Fast-Moving Sound Source Tracking With Relative Doppler Stretch

GUO CHEN AND YONGGANG LU[iD]

Institute of Systems Engineering, China Academy of Engineering Physics, Mianyang 621999, China

Corresponding author: Yonggang Lu (lygcaep@263.net)

ABSTRACT Tracking a fast-moving sound source in the air in an acoustic way has rarely been seen in the current literature. The speed of the source in the air is often not negligible compared to the speed of sound in the air. Time-varying propagation distance and noticeable Doppler effect also bring difficulties to traditional tracking methods with the time difference of arrival (TDOA) and direction of arrival (DOA). In this paper, we propose a particle filtering framework with relative Doppler stretch. We utilize the propagation delayed measurement (PDM) model and correct the posterior probability in traditional particle filter with propagation delayed state being a bridge. This method avoids the complex pre-processing of the raw acoustic signals. The simulation results show that the algorithm has an expected performance and is superior to the existing methods. The whole tracking process which starts from raw signals is shown for the first time, and the key factors affecting the pre-processing of the raw signals that are not mentioned in other articles are also discussed in this paper.

INDEX TERMS Moving sound source, relative Doppler stretch, particle filter, propagation delayed state.

I. INTRODUCTION

For most source tracking scenarios, radar is the most considered. However, for a low-altitude aircraft or a ground vehicle, such as an unmanned aerial vehicle, a helicopter or a tank, the requirements for radar devices will become expensive and complicated. If the source has the anti-radar capability, the performance will significantly reduce. Thus, a passive acoustic tracking method for the moving source is proposed in this paper as a supplement or an alternative to the above scenarios.

Acoustic methods have been widely used in stationary or slow-moving source tracking and localization [1]–[3]. The state-of-the-art deep learning is also used to adapt to the complex acoustic environment [4], [5]. However, when the velocity of a moving source cannot be ignored compared to the speed of sound in the air, the received acoustic signals radiated by the source will be greatly distorted due to Doppler effect. The time difference of arrival (TDOA), which is commonly used to locate a stationary or slow-moving sound source, will not work in such a situation with distorted signals. To extend this method to fast-moving source positioning and tracking, a new method was proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Jingchang Huang[iD].

to compensate for the TDOA error caused by the Doppler effect [6], [7]. In [6], [7], the wideband cross-ambiguity function based on continuous wavelet transform was used to jointly estimate the relative time scale and time delay between two sensors. However, due to the retardation effect, the signals received at the same time instant by different microphones correspond to the signals emitted in different positions for a moving sound source. Thus, observation time intervals for different microphones should be different so that the signals from the same source position can be included in the observation, which brings trouble to the sound source tracking. To this end, Genescà *et al.* [8] proposed a method that the signals from each microphone could be synchronized to compensate for the retardation effect. With this method, the relative Doppler stretches from seven microphones were directly calculated through the one-dimensional wideband cross-ambiguity function with equal observation time interval. Based on the previous work, Martín *et al.* [9], [10] used the Simple Genetic Algorithm (SGA) to search for the position of the source and carried out the simulation study with the takeoff-climb-out data extracted from the database of the FAA's Integrated Noise Model (INM) for a Boeing 737-400. Next, a radio-controlled airplane localization is experimentally tested in [11]. The simulations and experiments showed a good performance.

Similarly, considering that the acoustic signals are distorted in amplitude and frequency, if the Doppler effect can be removed, the signals can be used to estimate TDOA directly. This method was called de-Dopplerization [12]. So far, time domain and frequency domain de-Dopplerization were developed to recover signals, but the velocity should be known accurately before de-Dopplerization, which was difficult in passive tracking.

Besides TDOA, the direction of arrival (DOA) is another measurement to estimate the position of the sound source. In [13], the authors used a planar five-element array to estimate the angle and range of the aircraft in each time interval. In [14], the authors used distributed sensors to locate the vehicular sources with DOA. In [15], the authors proposed a method that could adapt to multi source tracking with DOA batches. In [16], the moving acoustic source was located by intersecting azimuth lines with double arrays. Nevertheless, the DOA of a fast-moving sound source is highly-dynamic. There were a few studies about the highly-dynamic DOA estimation with dynamic compressive sensing [17], [18]. Because dynamic compressive sensing is a difficult time-varying optimization problem, it is difficult to get satisfactory results. In [17], a weighted L_1 minimization algorithm is proposed to improve the performance by utilizing the DOA changing scope as a prior information. However, the preprocessing is still complex and time-consuming in tracking process.

In fact, there has rarely been a satisfactory acoustic tracking frame for a moving source in the air. The tracking frame contains two problems, one is that the speed of a moving source cannot be neglected, the other is that source motion and observation time do not match due to the time-varying distance. Both will cause numerous errors in tracking without compensation. The authors [19]–[21] proposed a propagation delayed measurement particle filter (PDM-PF) to find a solution. They changed the traditional state transition model and measurement model. The time delay and time delayed state were added to form an augmented state vector, then they predicted the current state from the delayed augmented state. The simulation experiment was carried out to show its performance with DOA, yet they assumed that the DOA estimation error was in a small range without building the relationship between the raw acoustic signals and tracking performance. Moreover, obtaining an accurate DOA for a moving source is not easy as mentioned above.

Therefore, this paper proposes an algorithm to overcome the above limitations. We use randomly distributed microphones to collect signals and directly calculate the relative Doppler stretch of the signals received by each pair of microphones without compensating for the retardation effect. Then a general particle filter framework based on the method of propagation delayed state is constructed to track the fast-moving source with relative Doppler stretch.

The structure of the rest of the article is as follows. Section II describes the acoustic characteristics of a moving sound source. Section III introduces the framework of the method proposed in the article. Section IV shows the basic settings of simulation. The results discussions are also included in this section. Finally, conclusions are placed in Section V.

II. ACOUSTIC CHARACTERISTICS OF A MOVING SOUND SOURCE

A. A SOUND FIELD GENERATED BY A MOVING SOUND SOURCE

When the size of the source is much less than the sound-propagation path-length, the size of the source is assumed to be ignored. Thus, the source can be regarded as a monopole sound source radiating noise in a spherical manner in the entire trajectory. In addition, the sound propagation model is considered an ideal propagation model without the effect of temperature.

Considering M randomly distributed microphones, according to [10], the sound pressure signals received by the microphone $m(m = 1, \ldots, M)$ at time t is

$$q_m(t) = \left(\frac{c}{c - \|\boldsymbol{v}\|_2 \times \cos(\theta_m(t - \tau_m))} \right) \times \left(\frac{s(t - \tau_m)}{4\pi r R_m(t - \tau_m)} \right) \quad (1)$$

where $\boldsymbol{v}$ is the velocity of the source, $\|.\|_2$ denotes the vector 2-norm, $s(t)$ is the sound source strength, τ_m is the propagation time of the acoustic signal to the microphone direction vector and the vector of the sound wave propagation direction pointing to the microphone m, R_m represents the sound-propagation path-length to the microphone m, and c is the speed of sound propagating in the air which is $344m/s$ in this article.

Let the distance from the initial position to the source be R_0, the sampled signals are

$$q_m(k\Delta_t) = \left(\frac{c}{c - |\boldsymbol{v}| \times \cos(\theta_m(k\Delta_t - \tau_m))} \right) \times \left(\frac{s(k\Delta_t - \tau_m)}{4\pi R_m(k\Delta_t - \tau_m)} \right), \left(k \geq \left\lceil \frac{R_0}{c\Delta_t} \right\rceil \right) \quad (2)$$

where k represents the number of samples, Δt is the sampling interval, and $\lceil \bullet \rceil$ denotes rounding up.

B. RELATIVE DOPPLER STRETCH (RDS)

As shown in Figure 1, the sampling interval and motion interval do not match due to the variation of the path-length R_m. In a short time period, T, the sound radiation characteristics and velocity $\boldsymbol{v}$ of the moving source can be assumed constant. From time t to time $t + T$, if the displacement of the moving source is much less than the sound-propagation path-length, i.e., $\|\boldsymbol{v}\|_2 \times T \ll R_m$, the angle θ_m can be considered constant.

Then, due to Doppler effect, the frequency f_m^d of the signals received by the microphone m satisfies

$$f_m^d = f \times \left(\frac{c}{c - \|\boldsymbol{v}\|_2 \times cos\theta_m} \right) \quad (3)$$
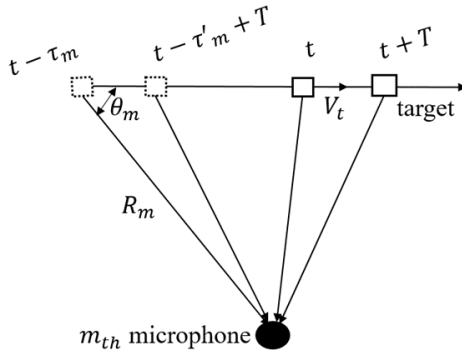
FIGURE 1. Time relations of the sound position at emission time and receiving time.

where f represents the frequency of the sound source. And θ_m is obtained as

$$cos\theta_m = \frac{v_{t-\tau_m}^T \left(r_m - P_{t-\tau_m}\right)}{\left\|v_{t-\tau_m}\right\|_2 \times \left\|r_m - P_{t-\tau_m}\right\|_2} \quad (4)$$

where r_m represents the position of the microphone m, $v_{t-\tau_m}$ and $P_{t-\tau_m}$ denotes the velocity and position of the source respectively at time $t - \tau_m$, and τ_m satisfies

$$\tau_m = \frac{1}{c} \times \left\|P_{t-\tau_m} - r_m\right\|_2 \quad (5)$$

Let $d_m = \frac{c}{c-\|v\|_2 \times cos\theta_m}$ represents the Doppler stretch, then the relationship between the source signals and the received signals can be written as

$$q_m \left(f_m^d\right) = A_m \times s \left(d_m f\right) e^{j2\pi f_m^d \tau_m} \quad (6)$$

where $q_m(f)$ and $s(f)$ is the Fourier transform of $q_m(t)$ and $s(t)$ respectively, A_m is the amplitude coefficient which is assumed constant in a short time interval. Thus, for a microphone pair (m, n) constructed by any two microphones, we can obtain

$$q_n \left(f\right) = A_{m,n} q_m \left(d_{m,n} f\right) e^{j2\phi(m,n)} \quad (7)$$

where $d_{m,n} = \frac{d_m}{d_n}$, and $\phi(m, n)$ represents the phase difference.

Then, according to [7], for the microphone pair (m, n), the cross-ambiguity function of the signals is defined as

$$X_p \left(\delta\right) = \sqrt{\delta} \times \int q_m \left(f\right) \times q_n \left(\delta f\right) e^{j\phi(m,n)} df \quad (8)$$

where δ is the relative frequency scale, p ($p = 1, \ldots, \frac{M(M-1)}{2}$) refers to the pth pair of microphones formed by microphone m and n.

As mentioned in [11], the ambiguity function will not be time domain restrictive if the signals have been synchronized. If without synchronization, we directly use two moduli $|q_m(f)|$ and $|q_n(\delta f)|$ ($\left|e^{j\phi(m,n)}\right| = 1$) and Equation (8) becomes:

$$X_p \left(\delta\right) = \sqrt{\delta} \times \int |q_m \left(f\right)| \times |q_n \left(\delta f\right)| df \quad (9)$$

The maximum value of $X_p(\delta)$ will correspond to $\delta = \frac{d_m}{d_n}$ which is also called relative Doppler stretch (RDS).

To calculate the value of $X_p(\delta)$, a 1-dimensional version of the discrete form is used.

Let

$$X_p \left(\delta\right) = \frac{\sqrt{\delta}}{L_N + 1} \sum_{l=0}^{L_N} |q_m \left(l\Delta f\right)| \times |q_n \left(\delta \times l\Delta f\right)| \quad (10)$$

where L_N is the available number of frequency bins satisfying Nyquist sampling theorem, then L_N is obtained as

$$L_N = \begin{cases} \left\lfloor \dfrac{f_s}{2\Delta f} \right\rfloor, & \delta < 1 \\ \left\lfloor \dfrac{f_s}{2\Delta f \times \delta} \right\rfloor, & \delta \geq 1 \end{cases} \quad (11)$$

where f_s is the sampling frequency, Δf is frequency resolution, and $\lfloor \bullet \rfloor$ denotes rounding down. For the discrete version, resampling of the frequency axis is needed, which is calculated by linear interpolation.

Additionally, with Equation (4), the relative Doppler stretch can also be written as

$$\delta = \frac{d_m}{d_n} = \frac{c - \frac{v_{t-\tau_n}^T (r_n - p_{t-\tau_n})}{\|r_n - p_{t-\tau_n}\|_2}}{c - \frac{v_{t-\tau_m}^T (r_m - p_{t-\tau_m})}{\|r_m - p_{t-\tau_m}\|_2}} \quad (12)$$

So far, we have only focused on subsonic sources. Assuming that the maximum speed of the low-altitude aircraft is v_{max}, we can define the search range of δ is $\left[\frac{c-v_{max}}{c+v_{max}}, \frac{c+v_{max}}{c-v_{max}}\right]$ in Equation (9).

For a 2D motion source, eight scalar unknowns appear in Equation (12) including the position and velocity of the source at two different time $t - \tau_m$ and $t - \tau_n$. Nevertheless, if we know its motion model, there are only four scalar unknowns. Thus, at least four microphones are required to form six microphone pairs to obtain more than four relative Doppler stretches. Besides, from Equation (12), the relative Doppler stretch obtained from Equation (9) is corresponding to two source states at two different moments if two received signals have not been synchronized. Thus, when using relative Doppler stretches, a correction of the tracking model is required.

C. PROPAGATION DELAYED MEASUREMENT (PDM) AND PROPAGATION DELAYED STATE (PDS)

For a tracking problem, the following state space model is considered

$$x_{T_k} = f_{T_{k-1},T_k} \left(x_{T_{k-1}}\right) + w_{T_{k-1},T_k} \quad (13)$$

where $x_{T_k} = \left[P_{T_k}^x, V_{T_k}^x, P_{T_k}^y, V_{T_k}^y\right]^T$ is the state vector, f_{T_{k-1},T_k} represents the state transition function, w_{T_{k-1},T_k} is Gaussian white noise satisfying $w_{T_{k-1},T_k} \sim N(w_{T_{k-1},T_k}; 0; S_{T_{k-1},T_k})$, S_{T_{k-1},T_k} is the covariance of the process

noise, w_{T_{k-1},T_k}. We assume that the transformation in Equation (13) is invertible which satisfies

$$x_{T_{k-1}} = f_{T_k,T_{k-1}}\left(x_{T_k}\right) + w_{T_k,T_{k-1}} \tag{14}$$

We consider a general measurement model

$$y_{T_k} = g(x_{T_k}) + n_{T_k} \tag{15}$$

where n_{T_k} is noise and g represents the function from state to measurement. According to section II.B, for the pth pair of microphones, when the relative Doppler stretch, δ_{p,T_k}, is used as the measurement, this model needs to be corrected as

$$\delta_{p,T_k} = g\left(x_{T_k-\tau_m}, x_{T_k-\tau_n}\right) + n_{T_k} \tag{16}$$

where δ_{p,T_k} is also called propagation delayed measurement (PDM), similarly, we define

$$x^D_{T_k} = \left[x_{T_k-\tau_m}, x_{T_k-\tau_n}\right]^{\mathrm{T}} \tag{17}$$

where $x^D_{T_k}$ is the propagation delayed state (PDS). The PDS is constructed by two different delayed states from the current state x_{T_k}. Therefore, the PDS establishes the relation between the current state and the PDM like a bridge.

With Equation (5) and (14), the problem of obtaining PDS is shown as

$$\begin{cases} x_{T_k-\tau_m} = f_{T_k,T_k-\tau_m}\left(x_{T_k}\right) + w_{T_k,T_k-\tau_m} \\ \tau_m - \dfrac{1}{c}\left\|\boldsymbol{P}_{T-\boldsymbol{\tau_m}} - \boldsymbol{r_m}\right\|_2 = 0 \end{cases} \tag{18}$$

where $\boldsymbol{P}_{T-\boldsymbol{\tau_m}}$ is composed of the position parameter in the state vector $x_{T_k-\tau_m}$. This kind of numeric root-finding problem can be solved with the Newton-Raphson method when the model is known in Equation (14).

III. PROPAGATION DELAYED STATE PARTICLE FILTER WITH RELATIVE DOPPLER STRETCH (PDS-PF-RDS)

As can be seen above, the tracking model is non-linear and non-Gaussian, thus, the particle filter is used here to provide the optimal solution.

The particle filtering algorithm is a form of recursive Bayesian filter. It is a non-linear filtering algorithm based on the Bayesian formula to realize state predictions and state updates. The main idea is to use a group of random samples (particles) to describe the posterior probability distribution, adjust the weight of each sample through the measurement model to approximate the actual probability distribution, and use the weighted average of the samples as an estimate of the current state, which is applicable for arbitrary non-linear, non-Gaussian systems.

A. SAMPLING IMPORTANCE RESAMPLING (SIR) PARTICLE FILTER

For a typical particle filter, for example [22], the sampling importance resampling (SIR), the sampling framework is expressed as follows:

1) Initialization. Sampling N particles $\{x^{(i)}_{T_k}, i = 1, 2, \ldots, N\}$ from the initial prior probability with uniform weight.

2) Prediction and updating. Predicts the new set of particles $\{x^{(i)}_{T_k}\}$ according to state transition model, the weight $\{w^{(i)}_{T_k}\}$ corresponding to $\{x^{(i)}_{T_k}\}$ is updated as

$$w^{(i)}_{T_k} = P\left(y_{T_k} \mid x^{(i)}_{T_k}\right) \tag{19}$$

3) Estimation. The current state estimate can be obtained as

$$\overline{X}_{T_k} = \sum_i^N \tilde{w}^{(i)}_{T_k} \times x^{(i)}_{T_k} \tag{20}$$

where $\tilde{w}^{(i)}_{T_k}$ is the normalized weight, i.e., $\tilde{w}^{(i)}_{T_k} = \dfrac{w^{(i)}_{T_k}}{\sum_i^N w^{(i)}_{T_k}}$.

B. PROPAGATION DELAYED STATE PARTICLE FILTER

In our case, when relative Doppler stretches are used, the measurements can be shown as

$$y_{T_k} = \left[\delta_1, \ldots, \delta_p, \ldots, \delta_{\frac{M(M-1)}{2}}\right]^{\mathrm{T}} \tag{21}$$

where δ_p represents the relative Doppler stretch of the pth pair of microphones. The relative Doppler stretches are considered to be independent of each other. Based on the SIR particle filtering framework, the weight $\{w^{(i)}_{T_k}\}$ of the particle i should be

$$w^{(i)}_{T_k} = \prod_{p=1}^{\frac{M(M-1)}{2}} P\left(\delta_p \mid x^{(i)}_{T_k}\right) \tag{22}$$

However, as mentioned above, the relative Doppler stretch δ_p is a propagation delayed measurement which is corresponding to the PDS $x^D_{T_k}$. Therefore, with PDS being a bridge, the posterior probability can be corrected as

$$P\left(\delta_p \mid x^{(i)}_{T_k}\right) = \int P\left(\delta_p \mid x^D_{T_k}\right) \times P\left(x^D_{T_k} \mid x^{(i)}_{T_k}\right) dx^D_{T_k} \tag{23}$$

where

$$P\left(x^D_{T_k} \mid x^{(i)}_{T_k}\right) \sim N\left(x^D_{T_k}; x^{D(i)}_{T_k}; S^{(i)}_{x_{T_k},x^D_{T_k}}\right) \tag{24}$$

$$x^{D(i)}_{T_k} = \left[x^{(i)}_{T_k-\tau_m}, x^{(i)}_{T_k-\tau_n}\right]^{\mathrm{T}} \tag{25}$$

$$S^{(i)}_{x_{T_k},x^D_{T_k}} = \begin{bmatrix} S^{(i)}_{T_k,T_k-\tau_m} & \mathbf{0} \\ \mathbf{0} & S^{(i)}_{T_k,T_k-\tau_n} \end{bmatrix} \tag{26}$$

Therefore, $P\left(\delta_p \mid x^{(i)}_{T_k}\right)$ is rewritten as

$$P\left(\delta_p \mid x^{(i)}_{T_k}\right) = \int P\left(\delta_p \mid x^D_{T_k}\right) \times N\left(x^D_{T_k}; x^{D(i)}_{T_k}; S^{(i)}_{x_{T_k},x^D_{T_k}}\right) dx^D_{T_k} \tag{27}$$

The non-linear transformation of probability can be solved by unscented transform(UT) [23]. UT can generate a set of sigma points $\left\{x^{D(i)}_{T_k}[j], weight\left(x^{D(i)}_{T_k}[j]\right)\right\}_j^J$ to represent the

probability distribution $N\left(x_{T_k}^D; x_{T_k}^{D(i)}; S_{x_{T_k}, x_{T_k}^D}^{(i)}\right)$. According to UT, the estimation of $p\left(\delta_p \mid x_{T_k}^{(i)}\right)$ is

$$P\left(\delta_p \mid x_{T_k}^{(i)}\right) \approx \sum_{j=1}^{J=2c+1} weight\left(x_{T_k}^{D(i)}[j]\right) \times P\left(\delta_p \mid x_{T_k}^{D(i)}[j]\right) \tag{28}$$

where c is the dimension of the state $x_{T_k}^{D(i)}$. It is difficult to calculate the complex probability $P\left(\delta_p \mid x_{T_k}^{D(i)}[j]\right)$ directly. We should use another likelihood function to describe it. The likelihood function need show the peak corresponds to likely PDS (i.e., a larger likelihood function value will be treated as a more likely PDS than a smaller value). If we use the Gaussian likelihood function $\left(\delta_p; \bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right); \sigma_p^2\right)$, the estimation of δ_p will be a time-consuming searching process due to the small searching step and repeated process of resampling in Equation (9) and σ_p can only be determined through iterating. Thus, a pseudo-likelihood function is proposed to describe the probability distribution $P\left(\delta_p \mid x_{T_k}^{D(i)}[j]\right)$. The function is

$$p\left(\delta_p \mid x_{T_k}^{D(i)}[j]\right) = \left\{X_p^D\left(\bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right)\right)\right\}^r \tag{29}$$

$$X_p^D\left(\bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right)\right) = \int |q_m(f)| \times \left|q_n\left(\bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right)\right)\right| df \tag{30}$$

where $p(x_{T_k}^{D(i)}[j])$ represents the relative Doppler stretch formed by the two delayed states in $x_{T_k}^{D(i)}[j]$, the exponential weight r is used to strengthen the main lobe peak and reduce the side lobe peak so that the likelihood function is closer to the true value. The closer to the true relative Doppler stretch (RDS), the greater value of the likelihood function and the greater weight. The pseudo-likelihood function does not require a search to find the peaks but imposes a weight on the possible state proportionally.

One cycle of the PDS-PF-RDS is depicted in Algorithm I based on the SIR particle filter.

IV. SIMULATION
This section describes the computer simulation of tracking a 2D moving sound source with four microphones, including basic parameter settings, simulation results and discussions.

A. IMPLEMENTATION
Four microphones are used to verify the tracking ability of the method in the 2D scene. For comparison, we choose the same simulation scene used in [19]–[21]. The distribution of these microphones and the source trajectory are shown in Figure 2, and their position coordinates are shown in TABLE 1. The size of the source is ignored. The source moves at a radius of $500m$ at a constant turn rate $w = -0.12 rad/s$. The initial position is $[-500m, 800m]^T$.

Algorithm 1 PDS-PF-RDS

1. Sampling N particles from initial probability distribution $P(x_{T_o})$ with uniform weight to form the particle set $\left\{x_{T_0}^{(i)}, w^{(i)}\right\}_i^N$ at time t_0.
2. For $T_1 : T_k$
   1) Sampling $\left\{x_{T_k}^{(i)}\right\}_i^N$ from $P(x_{T_k} \mid x_{T_{k-1}})$ to update the state.
   2) For the pth pair of microphones, obtain the PDS $x_{T_k}^{D(i)} = \left[x_{T_k - \tau_m}^{(i)}, x_{T_k - \tau_n}^{(i)}\right]^T$ corresponding to $x_{T_k}^{(i)}$, propagation time $\tau_m^{(i)}$ corresponding to the PDS and the covariance $S_{x_{T_k}, x_{T_k}^D}^{(i)}$.
   3) With UT transform, get the sigma points $\left\{x_{T_k}^{D(i)}[j], weight\left(x_{T_k}^{D(i)}[j]\right)\right\}_j^J$ and calculate the relative Doppler stretch $\bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right)$.
   4) Update $w^{(i)}$.
      - Calculate $X_p^D\left(\bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right)\right)$, then $P\left(\delta_p \mid x_{T_k}^{(i)}\right)$ is obtained as
        $$P\left(\delta_p \mid x_{T_k}^{(i)}\right) = \sum_{j=1}^{J=2c+1} weight\left(x_{T_k}^{D(i)}[j]\right) \times \left\{X_p^D\left(\bar{\delta}_p\left(x_{T_k}^{D(i)}[j]\right)\right)\right\}^r$$
      - Update the weight as
        $$w^{(i)} = \prod_{p=1}^{M(M-1)/2} P\left(\delta_p \mid x_{T_k}^{(i)}\right)$$
3. Normalize $w^{(i)}$ and get
   $$w^{(i)} = \frac{w^{(i)}}{\sum_i^N w^{(i)}}$$
4. Get the current estimation of source state:
   $$\bar{x}_{T_k} = \sum_i^N w^{(i)} \times x_{T_k}^{(i)}$$
5. Store the particles set $\left\{x_{T_k}^{(i)}, w^{(i)}\right\}_i^N$
6. End for

TABLE 1. Positions of the microphones with respect to the origin of coordinates (Location distance in meters).

m	1	2	3	4
x	0	-750	-600	200
y	0	750	200	600

1) SIGNAL PARAMETERS
The acoustic signals radiated by some moving sources, such as unmanned aerial vehicles or ground vehicles, are often dominated by a few harmonics [24], [25]. In this paper, the
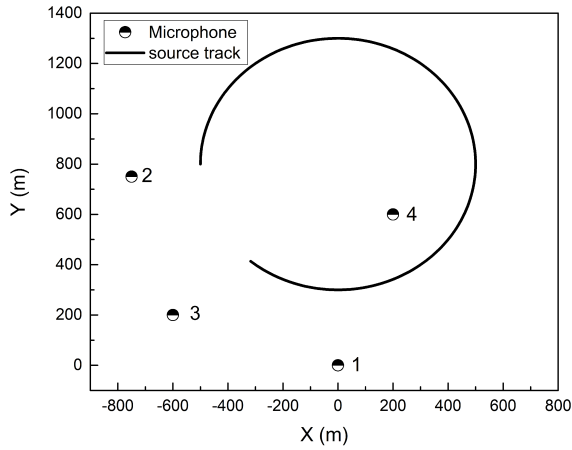
FIGURE 2. The source trajectory and microphones distribution.

single-frequency signal is superimposed to simulate an acoustic signal with a band [1000Hz−1500Hz]. This method is also used in [26], [27] for a moving sound source simulation. The sampling frequency of the microphone is 10KHz, the noise attached to the received signals is Gaussian white noise, and the signal-to-noise ratio is $SNR = 20$dB.

The signal frequency spectrums of the four microphones from $t = 4s$ to $t = 5s$ are shown in Figure 3. Figure 3 illustrates that the frequency spectrums of different microphones at different positions differ due to the different Doppler stretches and the unequal distance traveled.
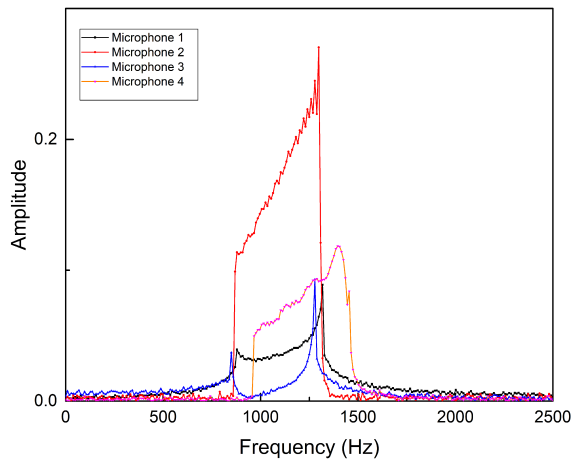
FIGURE 3. The frequency spectrums of the acoustic signals received by four microphones.

2) MODEL PARAMETERS

The source motion is modeled with a discretized coordinated turn model with an unknown constant turn rate and with Cartesian velocity. The state variable is $x_{T_k} = \left[P^x_{T_k}, V^x_{T_k}, P^y_{T_k}, V^y_{T_k}, w_{T_k} \right]^T$ and the discretized form of the model is given as

$$X_{T_{k+1}} = FX_{T_k} + w_{T_k, T_{k+1}} \qquad (31)$$

where F is state transition matrix, $w_{T_k, T_{k+1}}$ is the process noise with zero mean and covariance S_k, Let $T = T_{k+1} - T_k$, then F and S_k can be obtained as

$$F = \begin{bmatrix} 1 & \dfrac{\sin(wT)}{w} & 0 & \dfrac{1 - \cos(wT)}{-w} & 0 \\ 0 & \cos(wT) & 0 & \cos(wT) & 0 \\ 0 & \dfrac{1 - \cos(wT)}{w} & 1 & \dfrac{\sin(wT)}{w} & 0 \\ 0 & \sin(wT) & 0 & \cos(wT) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$S_k = \begin{bmatrix} \dfrac{T^3}{3}\sigma_x^2 & \dfrac{T^2}{2}\sigma_x^2 & 0 & 0 & 0 \\ \dfrac{T^2}{2}\sigma_x^2 & T\sigma_x^2 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{T^3}{3}\sigma_y^2 & \dfrac{T^2}{2}\sigma_y^2 & 0 \\ 0 & 0 & \dfrac{T^2}{2}\sigma_y^2 & T\sigma_y^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_w^2 T \end{bmatrix}$$

where σ_x, σ_y, represent the standard deviations of the velocity in x, y directions, σ_w is the standard deviations of the turn rate. we assume $\sigma_x = \sigma_y = 1m/s^2$, $\sigma_w = 0.01 rad/s^2$.

The configuration of the simulation parameter is as follows. The number of particles is $N = 1000$. We assume that the initial particle set is generated from the Gaussian distribution $N((x^{(i)}; \mu; diag([100^2, 10^2, 100^2, 10^2, 0.05^2]))$, and μ is generated from the Gaussian distribution $N((\mu; x_{t_0}; diag([100^2, 10^2, 100^2, 10^2, 0.05^2]))$ which represents the errors of obtaining the initial position. Considering the propagation time of the signals at the first moment, we start tracking from $t_0 = 4s$ to $t_{end} = 45s$ with the time interval $T = 1s$. A total number of 1000 Monte Carlo runs were made.

B. TRACKING PERFORMANCE METRICS

The Root Mean Square Error (RMSE) of position and velocity of the source are used as a reference to evaluate the performance of the method, which is defined as

$$RMSE = \frac{1}{Mc} \sum_{m_c=1}^{Mc} \sqrt{\left(\frac{1}{K}\right) \sum_{k=1}^{K} \left\| \bar{l}_{m,k} - l_{m,k} \right\|^2} \qquad (32)$$

where k is the estimation step, $\bar{l}_{m,k}$ and $l_{m,k}$ refers to the estimated value and real value respectively, and mc refers to the cycle index of the Monte Carlo simulation running.
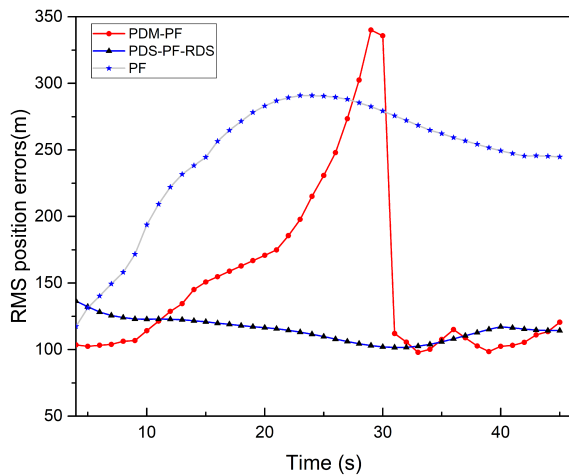
C. RESULT DISCUSSIONS

To assess the performance of the proposed method, PDS-PF, a comparative experiment with the existing moving sound source tracking methods with DOA is conducted, which is the PDM-PF [19]. In addition, the direct PF regardless of the propagation time delay is also conducted. The differences of the three methods are summarized in TABLE 2.

The RMS position and velocity errors are given in Figure 4 and Figure 5 respectively. The performance of the
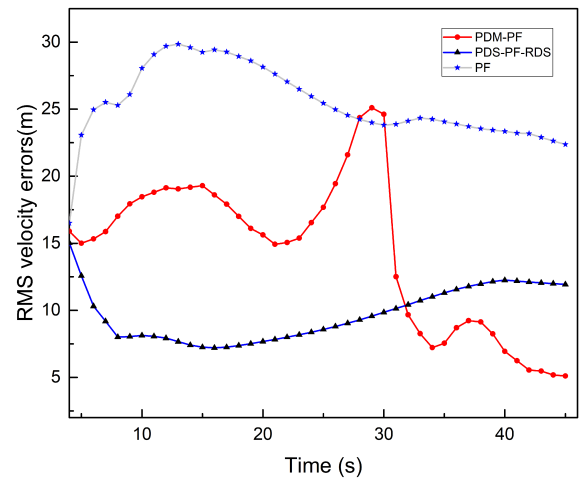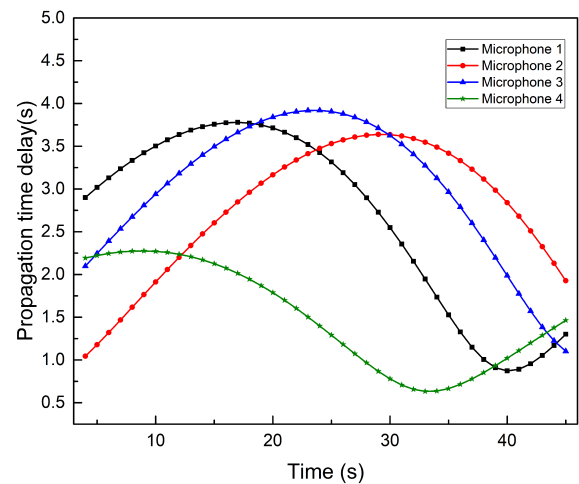
TABLE 2. Differences of three methods.

	PDM-PF	PF	PDS-PF-RDS
Ignore the delay	No	Yes	No
Measurement	DOA	Relative Doppler stretch	Relative Doppler stretch
Microphone array	One moving array	Randomly distributed microphones	Randomly distributed microphones

FIGURE 4. RMS-position errors of different methods.

FIGURE 5. RMS-velocity errors of different methods.

FIGURE 6. Time evolution of the propagation time delay of each microphone.

PDS-PF-RDS algorithm turns out to be superior to PDM-PF in terms of the position error and velocity estimation, the error of the PDS-PF-RDS can quickly reduce to a small value faster than other algorithms. And the error of the PF method increases fast, although it decreases after 25 seconds, it is still much bigger than the other two methods. In the whole tracking process, the performance of the PDS-PF-RDS is much more stable than PDM-PF and PF. The error of the PDM-PF increases after 10*s* and is even bigger than the error of the PF around 28*s* although it can reduce to a small value quickly after 30*s*.

In order to explain the performance differences of different algorithms, we show the true propagation delay in Figure 6. According to Equation (15) and (16), when the delay is neglected in PF, the current source states are not correspond-ing to the measurement, the relative Doppler stretch (RDS). When the delay is large, the bias between the current states and propagation delayed states (PDS) is not negligible. Thus, though the error also decreases when the delay is becoming smaller after 35*s*, the accumulation of the errors before 35*s* is too large to get a satisfactory performance with the PF method.

In [19], it is reported that when the delays change quite fast, the PDM-PF performance becomes inaccurate, although

the estimates can quickly recover after this period ends. In fact, the stability is affected by the change speed of the propagation delay. When the propagation delay changes fast, the corresponding propagation delayed state (PDS) also changes quickly. In order to keep up the fast changes, we need a smaller time interval to estimate the PDS, but our state transition model changes at a fixed time interval. However, in PDS-PF-RDS, the performance is always stable because randomly distributed microphones can reduce the influence of delay changes caused by the spatial position of a single microphone. Moreover, from Equation (12), relative Doppler stretch contains both the velocity information and position information of the source compared to DOA which is only related to the source position. Thus, when the relative Doppler stretch (RDS) is used as the measurement, the estimation of the posterior probability can be more accurate.

Next, we will carefully discuss some key impact parame-ters which influence the tracking performance of the propaga-tion delayed state particle filter with relative Doppler stretch.

1) EFFECT OF THE TIME INTERVAL

In section II.A, we assume that the frequency spectrum is constant in a short time interval. However, the time interval is not as short as possible. On one hand, a shorter time interval means more tracking steps which requires more computation time in the same tracking process, on the other hand, a shorter time interval results in lower frequency resolution. Therefore, we compare the relative Doppler stretch obtained by theory and measurement for different time intervals to make a compromise. Figure 7 illustrates the average errors of the relative Doppler stretch estimation with different time intervals.
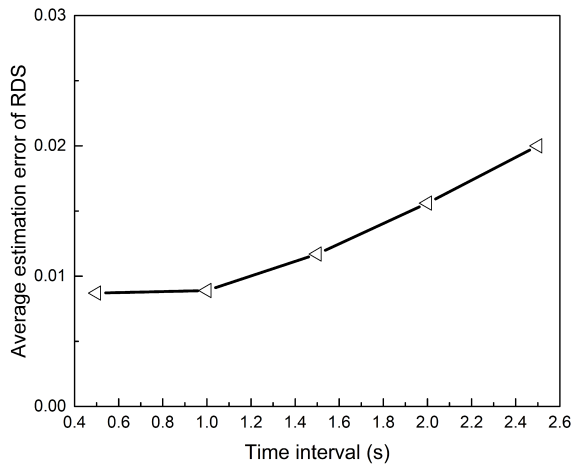
FIGURE 7. Average errors of the relative Doppler stretch with different time intervals.

From Figure 7, we can observe that the error is very small in a short time interval less than about $1s$ but will rise much faster with a larger interval. Thus, we choose the time interval is $T = 1s$ in our case. The real value and estimation of relative Doppler stretch is shown in Figure 8.

From Figure 8, the estimated relative Doppler stretch has errors but is similar to the true value. It means that Equation (9) and (10) can work well in obtaining the relative Doppler stretch with $T = 1s$. In [19], the reason for the interval chosen or the method for DOA estimation is not discussed, but it is worth studying to keep a balance between accuracy and efficiency as mentioned above.

2) EFFECT OF SIGNAL-TO-NOISE RATIO (SNR)

To evaluate the impact of the noise on the tracking performance, corresponding simulation experiments of different SNRs are carried out.

We discuss the performance of Equation (9) under different SNRs and give the limitations. The signals are from the microphone pair $(1, 2)$ from $t = 4s$ to $t = 5s$, and Figure 9 shows the results.

From Figure 9, we can observe that the Ambiguity function, $X(\delta)$, has a good performance when $SNR > 15dB$. However, the side lobe will be higher than the main lobe when $SNR < 10dB$. According to Figure 9, the pseudo-likelihood function, $X(\delta_p)$, in Equation (29) can only describe the
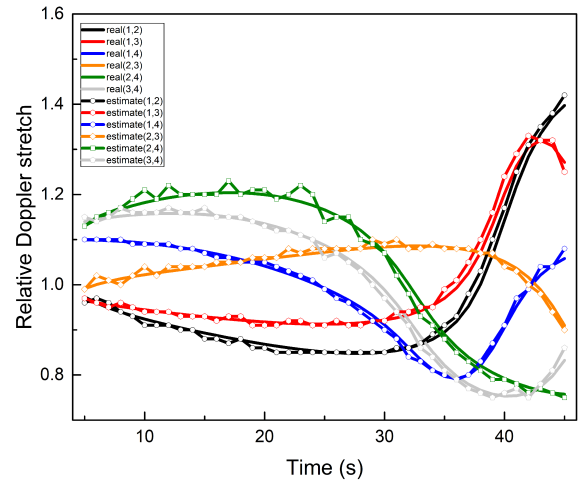

FIGURE 8. Real value and estimated value of the relative Doppler stretches (real(*m,n*) and estimate(*m,n*) represent the real and the estimated Doppler stretches respectively of the microphone pair(*m,n*)).
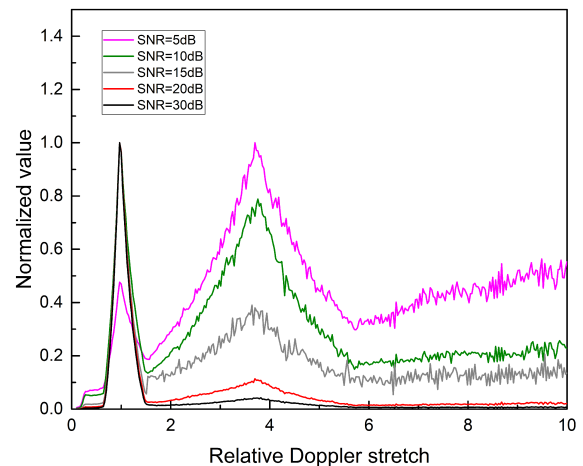

FIGURE 9. The value of $X(\delta)$ (Equation (9)) under different SNRs.

probability distribution $P\left(\delta_p \mid x_{T_k}^{D(i)}[j]\right)$ when $SNR > 15dB$. Though the exponential weight r can be used to strengthen the main lobe peak, there is still a wrong peak in the likelihood function under low SNR environment. When the particles weights update in our tracking frame, the particles related to the PDS corresponding to the wrong peak will get very big weights. As a result, the error between the true state and the weighted average of the particles will become large and the number of effective particles will decrease.

V. CONCLUSION

In this paper, we propose a propagation delayed state particle filter framework of acoustic tracking for a moving sound source with relative Doppler stretch (PDS-PF-RDS).

We avoid using complex microphone arrays but randomly distributed microphones to reduce the hardware difficulty in real-world application. We obtain the relative Doppler stretches of each pair of sensors directly by Equation (9) to form the observation vector. Compared to obtaining DOA or TDOA of a moving source, this method does not need

intricate preprocessing of the raw signals, such as using de-Dopplerization [12] or dynamic compressive sensing [17]. At the same time, we correct the observation model and construct the propagation delayed state (PDS) to be a bridge to get the corrected posterior probability. Moreover, we use the value $X(\delta)$ of the ambiguity function to approximate the weight of each particle instead of calculating the complex posterior probability directly. The simulation results show that the proposed tracking framework has smaller RMS position and velocity errors and can adapt to the changes of the propagation delay in the whole tracking process compared to other algorithms. In addition, we show the whole process of tracking a moving sound source with raw data received by sensors instead of ignoring the preprocessing of the signals.

As a limitation, the current method is only discussed in the subsonic scenes. When tracking a supersonic source, the received signal will contain both leading and lagging signals, and the measurement will no longer be reliable. However, it can still work well in tracking most subsonic sound sources and play an essential role in the multi-source information monitoring system. We will carry out field experiments in the next step, and the research on experiments will be presented in future work.

REFERENCES

[1] S. Gannot, M. Haardt, W. Kellermann, and P. Willett, "Introduction to the issue on acoustic source localization and tracking in dynamic real-life scenes," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 3–7, Mar. 2019, doi: 10.1109/jstsp.2019.2906923.

[2] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 826–836, Nov. 2003, doi: 10.1109/TSA.2003.818112.

[3] R. Wang, Z. Chen, and F. Yin, "Speaker tracking based on distributed particle filter and iterative covariance intersection in distributed microphone networks," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 76–87, Mar. 2019.

[4] J. Vera-Diaz, D. Pizarro, and J. Macias-Guarasa, "Towards end-to-end acoustic localization using deep learning: From audio signals to source position coordinates," *Sensors*, vol. 18, no. 10, p. 3418, Oct. 2018.

[5] M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, "Machine learning in acoustics: Theory and applications," *J. Acoust. Soc. Amer.*, vol. 146, no. 5, pp. 3590–3628, 2019.

[6] B. G. Ferguson, "Time-delay estimation techniques applied to the acoustic detection of jet aircraft transits," *J. Acoust. Soc. Amer.*, vol. 106, no. 1, pp. 255–264, Jul. 1999, doi: 10.1121/1.427054.

[7] B. G. Ferguson and K. W. Lo, "Passive wideband cross correlation with differential Doppler compensation using the continuous wavelet transform," *J. Acoust. Soc. Amer.*, vol. 106, no. 6, pp. 3434–3444, Dec. 1999, doi: 10.1121/1.428197.

[8] M. Genescà, J. Romeu, T. Pàmies, and A. Balastegui, "Passive acoustic method for tracking moving sound sources," *Acta Acustica United With Acustica*, vol. 97, no. 1, pp. 34–43, Jan. 2011.

[9] S. R. Martin, M. Genesca, J. Romeu, and R. Arcos, "Passive acoustic method for aircraft states estimation based on the Doppler effect," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 1330–1346, Apr. 2014.

[10] S. R. Martín, M. Genescà, J. Romeu, and T. Pàmies, "Aircraft tracking by means of the acoustical Doppler effect," *Aerosp. Sci. Technol.*, vol. 28, no. 1, pp. 305–314, Jul. 2013.

[11] S. R. Martín, M. Genescà, J. Romeu, and A. Clot, "Aircraft localization using a passive acoustic method. Experimental test," *Aerosp. Sci. Technol.*, vol. 48, pp. 246–253, Jan. 2016.

[12] H. Camargo and R. Burdisso, "A frequency domain technique to de-dopplerize the acoustic signal from a moving source of sound," in *Proc. 17th AIAA/CEAS Aeroacoustics Conf., 32nd AIAA Aeroacoustics Conf.*, Jun. 2011, p. 2768.

[13] H. Chen and J. Zhao, "On locating low altitude moving targets using a planar acoustic sensor array," *Appl. Acoust.*, vol. 64, no. 11, pp. 1087–1101, 2003, doi: 10.1016/S0003-682X(03)00073-2.

[14] M. R. Azimi-Sadjadi, S. K. Srinivasan, and S. Ahmadinia, "Acoustic localization of vehicular sources using distributed sensors," *J. Acoust. Soc. Amer.*, vol. 146, no. 6, pp. 4913–4925, Dec. 2019, doi: 10.1121/1.5138934.

[15] V. Cevher, R. Velmurugan, and J. H. McClellan, "Acoustic multitarget tracking using direction-of-arrival batches," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2810–2825, Jun. 2007, doi: 10.1109/TSP.2007.893962.

[16] J. Yin, C. Xiong, and W. Wang, "Acoustic localization for a moving source based on cross array azimuth," *Appl. Sci.*, vol. 8, no. 8, p. 1281, Aug. 2018, doi: 10.3390/app8081281.

[17] W. Wang and R. Wu, "High-dynamic doa estimation based on weighted L_1 minimization," *Prog. Electromagn. Res. C*, vol. 42, pp. 253–265, Aug. 2013, doi: 10.2528/PIERC13061410.

[18] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013, doi: 10.1109/TSP.2013.2273196.

[19] U. Orguner and F. Gustafsson, "Target tracking with particle filters under signal propagation delays," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2485–2495, Jun. 2011.

[20] U. Orguner and F. Gustafsson, "Target tracking using delayed measurements with implicit constraints," in *Proc. Int. Conf. Inf. Fusion*, 2008, pp. 1–8.

[21] U. Orguner and F. Gustafsson, "Distributed target tracking with propagation delayed measurements," in *Proc. Int. Conf. Inf. Fusion*, 2009, pp. 802–809.

[22] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002, doi: 10.1109/78.978374.

[23] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000, doi: 10.1109/9.847726.

[24] J. Tong, W. Xie, Y.-H. Hu, M. Bao, X. Li, and W. He, "Estimation of low-altitude moving target trajectory using single acoustic array," *J. Acoust. Soc. Amer.*, vol. 139, no. 4, pp. 1848–1858, Apr. 2016, doi: 10.1121/1.4944567.

[25] B. G. Ferguson and R. L. Culver, "Acoustics today," *Int. Student Challenge Problem Acoust. Signal Process.*, vol. 10, no. 2, pp. 26–29, 2014.

[26] F. Meng, Y. Li, B. Masiero, and M. Vorländer, "Signal reconstruction of fast moving sound sources using compressive beamforming," *Appl. Acoust.*, vol. 150, pp. 236–245, Jul. 2019, doi: 10.1016/j.apacoust.2019.02.012.

[27] F. Meng, G. Behler, and M. Vorländer, "A synthesis model for a moving sound source based on beamforming," *Acta Acustica United With Acustica*, vol. 104, no. 2, pp. 351–362, Mar. 2018, doi: 10.3813/AAA.919177.
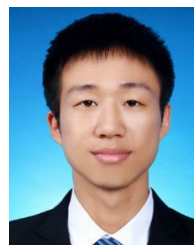
GUO CHEN received the B.Eng. degree in mechanical engineering from the Ha'erbin Institute of Technology, in 2013. He is currently pursuing the Ph.D. degree in acoustics engineering with the China Academy of Engineering Physics. His current research interests include acoustic signal processing and acoustic localization.

YONGGANG LU received the B.Eng. degree in engineering mechanics from Zhejiang University and the Ph.D. degree from the Beijing Institute of Technology.

He is currently a Professor and a Chief Engineer with the Institute of Systems Engineering, China Academy of Engineering Physics. He has presided over a number of national defense scientific research projects. His research interest includes systems engineering. He received the Military Progress Prize in Science and Technology for many times.

• • •