# Drone audition: implementation of an indoor multi-drone system for sound source tracking

Benjamin Yen*† and Kazuhiro Nakadai *
* Nakadai Lab, Tokyo Institute of Technology, Tokyo, Japan

***Abstract*—In this study, we introduce an indoor multi-drone system for tracking audible sound sources. The system utilises a number of Crazyflie drones, a mini-sized, open-source drone development platform that allows rapid prototyping and development using the Robotics Operating System 2 (ROS 2). Equipped with a microphone array, sound source tracking is realised by triangulating the spectral response from the Multiple Signal and Classification (MUSIC) algorithm combined with particle filtering. We evaluate the system using real-life demonstrations, and results show the system successfully tracks a single moving sound source in real-time.**

## I. Introduction

In many applications, acquiring and relaying information in a time-efficient manner is crucial. Drones, with the amount of development in recent years, have shown to be an ideal platform for acquiring a broad range of information [1]. Furthermore, it has shown to be especially useful in scenarios where navigating through difficult-to-reach environments is needed [2], which is commonly encountered in search and rescue [3] and military surveillance [4]. Amongst the range of sensing technologies available for drones, vision (i.e. via a video camera) is most commonly used. However, there are many situations where the effectiveness of vision information can be severely impeded, including low lighting, excessive vegetation and even collapsed buildings. On the other hand, audio information is immune to many such conditions and thus has the potential as a viable alternative [5], [6]. For this reason, drone audition, where drones are used as mobile acoustic sensors to perceive their surrounding environment and perform relevant tasks using audible sound, has gained increasing attention [7].

Amongst existing drone audition-related studies, many of them are carried out in areas such as sound source separation [8]–[11] and sound source localisation [12]–[15]. In this study of sound source localisation (and tracking), which is the main focus of this study, most studies utilise an array of microphones [16] that is mounted on the drone to utilise the spatial information that is unavailable with single microphone approaches. In addition, due to the high levels of rotor noise, which is one of the ongoing challenges in drone audition [10], utilising an array of microphones helps suppress the effects of such noise influences. For example, several studies utilised a noise correlation matrix (NCM) to reduce the influences from drone noise and, in turn, improve localisation performance based on the MUltiple SIgnal Classification (MUSIC) algorithm. Here, the NCM estimation ranges from simple frame averaging of the input signal's spectrogram [12], [17] to taking UAV input states and following a Gaussian process regression [18]. Other methods opted to directly reduce drone noise at the microphone inputs by utilising a Wiener postfilter (WF) [15], [19], beamforming [20], and DNN-based methods [14], [21].

As shown, most studies in sound source localisation for drones focused on single-drone (or single microphone array) approaches. While it allows the *direction* of the sound source to be tracked, it is not capable of tracking the exact *location* of the target in terms of 3D coordinates, which is often desired in practical scenarios. To address this, authors from [22], [23] made use of multiple microphone arrays as multiple drones with a single array attached to triangulate the localisation outputs of each drone to obtain the exact location of the target sound source. This includes triangulation of the target source directions from each drone that is processed through a Gaussian sum filter, triangulating the estimated directions [22] or triangulating the MUSIC spectrum from each drone followed by particle filtering, known as Particle Filtering with Integrated MUSIC (PAFIM) [23]. Recently, a method that allows the drones to autonomously navigate themselves such that the microphone arrays are in the optimal position for localising the sound source has also been proposed [24]. A unique aspect of these studies is that not only can the location of the sound source be obtained, but in the case that the sound source is moving, it can continuously track its whereabouts.

While these recent studies in multi-drone sound source tracking has shown promising results, most are evaluated under simulation. Several factors hinder the real-life implementation of such a system. First, due to the multi-drone nature, a system that allows time-synchronised inputs from multiple drones is required for the triangulation process to be carried out effectively. Implementing such a system outdoors is challenging. Furthermore, it is often difficult to gain access to an environment that allows the development of a system, which, on top of the implementation aspect, also includes extensive testing. In the case of Japan, airspace is bounded by various restrictions, and thus, many public spaces prohibit the presence of drones [25]. However, for development and evaluation of multi-drone drone audition-related research, it may not always be necessary to make use of a fully outdoor system. Instead, an indoor system that is free of such restrictions, where the drone's location can be readily tracked within a confined space,

†B. Yen is an International Research Fellow of the Japan Society for the Promotion of Science (JSPS) at the Nadakai Lab, Tokyo Institute of Technology.

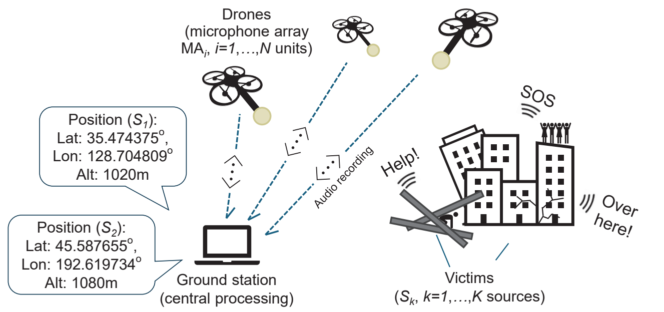


Fig. 1. Overview of multi-drone sound source tracking.

is a perfectly viable alternative for rapid development and real-life evaluation of such algorithms.

Fortunately, such a system can be achieved with the *Crazyflie* nano drone [26]. The small-sized Crazyflie drone can be equipped with a microphone array (known as the *audio deck*) developed specifically for the drone [27] to perform drone audition-related applications. Therefore, this study proposes an indoor multi-drone system utilising the Crazyflie drones to track static or moving sound sources in real-time. Utilising the *audio deck* combined with the *lighthouse positioning system* for location and orientation information for each drone, multi-drone 3D sound source tracking can be achieved. We note that while this study focuses on sound source tracking, such a system is also adaptable for the development of sound source separation or enhancement-related applications. To demonstrate the system's real-time capabilities, we provide experimental evaluation via video demonstration. In summary, the main contributions of this study are:

1) the implementation of an indoor multi-drone audition system.
2) the practical realisation of a multi-drone sound source tracking based on the indoor system.

As such, this study focuses on the implementation of multi-drone sound source tracking. Due to the unique nature of drone audition using multiple drones in an indoor setting, in this study, we describe the design challenges and solutions addressing these challenges.

## II. Problem Setup

This section describes the problem setup for this study. As shown in Fig. 1, the problem assumes $N$ microphone arrays $\mathrm{MA}_1, \ldots, \mathrm{MA}_N$ consisting of $M$ sensors, with each mounted to a drone. Each $i$th microphone array receives a number of mutually uncorrelated sound sources, including $K$ target sources $S_k(\omega, f)$, *spatially coherent* noise generated by $U$ rotors on the drone—with each $u$th rotor noise denoted as $N_u(\omega, f)$—and ambient *spatially incoherent* noise, such as air conditioning $\mathbf{v}(\omega, f)$, in an indoor environment. For the multi-drone scenario, we also consider the noise radiated by adjacent drones $N_{j \in N(j \neq i)}$ (i.e., all drones apart from the $i$th drone). Since drones are, at most times, a distance away from each other, we assume that the noise source of any neighbouring drone received by the $i$th microphone array is a point source, and the noise itself propagates from the $j$th drone to the $i$th microphone array according to the steering vector $\mathbf{a}_{i,j}(\omega, \phi_{i,j}, \theta_{i,j})$. $\omega$ and $f$ denote the angular frequency and frame index, respectively. Here, we assume all drones carry the same number of rotors $U$. Using the $N \times M$-channel noisy recordings, the system aims to localise and track the $K$ target source signals [22]. Assuming overdetermined cases (i.e., $M \geq U{+}K{+}N{-}1$ for each $i$th drone/microphone array), the short-time Fourier transform (STFT) of the $i$th microphone array's ($\mathrm{MA}_i$) input signals are expressed in vector form as

$$
\begin{aligned}
\mathbf{x}_i(\omega, f) &:= \left[X_{i,1}(\omega, f), \quad \cdots, \quad X_{i,M}(\omega, f)\right]^T \\
&= \sum_{k=1}^{K} \mathbf{a}_{i,S_k}(\omega, \phi_{i,S_k}, \theta_{i,S_k}) S_k(\omega, f) \\
&+ \sum_{u=1}^{U} \mathbf{a}_{i,u}(\omega, \phi_{i,u}, \theta_{i,u}) N_u(\omega, f) \\
&+ \sum_{j \in N(j \neq i)} \mathbf{a}_{i,j}(\omega, \phi_{i,j}, \theta_{i,j}) N_j(\omega, f) + \mathbf{v}(\omega, f), \quad (1)
\end{aligned}
$$

where $^T$ denotes the transpose, and $X_{i,m}(\omega, f)$ is the STFT of the $m$th microphone input signal. $\phi_{i,S_k}$ and $\theta_{i,S_k}$ are the azimuth and elevation directions (i.e., in spherical coordinates) from the $k$th target sound source to the centre of the microphone array $\mathrm{MA}_i$ in its own body coordinates, respectively. Likewise, $\theta_{i,u}$ and $\phi_{i,u}$ indicate the azimuth and elevation directions from the $u$th rotor $\mathrm{MA}_i$ in its own body coordinates, respectively. $\mathbf{a}_i(\omega, \phi_i, \theta_i)$ and $\mathbf{v}(\omega, f)$ are the steering vector between the source located at directions $\theta_i$, $\phi_i$, and each microphone $m$ in $\mathrm{MA}_i$, as well as the incoherent noise vector observed by the microphone array.

In addition to the microphone input signals, the state of each microphone array $\mathrm{MA}_n$ is described as

$$\boldsymbol{m}_i(f) = [\boldsymbol{m}_{i,xyz}(f)^T, \boldsymbol{m}_{i,\phi\theta\psi}(f)^T]^T, \quad (2)$$

$$\boldsymbol{m}_{i,xyz}(f) = [x_i(f), y_i(f), z_i(f)]^T, \quad (3)$$

$$\boldsymbol{m}_{i,\phi\theta\psi}(f) = [\phi_i(f), \theta_i(f), \psi_i(f)]^T, \quad (4)$$

where $x_i(f), y_i(f)$, and $z_i(f)$ indicate the center of $\mathrm{MA}_n$ in the three-dimensional coordinates, and $\phi_i(f), \theta_i(f)$, and $\psi_i(f)$ indicate the yaw, pitch, and roll angles of $\mathrm{MA}_i$. The location of the $k$th sound source is defined as

$$\boldsymbol{e}_k(f) = [x_k(f), y_k(f), z_k(f)]^T. \quad (5)$$

## III. System and Methodology

Fig. 2 shows a block diagram of the proposed multi-drone sound source tracking framework. Prior to introducing the proposed indoor multi-drone platform, including its hardware, we first present the method driving the multi-drone sound source tracking. The method is based on the MUSIC algorithm (Section III-A), combined with particle filtering (Section III-B) known as Particle Filtering with Integrated MUSIC (PAFIM) [23]. The following sections describe each component of the sound source tracking system.

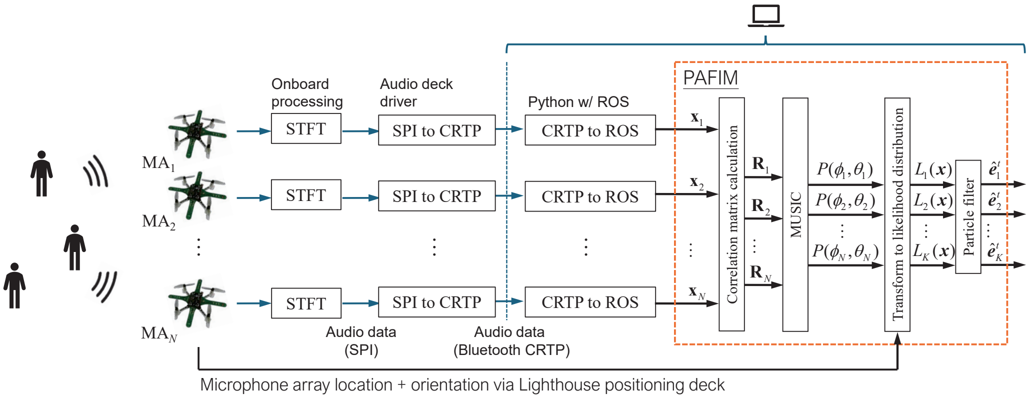


Fig. 2. Flowchart of indoor multi-drone system with sound source tracking using PAFIM.

### A. MUSIC

PAFIM, as apparent in the name, uses the MUSIC algorithm. MUSIC performs localisation by first performing Eigenvalue decomposition on the microphone array signal's correlation matrix $\mathbf{R}(\omega, t)$, calculated as

$$\begin{aligned}\mathbf{R}(\omega, t) &= \frac{1}{T_R} \sum_{\tau=f}^{f+T_R-1} \mathbf{x}(\omega, \tau)\mathbf{x}^H(\omega, \tau) \\ &= \mathbf{E}(\omega, t)\mathbf{\Lambda}(\omega, t)\mathbf{E}^{-1}(\omega, t), \end{aligned} \tag{6}$$

where $T_R$ is the number of frames to average to calculate $\mathbf{R}(\omega, t)$, and $^H$ denotes the Hermitian conjugate. $\mathbf{\Lambda}(\omega, t)$ and $\mathbf{E}(\omega, t)$ are the eigenvalue (diagonal) and eigenvector matrices, respectively. Using $\mathbf{E}(\omega, t)$ calculated from (6), the spatial spectrum of MUSIC is calculated as

By assuming orthogonality between the target sound source and the noise subspace, the spectrum indicating the likelihood of the target source's direction (at each time step $t$) is calculated using $\mathbf{E}$ and steering vector $\mathbf{a}(\omega, \phi, \theta)$ as

$$P(\phi, \theta, t) = \frac{1}{\omega_H - \omega_L + 1} \sum_{\omega=\omega_L}^{\omega_H} \frac{\|\mathbf{a}^H(\omega, \phi, \theta)\mathbf{a}(\omega, \phi, \theta)\|}{\|\mathbf{a}^H(\omega, \phi, \theta)\mathbf{E}_N(\omega, t)\|}. \tag{7}$$

Here, $\mathbf{a}(\omega, \phi, \theta)$ is the transfer function of a sound signal arriving from direction $(\phi, \theta)$. $\omega_L$ and $\omega_H$ indicate the minimum and maximum frequency bins, respectively. $\mathbf{E}_N(\omega, t)$ is the matrix of eigenvectors corresponding to the noise subspace in the microphone array signal's correlation matrix.

### B. Particle Filtering with MUSIC

Assuming an arbitrary location $\boldsymbol{x}$, its direction relative to drone location $\boldsymbol{m}_i$ is defined as $(\phi_i, \theta_i)$. Using this, once the MUSIC spectrum $P_i(\phi, \theta)$ for each $i$-th drone is calculated, a location likelihood $L(\boldsymbol{x})$ that triangulates $P_i(\phi, \theta)$ to $\boldsymbol{x}$ can be calculated as

$$L(\boldsymbol{x}) = \sum_i P_i(\tilde{\phi}_i^{\mathrm{round}}, \tilde{\theta}_i^{\mathrm{round}}), \tag{8}$$

$$\tilde{\phi}_i^{\mathrm{round}} = \mathrm{round}(\tilde{\phi}_i), \qquad \tilde{\theta}_i^{\mathrm{round}} = \mathrm{round}(\tilde{\theta}_i), \tag{9}$$

$$\begin{bmatrix} \cos\tilde{\phi}_i \cos\tilde{\phi}_i \\ \sin\tilde{\phi}_i \cos\tilde{\phi}_i \\ \sin\tilde{\phi}_i \end{bmatrix} = \boldsymbol{R}_n^{-1} \begin{bmatrix} \cos\phi_i \cos\phi_i \\ \sin\phi_i \cos\phi_i \\ \sin\phi_i \end{bmatrix}, \tag{10}$$

where $\mathrm{round}(\cdot)$ rounds the direction with respect to the resolution of transfer function $\boldsymbol{a}(\omega, \phi_S, \theta_S)$. $\boldsymbol{R}_i$ is the rotation matrix that converts between the $i$-th drone's body coordinates $\boldsymbol{m}_{i,\phi\theta\psi}$ and world coordinates.

Following, $L(\boldsymbol{x})$ can be used for calculating the weighting factor in a particle filter setting, assuming a distribution of potential locations $\boldsymbol{x}$. For each time step $t$, $I_{\mathrm{P}}$ number of particles is utilised. We define the state $\boldsymbol{x}_t^{i_{\mathrm{P}}} = [x_t^{i_{\mathrm{P}}}, y_t^{i_{\mathrm{P}}}, z_t^{i_{\mathrm{P}}} \dot{x}_t^{i_{\mathrm{P}}}, \dot{y}_t^{i_{\mathrm{P}}}, \dot{z}_t^{i_{\mathrm{P}}}]^T$, which includes the 3-dimentional location and velocity of the particle $i_{\mathrm{P}}$, and $w_t^{i_{\mathrm{P}}}$ denotes the weight of the particle. We then employ an excitation-damping model as the prediction model

$$\boldsymbol{x}_t^{i_{\mathrm{P}}} = \boldsymbol{F}\boldsymbol{x}_{t-1}^{i_{\mathrm{P}}} + \boldsymbol{H}\boldsymbol{v}, \tag{11}$$

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{I} & T\boldsymbol{I} \\ \boldsymbol{O} & a\boldsymbol{I} \end{bmatrix}, \qquad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{O} \\ b\boldsymbol{I} \end{bmatrix}, \tag{12}$$

$$\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{13}$$

where $\boldsymbol{I} \in \mathbb{R}^{3\times3}$ and $\boldsymbol{O} \in \mathbb{R}^{3\times3}$ denote the identity matrix and zero matrix, respectively. Parameter $a$ controls the ratio of velocity to carry on from the previous time step $t-1$ and parameter $b$ controls the level of excitation of particles to introduce in time step $t$. As mentioned, the location likelihood $L(\boldsymbol{x}_{t,pos}^{i_{\mathrm{P}}})$ is used to calculate the weighting of each particle, defined as

$$w_t^{i_{\mathrm{P}}} = w_{t-1}^{i_{\mathrm{P}}} \frac{L(\boldsymbol{x}_{t,pos}^{i_{\mathrm{P}}})}{\sum_i L(\boldsymbol{x}_{t,pos}^{i_{\mathrm{P}}})}, \tag{14}$$

where $\boldsymbol{x}_{t,pos}^{i_{\mathrm{P}}} = [x_t^{i_{\mathrm{P}}}, y_t^{i_{\mathrm{P}}}, z_t^{i_{\mathrm{P}}}]^T$. Finally, using the prediction model and the weightings obtained, the sound source location at time step $t$ is estimated as

$$\boldsymbol{e}(t) = \sum_{i_{\mathrm{P}}=1} w_t^{i_{\mathrm{P}}} \boldsymbol{x}_t^{i_{\mathrm{P}}}. \tag{15}$$

Further details regarding PAFIM, including its initialisation can be found in [23].

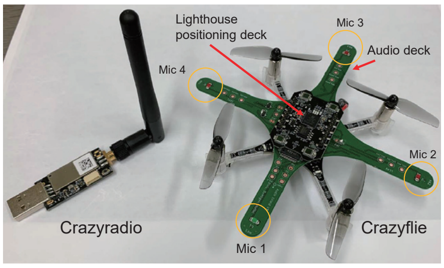


Fig. 3. Crazyradio (left) and Crazyflie with the audio deck and lighthouse positioning deck attached (right).

## IV. System Implementation

This section introduces the proposed implementation of an indoor multi-drone sound source tracking system. We first introduce the Crazyflie drone, including any additional sensors and hardware used. Following, we introduce the relevant equipment and their arrangement in the indoor space for real-time 3D sound source tracking in a stable and robust manner.

### A. Crazyflie Drone

The Crazyflie is a small-sized nano drone specialised for research development (see Fig. 3). One of its features is that the drones operate using the Robotics Operation System 2 (ROS 2), which is an open-source set of software libraries and tools that allows the development of algorithms and systems for robot-related applications [28]. This allows the development of real-time applications with time-synchronised data input from multiple sensors and subsequent processing.

Another feature of the Crazyflie drone is the *extension decks*, which allow various types of sensors to be easily equipped for the Crazyflie. Some examples of such extension decks include the greyscale camera with the *AI-deck* or the *Multi-ranger Deck* to give the drones obstacle avoidance capabilities. In this study, we utilise the *Lighthouse positioning deck*, which, operating along with the *Lighthouse V2 base station* motion capture camera, provides the Crazyflie drones with real-time location and orientation information with relatively high accuracy. As mentioned in Section I, although there are no microphone array extension decks available officially for the Crazyflie, a custom-designed open-source microphone array deck has been developed by authors in [27] (known as the *Audio deck*). The microphone array is equipped with four micro-electromechanical system (MEMS) microphones (see Fig. 3). To help offload some processing for the ground station PC, the audio deck is equipped with an STM32F446 microcontroller, which performs STFT on the recorded audio signals prior to data transmission.

### B. Indoor Multi-Drone Sound Source Tracking System

Fig. 4 shows an overview of the indoor multi-drone system platform. The system consists of a swarm of Crazyflie drones. Each drone communicates with the central ground station (i.e. laptop PC) via the Crazyradio (see Fig. 3). This includes transmission of the audio, position and orientation data from the Crazyflie to the ground station and controller command from the ground station to each of the drones for automated drone control.

TABLE I
Audio recording and PAFIM parameters.

| Variable | Value |
|---|---|
| Sampling rate | 48 kHz |
| Audio resolution | 16 bit |
| STFT length | 2048 |
| $\omega_L$ | 200 Hz |
| $\omega_H$ | 4000 Hz |
| $a$ | 0.9 |
| $b$ | 0.3 |
| $I_{\mathrm{P}}$ | 500 |

As mentioned in Section IV-A, the Lighthouse positioning deck requires the Lighthouse V2 base station in order for real-time position and orientation data to be provided. As such, four base stations are used, each played at a corner of the experiment environment (see Fig. 4).

As shown in Fig. 2, data transmission from the Crazyflie to the ground station is carried out by first passing the microphone signals via serial peripheral interface (SPI) protocol, where data is sent to the STM32F446 microcontroller. Following, the data is then set to the ground station via the Crazyradio with a custom Bluetooth protocol known as the Crazy RealTime Protocol (CRTP) [27]. Due to the data size of the audio data from the Crazyflie's microphone array in STFT form, each drone requires pairing with a single Crazyradio. Consequently, this increases the risk of interference between different radio signals. To combat this issue, the Crazyradios are separated as far away as possible (as shown in Fig. 4).

## V. Evaluation: Indoor demonstration

In this section, we evaluate the sound source tracking performance of the indoor multi-drone system as a proof of concept through demonstration, using the PAFIM method with array placement planning as the tracking algorithm. The experimental settings for the demonstration are described in Section V-A, with some discussions presented in Section V-B.

### A. Experimental Settings

Fig. 4 shows a top view of the experimental settings, including the initial locations of the target sound source and the drones. The setting consists of a single target sound source emitting male speech, with four drones placed around the sound source.

In this initial study, we evaluate the feasibility of the indoor sound source tracking system. Therefore, the target source was emitted loud enough for MUSIC to recognise the target as the primary sound source (i.e. the input signal-to-noise ratio (SNR) $\geq$ 5 dB). The target sound source moved around the room randomly for around $1-3$ minutes. The Crazyflie drones were flying at a fixed height of 1.2 m. Details regarding the audio recording and PAFIM-related parameters are shown in Table IV-B. A total of five demonstration runs were carried out.

For the ground station, we specifically use the ROS 2 Galactic with Python version 3.8.10 under the Ubuntu 20.04 Focal Rossa operating system.

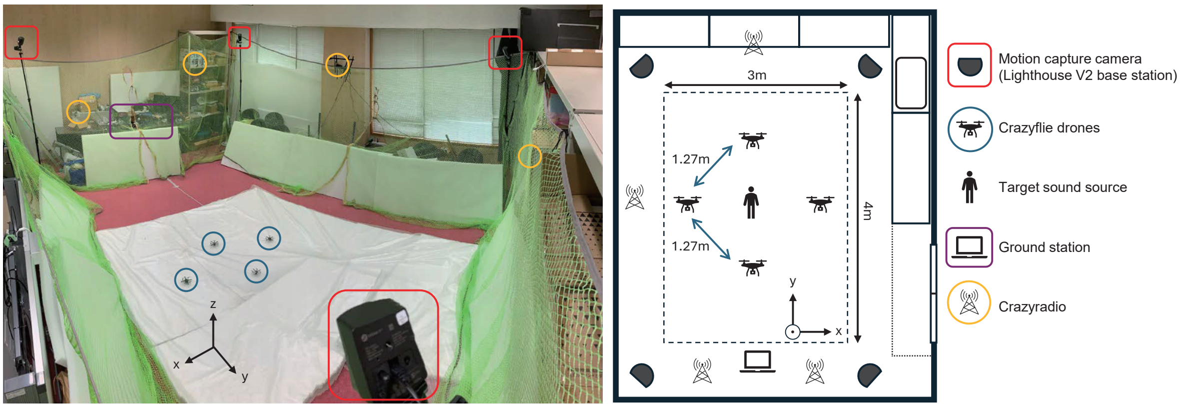


Fig. 4. Indoor multi-drone sound source tracking system (left) and experimental setting (right). Note that while the ground station computer is not shown in the figure, the location is highlighted by the purple box regardless.

### *B. Results and Discussion*

For this preliminary study, we evaluate the system through video demonstrations. Here, we discuss the performance using the following demonstration video. Overall, as shown in the video demonstration, PAFIM is able to successfully localise and track the target sound source accurately. In the case where the target sound source is continuously moving, whether quick or slow, PAFIM is able to track and update the correct target source location. There are occasions where the estimated location deviates from the actual location. However, as long as the target sound source continues to make sound, PAFIM will eventually converge to the correct location.

Due to the nature of the particle filter, after initialisation, it requires a few iterations for the particles to converge towards the target sound source position. As such, a minor lag occurs between the target source sounds making a sound and PAFIM updating the new target location. When the target sound source is moving, it also takes a few time frames to catch up to the target location. However, the lag is only apparent if the target source moves quickly. The current system is configured such that if it cannot detect the presence of the target sound source after five time frames, the particle filter will be reset. Therefore, as apparent in the demonstration, when the target stops making sound, the particle filter will reinitialise upon detecting the target sound again.

Despite the performance of the PAFIM algorithm, there are some factors that require addressing. As apparent from the video, the sound source tracking is only carried out in the xy plane, and the altitude of the target sound source is not estimated. This is due to the planar nature of the microphone array on the Crazyflies, where it is relatively difficult to distinguish the elevation direction of the target sound. Another issue includes the lack of explicit drone noise suppression for sound source tracking. Therefore, in this initial study, this is overcome by ensuring the target source being relatively loud. In addition, due to the planar nature of the audio deck, altitude tracking of the target source is sufficiently accurate. Addressing these issues, including a redesign of the microphone array to accommodate 3D sound source tracking and implementation of drone noise reduction algorithms, remains as future work.

The remaining demonstration videos, each showing similar levels of performance, can be found in video demonstrations.

## VI. Conclusion

This study introduces a practical implementation of an indoor multi-drone system based on the Crazyflie drone and ROS 2 for real-time sound source location tracking. By equipping each Crazyflie with a custom-designed microphone array and the sensors for positioning and orientation information, a system that tracks audible sound sources is realised. Evaluation through experimental demonstrations showed the system had successfully tracked a single moving sound source in real-time. Future work includes addressing the lack of a drone noise suppression algorithm and enabling tracking of the target sound source in all three spatial dimensions.

## Acknowledgment

This research was supported by the commissioned research fund provided by F-REI (JPFR23010102) and JSPS KAKENHI through Grant No. 22F22769 and Grant No. 22KF0141.

## References

[1] A. Koubaa and A. Azar, Eds., *Unmanned Aerial Systems*. Elsevier, 2021.

[2] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, “Dregon: Dataset and methods for uav-embedded sound source localization,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 1–8.

[3] S. Grogan, R. Pellerin, and M. Gamache, “The use of unmanned aerial vehicles and drones in search and rescue operations – a survey,” in *Proceedings of the PROLOG*, Sep. 2018, pp. 1–13.

[4] M. A. Ma’sum, M. K. Arrofi, G. Jati, *et al.*, “Simulation of intelligent unmanned aerial vehicle (UAV) For military surveillance,” in *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Sep. 2013, pp. 161–166.

[5] K. Hoshiba, O. Sugiyama, A. Nagamine, R. Kojima, M. Kumon, and K. Nakadai, "Design and assessment of sound source localization system with a uav-embedded microphone array," *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 154–167, 2017.

[6] S. V. Sibanyoni, D. T. Ramotsoela, B. J. Silva, and G. P. Hancke, "A 2-d acoustic source localization system for drones in search and rescue missions," *IEEE Sensors Journal*, vol. 19, no. 1, pp. 332–341, 2018.

[7] K. Nakadai and H. G. Okuno, "Robot audition and computational auditory scene analysis," *Advanced Intelligent Systems*, vol. 2, no. 9, p. 2 000 050, 2020.

[8] Z.-W. Tan, A. H.-T. Nguyen, and A. W.-H. Khong, "An efficient dilated convolutional neural network for UAV noise reduction at low input SNR," in *2019 Proceedings of Asia-Pacific Signal and Information Processing Association (APSIPA)*, Nov. 2019, pp. 1885–1892.

[9] L. Wang and A. Cavallaro, "A blind source separation framework for ego-noise reduction on multi-rotor drones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2523–2537, 2020.

[10] Y. Hioka, M. Kingan, G. Schmid, and K. A. Stol, "Speech enhancement using a microphone array mounted on an unmanned aerial vehicle," in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2016, pp. 1–5.

[11] B. Yen, Y. Li, and Y. Hioka, "Rotor noise-aware noise covariance matrix estimation for unmanned aerial vehicle audition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[12] K. Nakadai, M. Kumon, H. G. Okuno, *et al.*, "Development of microphone-array-embedded UAV for search and rescue task," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 5985–5990.

[13] K. Yamada, M. Kumon, and T. Furukawa, "Belief-driven control policy of a drone with microphones for multiple sound source search," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 5326–5332.

[14] T. Spadini, G. S. Imai Aldeia, G. Barreto, *et al.*, "On the application of SEGAN for the attenuation of the ego-noise in the speech sound source localization problem," in *2019 Workshop on Communication Networks and Power Systems (WCNPS)*, Oct. 2019, pp. 1–4.

[15] B. Yen and Y. Hioka, "Noise power spectral density scaled snr response estimation with restricted range search for sound source localisation using unmanned aerial vehicles," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, pp. 1–26, 2020.

[16] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications* (Digital Signal Processing). Springer, 2001. (visited on 06/26/2017).

[17] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2012, pp. 3288–3293.

[18] K. Furukawa, K. Okutani, K. Nagira, *et al.*, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2013, pp. 3943–3948.

[19] W. Manamperi, T. D. Abhayapala, J. Zhang, and P. N. Samarasinghe, "Drone audition: Sound source localization using on-board microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 508–519, 2022.

[20] D. Salvati, C. Drioli, G. Ferrin, and G. L. Foresti, "Acoustic source localization from multirotor UAVs," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8618–8628, 2019.

[21] L. Wang and A. Cavallaro, "Deep-learning-assisted sound source localization from a flying drone," *IEEE Sensors Journal*, vol. 22, no. 21, pp. 20 828–20 838, 2022.

[22] T. Yamada, K. Itoyama, K. Nishida, and K. Nakadai, "Sound source tracking by drones with microphone arrays," in *2020 IEEE/SICE International Symposium on System Integration (SII)*, IEEE, 2020, pp. 796–801.

[23] T. Yamada, K. Itoyama, K. Nishida, and K. Nakadai, "Assessment of sound source tracking using multiple drones equipped with multiple microphone arrays," *International journal of environmental research and public health*, vol. 18, no. 17, p. 9039, 2021.

[24] T. Yamada, K. Itoyama, K. Nishida, and K. Nakadai, "Placement planning for sound source tracking in active drone audition," *Drones*, vol. 7, no. 7, 2023.

[25] T. Ministry of Land Infrastructure and Tourism. "Flight rules for unmanned aircraft (drones and model aircraft, etc.)" (2023), [Online]. Available: https://www.mlit.go.jp/en/koku/uas.html (visited on 07/18/2024).

[26] W. Giernacki, M. Skwierczyński, W. Witwicki, P. Wroński, and P. Kozierski, "Crazyflie 2.0 quadrotor as a platform for research and education in robotics and control engineering," in *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*, 2017, pp. 37–42.

[27] F. Dümbgen, A. Hoffet, M. Kolundžija, A. Scholefield, and M. Vetterli, "Blind as a bat: Audible echolocation on small robots," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1271–1278, 2023.

[28] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot operating system 2: Design, architecture, and uses in the wild," *Science Robotics*, vol. 7, no. 66, eabm6074, 2022.