RESEARCH   **Open Access**

Classification, positioning, and tracking of drones by HMM using acoustic circular microphone array beamforming

Junfeng Guo, Ishtiaq Ahmad and KyungHi Chang[*]

Abstract

This paper addresses issues with monitoring systems that identify and track illegal drones. The development of drone technologies promotes the widespread commercial application of drones. However, the ability of a drone to carry explosives and other destructive materials may pose serious threats to public safety. In order to reduce these threats, we propose an acoustic-based scheme for positioning and tracking of illegal drones. Our proposed scheme has three main focal points. First, we scan the sky with switched beamforming to find sound sources and record the sounds using a microphone array; second, we perform classification with a hidden Markov model (HMM) in order to know whether the sound is a drone or something else. Finally, if the sound source is a drone, we use its recorded sound as a reference signal for tracking based on adaptive beamforming. Simulations are conducted under both ideal conditions (without background noise and interference sounds) and non-ideal conditions (with background noise and interference sounds), and we evaluate the performance when tracking illegal drones.

Keywords: Beamforming, Scanning, Classification, Tracking, Drone

1 Introduction

In recent years, the development of drones has received considerable attention due to their diverse applications. This accomplishes a reduction in drone manufacturing costs [1]. The advancements in drone technology have an established record for providing beneficial eye-in-the-sky services, but they have also increased serious apprehensions with respect to privacy and safety [2], such as the threat of chemical, biological, or nuclear attacks [3].

In order to eliminate threats by illegal drones, many authorities have been striving to achieve a solution in drone monitoring and drone-attack countermeasures. In [2], a system to combat unmanned aerial vehicles (UAVs) was designed based on wireless technology; it can realize detection, recognition, and jamming of UAVs. In [4], the concept of the low-altitude air surveillance control (LASC) system was presented. Moreover, technology based on the microphone array for sound-source positioning has been widely used in different scenarios [5, 6]. In [7], beamforming with a

circular microphone array was employed to localize environmental sources of noise from different directions. Zhang et al. [8] used a microphone array and acoustic beamforming to capture superior speech sounds and to localize the speakers in distributed meetings. Gebbie et al. [9] utilized a microphone array for small-boat localization.

In this paper, we design a monitoring system based on capturing acoustic sound signals to identify illegal drones. For detection, we use microphone arrays that do not depend on the size of the drone, but rather on the sound of the propeller, and can therefore serve as an effective means of detection and recognition, determining whether it is drone or not, and which can then track the drone. For detection and classification, the first step is feature extraction [10] in order to identify the components of the acoustic signal. Differences in system methodologies in the literature results in difficult to compare the proposed strategy with the other researches. In the literature, there are several techniques based on acoustic data for feature extraction, such as harmonic line association [2, 11], the wavelet transform [12], and the mel-frequency cepstral coefficient (MFCC) [13] method. The

* Correspondence: khchang@inha.ac.kr
Electronic Engineering Department, Inha University, Incheon, Korea

second step is classification, and for this, many mathematical models can be used, such as the support vector machine (SVM) [14], the Gaussian mixture model [15], and the hidden Markov model (HMM) [16]. The procedure for direction of arrival (DOA) estimation with drones is composed of beam-scan algorithms and subspace algorithms [17]. The beam-scan algorithms form a conventional beam, scan the appropriate region, and plot the magnitude squared output. Thus, minimum variance distortionless response (MVDR) [18], and root MVDR [19] are examples. Moreover, subspace algorithms comprise a set wherein the orthogonality between the signal and noise subspaces is exploited [17]. Thus, multiple signal classification (MUSIC), root-MUSIC, capon, and estimation of signal parameters via rotational invariance technique (ESPRIT) [20] are most efficient for estimating the DOA of the signals using array antennas. However, we use the recursive least squares (RLS) algorithm [21] based on minimum mean square error (MMSE) [22] criteria for estimating the DOA of drones. The RLS algorithm is a kind of non-blind adaptive algorithm, and it requires a reference signal [23] to find the target location. Kaleem and Rehmani presented schemes for drone localization and tracking [24]. Therefore, it is very difficult to compare the proposed acoustic-based scheme for positioning and tracking of illegal drones strategy with the other researches. Unlike the resource-allocation and interference-mitigation schemes [25–48], this paper addresses the positioning of drones with an HMM for classification and with beamforming for tracking, using an acoustic circular microphone array.

1.1 Main contributions
Our proposed framework is based on three major steps.

– First, we use microphones in a uniform circular array (UCA) to form a beam pattern to scan the sky and find sound sources.
– Second, we use the HMM for classification in order to recognize the sound source, and determine whether it is an illegal drone or something else.
– Finally, if it is an illegal drone, then we record its sound with the array's microphone elements (MEs) and use the recorded sound as a reference signal for tracking, based on RLS beamforming.
– Simulations are conducted under both ideal conditions (without background noise and interference sounds) and non-ideal conditions (with background noise and interference sounds), and we evaluate the performance when tracking illegal drones.

The rest of the paper is organized as follow. Section 2 provides the details of the system architecture that includes topological structure of a circular microphone array and array signal model. In Section 3, we describe the details of proposed acoustic signal-based procedure for drone positioning. The experiments and simulations are conducted in Section 4. Finally, we conclude the paper in Section 5.

2 System methodology for acoustic signal-based positioning of illegal drones
In this paper, we detect illegal drones based on sound recognition with an HMM for classification and with beamforming for tracking using a circular microphone array. We consider 32 MEs ($m = 1, 2, 3, ..., 32$) for sensing sounds. Figure 1 shows the system architecture for illegal-drone identification using acoustic signals, in which MEs are distributed uniformly on the circle, using an angle of $\phi_m = \frac{2(m-1)\pi}{12}$ between the MEs, and the radius of the circular array is about 0.18 m. $x_m(n)$ is the signal sample received by the mth element of the array, where n is the time index. The sampling rate of the ME is 44 KHz in the data acquisition process. The direction of arrival of an object in the air is calculated by azimuth and elevation angles. Thus, azimuth is on the x-y plane with respect to the x-axis and is denoted as ϕ, and elevation with respect to the z-axis is represented as θ.

At first, switched beamforming (SBF) is used for scanning the objects in the sky. Scanning is executed from 0° to 90° elevation and 0° to 360° azimuth. Indeed, the SBF is supposed to scan for illegal drones, but other objects can also be in the sky, such as birds and airplanes. Generally, airplanes fly at a very high altitude, so birds can be the main interference source while scanning for targets in the air. Our scenario considers not only an illegal drone (the target sound signal) but a bird (an interference sound signal), as shown in Fig. 1. Moreover, a uniform circular array (UCA) can provide 360° in azimuthal coverage and can estimate both azimuth and elevation simultaneously.

2.1 Circular microphone array method details
In this paper, we utilize a circular microphone array with a ring pattern for scanning the 3-D area, because it has uniform resolution throughout the entire azimuthal dimension, and it also provides the best performance when the exact position of the source is unknown [49]. There are usually six to 36 MEs used in a UCA for acoustic beamforming. In this paper, we consider 32 MEs because that number gives good-enough scanning accuracy and has the least complexity in our scenario. Figure 2 shows the orientation of the circular microphone array in which the 32 MEs are uniformly placed. The x-, y-, and z-axes represent the coordinates of the
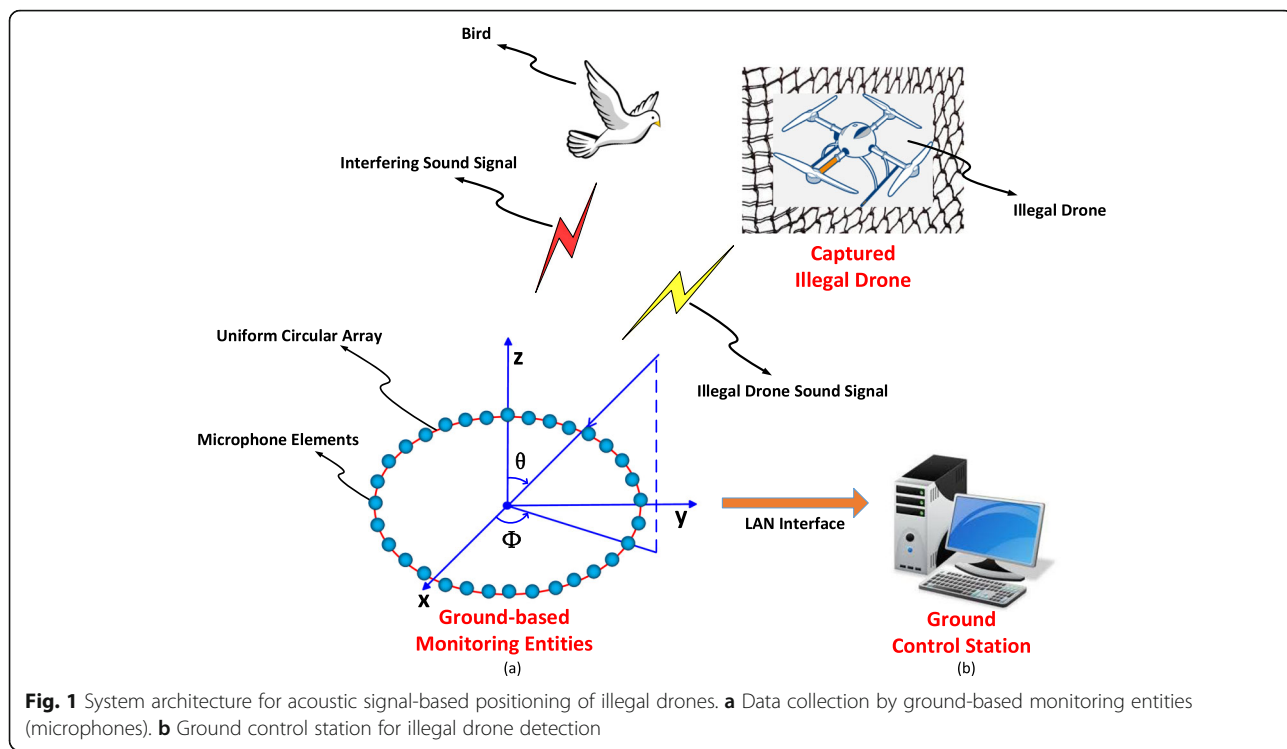
Fig. 1 System architecture for acoustic signal-based positioning of illegal drones. **a** Data collection by ground-based monitoring entities (microphones). **b** Ground control station for illegal drone detection

beamforming array in which the x-axis and y-axis denote the horizontal plane, and the z-axis indicates the height.

2.2 Detail of array signal model

Consider a signal source with angle (θ, ϕ) that impinges upon the MEs in an UCA, and let $F(\theta, \phi)$ denote the array factor. Each ME is weighted with a complex weight, $W(m)$ for $m = 0, 1, 2, 3, ..., M - 1$.
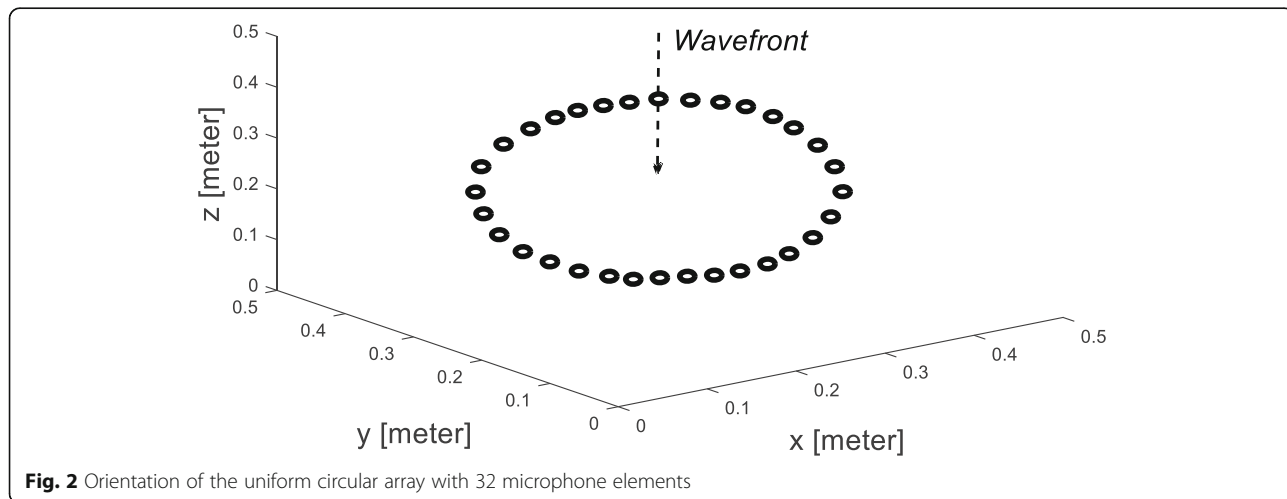
Since the M MEs are equally spaced around the UCA, with radius R, the azimuth angle ϕ_k of the mth ME is given as:

$$\phi_m = \frac{2m\pi}{M} \qquad (1)$$

The phase between the MEs is given as follows:

$$\beta_m = -aR \cos(\phi - \phi_m) \sin\theta \qquad (2)$$

where $a = \frac{2\pi}{\lambda}$ and λ = wavelength.

It follows that the array factor for a UCA with M equally spaced MEs is given as:

Fig. 2 Orientation of the uniform circular array with 32 microphone elements

$$F(\theta,\phi) = \sum_{m=0}^{M-1} A_m e^{j[\alpha_m - aR\cos(\phi-\phi_m)\sin\theta]} \qquad (3)$$

where A_m is the amplitude of the impinged signal at the mth ME, and hence, $A_m e^{j\alpha_m}$ represents the complex weight for the mth ME. In order to direct the main beam at angle (θ_0, ϕ_0) in space, the phase of the weight for the mth ME can be selected as:

$$\alpha_m = aR\cos(\phi-\phi_m)\sin\theta \qquad (4)$$

3 Proposed acoustic signal-based methodology for drone positioning

To track illegal drones based on sound recognition, in our proposed framework, we first scan the objects (sound sources) in the sky via SBF. Then, we use the HMM for classification to identify the sound source and determine whether it is an illegal drone or something else. Finally, if it is an illegal drone, its sound is recorded by the MEs, and this recorded sound is used as a reference signal for tracking, based on adaptive RLS beamforming.

Scanning with SBF is based on maximum output power criteria, and the scanning range for elevation is from 0° to 90°, and for azimuth, from 0° to 360°. When SBF completes the scan, the system detects the presence of the sound sources. It might be a plane or a bird, etc., or an illegal drone, and thus, in order to identify the sound source, the well-known HMM technique is employed. If the HMM classifier identifies the sound as an illegal drone, then adaptive beamforming requires reference signals, because RLS is non-blind beamforming. Hence, reference signal acquisition is based on the scanning and classification processes. Moreover, even if there are other interfering sound signals, we can still track the target by using the reference signal and updating the DOA estimation according to the target movements. Figure 3 shows the overall procedure of the proposed

acoustic signal-based scheme for illegal-drone positioning. The following subsections explain the scanning, classification, and tracking procedures in detail.

3.1 Method of scanning the objects in the sky
3.1.1 Details of switched beamforming based on maximum power criteria
The scanning or acquisition is done based on SBF. Sound source localization is achieved based on maximum output power. In the SBF scheme, the weight vector is given to the MEs in order to change the direction of the beam and scan the corresponding grid. Thus, output power is calculated in each grid to find the maximum value. This process is repeated until all the areas are scanned, and then, we compare the output power of each scanning result. The grid that gives the maximum output power indicates the location of the sound source in 3-D space, i.e., the peak of the beam coincides with the direction of the object. The output signal of the beamformer is given as:

$$\begin{aligned} y(n,\theta,\phi) &= \sum_{m=1}^{M} w_m^*(\theta,\phi)x_m(n) \\ &= W^H(\theta,\phi)X(n) \\ \text{where } & W(\theta,\phi) = [w_1(\theta,\phi), w_2(\theta,\phi), w_3(\theta,\phi)\cdots w_M(\theta,\phi)] \\ & X(n) = [x_1(n), x_2(n), x_3(n), \cdots, x_M(n)] \end{aligned} \qquad (5)$$

where $W(\theta,\phi)$ is a weight vector, and $X(n)$ is the output of the mth element at the nth snapshot.

The output power of each scanned area is calculated as:

$$\begin{aligned} P(\theta,\phi) &= E(|y(n,\theta,\phi)|^2) \\ &= W^H(\theta,\phi)E(X(n)X^H(n))W(\theta,\phi) \\ &= W^H(\theta,\phi)R(n)W(\theta,\phi) \end{aligned} \qquad (6)$$

where W^H represents the complex conjugate of the weight vector and $E(.)$ denotes the expectation. $R(n)$, the
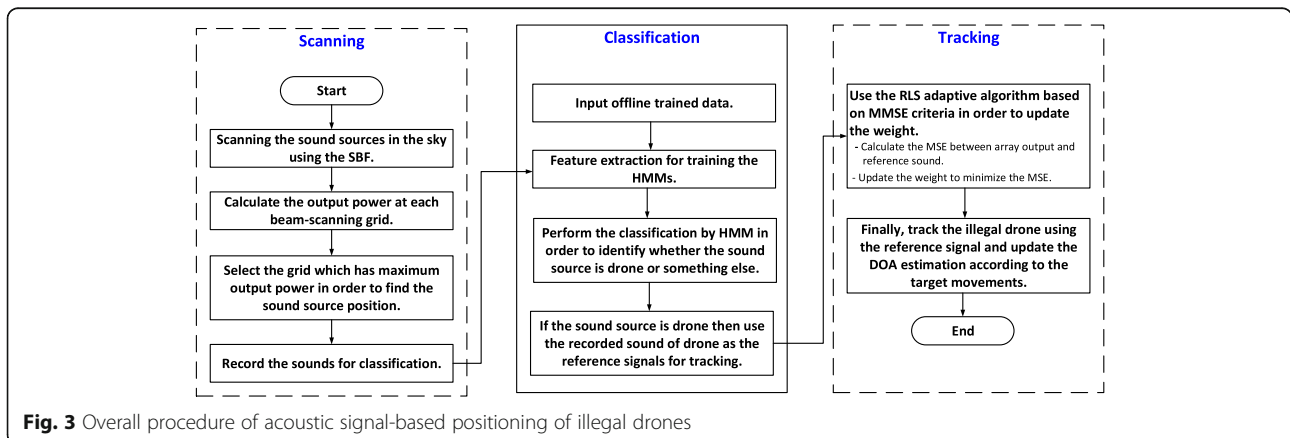

Fig. 3 Overall procedure of acoustic signal-based positioning of illegal drones

covariance matrix of the signal in the nth snapshot, is given as follows:

$$R(n) = X(n) \cdot X^H(n) \qquad (7)$$

Hence, we can calculate the output power of the beamformer according to (6) by changing θ and ϕ in order to find the maximum power, which initially identifies the target position.

3.1.2 Details of scanning accuracy by switched beamforming

In this subsection, we discuss the scanning accuracy of SBF. The simulations are performed under both ideal conditions (without background noise and interference) and non-ideal conditions (with background noise and interference). Figure 4 shows the beam-scanning route, which starts from 0° elevation and 0° azimuth. At first, beam scanning follows the route by increasing the elevation angle and azimuth angle with resolutions of 5° and 15°, respectively. The arrows indicate the movement of the beams scanning for sound sources in 3-D space, calculating the output power in each grid on the route. Finally, we compare the output power of each grid and select the grid that has the maximum output to determine whether there is a sound source or not.

We consider a single sound source (target) and calculate the output power following the beam-scanning route. Figure 5 shows the output signal in each beam-scanning grid according to the elevation and azimuth from SBF. From Fig. 5c, we can clearly see that the flying object is observed in that grid with elevation and azimuth of 45° and 180°, respectively, because the power of

the output signal is larger than the pre-defined threshold. Indeed, we also check the output power in other grids, but the output power of the signal is low, as shown in Fig. 5a, b, d. Moreover, Fig. 6 illustrates the different beam patterns for the flying-object directions based on varying the number of array elements. Indeed, the beam-scanning performance over the whole area (i.e., azimuth from 0° to 360° and elevation from 0° to 90°) depends on the radius of the UCA and the number of MEs. In this paper, we use a radius of 0.18 m and find the number of MEs that provides good-enough beam-scanning performance. From Fig. 6a, we can clearly see that 12 MEs are not enough, because the yellow region of the maximum peak (which indicates the location of the sound source) is almost equal to the peak power of the other grids. On the other hand, Fig 6b still does not provide good-enough scanning performance with 24 MEs, because the peak power of the yellow region is not high enough, compared to the peak power of the other grids in whole area.

However, Fig. 6c has the best scanning performance with 32 MEs, because the peak power of the yellow region is high enough, compared to the background. Thus, we select 32 MEs at a 0.18 m radius to scan for sound sources in the sky.

In addition, in order to verify the selection of 32 MEs with a 0.18 m radius, we check the peak to average power ratio (PAPR) versus the number of MEs at different radii of a circular array, such as 0.18 m, 0.2 m, 0.22 m, and 0.25 m, as seen in Fig. 7. PAPR is defined in (8), and thus, the maximum value of PAPR above the threshold indicates the direction of the flying object:
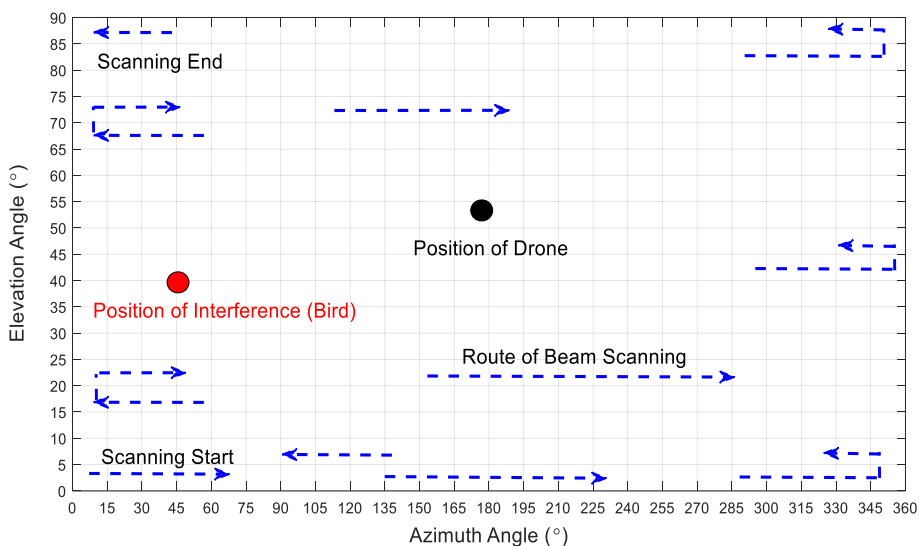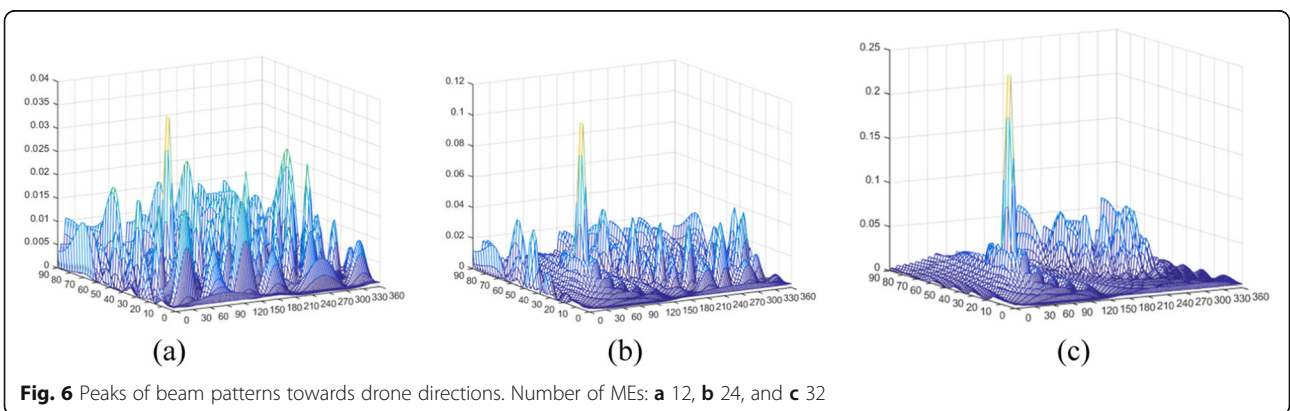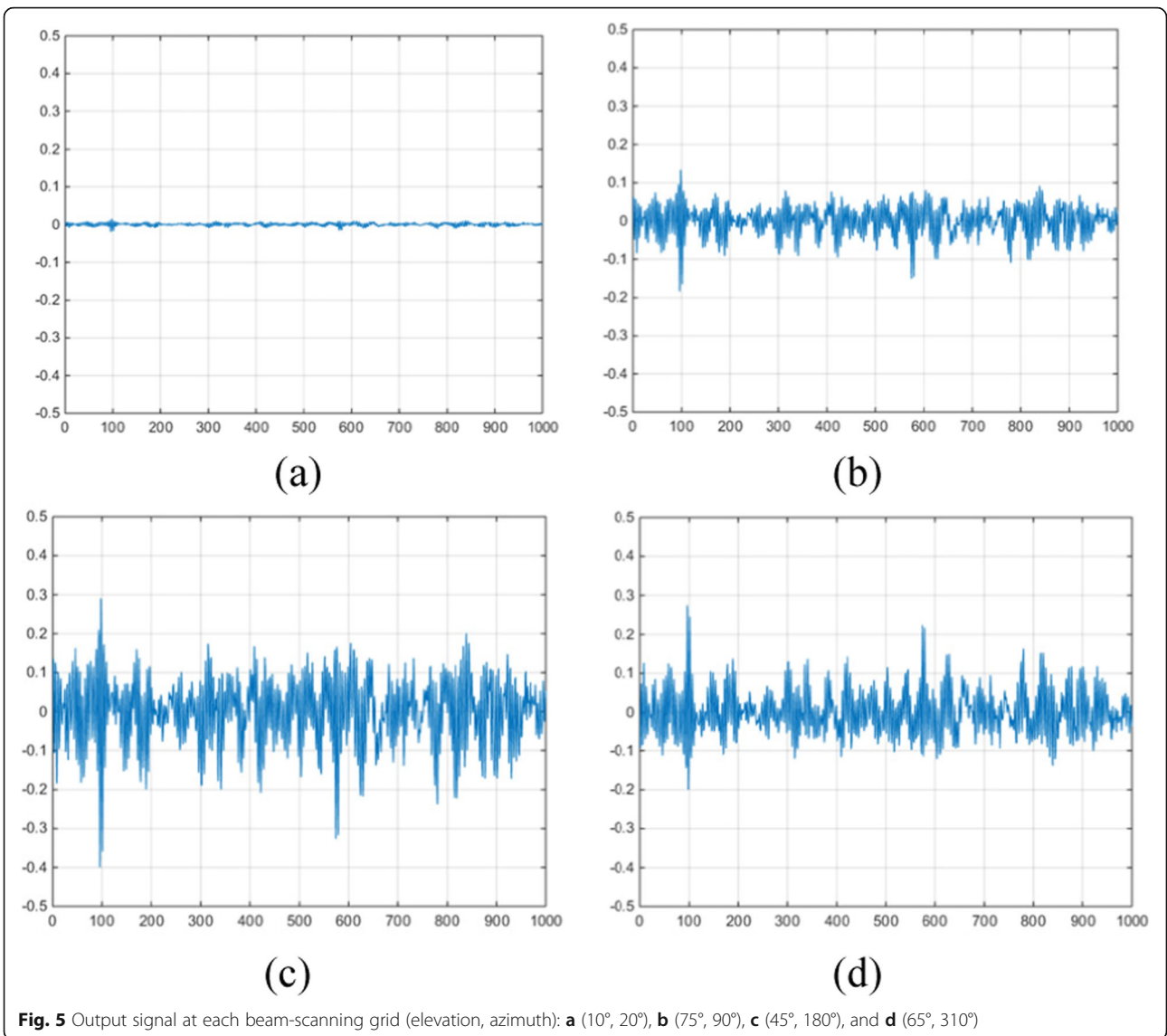

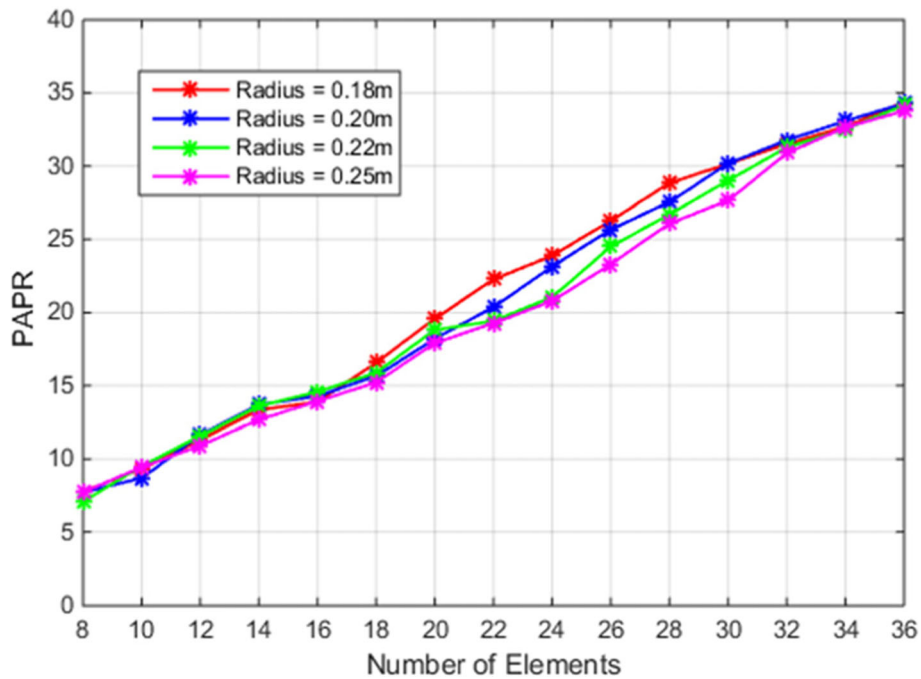**Fig. 4** Scanning resolution and the routes of sound source scanning by SBF

Fig. 5 Output signal at each beam-scanning grid (elevation, azimuth): **a** (10°, 20°), **b** (75°, 90°), **c** (45°, 180°), and **d** (65°, 310°)

Fig. 6 Peaks of beam patterns towards drone directions. Number of MEs: **a** 12, **b** 24, and **c** 32

Fig. 7 Peak to average power ratio versus different numbers of elements

$$\begin{aligned} &Peak\ to\ Average\ Power\ Ratio \\ &= \frac{Power\,(Peak)}{\sum\limits_{All\,Grids} Power\,(each\,Grid)/Nr.of\,Grids} \end{aligned} \quad (8)$$

It is obvious that PAPR increases by increasing the number of MEs and by decreasing the radius of the UCA. However, between eight and 20 MEs provide similar PAPR performance at different radii (0.18 m, 0.2 m, 0.22 m, and 0.25 m) of the circular array, as shown in Fig. 7. Using from 20 to 32 MEs at different radii (0.18 m, 0.2 m, 0.22 m, and 0.25 m) for the circular array has a significant impact on PAPR performance, and an array with a 0.18 m radius shows the best PAPR performance. Moreover, we can clearly observe that when we increase the number of MEs to more than 32, PAPR again becomes identical. Hence, we select 32 as the right number of MEs for the UCA in our scenario.

Now, we consider the environment when scanning for two targets in order to check the accuracy of beam-scanning results. Figure 8 shows a color map of the output power, in which the yellow region indicates the location of the targets. Similarly, we consider a scenario with one target and one interference source, and test the scanning; the results are in Fig. 9, in which the yellow part identifies the locations of the target and the interference. The accuracy of the scanning results is quite satisfactory, even in environments with multiple targets

and interference sources, and the error is less than 3°, which is acceptable.

3.2 Method of identification of sound sources using HMM classifier

For feature extraction, a 36 MFCC scheme is applied [50]. The recognition of the drone sound is accomplished despite background noise, and we evaluate the performance of the classifier. The feasibility and effectiveness of the proposed algorithm is seen in the experiment results.

3.2.1 Details of feature extraction of drone sounds

MFCC is a commonly used sound-signal feature extraction method that can extract features in the cepstral domain, and it is a mathematical trick to extract the envelope of the spectrum in the logarithm domain. The details of the MFCC procedure are described in Fig. 10.

At first, a short-time Fourier transform (STFT) is employed to transform the time domain signal into the frequency domain, including framing, windowing, and fast Fourier transform (FFT). Figure 11 shows a spectrogram of drone and bird sounds. The output signal after STFT is represented as follows:

$$X_i^{fft}(f) = FFT(x_i(n) \times w_i(n)), \quad 1 \le i \le I, 1 \le n \le N \quad (9)$$

where $x_i(n)$ and $w_i(n)$ are the acoustic data and window functions, respectively, used in the ith frame with total
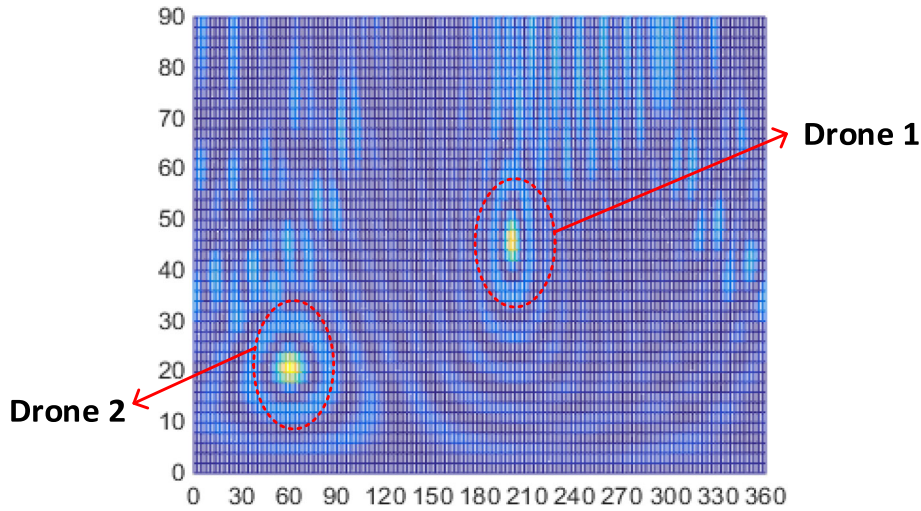
Fig. 8 Color map of output power [direction of drone 1 (45°, 200°); direction of drone 2 (20°, 60°)]

frame numbers *I*. *N* and $X_i^{fft}(f)$ are frame length and windowed signal in the frequency domain, respectively.

The Hamming N-point window function is written as:

$$w(n) = \alpha - \beta \, \cos\left(\frac{2\pi n}{N-1}\right), \quad \alpha = 0.54, \beta$$
$$= 0.46, \quad 0 \le n \le N-1 \quad (10)$$

A mel-scale filter bank is utilized, and is written as follows:

$$mel(f) = 1127 \, \ln\left(1 + \frac{f}{700}\right) \quad (11)$$

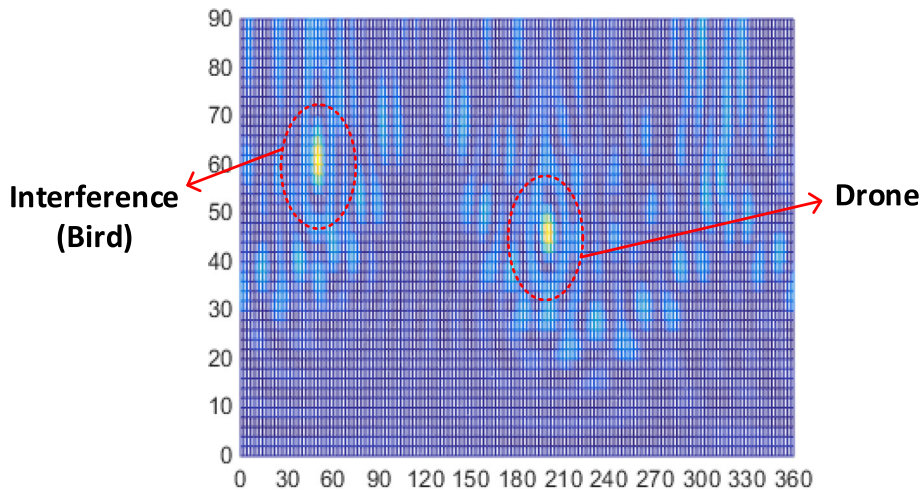Realizing the mth filter of the filter bank is denoted as:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \le k \le f(m) \\ \dfrac{f(m-1)-k}{f(m+1)-f(m)}, & f(m) \le k \le f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad 1 \le m \le M$$

$$(12)$$

where $f(\cdot)$ and M are mel-scale frequency and total number of filters, respectively.
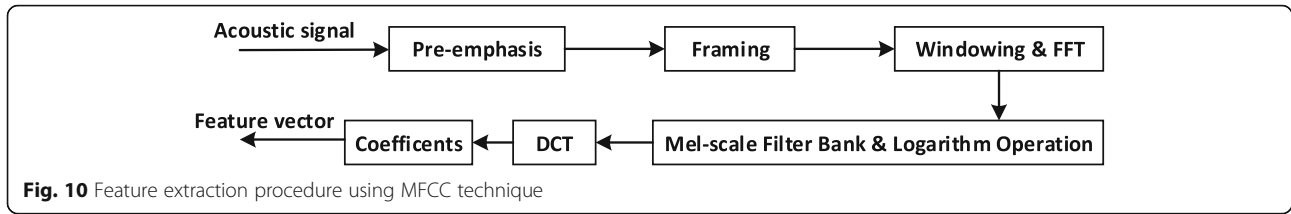
Take the logarithm of the mel spectrum using (13):

$$s(m) = \ln\left(\sum_{k=0}^{N-1} \left|X_i^{fft}(k)\right|^2 H_m(k)\right), \quad 1 \le m \le M$$
$$(13)$$

where *N* is the FFT length of $R_i^{fft}(k)$.

Fig. 9 Color map of output power [direction of drone (45°, 200°); direction of bird (60°, 50°)]

Fig. 10 Feature extraction procedure using MFCC technique

Then, a discrete cosine transform (DCT) is applied to get the nth cepstral coefficients, as follows:

$$c_n = \sum_{m=0}^{M-1} s(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right), \quad n = 1, 2, ..., N \tag{14}$$

where N represents the cepstral coefficients. Generally, Eq. (15) is used to calculate the delta coefficients in MFCCs:

$$d_n = \begin{cases} c_{n+1} - c_n, & n \leq L \\ \dfrac{\sum_{\delta=1}^{L} n(c_{n+\delta} - c_{n-\delta})}{\sqrt{2\sum_{\delta=1}^{L}\delta^2}}, & others \\ c_n - c_{n-1}, & n > N-L \end{cases} \tag{15}$$

In our research, we propose additional delta coefficients, as follows:

$$d_n = \begin{cases} c_{n+1} - c_n, & n \leq L \\ \dfrac{\sum_{\delta=1}^{L}(c_{n-\delta} + c_{n+\delta}) - 2LC_n}{2L}, & others \\ c_n - c_{n-1}, & n > N-L \end{cases}$$

$$d_n = \begin{cases} c_{n+1} - c_n, & n \leq L \\ \dfrac{\sum_{\delta=1}^{L}(c_{n-\delta} + c_{n+\delta}) - 2LC_n}{2L}, & others \\ c_n - c_{n-1}, & n > N-L \end{cases} \tag{16}$$

where δ is the step for calculating the difference of coefficients.

So, 36 MFCCs, including standard MFCCs and delta MFCCs for feature extraction, are applied in this paper [50].

3.2.2 Details of drone sound recognition using HMM

The HMM is a statistical model for an ordered sequence of variables, where states and inputs are hidden and observable, respectively. The sequence of observation vectors is denoted as:

$$O = \{o_t\}, \quad 1 \leq t \leq T \tag{17}$$

where T is the state sequence. Usually, the HMM model is presented as:

$$HMM = (N, M, A, B, \Pi) \tag{18}$$

where N, M, A, B, and Π, respectively, are the hidden states, distinct observations per state, a state transition matrix, the emission probability distribution per state, and the probability of initial state distribution, which is written as:

$$\lambda = (A, B, \Pi) \tag{19}$$

3.2.3 Details of training and test stages for classifier optimization

The parameters of the model are determined by the training data, and the input data for the model are the
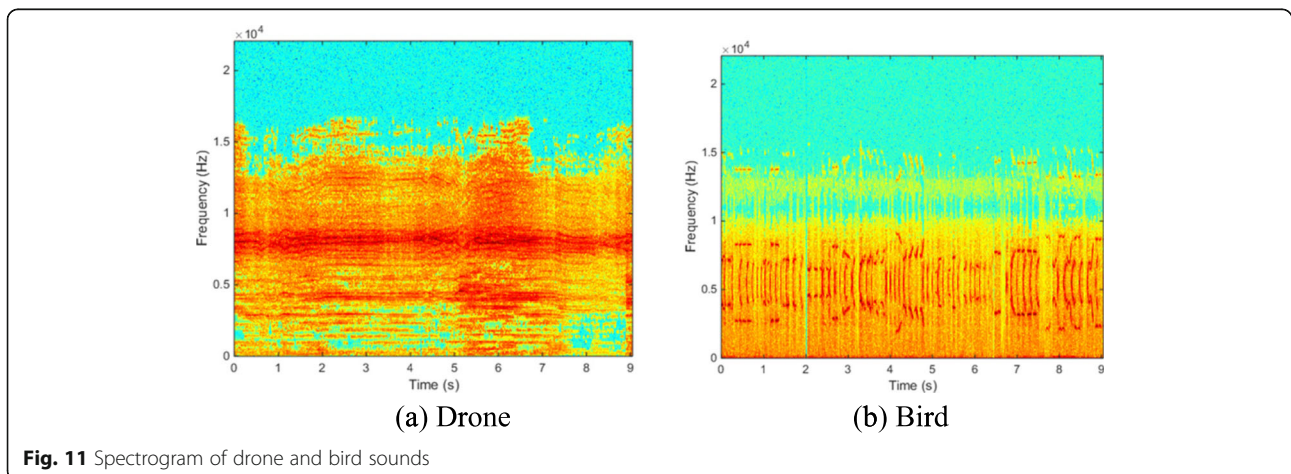
Fig. 11 Spectrogram of drone and bird sounds

(a) Drone   (b) Bird

Table 1 Five clusters of sounds in the training dataset

	D_{train}	Remarks
Cluster 1	Drone 1, Drone 2, Drone 3	Drone 1: SYMA X5SW
Cluster 2	Plane 1, Plane 2, Plane 3	Drone 2: BYROBOT Drone Fighter
		Drone 3: WLtoys Skywalker
Cluster 3	Car 1, Car 2, Car 3, Car 4, Car 5	Plane 1: Helicopter
		Plane 2: Boeing737
Cluster 4	Bird 1, Bird 2, Bird 3	Plane 3: Boeing747
Cluster 5	Rain 1, Rain 2, Rain 3	

extracted features of the training data. The trained models represent the most likely sound identity, and are used to evaluate new incoming acoustic data. In this paper, the training dataset is described as some clusters, each of which represents a certain type of sound (D_{train}). Table 1 shows the training data of each cluster. The procedure for the HMM-based drone sound recognition approach is shown in Fig. 12.

We use a drone, a plane, a car, a bird, and rain in the training dataset. Clusters 1 to 5 are for the sounds of drones, planes, cars, birds, and rain, respectively, as described in Table 1. For better performance, we used three kinds of sounds for drones, planes, birds, and rain clusters, but for the car cluster, five kinds of sounds are collected in order to keep the total sound-data length for each cluster equal, due to the shorter time durations for car sounds. Figure 13 shows the training procedure in the HMM where the training issue is solved with the Baum–Welch algorithm.

In the training stage, the classifier (with a mixture of five HMMs, λ_s) is trained, while the subsequent testing stage, where the Viterbi algorithm is applied, is to find the state sequence that maximizes the probability of the given sequence when the model is known.

The goal in a recognition process is to retrieve the input sound, which is represented by a sequence of feature vectors, O_{test}. The process is to find the HMM with the highest probability, given the sequence, i.e.,

$$g^* = \arg \max_{all\, s \in S} P(\lambda_s | O_{test}) \qquad (20)$$

And the model that gives the maximum probability is the one the test data belong to (i.e., the test data are classified in the cluster that is represented by the selected model).

Figure 14 is a block diagram of the testing procedure, given the trained HMMs and the test dataset.

4 Experiments and performance evaluation of the proposed acoustic signal-based methodology for drone positioning

We investigated drone-sound recognition with the 36-MFCC scheme in which 100 data samples were used for each cluster. The sound detection probably is defined as:

$$P_D = \Pr\left(\frac{\text{Incoming sound has been classified to } \lambda_s}{\text{Sound of cluster } s \text{ existence}}\right) \qquad (21)$$

The effect of training dataset suitability is examined by varying the number of sounds per cluster. Moreover, the power of the sound signal is normalized in order to avoid the effect of different sound energy. Normalized power is described as follows:
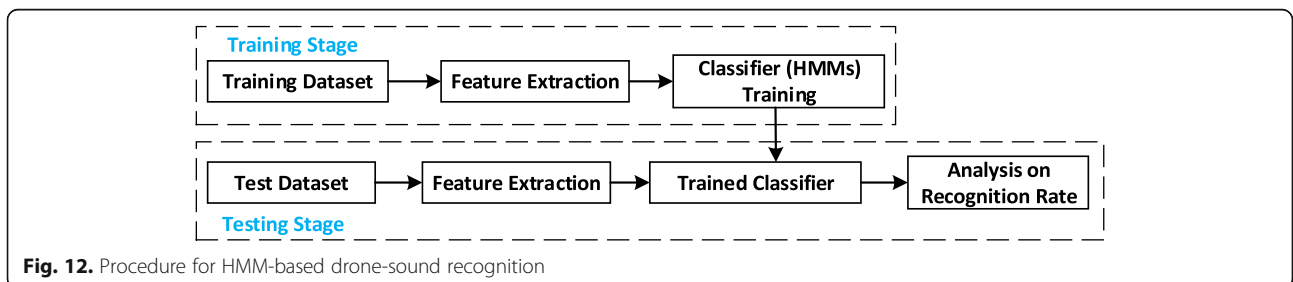
$$P = \frac{1}{N}\sum_{n=1}^{N} X(n)^2 \qquad (22)$$

where $X(n)$ is the signal and N is the number of samples.

Figure 15 shows the amplitude spectrum before normalization. Thus, a normalization factor (NF) is given as:

$$NF_{interference} = \sqrt{\frac{P_{interference}}{P_{drone}}} \qquad (23)$$

where $P_{interference}$ is the power of interference sounds (plane, car, bird, rain) and P_{drone} is the power of the drone sound.

Fig. 12. Procedure for HMM-based drone-sound recognition

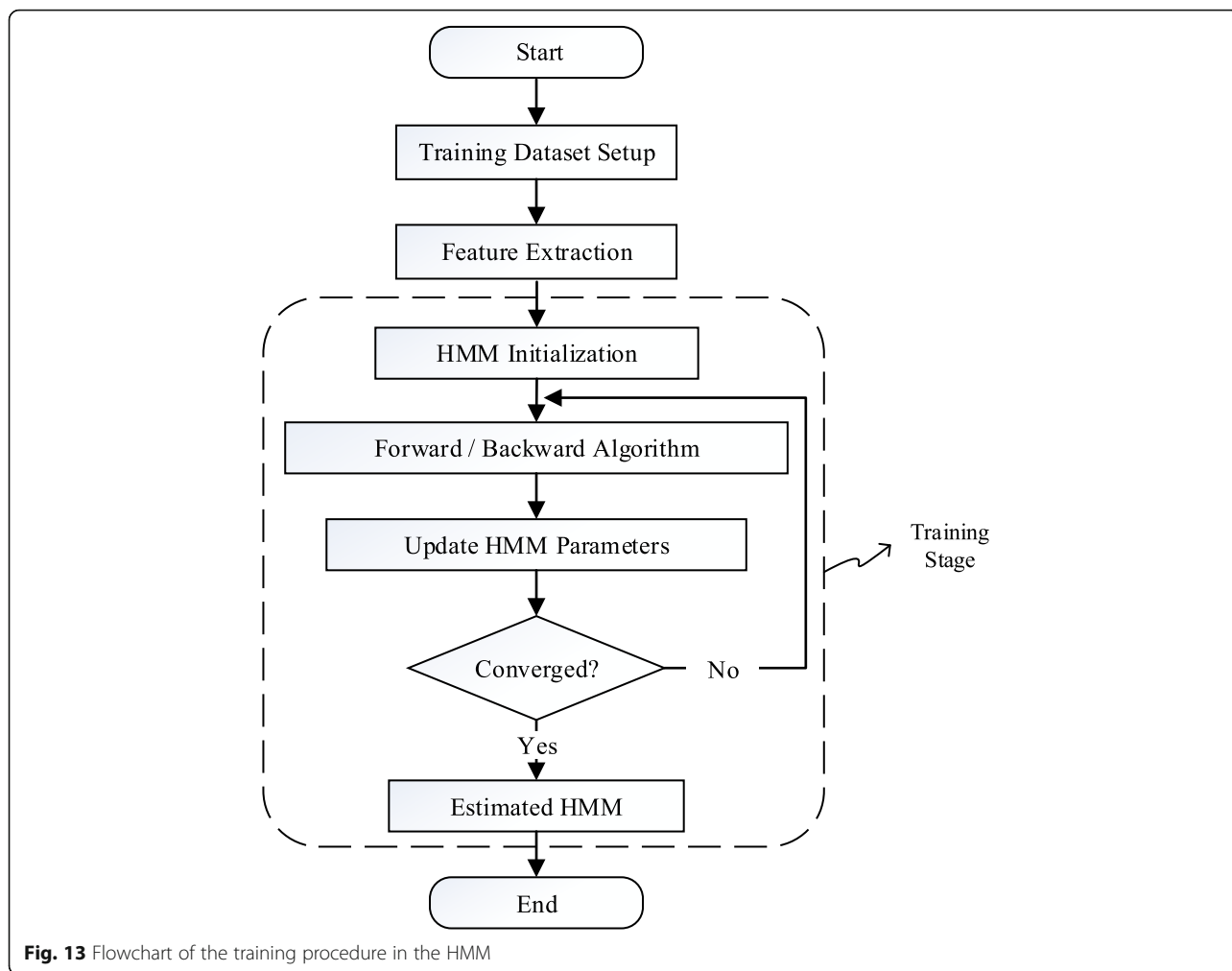**Fig. 13** Flowchart of the training procedure in the HMM

Figure 16 shows the amplitude spectrum after normalization. Hence, all the sounds have the same power as drone sounds after normalization.

Table 2 describes the results of the detection probability with 36 MFCCs. In ideal conditions (without background noise and interference), the detection probability of a drone can reach 100%, but in an actual environment, noise and interference are inevitable. Hence, we built background sounds by combining various interference sounds. Considering the power of each sound in a practical environment, the energy ratio is given as follows:

$$S_{bg} = 1 \times NF_p \times S_p + 1 \times NF_c \times S_c + 0.1 \\ \times NF_b \times S_b + 0.3 \times NF_r \times S_r \qquad (24)$$

where S_p is the sound of a plane, S_c is the sound of a car, S_b is the sound of a bird, and S_r is the sound of rain. It

Fig. 14. Block diagram of the testing procedure using HMMs
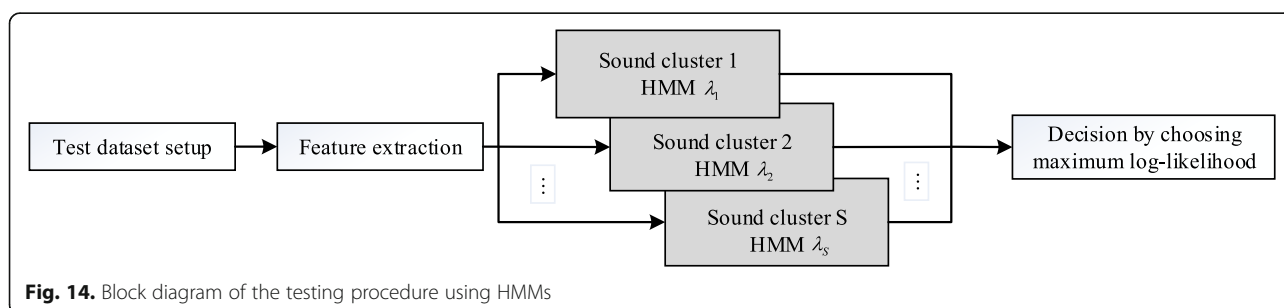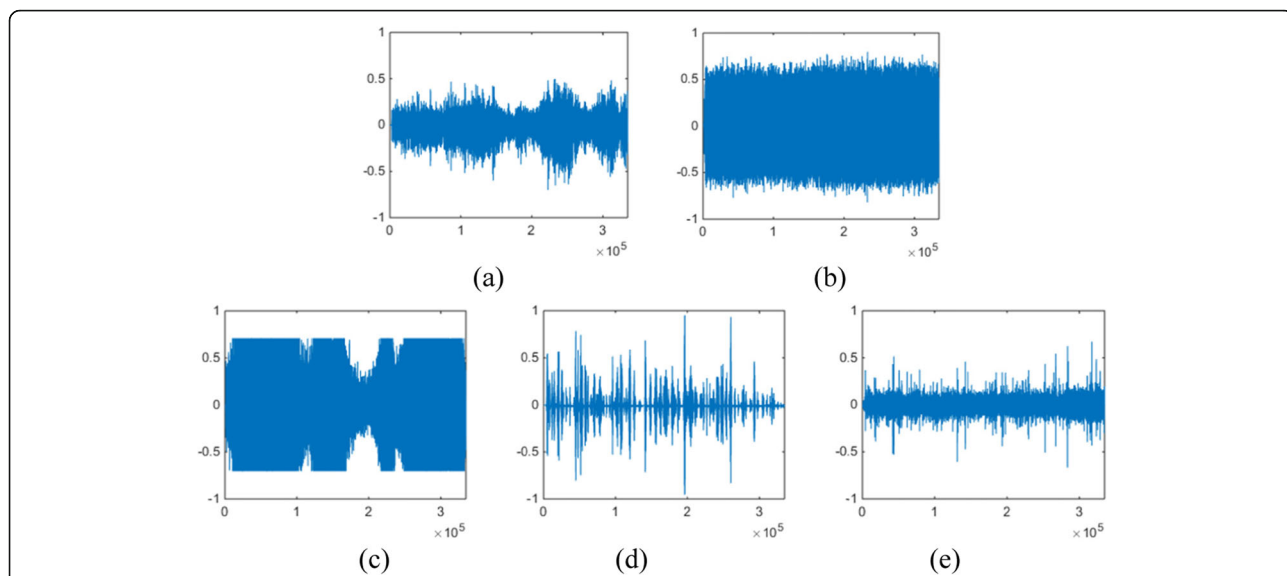
Fig. 15 Amplitude spectrum before normalization: **a** drone, **b** plane, **c** car, **d** bird, and **e** rain

means that background noise consists of each interference sound at a power ratio of 1 for a plane, 1 for a car, 0.1 for a bird, and 0.3 for rain. In reality, the power from a bird and from rain is usually less than that of a plane or a car.

Testing datasets S_1, S_2, and S_3 with various interference sounds are described as follows:

$$
\begin{aligned}
S_1 &= S_d + 0.1 \times S_{bg} \\
S_2 &= S_d + 0.5 \times S_{bg} \\
S_3 &= S_d + 1 \times S_{bg}
\end{aligned}
\quad (25)
$$

which represent an SNR of – 3.3 dB, 2.8 dB, and 16.8 dB, respectively.

Figure 17 shows detection probability versus interference power ratio. When interference power ratios are 0.1 and 0.5, then the detection probabilities for a drone sound are 100% and 90%, respectively. Moreover, when the power combining ratio is 1, the detection probability for a drone sound becomes 67% only. Hence, if there are fewer interference sounds, then the detection probability, of course, gets better.
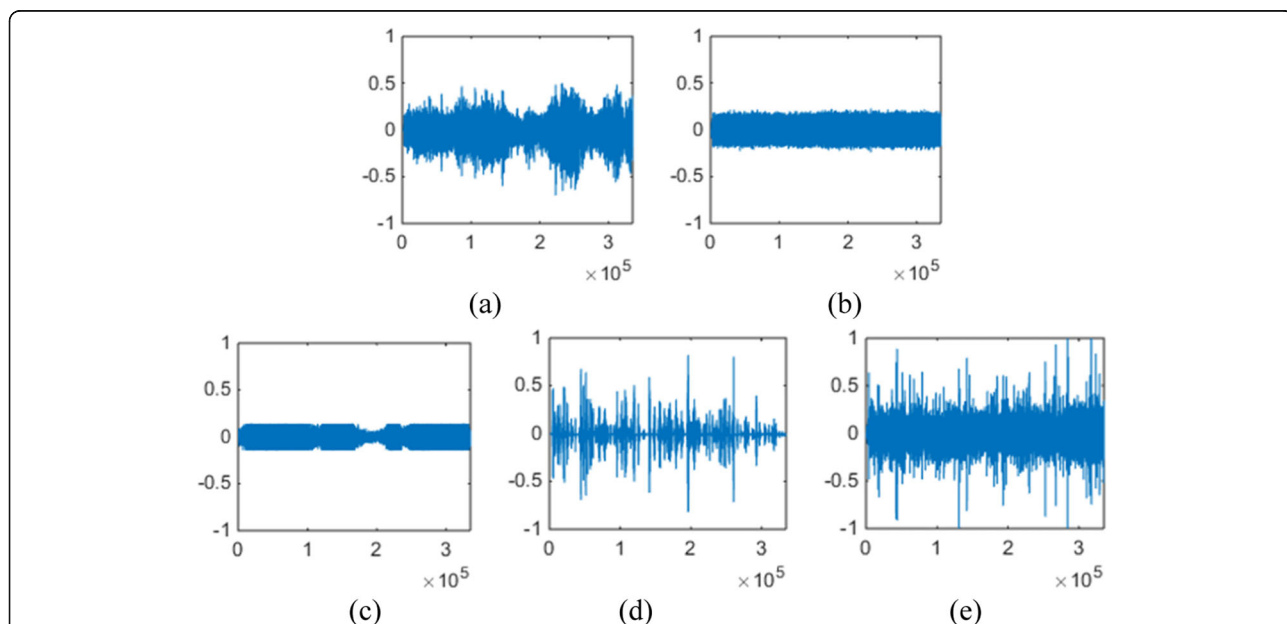
Fig. 16 Amplitude spectrum after normalization: **a** drone, **b** plane, **c** car, **d** bird, and **e** rain

Table 2 Experimental results for sound recognition without noise

Test dataset	Drone	Plane	Car	Bird	Rain
Detection probability	100% Cluster 1	84% Cluster 2	100% Cluster 3	100% Cluster 4	100% Cluster 5

4.1 Details of tracking of illegal drones with adaptive beamforming

4.1.1 Criteria for optimal weights

Since the location of the drone changes over time, the weight vector must be updated periodically. The data used to estimate the weight vector are influenced by noise, so it is suitable to utilize the current weight vector in order to find the next weight vector. The fundamental rule of adaptive beamforming technology is based on specific criteria to adjust the array weights in real time, which gives the best output signal. Generally, adaptive beamforming algorithms can be divided into two types: non-blind algorithms in which a reference signal is required, and blind algorithms in which a reference signal is not necessary. In this paper, we use non-blind algorithm-based adaptive beamforming in order to track illegal drones.

In the literature, there are several criteria for optimal weights, such as MMSE, maximum signal-to-interference ratio (MSIR), and minimum variance, and there are also many adaptive algorithms to update the weight in real time, such as the least mean squares algorithm (LMSE), direct sample covariance matrix inversion, and RLS. It is well-known that RLS offers a better convergence rate. In this paper, we use an RLS adaptive algorithm based on MMSE criteria for tracking illegal drones [22, 23].

The weights are chosen to minimize the mean squared error (MSE) between the beamformer output and the reference signal:

$$\varepsilon^2(t) = \left[d^*(t) - W^H x(t)\right]^2 \tag{26}$$

Taking the expected values for both sides of the equation, and carrying out some basic algebraic manipulation, we have the following:

$$E\{\varepsilon^2(t)\} = E\{d^2(t)\} - 2W^H r + W^H RW \tag{27}$$

where $r = E\{d^*(t)x(t)\}$ and $R = E\{x(t)x^H(t)\}$ are usually referred to as the covariance matrix. The MSE is given by setting the gradient vector of the previous equation (with respect to W) equal to zero:

$$\nabla W\left(E\{\varepsilon^2(t)\}\right) = -2r + 2RW \\ = 0 \tag{28}$$

It follows that the solution is $W_{opt} = R^{-1}r$, which is referred to as a Wiener-Hopf equation, or the optimum Wiener solution [51].

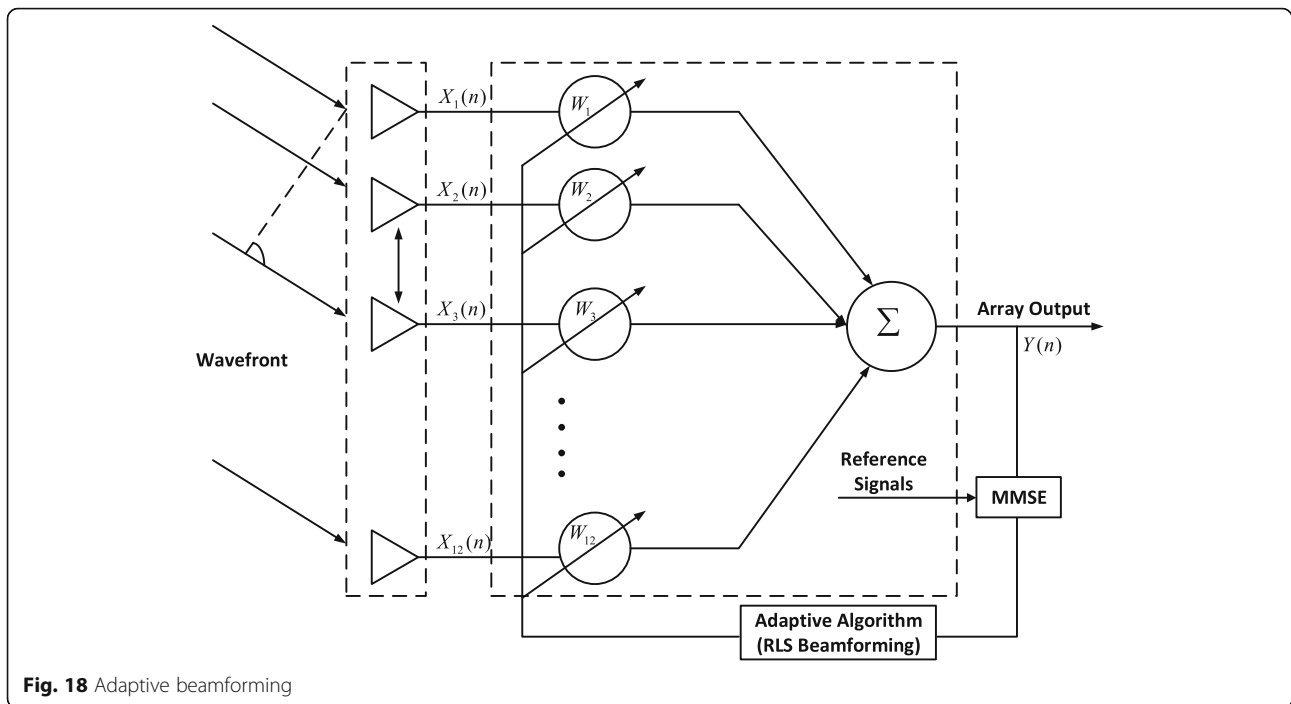#### 4.1.2 RLS algorithm application to update the weight

In the RLS algorithm, the correlation matrix and the correlation vector are calculated recursively [23]. The correlation matrix and the correlation vector are given as:

$$\tilde{R}(n) = \sum_{i=1}^{N} \gamma^{n-i} x(i) x^H(i) \tag{29}$$
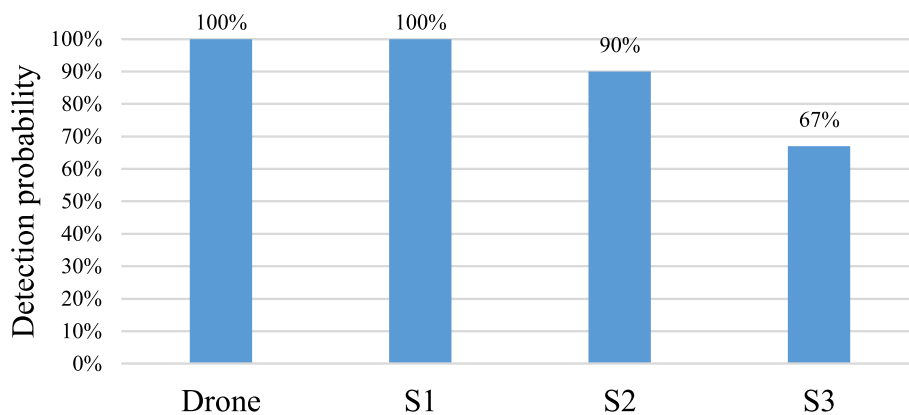
Fig. 17 Experimental results for sound recognition by varying the SNR

Fig. 18 Adaptive beamforming

$$\tilde{r}(n) = \sum_{i=1}^{N} \gamma^{n-i} d^*(i) x(i) \tag{30}$$

Factoring out the terms corresponding to $i = n$, we have the following recursion for updating both $\tilde{R}(n)$ and $\tilde{r}(n)$:

$$\tilde{R}(n) = \gamma \tilde{R}(n-1) + x(n) x^H(n) \tag{31}$$

$$\tilde{r}(n) = \gamma \tilde{r}(n-1) + d^*(n) x(n) \tag{32}$$

Using Woodbury's identity, we obtain the following recursive equation for deriving the inverse of the covariance matrix:

$$R^{-1}(n) = \gamma^{-1} \left[R^{-1}(n-1) - q(n) x(n) R^{-1}(n-1) \right] \tag{33}$$

where gain vector $q(n)$ is as follows:

$$q(n) = \frac{r^{-1} R^{-1}(n-1) x(n)}{1 + \gamma^{-1} x^H(n) R^{-1}(n-1) x(n)} \tag{34}$$

To develop the recursive equation for updating the least squares estimate, $W^{\wedge}(n)$, we use the equation $W_{opt} = R^{-1} r$ to express $W(n)$:

$$\hat{W}(n) = R^{-1}(n) r(n) \\ = \gamma^{-1} \left[R^{-1}(n-1) - q(n) x(n) R^{-1}(n-1) \right] \times \left[\gamma r(n-1) + d^*(n) x(n) \right] \tag{35}$$

Update the weight vector as follows:

$$\hat{W}(n) = \hat{W}(n-1) + q(n) \\ \times \left[d^*(n) - \hat{W}^H(n-1) x(n) \right] \tag{36}$$

Figure 18 shows the structure of adaptive beamforming in which $x_m(n)$ represents the output signal of each ME, and w_m is the weight of the mth element. Reference-signal acquisition is based on the scanning and classification processes mentioned in the previous section. Hence, the reference signals are used in adaptive beamforming for tracking illegal drones.

4.1.3 Tracking results for the direction of arrival of an illegal drone

The key idea is to use reference signals in adaptive beamforming while estimating the DOA of illegal drones. The simulations are performed under both ideal conditions (without background noise and interference sounds) and non-ideal conditions (with background

Table 3 Parameters for tracking

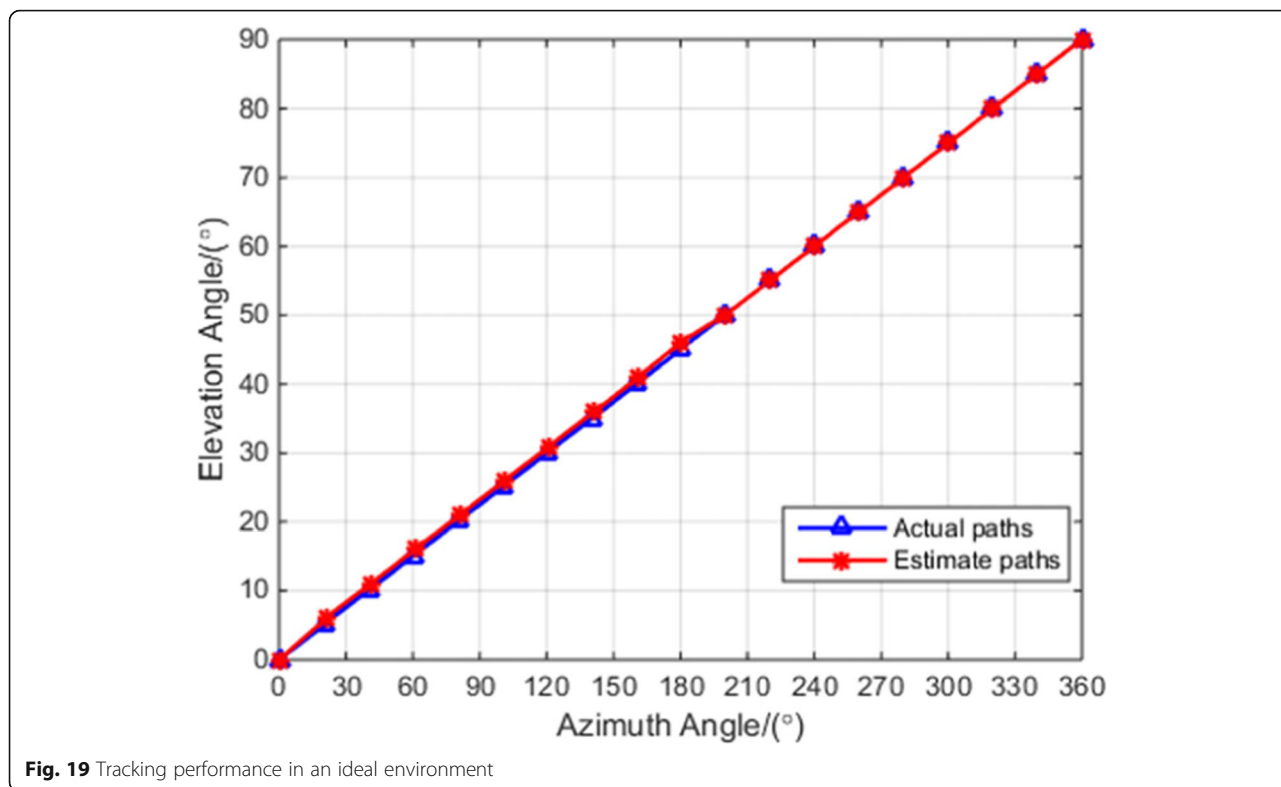| Desired signal	Snapshots	No. of array elements	Radius of array
Drone sounds	1000	32	0.18 m

Fig. 19 Tracking performance in an ideal environment

noise and interference sounds) in order to evaluate the MSE while tracking the illegal drone. The main simulation parameters are given in Table 3.

Error represents the difference between the actual path and an estimated path. Thus, elevation error, azimuth error, and mean squared error are calculated as follows:

$$\text{Elevation error} = \frac{1}{19}\sum_{i=1}^{19}(ae_i - ee_i) \tag{37}$$

$$\text{Azimuth error} = \frac{1}{19}\sum_{i=1}^{19}(aa_i - ea_i) \tag{38}$$

$$\text{Mean squared error} = \frac{1}{19}\sum_{i=1}^{19}\left(\sqrt{(aa_i - ea_i)^2 + (ae_i - ee_i)^2}\right) \tag{39}$$

where ae_i is actual elevation, ee_i is estimated elevation, aa_i is actual azimuth, and ea_i is estimated azimuth. Tracking is executed 19 times on different positions, as shown in Fig. 19.

Table 4 describes the tracking results based on an ideal environment (without noise and interference sounds). Actual elevation and actual azimuth refer to the actual direction of the drone, whereas estimated elevation and estimated azimuth describe the estimated results based on the simulation.

From Fig. 19, we can clearly see that tracking the drone position starts from a 0° elevation angle and a 0° azimuth angle and reaches 90° elevation and 360° azimuth. The estimation paths almost overlap the actual path of the drone. This ensures us that the RLS algorithm is more suitable for tracking of illegal drones owing to the high accuracy of the estimated path under an ideal environment. However, there are still some errors in tracking results; these might be due to several factors, such as the spectral content of the signal and computation error.

In order to check the robustness of the tracking procedure in practical noisy environments, we consider additive white Gaussian noise (AWGN) in our scenario. Figure 20 shows the performance of MSE versus SNR.

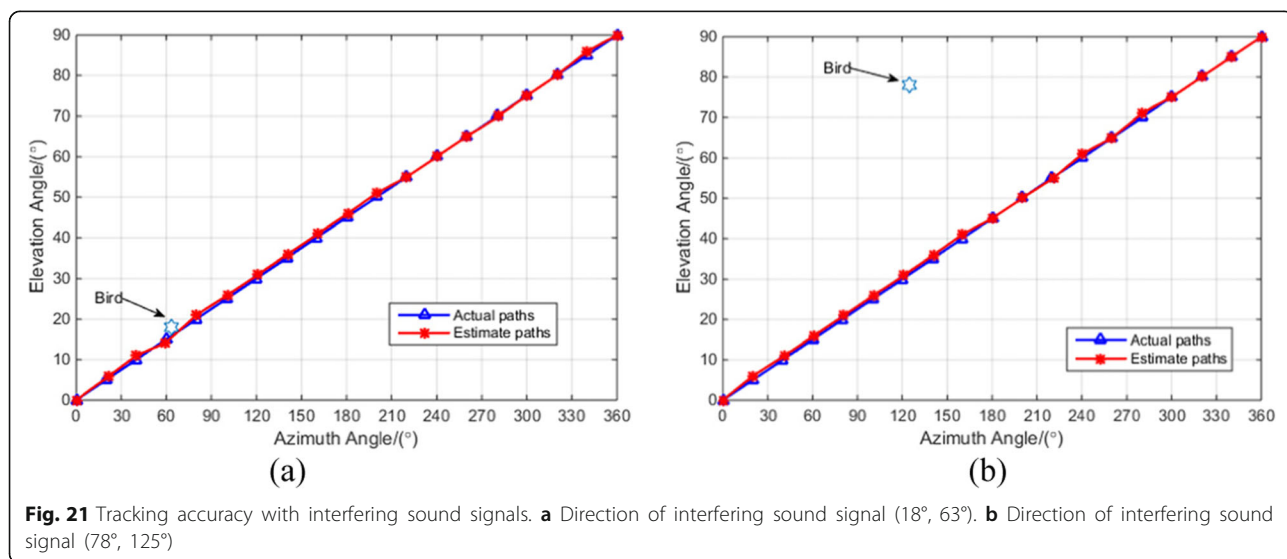**Table 4** Tracking results with an ideal environment

Actual elevation	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	Elevation error	MSE
Estimated elevation	0	6	11	16	21	26	31	36	41	46	50	55	60	65	70	75	80	85	90	0.0083 (rad)	0.0113 (rad)
Actual azimuth	0	20	40	60	80	100	120	140	160	180	200	220	240	260	280	300	320	340	360	Azimuth error	
Estimated azimuth	0	21	41	61	81	101	121	141	161	180	200	220	240	260	280	300	320	340	360	0.0073 (rad)	

Fig. 20 MSE vs. SNR in a practical noisy environment
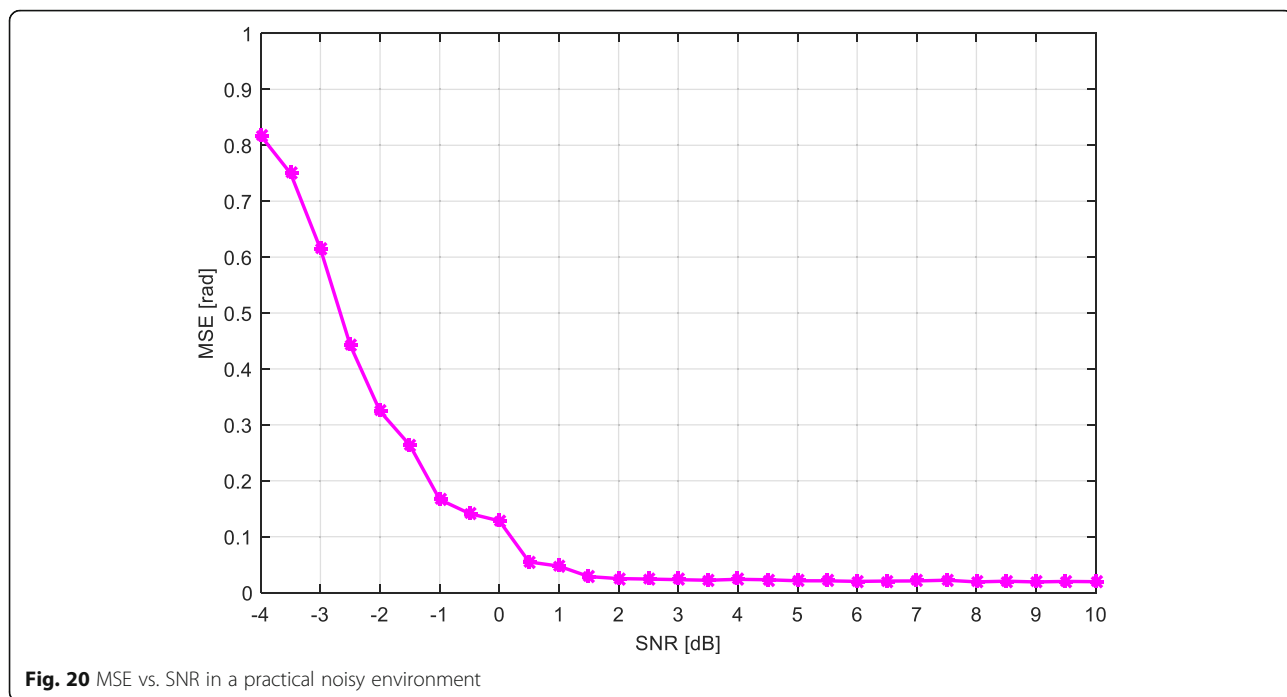
Moreover, the RLS algorithm generates mostly accurate DOA estimates at various SNR values. From Fig. 20, we can observe that by increasing the SNR, MSE lessens, and thus, tracking performance gets better. At a 2 dB SNR, if the MSE is 0.01 rad, tracking performance is almost similar to tracking performance under the ideal case. In addition, MSE of 0.05 rad is also acceptable, because the performance of the tracking system is still good enough, even in a noisy environment.

Similarly, we consider one target and one interference source in the environment in order to consider the effect of interference. Figure 21 analyzes the impact of an interfering sound signal on tracking system performance while tracking the target signal. The interfering sound signal degrades tracking system performance when it is near the target sound signal. Table 5 describes the MSE by varying the position of the interference. This results in errors in elevation angle and azimuth angle for the target's position. Indeed, the error in elevation is greater than the error in azimuth of the target signal's direction. Even though the drone is located in the interference region,

Fig. 21 Tracking accuracy with interfering sound signals. **a** Direction of interfering sound signal (18°, 63°). **b** Direction of interfering sound signal (78°, 125°)

Table 5 Error performance from varying the direction of the interference

	Elevation error (rad)	Azimuth error (rad)	MSE (rad)
Direction of interference (elevation, azimuth) = (18°, 63°)	0.0101	0.0083	0.0144
Direction of interference (elevation, azimuth) = (78°, 125°)	0.0092	0.0074	0.0122

the tracking system can still localize the target continuously.

5 Conclusions

In this paper, we design a monitoring system to detect and track illegal drones. The monitoring system combines sound-signal processing and array-signal processing technologies to scan for sound sources in the sky, and then identifies them to distinguish between drones or something else. In our simulation, we monitor illegal drones by considering both ideal conditions (without background noise and interference sounds) and non-ideal conditions (with background noise and interference sounds). Scanning is performed from 0° to 90° elevation and from 0° to 360° azimuth via SBF. The scanning identifies the direction of the sound sources by pointing the beam and recording the sounds of the objects. These recorded sounds are utilized in the classifier in order to identify the objects. The classifier is based on speech-detection technology in which an HMM model is used. The simulation results show that detection of the sound signal is accurate to around 95% in ideal environments. In addition, detection is more than 80% accurate even at a low SNR of 2 dB under non-ideal conditions. Moreover, the classifier not only identifies drones but also recognizes whether the sound source is a plane, car, bird, or rain. In practical environments, the drone is a moving object. Thus, it is necessary to use the adaptive beamforming technique to track the drone, relying on the reference signals that are acquired from the classifications. We also conducted a tracking simulation by considering a practical environment, such as AWGN and interference from birds. When the SNR increases, MSE becomes smaller, which enhances the tracking performance. From Fig. 20, at a 2 dB SNR, if the MSE is 0.01 rad, tracking performance is almost similar to tracking performance under ideal conditions. In addition, an MSE of 0.05 rad is also acceptable, because the performance of the tracking system is still good enough, even in a noisy environment. Even though the drone is located in an interference region, the tracking system can still localize the target continuously.

Abbreviations
DO: Direction of arrival; HMM: Hidden Markov model; ME: Microphone elements; MFCC: Mel-frequency cepstral coefficient; MMSE: Minimum mean square error; MSE: Mean squared error; MSIR: Maximum signal to interference ratio; PAPR: Peak to average power ratio; RLS: Recursive least squares; SBF: Switched beamforming; SNR: Signal to noise ratio; STFT: Short-time Fourier transform; SVM: Support vector machine; UCA: Uniform circular array

Authors' contributions
JG proposed an acoustic-based scheme for positioning and tracking of illegal drones. Our proposed scheme has three main focal points. He conducted the simulations under both ideal conditions (without background noise and interference sounds) and non-ideal conditions (with background noise and interference sounds). Moreover, he wrote some method aspects of the manuscript and performed the simulations. IA modified the abstract, introduction, conclusions parts, and corrected the sequence of the sections in the Manuscript as well. Moreover, he drew the system model figures and corrected all the English mistake in overall manuscript. In addition, he corrected technical issues related to the manuscript and proposed schemes as well. KC is the technical leader of this manuscript. He suggested the all the technical issues for the proposed acoustic-based scheme for positioning and tracking of illegal drones and for simulation aspects. In addition, he corrected all the simulation methodology of this manuscript and corrected all the mistakes in simulation environment as well as in the structure of overall manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Competing interests
The authors declare that they have no competing interests

References
1. G. Cai, J. Dias, and L. Seneviratne, A survey of small-scale unmanned aerial vehicles: recent advances and future development trends, World Scientific Publishing Company, 2(2), (2014).
2. Rohde and Schwarz, Signal monitoring of radio controlled civilian unmanned aerial vehicles and possible countermeasures, *Protecting the Sky Whitepaper*, 2(2015), (2015).
3. M. Zohaib, A. Jamalipour, Machine learning inspired sound-based amateur drone detection for public safety applications. IEEE Trans. Veh. Technol. 68(3), 2526–2534 (2019)
4. I. Tchouchenkov, F. Segor, and T. Bierhoff, Detection, recognition and counter measures against unwanted UAVs, in Proceeding 10th Future Security Research Conference, (Berlin, Germany, 2015), pp.15-17.
5. A. Zelnio and B. Rigling, Low-Cost Acoustic Array for Small UAV Detection and Tracking, in Proceeding IEEE National Aerospace and Electronics, (Dayton, USA, 2008), pp.110-113.
6. M. Peacock and M. Johnstone, Towards detection and control of civilian unmanned aerial vehicles, in Proceeding of 14 th Australian Information Warfare Conference, (Perth, Australia, 2013) pp. 99-103.
7. E. Tianaroig, F. Jacobsen, Beamforming with a circular microphone array for localization of environmental noise sources. The Journal of the Acoustical Society of America 128(6), 3535–3542 (2011)
8. C. Zhang, D. Florencio, Z. Zhang, Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings. IEEE Transactions on Multimedia 10(3), 538–548 (2008)

9. J. Gebbie, M. Siderius, and J. Giard, Small boat localization using adaptive three-dimensional beamforming on a tetrahedral and vertical line array, Journal of the Acoustical Society of America, 19(1): 2013.

10. X. Zhuang and X. Zhou, Feature analysis and selection for acoustic event detection, in Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, (Las Vegas, USA, 2008), pp.17-20.

11. E. William and M. Hoffman, Classification of military ground vehicles using time domain harmonics' amplitudes, *IEEE Transactions on Instrumentation and Measurement*, 60(11), pp.3720-3731, (2011).

12. A. Averbuch, A. Zheludev, Wavelet-based acoustic detection of moving vehicles. *Journal of Multidimensional Systems and Signal Processing* 20(1), 55–80 (2009)

13. E. Chaves, M. Travieso, and A. Camacho, Katydids acoustic classification on verification approach based on MFCC and HMM, in Proceeding of IEEE Conference on Intelligent Engineering Systems, (Lisbon, Portugal, 2012), pp. 561-566.

14. C. Lin, H. Chen, Audio classification and categorization based on wavelets and support vector machine. IEEE Transactions on Speech and Audio Processing 13(5), 644–651 (2005)

15. I. Sen, M. Saraclar, P. Kahya, A Comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds. *IEEE Trans. Biomed. Eng.* 62(7), 1768–1776 (2015)

16. A. Aljaafreh and L. Dong, Ground vehicle classification based on hierarchical hidden Markov model and Gaussian mixture model using wireless sensor networks, in Proceeding of IEEE International Conference on Electro/ Information Technology, (Illinois, USA, 2010), pp 1-4.

17. H. Van, Optimum Array Processing. New York: Wiley, 2nd ed. Part IV of "Detection, estimation and modulation theory", 2002.

18. F. Akbari, S. Moghaddam, and T. Vakili, MUSIC and MVDR DOA estimation algorithms with higher resolution and accuracy, in Proceeding of International Symposium on Telecommunications, (Tehran, Iran, 2010), pp. 76-81.

19. S. Chen, C. Meng, A. Chang, DOA and DOD estimation based on double 1-D root-MVDR estimators for bistatic MIMO radars. *Wirel. Pers. Commun.* 86(3), 1321–1332 (2016)

20. R.J. Weber and Y. Huang, Analysis for Capon and MUSIC DOA estimation algorithms, in Proceeding of IEEE Antennas and Propagation Society International Symposium. (Charleston, USA, 2009), pp. 1-4.

21. D.N. Patel, B.J. Makwana, and P.B. Parmar, Comparative analysis of adaptive beamforming algorithm LMS, SMI and RLS for ULA smart antenna, in Proceeding of 2016 International Conference on Communication and Signal Processing (ICCSP), (Melmaruvathur, India, 2016), pp. 1029-1033.

22. R. Islam, F. Hafriz, and M. Norfauzi, Adaptive beamforming with 16 element linear array using MaxSIR and MMSE algorithms, in Proceeding of IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, (Penang, Malaysia, 2007), pp.165-170.

23. B. Pattan, Robust modulation methods & smart antennas in wireless communications, electromagnetic Theory and antennas electromagnetic theory and antennas, pp.1149, (2000).

24. Z. Kaleem and M. Rehmani, "Amateur drone monitoring: state-of-the-art architectures, key enabling technologies, and future research directions," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 150-159, May. 2018. DOI: 10. 1109/MWC.2018.1700152

25. I. Ahmad, W. Chen, and K. H. Chang, "Co-channel interference analysis using cooperative communication schemes for the coexistence of PS-LTE and LTE-R networks," in *Proc. of IEEE Communication and Electronics Special Session on LTE Technologies and Services*, Jul. 2016, pp. 181-182.

26. I. Ahmad, Z. Kaleem, and K. H. Chang, "Block error rate and UE through- put performance evaluation using LLS and SLS in 3GPP LTE downlink," in *Proc. of Korean Institute of Communication and Information Sciences*, Feb. 2013, pp. 512-516.

27. I. Ahmad, W. Chen, K.H. Chang, LTE-railway user priority-based cooperative resource allocations schemes for coexisting public safety and railway networks. *IEEE Access* 5, 7958–8000 (2017). https://doi.org/10.1109/ACCESS.2017.2698098

28. Z. Kaleem, M.Z. Khaliq, A. Khan, T.Q. Duong, PS- CARA: context-aware resource allocation scheme for mobile public safety networks. *Journal of Sensors* 18(5), 1–17 (2018). https://doi.org/10.3390/s18051473

29. Z. Kaleem, Y. Li, K.H. Chang, Public safety users priority-based energy and time-efficient device discovery scheme with contention resolution for ProSe

30. I. Ahmad, Z. Kaleem, and K. H. Chang, "Uplink power control for interference mitigation based on users priority in two-tier femtocell network," in *Proc. of IEEE International Conference on ICT Convergence*, Oct. 2013, pp. 474-475.

31. I. Ahmad, K. H. Chang, "Analysis on MIMO transmit diversity and multiplexing techniques for ship ad-hoc networks under a maritime channel model in coastline areas," in *Proc. of IEEE International Conference on ICT Convergence*, Oct. 2017, pp. 18-20.

32. I. Ahmad, K.H. Chang, Analysis on MIMO transmit diversity techniques for ship ad-hoc network under a maritime channel model in coastline areas. *Journal of Korean Institute of Communications and Information Sciences* 42(2), 383–385 (2017). https://doi.org/10.1109/ICTC.2017.8190820

33. I. Ahmad, Z. Kaleem, K.H. Chang, QoS priority based femtocell user power control for interference mitigation in 3GPP LTE-A HetNet. *Journal of Korean Institute of Communications and Information Sciences* 39(2), 61–74 (2014) https://doi.org/10.7840/kics.2014.39B.2.61

34. W. Chen, I. Ahmad, K.H. Chang, Co-channel interference management using eICIC/FeICIC with coordinated scheduling for the coexistence of PS-LTE and LTE-R networks. *EURASIP Journal on Wireless Communications* 2017(34), 1–14 (2017). https://doi.org/10.1186/s13638-017-0822-6

35. Ahmad, L. D. Nguyen, and D. B. Ha, "Quality-of-service aware game theory-based uplink power control for 5G heterogeneous networks," *Mobile Networks and Applications*, Vol. 24, No. 2, pp 556–563 pp. 1-8, 2019. DOI: 10. 1007/s11036-018-1156-2

36. I. Ahmad, K.H. Chang, Effective SNR mapping and link adaptation strategy for next-generation underwater acoustic communications networks: a cross-layer approach. *IEEE Access* 7, 44150–44164 (2019). https://doi.org/10.1109/ACCESS.2019.2908018

37. Z. Kaleem, M. Yousaf, A. Qamar, A. Ahmad, Trung Q, Duong, W, Choi, A. Jamalipour, "UAV-Empowered Disaster-Resilient Edge Architecture for Delay-Sensitive Communication", *IEEE Netw.*, Vol. 99, pp. 1-9, 2019. DOI: https://doi.org/10.1109/MNET.2019.1800431

38. I. Ahmad and K. H. Chang, "Design of system-level simulator architecture for underwater acoustic communications and networking," in Proc. ICTC, Oct. 2016, pp. 384-386.

39. Z. Kaleem, I. Ahmad, and C. Lee, "Smart and energy efficient LED street light control system using zigbee network," in *Proc.* FIT, Islamabad, Pakistan, 2014, pp: 361-365.

40. W. Chen, I. Ahmad and K. H. Chang, "Analysis on the co-channel interference for the coexistence of PS-LTE and LTE-R networks," in *Proc. Conference of Korean Institute of Communications and Information Sciences (KICS)*, June 2016, Jeju, Korea, pp: 202-203.

41. Alamgir, I. Ahmad and K. H. Chang, "On the underwater channel model and network layout," in *Proc. Conference of Korean Institute of Communications and Information Sciences (KICS)*, Jan. 2018, Korea, pp: 202-203.

42. J. Xiao, I. Ahmad and K. H. Chang, "eMBMS and V2V communications for vehicle platooning in eV2X system," in *Proc.* Conference of Korean Institute of Communications and Information Sciences (KICS), Jun. 2018.

43. U. A. Mughal, I. Ahmad and K. H. Chang, "Virtual cells operation for 5G V2X communications," in *Proc. Conference of Korean Institute of Communications and Information Sciences (KICS)*, Jan. 2019, Korea, pp: 1-2.

44. U. A. Mughal, I. Ahmad and K. H. Chang, "Cellular V2X communications in unlicensed spectrum: compatible coexistence with VANET in 5G Systems," in *Proc. JCCI*, May. 2019, Korea, pp: 1-2.

45. I. Ahmad, K.H. Chang, Mission critical user priority-based random-access scheme for collision resolution for coexisting PS-LTE and LTE-M networks. *IEEE Access* 7, 115505–115517 (2019)

46. I. Ahmad and K. H. Chang, Downlink power allocation strategy for next-generation underwater acoustic communications networks," *Electronics*, vol. 8, pp: 1-14, 2019.

47. I. Ahmad and K. H. Chang, "Mission-critical user priority–based cooperative resource allocation schemes for multi-layer next-generation public safety networks," Physical Communication, Nov. 2019.

48. Y. He, I. Ahmad, L. Shi, and K. H. Chang, "SVM-based drone sound recognition using the combination of HLA and WPT techniques in practical noisy environment", *KSII Trans. on Internet and Information System*, vol. 13, no. 10, pp: 5078-5094, May. 2019.

49. J. Litva, T.K. Lo, *Digital beamforming in wireless communications* (Artech House, Boston, 1996), pp. 13–27

50. L. Shi, I. Ahmad, Y. He, K.H. Chang, Hidden Markov model-based drone sound recognition using MFCC technique in practical noisy environments. J. Commun. Netw. **20**(5), 509–518 (2018). https://doi.org/10.1109/JCN.2018.000075

51. S. Haykin, Adaptive Filter Theory. 4th ed. Englewood Cliffs, Prentice Hall, 2002.

Publisher's Note