18. Zhong X, Mohammadi A, Wang W; et al. Acoustic source tracking in a reverberant environment using a pairwise synchronous microphone network. In Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey, 9–12 July 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 953–960.
19. Wang, R.; Chen, Z.; Yin, F. Speaker tracking based on distributed particle filter and iterative covariance intersection in distributed microphone networks. *IEEE J. Sel. Top. Signal Processing* **2019**, *13*, 76–87.
20. Tian, Y.; Chen, Z.; Yin, F. Distributed iterated extended Kalman filter for speaker tracking in microphone array networks. *Appl. Acoust.* **2017**, *118*, 50–57.
21. Tian, Y.; Chen, Z.; Yin, F. Distributed IMM-unscented Kalman filter for speaker tracking in microphone array networks. *IEEE/ACM Trans. Audio Speech Lang. Processing* **2015**, *23*, 1637–1647.
22. Thomas, T.; Sreeja, S. Comparison of Nearest Neighbor and Probabilistic Data Association Filters for Target Tracking in Cluttered Environment. In Proceedings of the 2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA), Arad, Romania, 17–19 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 272–277.
23. Zhang, Q.; Zhang, W.; Feng, J.; Tang, R. Distributed Acoustic Source Tracking in Noisy and Reverberant Environments With Dis-tributed Microphone Networks. *IEEE Access* **2020**, *8*, 9913–9927.
24. Wang, R.; Chen, Z.; Yin, F. Distributed Multiple Speaker Tracking Based on Unscented Particle Filter and Data Association in Microphone Array Networks. *Circuits Syst. Signal Processing* **2022**, *41*, 933–955.
25. Zhang, J.; Gao, S.; Zhong, Y.; Qi, X.; Xia, J.; Yang, J. An advanced cubature information filtering for indoor multiple wideband source tracking with a distributed noise statistics estimator. *IEEE Access* **2019**, *7*, 151851–151866.
26. Woźniak, S.; Kowalczyk, K. Passive joint localization and synchronization of distributed microphone arrays. *IEEE Signal Processing Lett.* **2018**, *26*, 292–296.
27. Knapp, C.; Carter, G. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech Signal Processing* **1976**, *24*, 320–327.
28. Arasaratnam, I.; Haykin, S. Cubature kalman filters. *IEEE Trans. Autom. Control* **2009**, *54*, 1254–1269.
29. Kirubarajan, T.; Bar-Shalom, Y. Probabilistic data association techniques for target tracking in clutter. *Proc. IEEE* **2004**, *92*, 536–557.
30. Bar-Shalom, Y.; Li, X.R. *Multitarget-Multisensor Tracking: Principles and Techniques*; YBs: Storrs, CT, USA, 1995.
31. Souden, M.; Kinoshita, K.; Nakatani, T. An integration of source location cues for speech clustering in distributed microphone arrays. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 111–115.
32. Hodson, T.O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev. Discuss.* **2022**, *15*, 5481–5487.
33. Lehmann, E.A.; Johansson, A.M.; Nordholm, S. Reverberation-time prediction method for room impulse responses simulated with the image-source model. In Proceedings of the 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 21–24 October 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 159–162.
34. Xiao, L.; Boyd, S.; Lall, S. A scheme for robust distributed sensor fusion based on average consensus. In Proceedings of the IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, Boise, ID, USA, 15 April 2005; IEEE: Piscataway, NJ, USA, 2005; pp. 63–70.