Indoor Drone Localization and Tracking Based on Acoustic Inertial Measurement

Yimiao Sun, *Student Member, IEEE,* Weiguo Wang, *Student Member, IEEE,*
Luca Mottola, *Senior Member, IEEE,* Jia Zhang, *Student Member, IEEE,*
Ruijin Wang, *Member, IEEE,* Yuan He, *Senior Member, IEEE*

Abstract—We present Acoustic Inertial Measurement (AIM), a one-of-a-kind technique for indoor drone localization and tracking. Indoor drone localization and tracking are arguably a crucial, yet unsolved challenge: in GPS-denied environments, existing approaches enjoy limited applicability, especially in Non-Line of Sight (NLoS), require extensive environment instrumentation, or demand considerable hardware/software changes on drones. In contrast, AIM exploits the acoustic characteristics of the drones to estimate their location and derive their motion, *even in NLoS* settings. We tame location estimation errors using a dedicated Kalman filter and the Interquartile Range rule (IQR) and demonstrate that AIM can support indoor spaces with arbitrary ranges and layouts. We implement AIM using an off-the-shelf microphone array and evaluate its performance with a commercial drone under varied settings. Results indicate that the mean localization error of AIM is 46% lower than that of commercial UWB-based systems in a complex $10\mathrm{m}\times10\mathrm{m}$ indoor scenario, where state-of-the-art infrared systems would not even work because of NLoS situations. When distributed microphone arrays are deployed, the mean error can be reduced to less than $0.5\mathrm{m}$ in a $20\mathrm{m}$ range, and even support spaces with arbitrary ranges and layouts.

Index Terms—Drone, Indoor Tracking, Microphone Array, Acoustic Signal

1 Introduction

Location information is crucial for drone operation [1], [2], regardless of the application and target deployment environment [3], [4], [5]. For example, in an indoor warehouse like the one of Fig. 1, a drone for cargo inventory needs location information to determine the position of the cargo relative to its own. When performing cargo deliveries, a drone must follow the predefined route and land at the right target location for the drop-off.

Location information must be *accurate*. Errors in location estimates may not just degrade system performance, but represent a safety hazard as the drone's own movements are largely determined by location information. In outdoor settings, GPS is arguably the mainstream to provide accurate location. The indoor setting, however, represents a completely different ballgame.

There have been many different approaches and solutions for drone localization and tracking [6], [7], [8], [9], [10]. Radar-based approaches [7], [11], for example, work both indoors and outdoors. Their spatial resolution is limited so that it is generally difficult to localize small-size drones. Further, objects in the target environment easily interfere with the radar signals, degrading the accuracy. RF-based localization approaches [9], [12] require installing wireless transceivers on the drone and reengineering the flight controller. Inertial measurement methods [13], [14] are useful when absolute localization is unavailable, but the accumulation of errors likely becomes an issue. Infrared-based systems require dedicated hardware and corresponding software changes on both drones and control stations [10].

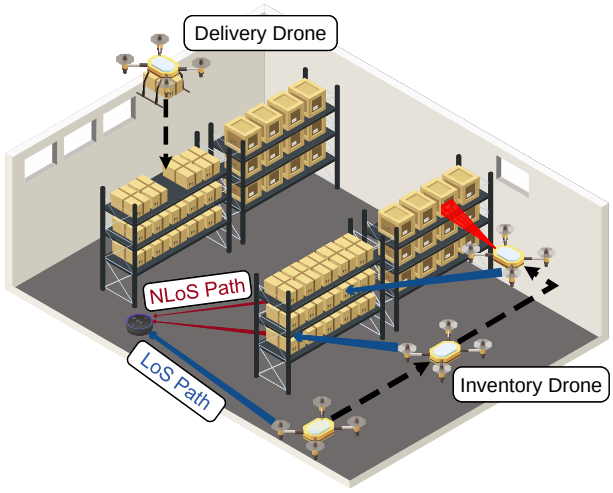


Figure 1: An example of AIM's application scenario.

A *low-cost* and *accurate* localization approach is arguably still missing on drones. Inspired by our observation on the dynamics of drones [15], [16], [17] and the existing work

- *Y. Sun, W. Wang, J. Zhang, Y. He are with the School of Software and BNRist, Tsinghua University, Beijing, China. E-mail: {sym21, wwg18, j-zhang19}@mails.tsinghua.edu.cn, heyuan@mail.tsinghua.edu.cn.*
- *L. Mottola is with Politecnico di Milano (Italy), RI.SE Sweden, and Uppsala University (Sweden). E-mail: luca.mottola@polimi.it.*
- *R. Wang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. E-mail: ruijinwang@uestc.edu.cn.*

This work is partially supported by the National Science Fund of China under grant No. U21B2007 and No. 62271128, the R&D Project of Key Core Technology and Generic Technology in Shanxi Province under grant No. 2020XXX007, the Swedish Science Foundation (SSF), the Digital Futures programme (project Drone Arena), the Swedish Research Council under grant 2018-05024, and KAW project UPDATE.
(Corresponding author: Yuan He.)

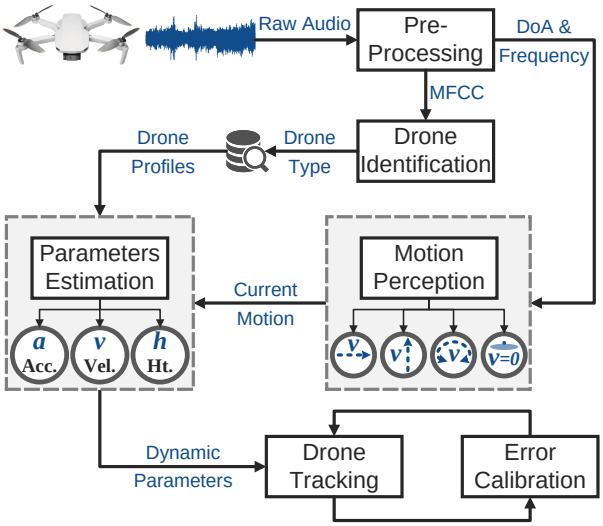


Figure 2: AIM workflow.

that utilizes the propellers to produce audio [18], we present Acoustic Inertial Measurement (AIM), a *completely passive* approach to localize the drones with a single microphone array. The term *passive* means AIM requires *no* additional hardware and *no* software changes on the drones, only using the acoustic signals naturally produced by the drone itself. AIM works with only a single microphone array but may be extended with ease to support spaces with arbitrary ranges and layouts by deploying distributed arrays.

To achieve this, we must tackle three key challenges:

1) A single microphone array can only acquire one direction of arrival (DoA), which denotes the drone's direction relative to the array; this information alone is insufficient for location calculation.
2) The only input to AIM is the propellers' sound of the drone; how to infer the drone's location and motion from this single acoustic signal is an open problem.
3) In complex indoor environments, the acoustic channel between the drone and the microphone array is easily interfered by ambient noise and obstacles, or travels along NLoS paths as Fig. 1 illustrates.

AIM. We address these issues based on the fundamental observation that the rotating propellers create a *dual acoustic channel*: from the microphone array's view, the propellers are regarded as the sound source, so the DoA of sound denotes the *orientation* of the drone. At the same time, the propellers are also high-speed rotating machinery, so the frequency properties of the sound actually correspond to the rotating state of the propellers, which in turn determines the drone's *motion*. Obtaining *orientation and motion* information allows us to track the drone's location continuously.

Fig. 2 illustrates AIM's workflow. The raw acoustic signal captured by the microphone array is first pre-processed to extract the characteristics of the acoustic signal, for example, DoA, frequencies, and Mel-Frequency Cepstral Coefficients (MFCC). DoA and frequencies help deduce the drone's current motion, whereas MFCC is utilized for identifying the specific drone structure, for example, a quadcopter as opposed to an octocopter, and then loading the corresponding profile information (*e.g.*, mass) from a database.

By feeding the drone's profiles into a set of dynamic equations we formulate, we estimate its dynamic parameters, that is, acceleration and velocity. The drone's location is calculated consequently. To reduce error, we adopt a dedicated Kalman filter and the Interquartile Range rule (IQR). We further show how AIM can be extended to support indoor spaces with arbitrary ranges and layouts by deploying distributed microphone arrays.

Our contribution can be summarized as follows:

1) We design AIM, a *completely passive* drone tracking approach that can work with a single microphone array. At the core of AIM is exploiting the dual acoustic channel to perceive the drone's motion and estimate its location.
2) We exploit the acoustic characteristics of the drones to derive their motion and estimate their location, *even in NLoS* settings. We combine this with a dedicated Kalman filter and the Interquartile Range rule (IQR) to reduce the error, and demonstrate that AIM can support indoor spaces with arbitrary ranges and layouts.
3) We implement AIM using off-the-shelf microphone arrays and perform an evaluation using a commercial drone under varied settings. Results indicate that the mean localization error of AIM in a complex $10\text{m}\times10\text{m}$ indoor scenario is 1.89m, 46% lower than that of commercial UWB-based systems, where state-of-the-art infrared systems would not even work. Further, AIM can be extended to support indoor spaces with arbitrary ranges and layouts by deploying distributed microphone arrays.

Works close to our efforts are summarized in Sec. 2. Sec. 3 introduces the unique acoustic features of different drone motions. Then, Sec. 4 presents methods to distinguish different drone motions and drone structures. Sec. 5 elaborates on the core algorithm of AIM for drone trajectory tracking and Sec. 6 unfolds how to use distributed microphone arrays to extend the operating range. The implementation and evaluation results are presented in Sec. 7. We discuss practical issues in Sec. 8 and conclude the paper in Sec. 9.

2 Related Work

The distinctive feature of our work is to perform drone localization and tracking using acoustic signals. We briefly survey existing efforts in either field.

2.1 Drone Localization and Tracking

RF-based methods. RF signals are extensively explored for drone localization [9], [19], [20], [21], [22]. In outdoor scenarios, mmWave and WiFi are usually used. For example, mmHawkeye [20] exploits commercial mmWave radars to capture the feature of drone's periodic micro-motion (PMM) and achieve less than 10cm tracking error within 30m. Nguyen et al. [9] explore a passive approach to localize both the drone and its controller in 2.4GHz WiFi frequency channel. They show an average error of around 10m in the 30m to 150m distance.

In indoor scenarios, Ultra Wide Band (UWB)-based approaches are mainstream. UWB techniques [21], [23] achieve decimeter accuracy for drone tracking. To improve accuracy,

UWB may integrate with other techniques, such as visual SLAM [24], RGB-D camera [25] and optical flows [26]. The errors of these methods are usually lower than 20m. However, the performance of RF-based methods will degrade in complex NLoS scenarios, especially in the presence of equipment that absorbs or scatters RF signal [27].

Acoustics-based methods. AIM enjoys the fact that acoustic signals may be fruitfully employed also in NLoS settings [28], [29]. For example, Mao et al. [30] attach two speakers on the drone to emit Frequency-Modulated Continuous-Wave (FMCW) signals, used to estimate the distance between the drone and a mobile phone. As for AIM, it does not install any extra equipment on the drone. Other efforts [31], [32] only regard the drone as a mobile sound source and deploy 3D or large microphone arrays to estimate its location. Compared with these techniques, we explore the theoretical connection between the drone's sound and its motions, deduce the drone's dynamic parameters, such as velocity and acceleration, from its sound and track the drone by using only a small 2D microphone array.

Data-driven methods and other. AIM is a model-driven technique for drone tracking and localization. Various data-driven methods exploiting machine learning or deep learning exist [33], [34], [35], [36]. However, these methods may require complex algorithms and pose challenges in transferring a specific model to another drone or environment, which makes them arguably impractical.

GPS is a mature approach widely used for drone localization and offers meter-level accuracy, but its application indoors is extremely difficult [37]. Methods based on optics and vision [10], [38], [39], [40] can provide much more accurate results for indoor drone localization, whose errors are even less than 1mm as reported [38]. However, these methods vastly assume line-of-sight (LoS) conditions and are sensitive to lighting conditions.

2.2 Acoustics-based Tracking

Indoor tracking. Several works demonstrate the use of acoustic signals for localization and tracking [41], [42]. With a single microphone array, Voloc [43] aligns the multi-path DoA estimation for accurate localization of indoor acoustic sources; Symphony [44] extends this method to localize multiple sources by leveraging the prior-known layout of the array. PACE [45] localizes multiple mobile users simultaneously by leveraging structure-borne and air-borne footstep impact sounds. These works assume that the localization target and the microphone array are on the same plane or that the target's altitude is known, to solve a bi-dimensional localization problem. Differently, we exploit the signal feature in both the spatial and frequency domains, achieving *three-dimensional* localization with a single array.

Short-range tracking. Recent works adopt wearable devices for tracking, such as smartwatches and earphones. SoM [46] tracks the wrist using a smartwatch with IMUs and employs the smartphone to send beacons for error calibration. EarAR [47] uses the IMU in earphones and smartphones to track the indoor user's location and gazing orientation. When the embedded microphone and speaker in the wired or wireless earphones have already formed a transceiver pair, EarphoneTrack [48] proposes to track either the microphone or speaker with this pair. Unlike what we do with AIM, these approaches are effective only in the short range, specifically between wearable devices and users' smartphones.

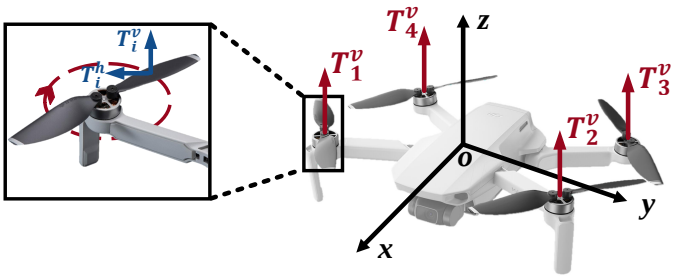


Figure 3: Quadcopter drone structure.

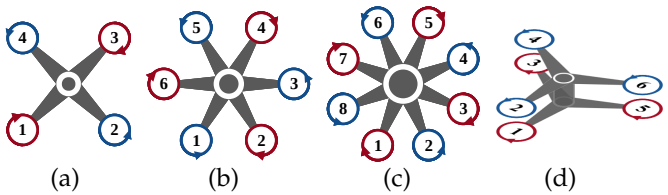


Figure 4: Typical structures of four drone types: (a) quadcopter; (b) hexacopter; (c) octocopter; (d) Y6. Different colors represent different directions of rotation.

3 The Sound of Drones

In this section, we explore the features of a drone's sound signals and how they relate to motion.

3.1 Key Features

Drone propellers are designed to displace the air around them. The resulting pressure gradient creates a force vector. We model the connection between the sound of the drone's propellers and its physical structure.

Fig. 3 illustrates the most common drone structure, that is, a quadcopter composed of two orthogonal arms. A propeller is mounted at either end of each arm. The force vector obtained by the propeller rotation can be decomposed into a vertical component T_i^v and a horizontal component T_i^h.

The vertical component lifts the drone and can be calculated as $T_i^v = k^v f_i^2$, where f_i is the rotation frequency of the i^{th} propeller and k^v is a constant related to the lift coefficient. The drag force T_i^h horizontally controls the rotation of the body and can be calculated as $T_i^h = k^h f_i^2$, where k^h is a constant related to drag coefficient [49]. The lift forces of all propellers follow the same direction, while the drag forces of adjacent propellers are opposite to compensate for the torque otherwise generated, which induces spinning.

The sound produced by the propellers is highly correlated with the frequency f_i of each motor. Because each propeller has multiple blades, two in most cases, the fundamental frequency of the sound is not the rotation frequency f_i, but the blade passing frequency (BPF). The BPF is defined as $f_i^{BPF} = nf_i$, where n is the number of blades. In addition

to the BPF, harmonic frequencies may also be observed as an integer multiple of the BPF [50].

If we can capture the drone's sound and obtain the BPF as well as its harmonics, we may then estimate the rotation frequencies f_i, and thus the forces exerted by each propeller. Using a model of the drone's physical dynamics, which is necessarily a function of its mechanical structure, we may also estimate its direction and motion. This is the essence of the frequency-based localization and tracking in AIM.

3.2 Sound and Motion

We analyze here the inner relationship between the drone's sound and its physical motion.

We theoretically analyze the acoustic properties of four common drone structures, shown in Fig. 4. Drone flights are composed of four basic motions: hovering, yaw, horizontal linear motion and vertical linear motion, as depicted in Fig. 5. Interestingly, we find that these basic motions exhibit different acoustic properties in the frequency domain because they are performed by changing each motor's rotation frequency f_i differently. In the following, $N = 4, 6$ or 8 depending on the drone structure among the ones in Fig. 4.

Hovering: in the absence of environmental effects requiring compensation, all propellers rotate at the same frequency to maintain the vertical and horizontal balance, so the drone remains stationary. Therefore, we have $f_i = f_j, 1 \leq i, j \leq N$.

Yaw: propellers operate in pairs, shown by different colors in Fig. 4. Each pair rotates at the same frequency, creating a rotational momentum while maintaining the vertical balance, which makes the drone rotate around the center. Thus, we have $f_{2i-1} = f_{2j-1} \neq f_{2i} = f_{2j}, 1 \leq i, j \leq \frac{N}{2}$.

Horizontal motion: propellers operate in pairs again, this time to tilt the body while maintaining the vertical balance. Then the drone moves horizontally. We use parentheses to indicate equal frequencies for brevity. When the drone tilts forwards or backwards, that is, it pitches, we have $(f_1 f_2)$ $(f_3 f_4)$ for quadcopters, $(f_1 f_2)$ $(f_3 f_6)$ $(f_4 f_5)$ for hexacopters, $(f_1 f_2)$ $(f_3 f_8)$ $(f_4 f_7)$ $(f_5 f_6)$ for octocopters and $(f_3 f_4 f_5 f_6)$ $(f_1 f_2)$ for Y6 structures. Symmetric observations apply when the drone tilts leftwards or rightwards, that is, it rolls.

Vertical motion: all propellers rotate at the same speed to generate thrust greater or lower than the force of gravity on the drone. Accordingly, the drone moves upwards or downwards, so we have $f_i = f_j, 1 \leq i, j \leq N$.

In the following, we illustrate how these observations may be a stepping stone to achieving accurate drone localization and tracking.

4 Motion and Structure

We use the features of the sound signal in the frequency, spatial, and time domains to estimate the drone's motion and identify its structure. These two components are the basis of our system.

4.1 Motion Detection

Based on the analysis of Sec. 3, we conduct a proof-of-concept experiment to check whether the four basic motions can be distinguished by the sound characteristics. In this experiment, we use a DJI Mini 2 quadcopter and a microphone to receive the acoustic signal.

Fig. 6 shows the spectrum of the acoustic signal corresponding to the motions of Fig. 5 and conforms to our understanding of the drone's dynamics. Specifically, we observe two peak fundamental frequencies in the case of yaw and horizontal motion. In comparison, there is only one peak fundamental frequency in the case of hovering and vertical motions.

Exclusively based on frequency domains, we can only classify the four motions into two categories, depending on the number of peak fundamental frequencies. To resolve this ambiguity, we leverage the spatial information of the sound. Crucially, we note that the drone spatial coordinates are stable during hovering or yaw, while they change during vertical or horizontal motion. The change in position may be detected by the sound's DoA, as elaborated in Sec. 5.1. By combining the information obtained from the number of peak fundamental frequencies and DoA as shown in Tab. 1, AIM can correctly discern the four basic motions.

Detecting the four basic motions is vastly sufficient to localize and track drones in a multitude of indoor drone applications, including most of those we mention in the Introduction. In indoor settings, for example, warehouses or smart factories, planning of robot movements—not just drones—is most often achieved by sequentially combining the four basic motions. This is beneficial in at least two respects: *i)* it matches the regular physical layout of the target deployment scenarios; in a warehouse, for example, shelves are side-by-side horizontally laid and goods are stacked vertically; and *ii)* it greatly simplifies path planning, yielding much more scalable systems.

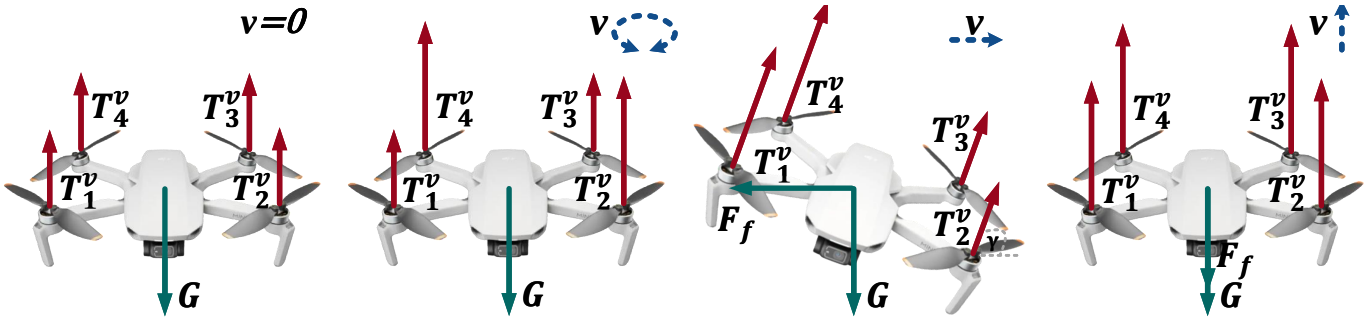
To further improve the accuracy in detecting the four basic drone motions, we further observe that high-frequency harmonics share similar characteristics with the fundamental frequencies. Because the noise in the low-frequency band is usually stronger than that in the high-frequency band, the harmonics may experience less noise than the original BPF. Thus, we estimate the BPF from the weighted average of both the fundamental frequencies and the harmonics, which are weighted by their amplitudes. For hovering, a single band is present on the spectrogram.

4.2 Drone Structure Identification

There exist several types of drones apt to support distinct applications. For instance, drones with high load-carrying capacity can be designated to transport goods, while drones

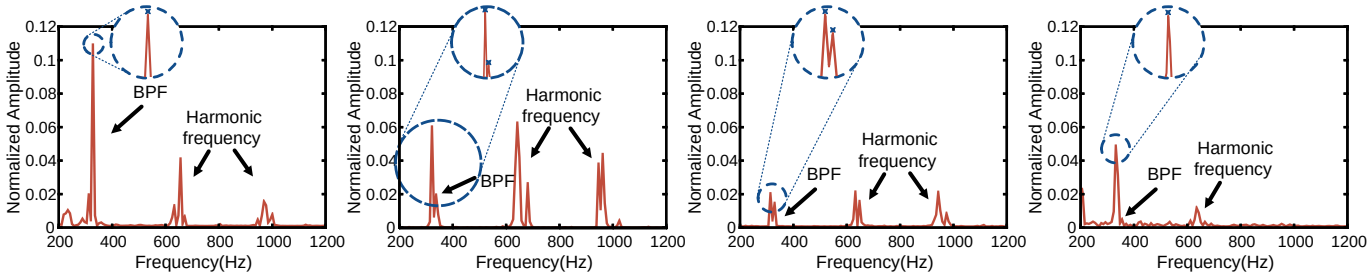
Table 1: Classification scheme of the four motions.

	Single-Peak	**Multiple**-Peak
Unstable DoA	Vertical linear motion	Horizontal linear motion
Stable DoA	Hovering motion	Yaw motion



(a) Hovering motion. (b) Yaw motion. (c) Horizontal linear motion. (d) Vertical linear motion.

Figure 5: Force analysis of basic drone motions.



(a) Hovering motion. (b) Yaw motion. (c) Horizontal linear motion. (d) Vertical linear motion.

Figure 6: Acoustic spectrum of basic drone motions.

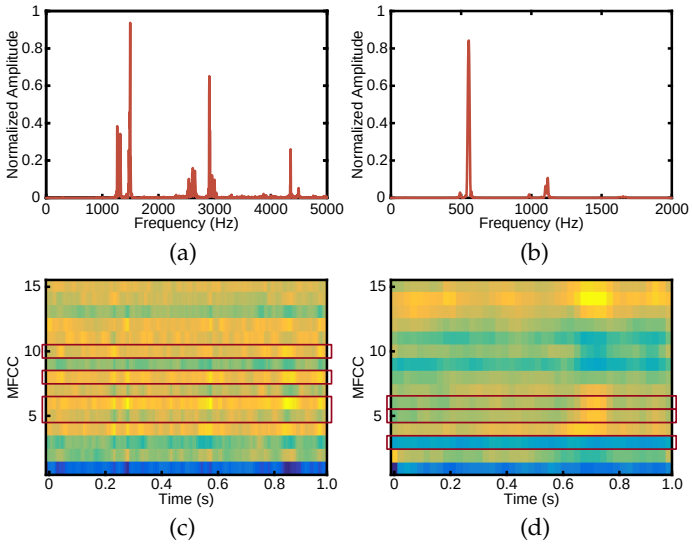


Figure 7: MFCC of different drones: (a) spectrum of DJI Avata in yaw motion; (b) spectrum of DJI FPV in hovering motion; (c) MFCC of a yawing DJI Avata; (d) MFCC of a hovering DJI FPV.

with large-capacity of batteries can be employed for environmental surveillance. Each such type of drone uses a different physical structure, expressly designed to optimize the aerodynamics features required to carry out a specific task. For AIM to work accurately, it is crucial to precisely recognize the particular drone structure once it is detected by the microphone array. In the following, we illustrate a technique to do so, even in case different drone types coexist in the same area.

We design specific band-pass filters for each type of drone, based on their distinctive BPF and harmonic frequencies. For example, the BPF of the DJI Avata, which uses five blades rotating at $\approx$ 300Hz, is approximately 1500Hz, while that of the DJI FPV, which uses three blades rotating at $\approx$ 185Hz, is around 555Hz, as illustrated in Fig. 7(a) and Fig. 7(b), respectively.

We first process the captured acoustic signal through the band-pass filters of each possible drone structure, to form multiple filtered narrow-band acoustic signals. Then, we calculate the Mel-Frequency Cepstral Coefficients (MFCC) for each filtered signal. MFCC carries information that can effectively represent a drone's sound characteristics in both frequency and time domains [51], so we utilize it to differentiate between drones. Fig. 7(c) and Fig. 7(d) demonstrate the distinct MFCC features of the DJI Avata and FPV, whose energy distributions vary among MFCC vectors, especially where their BPF and harmonic frequencies are located, as shown by the MFCC vectors in red frames.

Finally, we normalize the MFCC vectors of all the filtered signals and borrow the method proposed in DronePrint [51] to train a Long Short-Term Memory (LSTM) neural network for drone identification. If multiple drones are located in the same area, we can identify them according to their corresponding filtered signals.

The profiles of drone structures that cater to a warehouse are pre-archived in a database. Upon identification of a drone, the corresponding profile is fed to dynamics equations for position estimation, which we discuss next.

5 Drone Trajectory Tracking

We articulate here how to combine information from the drone dynamics with the input from acoustic signals to

achieve accurate drone localization and tracking. We further illustrate our system's operation in NLoS settings and how we use a dedicated Kalman filter to tame tracking errors.

5.1 Tracking Model

We first derive a dynamic drone model, which we use as a basis for tracking. We consider a quadcopter as an example for intuitive analysis, but the analytical process would be exactly the same for other drone structures.

Yaw. In this case, $(T_1^h + T_3^h) - (T_2^h + T_4^h) \neq 0$, which causes the rotation of the fuselage, as shown in Fig. 5(b), and two BPF peaks. During the rotation process, the moment of inertia I reflects the magnitude of inertia and is regarded as a constant. We can thus obtain the angular acceleration β_t at time t by solving the equation:

$$\frac{k^h}{n^2}\left|\sum_{i=1}^{N/2}(f_{2i-1}^{BPF})^2 - \sum_{i=1}^{N/2}(f_{2i}^{BPF})^2\right| = I\beta_t \tag{1}$$

Thus, in a known time interval τ, the rotation angle $\Delta\psi = \int_0^\tau \beta_t t\, dt$. However, as mentioned in Sec. 4.1, ambiguity exists if we only rely on the frequency characteristics. To solve this ambiguity, we regard the drone as a mobile sound source and leverage the microphone array to obtain spatial information. Due to the limited resolution of commercial microphone arrays, the drone is always in the far-field [44], so that we can hardly obtain accurate location information but only a DoA, including azimuth α and elevation ϕ. Even in this case, DoA information is sufficient for AIM to function. For instance, DoA information captured by a uniform 4-microphone array in a squared configuration is

$$\begin{cases}\tan\alpha = \frac{\tau_{42}^*}{\tau_{31}^*} \\ \sin\phi = \frac{c}{2d}\sqrt{{\tau_{42}^*}^2 + {\tau_{31}^*}^2}\end{cases} \tag{2}$$

where c is the sound velocity and τ_{ij}^* is the time delay between microphones M_i and M_j. We calculate the latter with the GCC-PHAT algorithm [52].

Horizontal motion. The rotation frequencies of two motors on the same side increase simultaneously to generate a lift force, for example T_1^v and T_4^v in Fig. 5(c), so that the sound contains two groups of BPF peaks, $f_1^{BPF} = f_4^{BPF}$ and $f_2^{BPF} = f_3^{BPF}$. Then the drone tilts with an angle γ, as shown in Fig. 5(c), so that we can decompose T_i^v into vertical and horizontal directions. The vertical component of T_i^v is balanced with the drone's gravity, so we can solve γ with the knowledge of the drone's mass m and the acceleration of gravity g, which are known. The horizontal component of T_i^v works against the resistance $F_f = \lambda^h(v_t^h)^2$ to make the drone move horizontally, where λ^h can be regarded as a constant related to γ. We solve the horizontal velocity v_t^h and acceleration a_t^h at time t with the γ by the following dynamics equations:

$$\begin{cases}\frac{k^v}{n^2}\sum_{i=1}^{N}(f_i^{BPF})^2\sin\gamma = mg \\ \frac{k^v}{n^2}\sum_{i=1}^{N}(f_i^{BPF})^2\cos\gamma - \lambda^h(v_t^h)^2 = ma_t^h\end{cases} \tag{3}$$

Vertical motion. Consider the case of climbing as an example: $f_i, i = 1, 2, 3, 4$ increase simultaneously to work against the gravity and downward resistance $F_f = \lambda^v(v_t^v)^2$, where λ^v can be regarded as a constant, illustrated in Fig. 5(d). Thus, only one BPF peak is captured. Vertical velocity v_t^v and acceleration a_t^v at time t can be determined by solving the equation:

$$\frac{k^v}{n^2}\sum_{i=1}^{N}(f_i^{BPF})^2 - mg - \lambda^v(v_t^v)^2 = ma_t^v \tag{4}$$

Finding coordinates. Consider the situation shown in Fig. 9, where a drone flies from S_t to S_{t+1}. A single 4-microphone array with elements M_1, M_2, M_3, M_4 is deployed to capture the acoustic signals. The coordinate of the drone at time t are $S_t(h_t\tan\phi_t\cos\alpha_t,\ h_t\tan\phi_t\sin\alpha_t, h_t)$, where the height h_t is now the only unknown quantity. Fortunately, determining h_t is not difficult. For two adjacent coordinates S_t and S_{t+1}, in the case of horizontal motion, $h_t = h_{t+1}$, so that

$$|h_{t+1}\tan\phi_{t+1} - h_t\tan\phi_t| = v_t^h\tau + \frac{1}{2}a_t^h\tau^2 \tag{5}$$

where τ is a predefined interval for location updating. In the case of vertical motion, we have

$$|h_{t+1} - h_t| = v_t^v\tau + \frac{1}{2}a_t^v\tau^2 \tag{6}$$

We solve these equations in h_t and determine the complete coordinates of the drone during the flight.

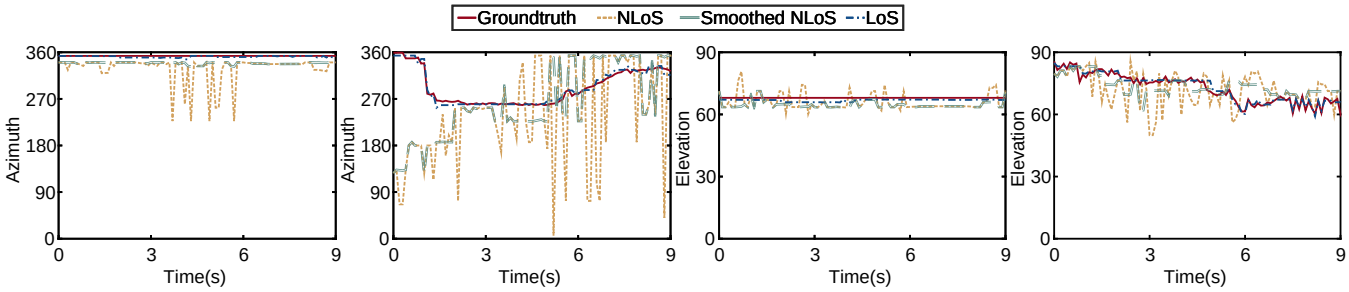
5.2 Tracking in NLoS

Indoor scenarios likely include objects that create NLoS settings, for example, in busy warehouses. Here, the DoA information captured by the microphone array may be deviated. For instance, the yellow dashed curves in Fig. 8 depicts the estimated DoA information in NLoS settings. The severe deviation occurs in NLoS no matter whether the drone moves. In this case, traditional triangulation with distributed microphone arrays cannot work, yet alternative indoor localization systems such as UWB- and infrared-based systems may be equally prevented from working altogether in such settings.

In contrast to the state of the art, AIM can recognize if the LoS is blocked and continue to track the drone in NLoS. Despite a few outliers, the dominated diffraction or reflection path with the highest signal energy is stable when the location of the drone is unchanged, while it is irregular when the drone moves. Thus, we employ the Interquartile Range rule (IQR) [53] to eliminate outliers and smooth the estimated DoA information in a sliding window.

When the drone is hovering or yawing, the estimated DoA is smooth, as in Fig. 8(a) and Fig. 8(c), even if the observations slightly deviate from the ground truth. Instead, the smoothed DoA information is erratic when the drone is moving, as in Fig. 8(b) and Fig. 8(d). As described in Tab. 1, we use the stability of DoA information rather than the absolute values to determine the kind of drone motion in LoS. Fig. 8 provides evidence that we can employ the same criteria for the NLoS case.

To detect the NLoS setting in the first place, AIM sets a threshold to evaluate the variance of smoothed azimuth



(a) Stable azimuth. (b) Unstable azimuth. (c) Stable elevation. (d) Unstable elevation.

Figure 8: DoA estimation results in LoS and NLoS.

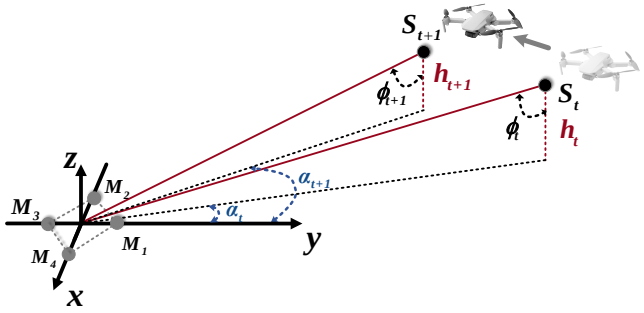


Figure 9: Schematic diagram of AIM in action.

information in a time window. If the variance is beyond the threshold, we consider the LoS to be blocked, because even if smoothed, the DoA in NLoS is still unstable, which is especially evident in azimuth estimation, as shown by the green curve in Fig. 8(b).

5.3 Error Calibration

We employ a dedicated Kalman filter to tame the inaccuracies in the estimation of orientation after yawing and in absolute localization following horizontal or vertical motion.

The drone location is described by a state vector $A_t = [x_t,\ y_t,\ z_t]^T$, with A_0 being initialized with the first few points at the beginning of the flight. Then processing unfolds as follows:

1) We predict the subsequent state vector $\hat{A}_t^-$, that is, the a priori state estimate, according to the state transition matrix;
2) We estimate the drone's current motion following the rules in Tab. 1 as well as the current coordinate according to the dynamic equations and identified motion;
3) Based on the variance of the smoothed azimuth, we identify whether the LoS exists. If not, the estimated DoA information is discarded;
4) With yaw motion, possible trajectories caused by the ambiguous orientations are tracked until the LoS is regained. If the LoS exists now, the current coordinates can be updated with DoA, eliminating the ambiguity;
5) No matter whether in LoS or NLoS, the measured coordinates are fused with $\hat{A}_t^-$ to output the optimal estimate $\hat{A}_t$, that is, the a posteriori state estimate.

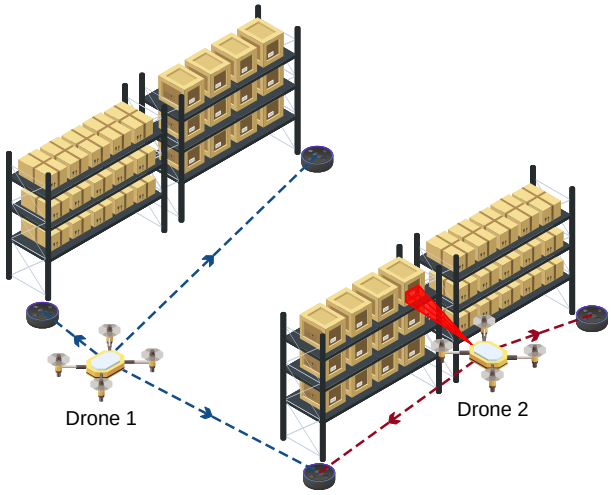


Figure 10: Example scenarios of tracking with multiple arrays.

6 Extending Operating Range

Despite the ability of AIM to operate with a single microphone array, in realistic indoor settings such as a warehouse, the coverage may be insufficient. As a result, we extend our tracking scheme using distributed microphone arrays to accommodate indoor environments with variable ranges and configurations.

6.1 Basic Model

An example of a warehouse employing distributed microphone arrays is depicted in Fig. 10. In this scenario, the arrays are positioned at regular intervals among the shelves to facilitate tracking of the drone through relaying. As the drone traverses these zones, we use neighboring microphone arrays to calculate its location and subsequently refine the results reported by acoustic inertial measurement, thereby enhancing localization accuracy.

Our approach involves computing the time difference of arrival (TDoA) between each pair of microphone arrays. We uniformly orient all arrays in the same direction and number their elements according to consistent rules. If we designate the m-microphone arrays Arr_p and Arr_q to have elements $M_1^p ... M_m^p$ and $M_1^q ... M_m^q$, respectively, the TDoA T_{pq} between the two arrays is determined as:

$$T_{pq} = \frac{\Sigma_{i=1}^{m} \tau^*(M_i^p, M_i^q)}{m} \tag{7}$$

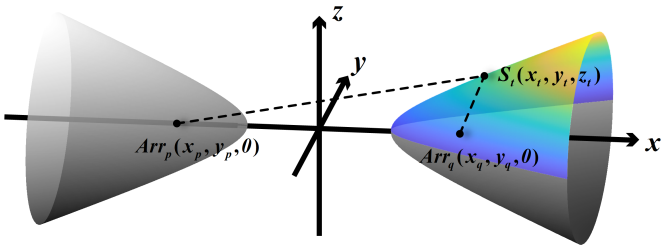


Figure 11: TDoA between two microphone arrays.

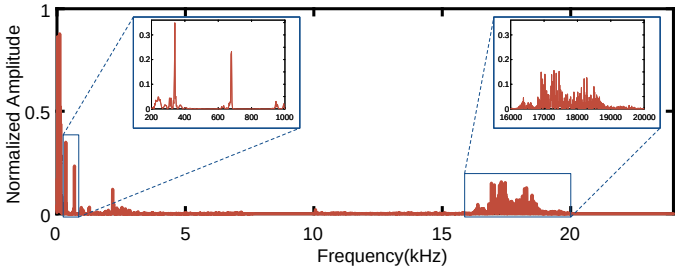


Figure 12: Spectrum when the beacon and drone sound exist simultaneously.

where $\tau^*(M_i^p, M_i^q)$ is the time delay between the corresponding elements of two arrays.

It follows that the locations of the drone that satisfy this TDoA form a hyperboloid, as depicted in Fig. 11. Here, we denote the drone's location at time t as $S_t(x_t, y_t, z_t)$ and the positions of the two arrays as $Arr_p(x_p, y_p, 0)$ and $Arr_q(x_q, y_q, 0)$. The shape of the hyperboloid is derived from the calculated TDoA as follows:

$$F(Arr_p, Arr_q) = \frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} - 1 \tag{8}$$

where $a = \frac{1}{2} \cdot \text{abs}\left(||S_t\ Arr_q|| - ||S_t\ Arr_p||\right) = \frac{1}{2}c \cdot T_{pq}$ and $b = c = \sqrt{\frac{1}{2} \cdot ||Arr_p\ Arr_q||^2 - a^2}$.

With at least three microphone arrays, say Arr_p, Arr_q, and Arr_s, we can estimate the drone's location at time t by solving the following set of equations:

$$S_t(x_t, y_t, z_t) = \begin{cases} F(Arr_p, Arr_q) = 0; \\ F(Arr_p, Arr_s) = 0; \\ F(Arr_q, Arr_s) = 0; \\ \vdots \end{cases} \tag{9}$$

To synchronize the microphone arrays involved in location estimation, we employ commercial speakers to intermittently emit an acoustic beacon, which consists of a pre-defined pseudo-random noise. During drone tracking, the microphone arrays detect this beacon to align with one another [54]. The beacon frequency ranges from 16kHz to 20kHz, as depicted in Fig. 12, and is distinct from the signals used for localization, making it separable via band-pass filters. As shown in Fig. 13(a) and Fig. 13(b), the beacon is accurately detected also when the drone is present.

There may be cases where a drone can establish line of sight with only two microphone arrays, as for drone 2 in Fig. 10. If so, Eq. (9) becomes negative definite or non-full rank, rendering it unsolvable and it becomes impossible to obtain the 3D coordinates of the drone. To address this issue, we no longer treat each microphone array as a whole, as in Eq. (8), and instead choose multiple individual microphone elements for localization.

Say we can only rely on two microphone arrays[1]. In this case, we choose two elements from each array, respectively, and consider each of them as a new two-microphone array. The distance between two elements in the same array must be the largest, and furthermore, the selected four elements must not be collinear. We can then employ the model in Sec. 6.1 to calculate the drone's location using the four selected microphone elements.

1. If there are multiple arrays arranged on a single line, we select the two nearest to the drone.

6.2 Selecting Microphone Arrays for Localization

Although there may be several microphone arrays located near the drone that can receive the acoustic signal with a high amplitude, some of them may be in NLoS or surrounded by multiple reflectors. If these arrays are chosen to perform TDoA and location calculation, the resulting localization information may be inaccurate. To mitigate this issue, we execute a dedicated array selection algorithm, which is depicted in Algorithm 1.

We perform a preliminary screening using the method outlined in Sec. 5.2 to filter out microphone arrays that report unstable or inaccurate results. If the number of the remaining arrays is enough to determine the 3D coordinates of the drone, that is, there are at least 3 arrays that show reliable DoA estimation, we proceed to the following fine-grained selection process. Otherwise, if all arrays are in a line or only two arrays can be used, we calculate the location as described above.

Next, we apply an additional filtering process to further refine the selected microphone arrays. Let d_{pq} denote the distance between two microphone arrays Arr_p and Arr_q. We select the first three microphone arrays as the initial set, where the product of the distance between them is the largest. This is because TDoA estimates tend to be more accurate when microphone arrays are more dispersed.

With the selected three microphone arrays, we obtain an initial estimation of the drone's coordinates. This estimation may not be stable enough as it is based on only three microphone arrays. If there is any other candidate microphone array providing preferable DoA estimation, we add this to the processing to improve the accuracy, choosing the one with the most stable DoA estimations. To further enhance the accuracy, results reported by the distributed microphone arrays are fused with those obtained from acoustic inertial measurement, as explained next.

6.3 Fusing Data

After obtaining the drone location from Eq. (9), we fuse this result with that of the acoustic inertial measurement. We use the complementary filter for this, because of two reasons. First, location estimations output by distributed microphone arrays can exhibit jitter, which results in high-frequency noise, while estimations of acoustic inertial measurement

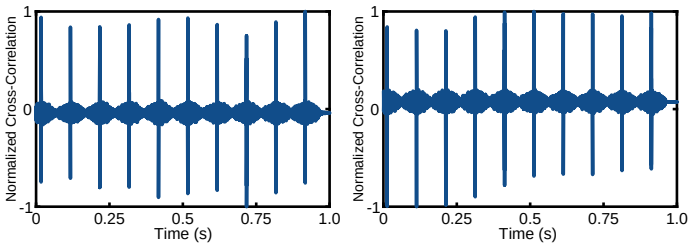


Figure 13: Detecting the presence of the beacon: (a) when the drone does not take off; (b) when the drone is hovering.

Algorithm 1: Array Selection Algorithm.

for $t = 1, 2, 3, ...$ **do**
  Determine the set of microphone arrays C_{Arr};
  if *COUNT*(C_{Arr})> 3 **then**
    for $Arr_p \in C_{Arr}$ **do**
      if *Variance* $Var_{\hat{\alpha}_p} < Threshold$ **then**
        Add Arr_p to the candidate set $C_{selected}$;
      end
    end
    Load the distance d_{pq} between every microphone array in $C_{selected}$;
    Initial set $C_{initial} = \arg\max_{(p,q,s)} d_{pq} * d_{ps} * d_{qs}$;
    for $Arr_p \in C_{selected} - C_{initial}$ **do**
      Find the $Arr_p^{'}$ with the minimum $Var_{\hat{\alpha}_p}$;
    end
    Add $Arr_p^{'}$ to $C_{initial}$;
  end
  Output $C_{initial}$;
end

are smooth in a short period of time, which can therefore effectively compensate for this problem. On the other hand, both distributed microphone arrays and acoustic inertial measurements produce fairly accurate results so we can employ the lightweight complementary filter to avoid nesting of two Kalman filters, greatly reducing processing times.

Let $s(\Delta t)$ denote the true trajectory of the drone over a time period Δt, so we have

$$\begin{aligned} z_M(\Delta t) &= s(\Delta t) + n_1(\Delta t) \\ z_A(\Delta t) &= s(\Delta t) + n_2(\Delta t) \end{aligned} \tag{10}$$

where $z_M(\Delta t)$ and $n_1(\Delta t)$ are the estimation results and noise of the distributed microphone arrays, and $z_A(\Delta t)$ and $n_s(\Delta t)$ are those of the acoustic inertial measurement. Then we perform a data fusion process based on

$$\hat{S}(f) = Z_M(f)G(f) + Z_A(f)\left[1 - G(f)\right] \tag{11}$$

where $\hat{S}(f)$ is the Fourier transform of the fused result $\hat{s}(\Delta t)$, $Z_M(f)$ and $Z_A(f)$ are the Fourier transform of $z_M(\Delta t)$ and $z_A(\Delta t)$, and $G(f)$ and $1 - G(f)$ is the low-pass filter and the complementary high-pass filter.

Finally, we can obtain the fused result $\hat{s}(\Delta t)$ by performing inverse Fourier transform for $\hat{S}(f)$.

7 Evaluation

We report evaluation results of AIM using off-the-shelf microphone arrays and a commercial drone. We describe first the implementation and evaluation settings in Sec. 7.1. Next, our investigation of AIM performance is two-pronged: Sec. 7.2 compares our system with the state-of-the-art indoor drone tracking systems and reports on their performance under different scenarios; in Sec. 7.3, we dissect the impact on tracking accuracy of environment noise, flight range and velocity, as well as of the deployment configurations of distributed microphone arrays and of the beacon volume. We discuss the real-world performance of AIM in Sec. 7.4.

Our results indicate that:

1) The mean localization error of AIM in NLoS settings, arguably most realistic for indoor drone applications, is 46% lower than a UWB-based baseline;
2) Unlike an infrared-based baseline, AIM constantly provides location updates, even in NLoS settings;
3) AIM is robust to moderate noise sources in the environment, such as someone speaking;
4) Flight range and velocity of the drone influence AIM's performance differently, yet the absolute accuracy never degrades drastically.
5) With distributed microphone arrays, AIM can be extended to support indoor spaces with arbitrary ranges and layouts without loss of accuracy.

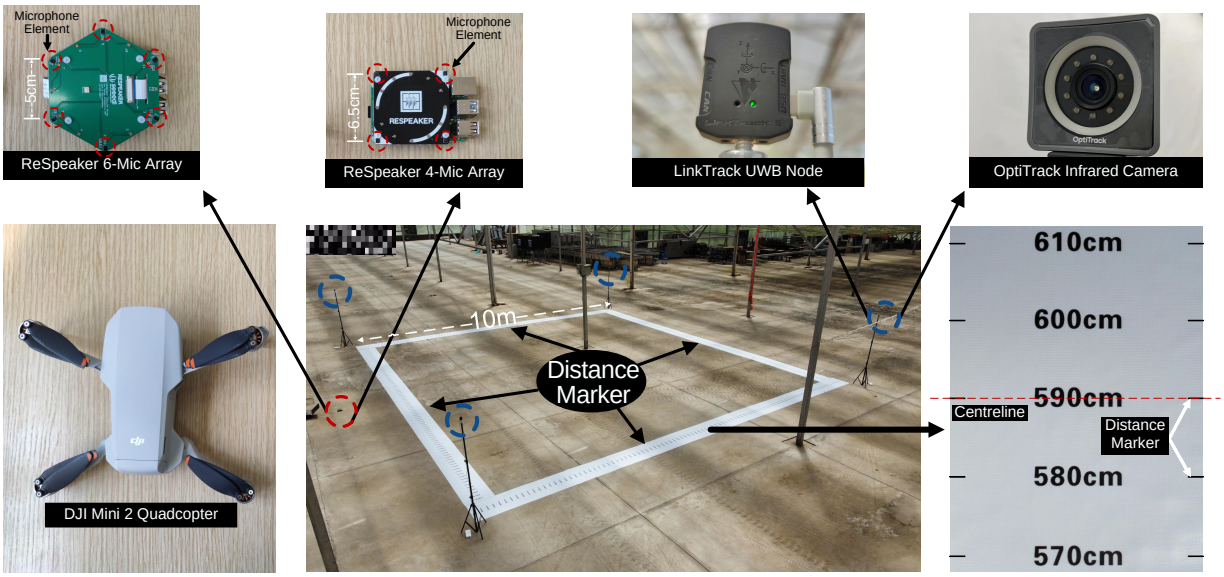
7.1 Implementation and Settings

AIM works with any layout of bidimensional microphone array to track drones of various structures. Without loss of generality, here we consider a quadcopter and two types of microphone arrays.

Drones and microphone arrays. We use a DJI Mini 2 quadcopter [55], shown in Fig. 14(a). The DJI Mini 2 weighs 249g; as such, flying the DJI Mini 2 in most countries does not require a professional drone piloting license, which makes it ideal for indoor use. Each propeller is equipped with two blades. When the drone is hovering, the sound pressure level measured at a 1m distance is empirically determined to be around 77dB and motors run at 164Hz, so the BPF is around 328Hz. By default, the DJI Mini utilizes the built-in GPS for horizontal localization and an infrared time of flight (ToF) sensor to obtain vertical altitude. However, in the indoor experimental environment we use, shown in Fig. 14(b), GPS cannot work and only the ToF sensor provides useful altitude information.

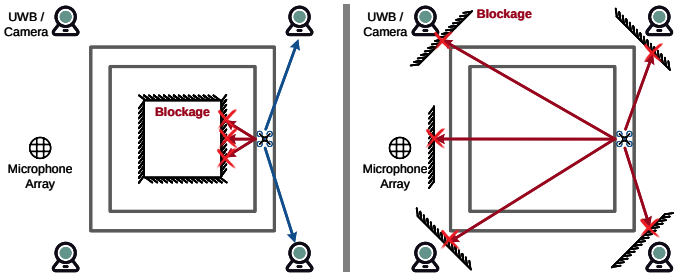
We use two types of commercial off-the-shelf microphone arrays for our AIM prototype: a Seeed Studio ReSpeaker 6-mic circular array [56] and Seeed Studio ReSpeaker 4-mic array [57], shown on the upper left of Fig. 14. The inter-distance between two single microphones is 5cm and 6.5cm, respectively. Each microphone array is set on a Raspberry Pi 4 Model B, using a 48kHz sampling rate. Unless stated otherwise, the results we discuss next are obtained with the 6-mic circular microphone array.

Baselines. To obtain ground-truth information, we take the readings of the built-in ToF sensor on the DJI Mini 2 as vertical altitude. As for the horizontal coordinates, we



(a) Drone. (b) Experiment area. (c) Distance marker.

Figure 14: Experiment settings.



(a) PLoS (b) NLoS

Figure 15: Experiment scenarios.

employ a method often used in indoor drone testbeds [58]: we lay down distance markers on the ground at intervals of 10cm, as shown in Fig. 14(b) and Fig. 14(c). Using the downward-facing camera of the drone, we examine its view of the ground-level markers during the flight. Fig. 14(c) shows an example image captured by the drone during the experiments. Once the tick of the marker matches the centerline of the image, this reading of the corresponding maker is regarded as the real-time horizontal coordinates.

We compare AIM with LinkTrack [59], an UWB-based indoor localization system, and OptiTrack [38], an infrared-based motion tracking system, both of which are shown on the upper right of Fig. 14. LinkTrack localizes the target via triangulation. We fix a UWB tag on the drone and four UWB anchors on four tripods, then record the tracking results on a base station. OptiTrack localizes the target by converting the drone positions in bidimensional photos captured at high frequency by multiple infrared cameras to three-dimensional coordinates. We fix reflective markers on the drone and four infrared cameras on four tripods, and also record the tracking results on a base station. Whenever the drone carries a UWB tag or reflective markers, we accordingly update its tracking model and dynamic parameters.

Note that the OptiTrack system is vastly considered as state of the art in indoor testbeds. Because of its cost, difficulty in installation, and inability to work in NLoS settings, however, it is rarely employed for real applications [58].

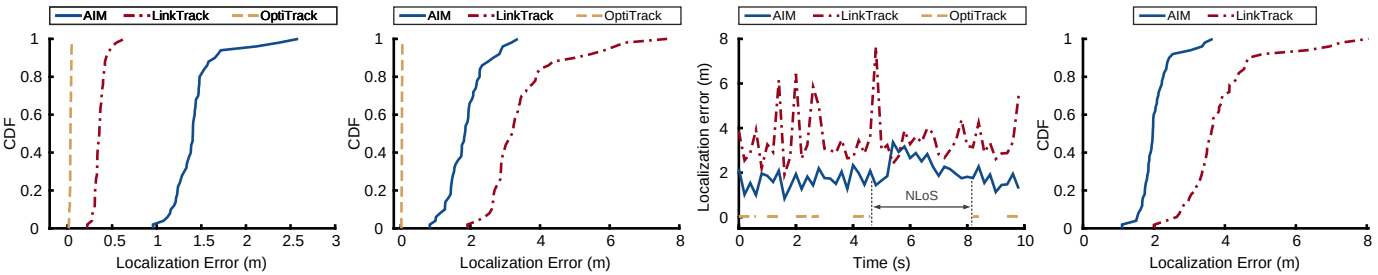
Scenarios and drone mobility. We select three scenarios. In *Line-of-Sight* (LoS), nothing is deployed in the middle of the experiment area shown in Fig. 14(b) and every device involved in localization can establish LoS with each other and with the drone. Note how this scenario, while common in indoor drone testbeds that are in fact designed to isolate drones from their surroundings, is quite unlikely in real applications. In *Partial Line-of-Sight* (PLoS), several steel shelves stacked with various objects such as books and bricks are deployed in the middle of the experiment area. As shown in Fig. 15(a), depending on the relative position of the drone with respect to the rest of the experiment area, the LoS is blocked at times. In *None-Line-of-Sight* (NLoS), the shelves are deployed in front of every tripod hosting infrastructure node for localization. Every LoS path is thus blocked, as depicted in Fig. 15(b). No matter where the drone flies in the experiment field, it can not establish LoS connection to any device on any of the tripods.

We tested varied combinations of drone motions. For *horizontal motions*, we control the drone to fly along the distance maker, shown in Fig. 14(c), and keep vertical coordinates unchanged. For *vertical motions*, once the drone is hovering, we control the drone to climb or descent to a certain height, while keeping horizontal coordinates unchanged.

7.2 General Performance

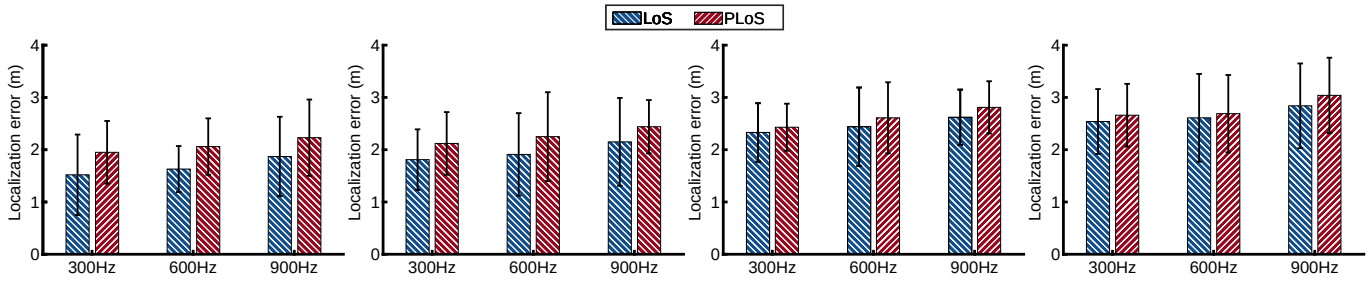
We fly a 10m×10m squared trajectory comparing AIM with LinkTrack and OptiTrack in LoS, PLoS and NLoS scenarios. Fig. 16 reports the performance of the three systems.

Fig. 16(a) indicates that in LoS scenarios, the mean error of AIM is 1.43m while those of LinkTrack and OptiTrack



(a) LoS scenarios. (b) PLoS scenarios. (c) Variation of errors. (d) NLoS scenarios.

Figure 16: Performance comparison.



(a) 50dB. (b) 55dB. (c) 60dB. (d) 65dB.

Figure 17: Impact of environment noise on accuracy.

are 0.37m and 0.03m, respectively[2]. AIM is, therefore, the least accurate system in LoS scenarios, which are, however, arguably rare in real applications.

Fig. 16(b) illustrates the performance in PLoS scenarios. Here AIM outperforms LinkTrack with a mean error of 1.89m, which is 46% less than LinkTrack. The increase of error is caused by the lack of DoA calibration for AIM and by signal attenuation for LinkTrack. In that case, AIM can only calibrate the location with the opportunistic clean DoA.

Fig. 16(c) offers a closer view on this specific experiment by showing an accuracy comparison during a 10s flight, including about 2s of NLoS. LinkTrack is heavily influenced by the obstacles, which absorb UWB signals. When the LoS is obstructed, OptiTrack simply does not work and produces no output. Thus, although its mean error does not increase in PLoS scenarios, OptiTrack is plainly inapplicable as completely losing the drone position even for a short among of time would be unacceptable for safe and dependable operation. Instead, the localization error of AIM suddenly increases at the beginning of the NLoS sting, but gradually decreases later, without ever losing the target.

In NLoS scenarios, shown in Fig. 16(d), we only compare AIM with LinkTrack because OptiTrack produces no output for the entire duration of the experiments, because of the aforementioned reasons. The mean error of AIM increases to 2.08m but it is still lower than that of LinkTrack, which is almost twice as much at around 4m.

Note how the progression through different scenarios in our discussion, from LoS in Fig. 16(a) to NLoS in Fig. 16(d), reflects increased realism in indoor drone applications. NLoS settings are indeed expected to abound when drones fly in complex physical environments. These settings are precisely where AIM reaps the greatest benefits compared to the baselines: its performance degradation, indeed, is much less pronounced compared to LinkTrack, whereas it can supply continuous location updates, unlike OptiTrack.

7.3 Factors Influencing Accuracy

We analyze the impact of three different factors on localization accuracy, that is, noise in the environment, the flight range and velocity, and the number of microphones.

Environment noise. We examine the performance of AIM in noisy conditions. We place a noise source 2m away from the microphone array. To study different degrees of interference, we set the volume of the noise source to 50dB, 55dB, 60dB and 65dB. We broadcast Gaussian white noise with 100Hz bandwidth in three different center frequencies, that is, at 300Hz, 600Hz and 900Hz, to simulate interference on the BPF and its harmonic frequency.

The results in Fig. 17 indicate that, as expected, the localization accuracy degrades as the frequency of the noise or the SPL of the noise increases. This is because AIM weights the BPF and its harmonics according to their amplitude and sums them up to obtain the final frequency, which is the input of dynamic equations. In general, BPF and lower harmonics exhibit higher energy and thus are given higher weights. However, if the noise is at high frequency, peaks in this frequency band gain much higher weights. Therefore, the results are polluted.

Importantly, results show that AIM still maintains relatively stable performance under noisy conditions, which is sufficient to deal with common noise environments such as

2. Note that for OptiTrack, we note a difference between the error measured in our experiments and what is advertised by the manufacturer, which is below 1mm. The reason for this is that OptiTrack sometimes temporarily recognizes LEDs on the drones as the markers, affecting the measurements. We cannot turn off or cover these LEDs, as the drone would refuse to take off, raising exceptions in the control software.

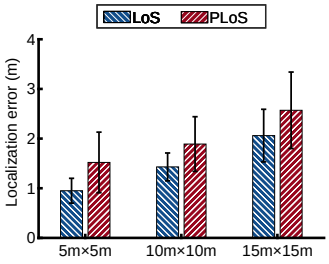


Figure 18: Flight range.

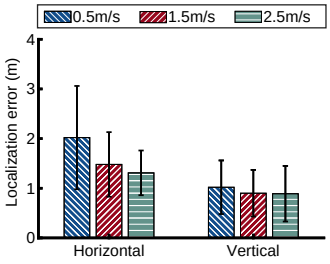


Figure 19: Flight velocity.

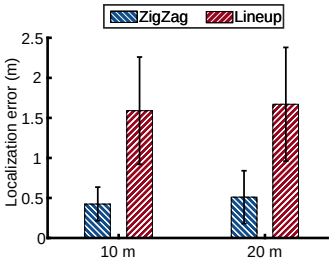


Figure 20: Deployment of mics.

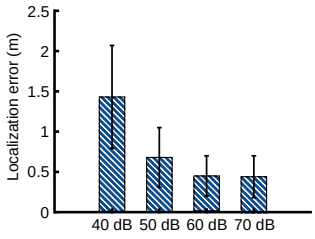


Figure 21: Volume of beacon.

someone speaking, which is around 53.7dB at 1m distance. We also demonstrate that AIM can cope with narrowband noise, whose frequency band does not violate all the BPF and harmonic frequencies simultaneously. Even faced with broadband noise (*e.g.*, music), AIM still provides accurate localization results as long as the noise intensity is lower than that of the drone signals. If not, multiple options exist to resist noise in practice. We may, for example, introduce a band-pass filter to filter out the noise band and continue tracking using the uncontaminated frequency band. AIM is also flexible in the deployment of the microphone array, as no specific requirements must be fulfilled to during installation. We may simply alter its position to lessen the impact of nearby noise sources.

Flight range and velocity. First, we investigate the performance of AIM depending on the distance between the drone and the microphone array. We specifically test three flight paths, composed of 5m×5m, 10m×10m, and 15m×15m square trajectories. The drone is controlled to fly at a velocity of 1.5m/s in both horizontal and vertical motions. Fig. 18 shows the results.

When the drone flies along the 5m×5m square, the mean errors are 0.95m in LoS and 1.52m in PLoS. When the drone flies along the 10m×10m square, the mean errors are 1.43m in LoS and 1.89m in PLoS. If the drone flies over a larger area, the signal attenuation worsens so the error increases. Correspondingly, the results show that the mean errors in both LoS and PLoS are over 2m as the drone flies along a 15m×15m field.

Based on these results, we define 10m as the *operational range* for the pair DJI Mini 2/ReSpeaker 6-mic. The operational range is an empirical value, which sets a limit on the acceptable tracking error. Note that this value may be different between different drones and microphone arrays, as it is mainly determined by the SPL of the sound produced by the drone's propellers and the sensitivity of the microphone array. The higher the drone's SPL and the array's sensitivity, the lower the tracking error in a given field and the larger the operational range.

We also conduct experiments to evaluate if the drone's velocity has an impact on accuracy. These experiments are conducted in the LoS scenario, and both horizontal and vertical motion are evaluated, respectively. In the horizontal motion, we control the drone to fly along a 10m×10m square. The results are shown in Fig. 19. For horizontal motion, the drone's velocity influences the accuracy in that the mean error decreases as the velocity increases, while for vertical motion, the change of velocity does not significantly impact accuracy. The reason is two-fold. On the one hand, two frequency peaks must be captured for horizontal motion. Higher velocity results in larger intervals between the two frequency peaks, hence they are easier to separate out. In contrast, only one peak must be captured during vertical motion. On the other hand, every two propellers contribute to the energy of one frequency peak with horizontal motion, while all propellers generate the signal at the same frequency with vertical motion. The energy of the frequency peak in vertical motion is higher than that in horizontal motion and, therefore, results in more stable performance.

Deployment of microphone arrays and beacon volume. We evaluate the localization accuracy in continuous drone tracking by varying the deployment of microphone arrays and the volume of the beacon.

Firstly, we deploy several microphone arrays in two different configurations: ZigZag and straight lines. Then, we compare the localization accuracy of these two deployments in the 10m and 20m range. In the straight line setting, we place several arrays in a line, and to simulate the corner of the warehouse, we also place one array at the end of the line that is not colinear with the others. This arrangement provides the opportunity to perform error calibration with at least three microphone arrays. In the ZigZag setting, the arrays are placed in two lines as a form of ZigZag and the distance between the two lines is 10m. The drone is controlled to fly along the center line of two lines, so the horizontal distance between the drone and each microphone is around 5m. The drone velocity is 1.5m/s and all microphone arrays can establish a LoS with the drone.

The results in Fig. 20 show that the ZigZag configuration provides much better accuracy, with errors less than 0.5m, in both the 10m and 20m range. As the horizontal distance between the drone and each microphone array during flight is around 5m, and the flight height is 2m, the relative error in this setting is less than 9.28% $\left(0.5/\sqrt{5^2+2^2}\right)$. In contrast, the errors in the straight line setting are around 1.5m, even with the opportunity for calibration. Thus, we recommend deploying distributed microphone arrays as in the ZigZag configuration for better performance, if conditions permit.

We also investigate the impact of varying the volume of the time synchronization beacon. The experiments are conducted with microphone arrays deployed in the ZigZag configuration, while the drone flies in the 10m range with the velocity of 1.5m/s. As Fig. 21 shows, increasing the volume of the beacon leads to a reduction in localization error. Specifically, the error decreases from 1.43m at 40dB to 0.45m and 0.44m, at 60dB and 70dB, respectively. How-

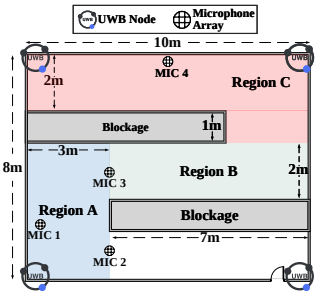


Figure 22: Warehouse layout.

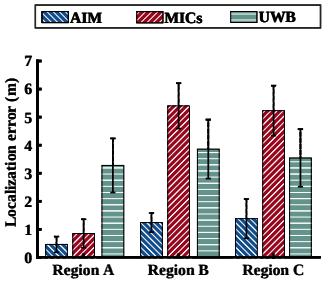


Figure 23: Accuracy.

ever, noise can also exist in the band of the beacon, even with high frequency, and therefore, higher volumes may not always result in better performance. Moreover, some industrial settings may have strict regulations on sound volume, including those in the frequency range that is not audible to humans. To address these limitations, we may extend the length of the beacon, instead of increasing the volume, which can compensate for the reduction in volume without affecting the performance.

7.4 Performance in Realistic Settings

We offer further evidence on the real-world applicability of AIM. The instrument we use to this end is a real deployment in a warehouse, whose layout is shown in Fig. 22. At three different regions in the warehouse we compare the localization accuracy of AIM with that of LinkTrack and triangulation using distributed microphone arrays, which is usually used in many acoustics-based localization methods [60], [61], [62].

Fig. 23 reports the results. In Region A, triangulation achieves a fair accuracy with a mean error of 0.85m. In comparison, AIM reports shows more accurate results with a mean error of 0.46m. The reason is that AIM can fuse the results from distributed microphone arrays to output more precise and stable results. When the drone enters Region B and Region C, triangulation becomes inapplicable, as it returns an error above 5m, but AIM's performance is not affected. This is because our system only requires one LoS to disambiguate or not even that, whenever the drone does not perform yaw motion in NLoS. In contrast, for triangulation to work, LoS from all microphone arrays is mandatory.

As for LinkTrack, we set the four UWB anchors at the corners of the area to cover the whole warehouse, as shown in Fig. 22. In such a deployment configuration, LinkTrack performs poorly in all three regions because of the signal loss caused by the obstacles in the warehouse.

8 Discussion

We complete the discussion of AIM by articulating practical issues of applicability and general use. Two aspects are worth considering here.

Sensor fusion for indoor tracking. Different techniques have their unique advantages and disadvantages. Multiple techniques could be combined to improve performance.

Most existing commercial drones are already equipped with multiple sensors, including ToF, IMU and cameras, for accurate indoor localization. In the context of passive drone tracking, sensor fusion is also feasible. For example, one may deploy UWB nodes or cameras at the corner to calibrate the drone's location, while exploiting microphone arrays in other places to reduce cost. Besides, in PLoS indoor scenarios like Fig. 15(a), the drone can establish LoS paths with at least two sensors in most cases. Therefore, a real-time sensor fusion algorithm can be applied to achieve accurate localization results. However, strict time synchronization between different sensors and quick identification of LoS paths are required.

Multi-drone tracking. When multiple drones enter the same area, AIM can still track them separately if their BPF are different. Otherwise, frequency aliasing happens. We may handle this problem by borrowing ideas from existing works to discriminate different sound sources along different propagation paths [44] or to modulate the unique acoustic signature in the drone motor sound [18].

9 Conclusion

We presented AIM, a one-of-a-kind passive indoor drone tracking technique that works with a single microphone array, but may also be extended to support spaces with any range and layout by deploying distributed microphone arrays. AIM innovates the acoustic tracking technique in that it fully exploits the dual acoustic channel from the drone to the microphone array, based on an in-depth understanding of the drone's dynamics and the characteristics of its acoustic signal. Through extensive experiments, we demonstrate that AIM offers strikingly better performance than state-of-the-art solutions, especially in NLoS settings, and enjoys stable performance across complex indoor environments.

References

[1] A. El Yaacoub, L. Mottola, T. Voigt, and P. Rümmer, "Scheduling Dynamic Software Updates in Safety-Critical Embedded Systems-the Case of Aerial Drones," in *Proceedings of ACM/IEEE ICCPS*, 2022.

[2] K. Mahima, M. Weerasekara, K. De Zoysa, C. Keppitiyagama, M. Flierl, L. Mottola, and T. Voigt, "Fighting Dengue Fever with Aerial Drones," in *EWSN*, 2022.

[3] M. Nagai, T. Chen, R. Shibasaki, H. Kumagai, and A. Ahmed, "UAV-Borne 3-D Mapping System by Multisensor Integration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 701–708, 2009.

[4] A. Bekar, M. Antoniou, and C. J. Baker, "Low-Cost, High-Resolution, Drone-Borne SAR Imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[5] E. Frachtenberg, "Practical Drone Delivery," *IEEE Computer*, vol. 52, no. 12, pp. 53–57, 2019.

[6] Y. He, W. Wang, L. Mottola, S. Li, Y. Sun, J. Li, H. Jing, T. Wang, and Y. Wang, "Acoustic Localization System for Precise Drone Landing," *IEEE Transactions on Mobile Computing*, 2023.

[7] M. Ezuma, O. Ozdemir, C. K. Anjinappa, W. A. Gulzar, and I. Guvenc, "Micro-UAV Detection with a Low-Grazing Angle Millimeter Wave Radar," in *Proceedings of IEEE RWS*, 2019.

[8] Y. Sun, W. Wang, L. Mottola, R. Wang, and Y. He, "AIM: Acoustic Inertial Measurement for Indoor Drone Localization and Tracking," in *Proceedings of ACM SenSys*, 2022.

[9] P. Nguyen, T. Kim, J. Miao, D. Hesselius, E. Kenneally, D. Massey, E. Frew, R. Han, and T. Vu, "Towards RF-Based Localization of a Drone and Its Controller," in *Proceedings of ACM DroNet*, 2019.

[10] Bitcraze, "Lighthouse Positioning System," 2023. [Online]. Available: https://www.bitcraze.io/documentation/system/positioning/ligthouse-positioning-system/

[11] D. Solomitckii, M. Gapeyenko, V. Semkin, S. Andreev, and Y. Koucheryavy, "Technologies for Efficient Amateur Drone Detection in 5G Millimeter-Wave Cellular Infrastructure," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 43–50, 2018.

[12] S. Basak and B. Scheers, "Passive Radio System for Real-Time Drone Detection and DoA Estimation," in *Proceedings of IEEE ICMCIS*, 2018.

[13] H. D. K. Motlagh, F. Lotfi, H. D. Taghirad, and S. B. Germi, "Position Estimation for Drones Based on Visual SLAM and IMU in GPS-Denied Environment," in *Proceedings of IEEE ICRoM*, 2019.

[14] Z. Li and Y. Zhang, "Constrained ESKF for UAV Positioning in Indoor Corridor Environment Based on IMU and WiFi," *MDPI Sensors*, vol. 22, no. 1, p. 391, 2022.

[15] J. Kim, C. Park, J. Ahn, Y. Ko, J. Park, and J. C. Gallagher, "Real-Time UAV Sound Detection and Analysis System," in *Proceedings of IEEE SAS*, 2017.

[16] L. Mottola and K. Whitehouse, "Fundamental Concepts of Reactive Control for Autonomous Drones," *Communications of the ACM*, vol. 61, no. 10, pp. 96–104, 2018.

[17] E. Bregu, N. Casamassima, D. Cantoni, L. Mottola, and K. Whitehouse, "Reactive Control of Autonomous Drones," in *Proceedings of ACM MobiSys*, 2016.

[18] A. Bannis, H. Y. Noh, and P. Zhang, "Bleep: Motor-Enabled Audio Side-Channel for Constrained UAVs," in *Proceedings of ACM MobiCom*, 2020.

[19] M. Meles, A. Rajasekaran, K. Ruttik, R. Virrankoski, and R. Jäntti, "Measurement Based Performance Evaluation of Drone Self-Localization Using AoA of Cellular Signals," in *Proceedings of IEEE WPMC*, 2021.

[20] J. Zhang, X. Na, R. Xi, Y. Sun, and Y. He, "mmHawkeye: Passive UAV Detection with a COTS mmWave Radar," in *Proceedings of IEEE SECON*, 2023.

[21] Bitcraze, "Loco Positioning System," 2022. [Online]. Available: https://www.bitcraze.io/documentation/system/positioning/loco-positioning-system/

[22] J. Zhang, R. Xi, Y. He, Y. Sun, X. Guo, W. Wang, X. Na, Y. Liu, Z. Shi, and T. Gu, "A Survey of mmWave-Based Human Sensing: Technology, Platforms and Applications," *IEEE Communications Surveys & Tutorials*, 2023.

[23] M. Khalaf-Allah, "Particle Filtering for Three-Dimensional TDoA-Based Positioning Using Four Anchor Nodes," *MDPI Sensors*, vol. 20, no. 16, p. 4516, 2020.

[24] J. Tiemann, A. Ramsey, and C. Wietfeld, "Enhanced UAV Indoor Navigation through SLAM-Augmented UWB Localization," in *Proceedings of IEEE ICC workshops*, 2018, pp. 1–6.

[25] F. J. Perez-Grau, F. Caballero, L. Merino, and A. Viguria, "Multi-Modal Mapping and Localization of Unmanned Aerial Robots Based on Ultra-Wideband and RGB-D Sensing," in *Proceedings of IEEE IROS*, 2017.

[26] J. Tiemann and C. Wietfeld, "Scalable and Precise Multi-UAV Indoor Navigation Using TDOA-Based UWB Localization," in *Proceedings of IEEE IPIN*, 2017.

[27] J. F. Coll, P. Ängskog, C. Karlsson, J. Chilo, and P. Stenumgaard, "Simulation and Measurement of Electromagnetic Radiation Absorption in a Finished-Product Warehouse," in *Proceedings of IEEE International Symposium on Electromagnetic Compatibility*, 2010.

[28] W. Wang, L. Mottola, Y. He, J. Li, Y. Sun, S. Li, H. Jing, and Y. Wang, "MicNest: Long-Range Instant Acoustic Localization of Drones in Precise Landing," in *Proceedings of ACM SenSys*, 2022.

[29] W. Wang, Y. He, M. Jin, Y. Sun, and X. Guo, "Meta-Speaker: Acoustic Source Projection by Exploiting Air Nonlinearity," in *Proceedings of ACM MobiCom*, 2023.

[30] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun, "Indoor Follow Me Drone," in *Proceedings of ACM MobiSys*, 2017.

[31] T. Blanchard, J.-H. Thomas, and K. Raoof, "Acoustic Localization and Tracking of a Multi-Rotor Unmanned Aerial Vehicle Using an Array with Few Microphones," *The Journal of the Acoustical Society of America*, vol. 148, no. 3, pp. 1456–1467, 2020.

[32] H. Madokoro, S. Yamamoto, K. Watanabe, M. Nishiguchi, S. Nix, H. Woo, and K. Sato, "Prototype Development of Cross-Shaped Microphone Array System for Drone Localization Based on Delay-and-Sum Beamforming in GNSS-Denied Areas," *MDPI Drones*, vol. 5, no. 4, p. 123, 2021.

[33] G. Chi, Z. Yang, J. Xu, C. Wu, J. Zhang, J. Liang, and Y. Liu, "Wi-Drone: Wi-Fi-Based 6-DoF Tracking for Indoor Drone Flight Control," in *Proceedings of ACM MobiSys*, 2022.

[34] S. Jung, S. Hwang, H. Shin, and D. H. Shim, "Perception, Guidance, and Navigation for Indoor Autonomous Drone Racing Using Deep Learning," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2539–2544, 2018.

[35] D. Palossi, A. Loquercio, F. Conti, E. Flamand, D. Scaramuzza, and L. Benini, "A 64-mW DNN-Based Visual Navigation Engine for Autonomous Nano-Drones," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8357–8371, 2019.

[36] P. Chhikara, R. Tekchandani, N. Kumar, V. Chamola, and M. Guizani, "DCNN-GA: A Deep Neural Net Architecture for Navigation of UAV in Indoor Environment," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4448–4460, 2020.

[37] G. De Croon and C. De Wagter, "Challenges of Autonomous Flight in Indoor Environments," in *Proceedings of IEEE IROS*, 2018.

[38] OptiTrack, "OptiTrack Motion Tracking System," 2023. [Online]. Available: https://optitrack.com/

[39] C. Ruiz, S. Pan, A. Bannis, X. Chen, C. Joe-Wong, H. Y. Noh, and P. Zhang, "Idrone: Robust Drone Identification through Motion Actuation Feedback," *Proceedings of ACM IMWUT*, vol. 2, no. 2, pp. 1–22, 2018.

[40] J. Wang and E. Olson, "AprilTag 2: Efficient and Robust Fiducial Detection," in *Proceedings of IEEE IROS*, 2016, pp. 4193–4198.

[41] C. Cai, R. Zheng, and J. Luo, "Ubiquitous Acoustic Sensing on Commodity IoT Devices: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 432–454, 2022.

[42] D. Li, J. Liu, S. I. Lee, and J. Xiong, "FM-Track: Pushing the Limits of Contactless Multi-Target Tracking Using Acoustic Signals," in *Proceedings of ACM SenSys*, 2020.

[43] S. Shen, D. Chen, Y.-L. Wei, Z. Yang, and R. R. Choudhury, "Voice Localization Using Nearby Wall Reflections," in *Proceedings of ACM MobiCom*, 2020.

[44] W. Wang, J. Li, Y. He, and Y. Liu, "Symphony: Localizing Multiple Acoustic Sources with a Single Microphone Array," in *Proceedings of ACM SenSys*, 2020.

[45] C. Cai, H. Pu, P. Wang, Z. Chen, and J. Luo, "We Hear Your Pace: Passive Acoustic Localization of Multiple Walking Persons," *Proceedings of ACM IMWUT*, vol. 5, no. 2, pp. 1–24, 2021.

[46] T. Zheng, C. Cai, Z. Chen, and J. Luo, "Sound of Motion: Real-Time Wrist Tracking with a Smart Watch-Phone Pair," in *Proceedings of IEEE INFOCOM*, 2022.

[47] Z. Yang, Y.-L. Wei, S. Shen, and R. R. Choudhury, "Ear-AR: Indoor Acoustic Augmented Reality on Earphones," in *Proceedings of ACM MobiCom*, 2020.

[48] G. Cao, K. Yuan, J. Xiong, P. Yang, Y. Yan, H. Zhou, and X.-Y. Li, "EarphoneTrack: Involving Earphones into the Ecosystem of Acoustic Motion Tracking," in *Proceedings of ACM SenSys*, 2020.

[49] T. Luukkonen, "Modelling and Control of Quadcopter," *Independent Research Project in Applied Mathematics*, vol. 22, no. 22, 2011.

[50] H. Bi, F. Ma, T. D. Abhayapala, and P. N. Samarasinghe, "Spherical Array Based Drone Noise Measurements and Modelling for Drone Noise Reduction via Propeller Phase Control," in *Proceedings of IEEE WASPAA*, 2021, pp. 286–290.

[51] H. Kolamunna, T. Dahanayaka, J. Li, S. Seneviratne, K. Thilakaratne, A. Y. Zomaya, and A. Seneviratne, "DronePrint: Acoustic Signatures for Open-set Drone Detection and Identification with Online Data," *Proceedings of ACM IMWUT*, vol. 5, no. 1, pp. 1–31, 2021.

[52] C. Knapp and G. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[53] G. Barbato, E. Barini, G. Genta, and R. Levi, "Features and Performance of Some Outlier Detection Methods," *Taylor & Franci Journal of Applied Statistics*, vol. 38, no. 10, pp. 2133–2149, 2011.

[54] W. Wang, J. Li, M. Jin, and Y. He, "ChordMics: Acoustic Signal Purification with Distributed Microphones," in *Proceedings of IEEE ICCCN*, 2020.

[55] DJI, "DJI Mini 2 Quadcopter," 2023. [Online]. Available: https://www.dji.com/cn/mini-2

[56] Seeed Studio, "ReSpeaker 6-Mic Circular Array," 2023. [Online]. Available: https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/

[57] ——, "ReSpeaker 4-Mic Array," 2023. [Online]. Available: https://wiki.seeedstudio.com/ReSpeaker_4_Mic_Array_for_Raspberry_Pi/

[58] M. Afanasov, A. Djordjevic, F. Lui, and L. Mottola, "FlyZone: A Testbed for Experimenting with Aerial Drone Applications," in *Proceedings of ACM MobiSys*, 2019.

[59] Nooploop, "LinkTrack UWB Tracking System," 2023. [Online]. Available: https://www.nooploop.com/
[60] X. Chang, C. Yang, J. Wu, X. Shi, and Z. Shi, "A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays," in *Proceedings of IEEE SAM*, 2018.
[61] Z. Shi, X. Chang, C. Yang, Z. Wu, and J. Wu, "An Acoustic-Based Surveillance System for Amateur Drones Detection and Localization," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2731–2739, 2020.
[62] A. Famili and J.-M. J. Park, "ROLATIN: Robust Localization and Tracking for Indoor Navigation of Drones," in *Proceedings of IEEE WCNC*, 2020.

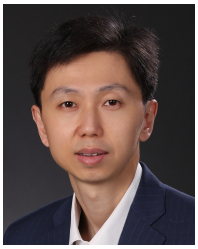
Yimiao Sun is currently a PhD. student in Tsinghua University. He received his B.E. degree in the University of Electronic Science and Technology of China (UESTC). His research interests include mobile computing and wireless sensing.

Weiguo Wang received his PhD. degree in Tsinghua University, and his B.E. degree in the University of Electronic Science and Technology of China (UESTC). His research interests include acoustic sensing and mobile computing.

Luca Mottola is a Full Professor at Politecnico di Milano (Italy) and a Senior Researcher at RI.SE Sweden. He is past General Chair for ACM/IEEE CPS-IoT Week 2022 and past PC chair for ACM MOBISYS, ACM SENSYS, ACM/IEEE IPSN, and ACM EWSN. He received the ACM SENSYS Test of Time Award in 2022, two ACM SigMobile Research Highlights, and is a Google Faculty Award winner. He holds or held visiting positions at Uppsala University, NXP Technologies, TU Graz, and USI Lugano.

Jia Zhang is currently a PhD. student in the School of Software and BNRist of Tsinghua University. He received his B.E. degree in Tsinghua University in 2019. His research interests include Internet of Things and wireless sensing.

Ruijin Wang is an associate professor of University of Electronic Science and Technology of China. He is the secretary general of ACM Chengdu Chapter, senior member of CCF and member of Asia Pacific Association for Artificial Intelligence (AAIA). He is a visiting scholar at Northwestern University. His research interests include cloud-edge intelligent computing, artificial intelligence, security, and blockchain.



Yuan He is an associate professor in the School of Software and BNRist of Tsinghua University. He received his B.E. degree in the University of Science and Technology of China, his M.E. degree in the Institute of Software, Chinese Academy of Sciences, and his PhD degree in Hong Kong University of Science and Technology. His research interests include wireless networks, Internet of Things, pervasive and mobile computing. He is a senior member of IEEE and a member ACM.