Swarm Active Audition System with Robots and Drones for a Search and Rescue Task

Kazuhiro Nakadai*, Makoto Kumon†, Yoko Sasaki‡, Kotaro Hoshiba*, and Benjamin Yen*
* Tokyo Institute of Technology
E-mail: {nakadai@ra.sc, hoshiba@rmsv.mech, benjamin@ra.sc}.e.titech.ac.jp
† Kumamoto University
E-mail: kumon@gpo.kumamoto-u.ac.jp
‡ The National Institute of Advanced Industrial Science and Technology (AIST)
E-mail: y-sasaki@aist.go.jp

Abstract—**This paper addresses the design of system based on drone audition and its necessity to search for rescuers. We propose a swarm active audition system (SAAS) that employs robots and drones as an efficacious means of identifying individuals in need of rescue during the initial stages of search and rescue operations. The concept and system architecture of SAAS is presented, and the effectiveness of drone audition functionality is demonstrated on a case study basis. Furthermore, interviews will be conducted with relevant personnel involved in search and rescue activities at disaster sites to ascertain the necessity of SAAS in the field. The necessity and lack of SAAS functions will be discussed in light of the analysis of these interviews.**

I. Introduction

In the context of disaster relief, the search for individuals in need of rescue represents the primary and most crucial objective, given that a failure to rescue a person in need within 72 hours of a disaster's occurrence has been linked to a significantly reduced probability of survival. In disaster areas, the disruption to the traffic network is such that even emergency vehicles are unable to pass through with ease, thereby rendering drone searches an effective solution. While mounting a camera on a drone and conducting a search based on the camera information is an effective approach, its effectiveness is limited when individuals are obscured by debris or when visibility is constrained, such as in foggy weather.

Considering this situation, "drone audition" has started as an extension of robot audition [1], in which a drone equipped with a microphone array is used to search for rescuers in need. Okutani et al. [2] introduced a method of multiple signal classification based on generalized singular value decomposition (GSVD-MUSIC), a technique developed for robot audition [3]. In another early study, Basiri et al. proposed a 4 ch microphone array that can localize sound sources in both azimuth and elevation [4]. Ohata et al. proposed iGSVD-MUSIC, an augmented variant of GSVD that incorporates an incremental estimation of the dynamically-changing drone noise [5]. Additionally, Furukawa et al. proposed a further noise estimation method based on a Gaussian process, with the objective of improving sound source localization from the sky [6].

In 2014, the Japanese Cabinet Office launched the ImPACT (Impulsing Paradigm Change through Disruptive Technologies Program), which included the Tough Robotics Challenge (TRC), led by Professor Tadokoro of Tohoku University. The field of drone audition has made notable advancements in Japan with the backing of this initiative. The project encompassed extensive research into drone audition, particularly in the areas of sound source localization and tracking. This was based on the premise that the primary objective of drone audition is to locate individuals in disaster zones. Once the fundamental offline technology had been established, attention was directed towards the development of a technology that would be robust in the real environment and operational online. This involved the design of an accessible microphone array [7], the implementation of 3D positional sound source localization [8], and the handling of multiple sound sources [9]. At last, the online demonstration has been achieved [10], [11].

With regard to groups engaged in drone audition research on source localization and tracking outside of Japan or ImPACT TRC during this period, for instance, Wang et al. [12] reported the use of time-frequency masks for source localization, a technique frequently employed in acoustic signal processing. Lauzon et al. [13] reported a study in which ground-based microphones were used to localize drones, as opposed to drone-mounted microphones, for the purpose of sound source localization. However, there were few reports of drone audition studies, in part due to the absense of a representative drone audition dataset, such as the DREGON dataset [14].

Sound source separation and speech enhancement have been studied more widely. The majority of studies were based on signal processing [15]–[20], with a few employing deep learning techniques [21]. In ImPACT TRC, research has been conducted on the integration of sound source separation and classification as a subsequent step following sound source localization and tracking [22]–[24]. Many of the methods proposed in these studies employ a combination of signal processing and deep learning [22], [24], or solely deep learning [23], in part due to the advent of deep learning.

In recent years, there have been ambitious attempts to perform automatic speech recognition for indoor use [25].

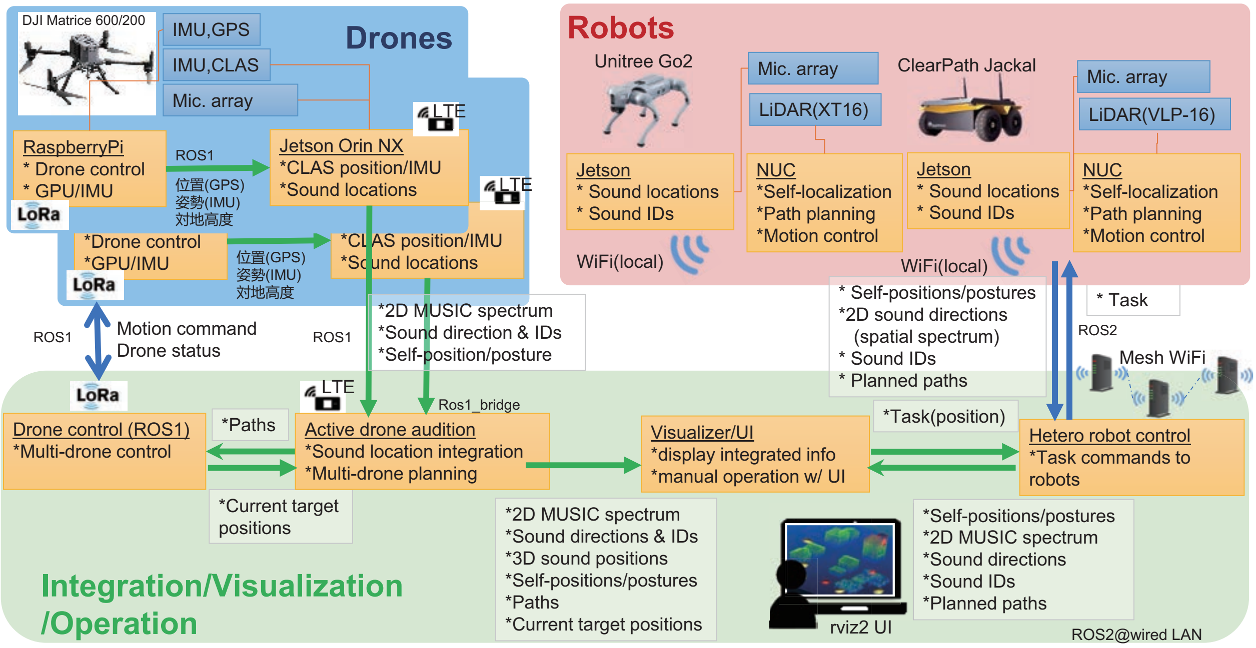


Fig. 1. The Architecure of Swarm Active Audition System

Additionally, studies have been conducted to explore the potential of employing the stop-perceive-act strategy in robot audition [1]. This study presents a unique and practically effective approach that attempts to address the noise issue by utilizing a kiteplane that can glide with its rotors stopped and listen to sounds during quiet periods when the rotors are stopped [26].

For sound source localization and tracking, a method for estimating the 3D position of a sound source has been reported. However, this method is limited by two assumptions: first, that the source is on the ground [10], [11]; and second, that multiple observations are required to localize sound sources [9], [26]. The aforementioned assumption is overly simplistic and therefore unsuitable for application in search and rescue situations, given that a person in need of rescue may be located on an upper floor of a building. The latter approach presents two significant challenges. Firstly, the time required to locate a person in need of rescue is considerable. Secondly, the search area narrows considerably with the application of this approach.

To address these challenges, research has been conducted employing the use of multiple drones simultaneously [27]–[29]. The use of multiple drones allow for the application of triangulation to achieve 3D localization without the aforementioned assumptions. Yamada et al. identified a limitation of the method, namely that the source directions obtained from multiple observations do not necessarily intersect at a single point. They proposed a solution, which involved the use of a Gaussian filter [27]. This was enhanced by replacing it with particle filtering with integrated MUSIC (PAFIM) [28], and by introducing Wiener filtering [29]. Additionally, they proposed the concept of "active drone audition," in which multiple drones proactively undertake optimal positioning to enhance target sound source localization [28]. However, these studies evaluated the efficacy of the proposed methods exclusively through numerical simulations using audio signals synthesized based on the recorded audio data.

In light of the aforementioned limitations, we propose the swarm active audition system (SAAS), which employs a heterogeneous fleet of drones and robots to facilitate human intervention, as an extension and realization of the PAFIM-based sound source localization and tracking method. Our objective is not merely to ascertain the functionality of SAAS in a real-world setting; We are also interested in its potential utility in a real-world disaster scenario. We aim to ensure that our research is not merely theoretical but has tangible, practical applications, this paper aims to contribute to the following discussions:

- The proposal of SAAS for the search and rescue of individuals in need in a disaster setting.
- An overview of the advancements made in SAAS, with a particular emphasis on its drone audition functionality.
- A survey of the system requirements necessary for effective operation in actual disaster areas, and an examination of the extent to which the proposed SAAS system meets those requirements.

II. Swarm Active Audition System

Figure 1 illustrates the architecture of the swarm active audition system (SAAS). It consists of three principal components: **Drones**, **Robots** and Integration/Visualization/Operation (**IVO**).

Drones ensures the *high speed* and *wide area* of the search by enabling 3D sound source localization through the use of multiple drones. This component receives PAFIM results implemented in the active drone audition module and optimally positions the drones. In addition to sound source localization and tracking, each drone is also responsible for sound source classification in conjunction with sound source separation.

Robots ensures the *accuracy* of sound source localization and the feasibility of *runnability*. **Drones** is capable of traversing distances at high speeds, yet their search patterns are relatively coarse. In contrast, **Robots** moves at slower speeds, employing a stop-perceive-act strategy to focus intensively on areas where rescuers are likely to be located with high *accuracy*. In addition, a wheeled robot and a quadruped robot are employed together, with the optimal robot selected according to the area to be searched to enhance *runnability*. Furthermore, **Robots** play a pivotal role in updating the 3D map in real-time using onboard LiDAR [30], and in performing path planning on the updated map through a graph search-based method [31].

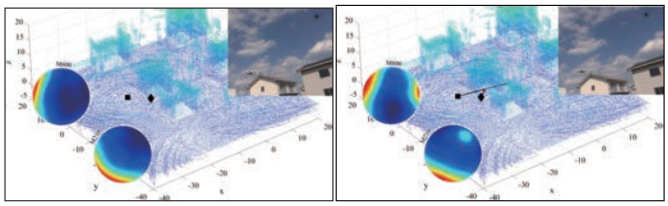
IVO guarantees *system robustness* and *user friendliness*. For *system robustness*, three measures must be taken to achieve low latency and robust data communication. First, a combination of multiple wireless communication paths is employed. Specifically, for communication between **Drones** and **IVO**, multiple communication methods, including WiFi, LTE, and LoRa, are employed in conjunction to ensure the robustness of communication and connection. In order to facilitate communication between **Robots** and **IVO**, Mesh WiFi is adopted to ensure transparent communication over a broad area. For stable communication within **IVO**, wired connections are employed. Furthermore. the benefits of ROS2 are maximized through the use of distributed modules across multiple PCs. Second, data compression is performed with consideration of the bandwidth limitations inherent to each communication path. For example, there is no requirement for **IVO** to convey precise information. In the event that the objective is solely for display purposes, the quantity of information may be diminished to the extent that it does not impede the user's comprehension. Furthermore, discretization does not necessarily need to be performed in an even manner. Intelligent data compression should be performed by widening the range of discretization in the range where information degradation is not a problem, and by narrowing the range of discretization in the range where information degradation is a problem. Finally, ROS2 is selected due to its stability for large-scale systems through the utilization of the data distribution service (DDS). However, initially, ROS1 will be used for communication within **Drones** and between **Drones** and **IVO**, due to the constraints of the embedded system of drones.

To enhance *user friendliness*, the user interface (UI) will be designed to facilitate comprehension of the situation at a glance for users who are not familiar with the system. They will be achieved by eliminating unnecessary sound sources, such as drone noise, and by employing color-coding to differentiate between the various sound sources displayed on the 3D map according to their respective types. Futhermore, the sound sources are displayed as ellipsoids that take the confidence level into account, thereby providing information on the sound sources that should be prioritized. Additionally, UI allows the human operator to intervene with the system at any time as needed based on such information, for example, to specify areas to be searched intensively.

Fig. 2. DJI Matrice 600 Pro(left) and DJI Matrice 210 RTK V(right). Each has a 16 ch spherical microphone array (white) at the end of the pole which is covered by windshield made of special urethane with light weight.

Fig. 3. Clearpath JACKAL(left) and Unitree Go2(right). Each has a 16 ch cylindrical microphone array and LiDAR.

Fig. 4. Generated 3D Point Cloud for Fukushima RTF town area field



a) Two drones are hovering (black square/diamond)

b) A sound source at 2F is successfully detected (red dot)

Fig. 5. Sound Source Localization with Multiple Drones

The construction of SAAS is scheduled for completion by November 2024, along with the online demonstration. The following demonstration will be taken into consideration. The two drones depicted in Fig. 2 and the two robots depicted in Fig. 3 are utilized. (1) The two drones are capable of performing 3D multiple sound source localization broadly over a wide area. (2) Each type of sound source detected is displayed on a 3D point cloud map as an ellipsoid of a different color, with the diameter corresponding to the error margin. (3) The wheeled and legged robots initiate high-precision sound localization and autonomously move to the location of the sound source to identify the individual in need of rescue.

A. System Performance

At the time of writing this paper, the system construction is still in progress. However, the implementation of each component has been completed, and the communication between the three components has been verified. This section presents illustrative examples of the operation of each component on a case study basis.

Figures 2 and 3 illustrate the drones and robots used in the test. A variety of drones and robots with disparate characteristics were selected for inclusion in the study. These included the DJI Matrice 600 Pro (six propellers), the DJI Matrice 210 RTK V (four propellers), the Clearpath JACKAL (four wheels), and the Unitree Go2 (four legs). Figure 4 depicts a 3D point cloud map of the town area at Fukushima Robot Test Field (RTF), which serves as the test site. This 3D point cloud map was generated in advance using the two robots used in **Robots**. Despite the town area comprising several blocks, including residential and debris areas, and an intricate internal structure, the map was successfully generated.

The objective of the test was to ascertain the ability of the two drones to detect calls for assistance from the second floor of House A, which is included in the target scenario. Figure 5 shows the situation of the test. As illustrated in Figure 5a), the two drones are hovering around 10 m from the sound source. The drones are shown as photos in the upper right view and also as a black square and diamond on the 3D point cloud map. The black square and diamond correspond to the DJI Matrice 600 Pro and 210 RTK V, respectively. The two circles on the left represent sound intensity maps, with red indicating high sound pressure levels from the corresponding direction, and blue indicating the opposite. The red areas within the circles in this figure are caused by drone propeller noise during hovering.

In Figure 5b), the individual on the second floor initiated a call for help. The two circles exhibited red and light blue areas, which were caused by the calls. At the same time, two black lines emerged from the black square and diamond, indicating the direction of the sound source as observed from the drone's perspective. In general, the two lines do not intersect in 3D space. However, PAFIM was able to successfully identify and estimate the position of the sound source, marked as a red dot, based on particle filtering.

III. Survey from Rescue Workers at Disaster Site

Drone audition has been studied in a seed-driven manner, with the interest in the effectiveness of acoustic signal processing and machine learning technologies in harsh environments. This strongly influences the statement that the application of drone audition technology is to search for people in need of rescue. It remains unclear whether drone audition technology and SAAS are genuinely beneficial or necessary in such contexts. This section presents the identified needs within the field.

We, thus, investigated the expectations for drone audition technology and its segregation from other technologies through interviews with the results of drone utilization at disaster sites to date and expectations for drones at disaster sites.

TABLE I
INTERVIEW ITEMS (IN EXCERPTS)

Class	Content
Overview	Purpose of using drones in S&R operations Cases where drone audition would have contributed to S&R? Provide info on the status of drone operations in the field
Data	Who will use it for what and how? Challenges in camera and thermography? Cases already using auditory technology?
Env.	Provide the surrounding environment in the field Challenges in data communication? Any other sound sources?
Use case	Any use of drones in peacetime? Any chances for drone audition except for S&R

The four parties interviewed were two prefectural fire departments and two companies that engage in drone operation. All four were involved in rescue operations following the Noto Peninsula earthquake in January 2024.

Table I enumerates the selected interview items, which were designed by considering issues at disaster sites, the environment at disaster sites, the needs of robot audition technology, and comparison with other technologies. The interviews were conducted in approximately one hour per party, allowing for a comprehensive examination of the specific details, rather than a mere affirmation or negation.

a) Issues at disaster sites: Only one of the four parties had prior experience in searching for individuals in need of rescue. This is due to the fact that the request for rescue was made more than two weeks after the disaster occurred, and the timing for searching for those in need of rescue was missed. In practice, most the drone applications are mainly for creating orthoimages to assess the situation. The party with the experience indicated that the search for rescuers was conducted manually, rather than by drone. The first priority in the search for rescuers is the safety of the search party members of the search party, and it takes time to ensure their safety before conducting the search.

b) Environment at disaster site: While environmental conditions may vary considerably from one location to another, the following examples are representative of the typical circumstances encountered.

- In order to generate ortho imaging, the drone is often required to undertake a flight of between 10 and 20 minutes in duration, at a speed of between 2 and 3 m/s and an altitude of between 50 and 100 m.
- There are a variety of sound sources at the site, and the major loud sources include wind, people, heavy machinery, helicopters.
- In the case of a fire, the sounds of flames and water spray are dominant enough to drown out the drone noise.

c) Needs of robot audition technology: Positive comments are as follows:

- It would be an efficacious strategy to utilize the drone audition in conjunction with the drone-equipped loudspeaker to call out to individuals.
- In the context of the transportation of medical supplies, it is essential to ascertain whether the intended recipient has indeed received the items in question. As an extension of

the drone audition, it would be advantageous if mutual communication via verbal means were possible.
- In mountain rescue contexts, it is not uncommon for rescuers to be difficult to detect victims, even when equipped with cameras or infrared cameras, due to the presence of trees and other obstructions. The ability to detect the voice of a person in need of rescue would therefore be a valuable addition to current technology. Furthermore, the potential for mutual communication by voice would enhance the efficacy of this system.
- Ultimately, the rescue of individuals from collapsed buildings relies on human hearing. Given that a person cannot commence a search until they have ensured their own safety, it would be beneficial if drone audition could be employed to augment the search process.

There was a negative option that rescuers may have difficulty speaking loudly, given that they are often in a debilitated state. It seems unlikely that a low volume voice can be detected by drone audition.

d) Comparison with other technologies: In practice, the search process heavily relies on the human senses, including hearing. Nevertheless, they are willing to adopt new technologies that are expected to be effective. In fact, imaging and electromagnetic wave search devices have been employed to search for rescuers. However, all of these devices have their respective advantages and disadvantages, including the potential for delays and a restricted search area. It is notable that no single method has been definitively identified as the optimal solution.

IV. Discussions

This section presents the findings of the survey and an analysis of the extent to which the proposed SAAS aligns with user requirements. It is also necessary to determine whether the functions provided by SAAS are essential or whether there are functions currently lacking in SAAS.

A. Discussion on the survey results

In the context of disaster site operations, insights were gleaned from a number of valuable and informative sources. With regard to the issues and comparisons presented, it is evident that the use of drones for the purpose of searching for rescuers is not yet a common practice, regardless of the type of sensors employed. Moreover, even when searching for individuals in need of assistance, the search is still conducted manually. This is largely attributable to structural deficiencies in search and rescue operations and to the awareness for the search activities. Furthermore, the sensing devices currently available for search and rescue operations are still in their infancy, necessitating the urgent establishment and deployment of technologies for locating individuals in need of rescue. It is similarly important to familiarise people with the technology in question, thereby enabling them to utilise it for available tasks during normal times.

In the context of disaster site environments, it is important to recognise that there are a multitude of sound sources present, beyond those originating from targets. This indicates that, in addition to sound source localization and tracking, sound source classification is an effective approach. In comparison to orthoimagery, drone audition searches for rescuers are likely to necessitate the implementation of lower altitudes. To guarantee the security of personnel operating drones at such altitudes, it is imperative to enhance automated control systems and technologies designed to assist drone operators, including dynamic mapping and path planning.

With regard to the requirements of drone audition technology, there is a consensus among the relevant parties that the drone audition function would be beneficial in a disaster scenario. Furthermore, the participants expressed greater enthusiasm for the potential of the drone audition function to facilitate communication between the drone and rescue personnel. Conversely, there is a concern that the drone may not be effective when the volume of the voice from the rescuers is low.

B. Discussions on the SAAS functions

In light of the aforementioned discussion, an examination will be conducted to ascertain whether the current SAAS comprises the requisite functions for those working at disaster sites, and whether any functions are absent.

The initial step is to examine the function of drone audition. The principal functions of **Drones** are 3D sound localization, high-speed and wide-area search, and sound source classification. The initial two functions will be invaluable in the field for two key reasons: One factor is that the circumstances at the site of a search and rescue operation are known to be variable, and the location of the rescuers cannot be assumed to be known in advance. The second reason pertains to the fact that, given the current reliance on human resources for search operations, the safety of the searchers takes precedence over the search itself. The third function is also essential for an efficient search, given the presence of a multitude of potent sound sources at the site. Conversely, it is postulated that the issue of detecting low voices can be resolved by flying the drone at lower altitudes and utilizing **Robots** as previously introduced in SAAS. It may be necessary in the future to introduce a stop-perceive-act strategy for **Drones** as well.

Robots has two main functions; enhancement of the runnability using heteogeneous robots and dyamic map generation with path planning. The first function is a suitable option, as no assumption about the circumstances can be put in disaster site. The second function is advantageous in situations where the specific circumstances of the disaster site are not immediately apparent. Moreover, it is beneficial for the secure operation of drones at low altitude, which may be necessary for drone audition tasks. One concern is how well the robot can detect low-volume that cannot be detected by the drone. A detailed evaluation will be necessary.

IVO is designed to enhance its robustness through the integration of diverse wireless communication channels, the implementation of data compression techniques that account for the constraints of communication bandwidth, and the incorporation of ROS2. No opinions were identified that would impede the implementation of these measures. Since a variety of wireless channels are in use at disaster sites, the ability to secure multiple wireless communication paths is considered to be advantageous. As we felt at the time of the interviews, the people working at the disaster site are basically not experts in engineering technology. It is therefore essential to provide a UI that allows the users to grasp the situation at a glance. Several requests were received for mutual voice communication with the rescuers. This is not currently planned for SAAS and is an item that should be seriously considered for UI implementation in the future.

V. Conclusion

In this paper, we propose a swarm active audition system (SAAS) in which multiple drones and robots can work together to search for people in need of rescue, and demonstrate its effectiveness, focusing on sound source localization by multiple drones. In addition, to verify whether SAAS is useful in actual disaster sites, we interviewed parties who have engaged in disaster site work and discussed its usefulness by contrasting the results of the interviews and the functions of SAAS. As a result, the SAAS functions are all in line with user requests although a concern if detecting low volume voices raised, which can be solved by total SAAS not only by drone audition. In addition, there were strong requests for a voice communication function as a function desired in the future. We will consider introducing this function in the future.

Acknowledgment

This work was supported by KAKENHI JP22F22769 and JP22KF0141, JST CREST JPMJCR19K1, and F-REI (JPFR23010102).

References

[1] K. Nakadai *et al.*, "Active audition for humanoid," in *AAAI*, AAAI, 2000, pp. 832–839.
[2] K. Okutani *et al.*, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *IROS*, IEEE, 2012, pp. 3288–3293.
[3] K. Nakamura *et al.*, "Real-time super-resolution sound source localization for robots," in *IROS*, IEEE, 2012, pp. 694–699.
[4] M. Basiri *et al.*, "Robust acoustic source localization of emergency signals from micro air vehicles," in *IEEE/RSJ IROS*, 2012, pp. 4737–4742.
[5] T. Ohata *et al.*, "Improvement in outdoor sound source detection using a quadrotor-embedded microphone array," in *IROS*, 2014, pp. 1902–1907.
[6] K. Furukawa *et al.*, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *IROS*, IEEE, 2013, pp. 3943–3948.
[7] T. Ishiki and M. Kumon, "Design model of microphone arrays for multirotor helicopters," in *IROS*, 2015, pp. 6143–6148.
[8] K. Washizaki *et al.*, "Position estimation of sound source on ground by multirotor helicopter with microphone array," in *IROS*, 2016, pp. 1980–1985.
[9] K. Yamada *et al.*, "Belief-driven control policy of a drone with microphones for multiple sound source search," in *IROS*, 2019, pp. 5326–5332.
[10] K. Nakadai *et al.*, "Development, deployment and applications of robot audition open source software HARK," *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 16–25, 2017.
[11] K. Hoshiba *et al.*, "Design and assessment of sound source localization system with a UAV-embedded microphone array," *JRM*, vol. 29, no. 1, pp. 154–167, 2017.
[12] L. Wang and A. Cavallaro, "Time-frequency processing for sound source localization from a micro aerial vehicle," in *ICASSP*, 2017, pp. 496–500.
[13] J.-S. Lauzon *et al.*, "Localization of RW-UAVs using particle filtering over distributed microphone arrays," in *IROS*, 2017, pp. 2479–2484.
[14] M. Strauss *et al.*, "DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization," in *IROS*, IEEE, 2018, pp. 5735–5742.
[15] Y. Hioka *et al.*, "Speech enhancement using a microphone array mounted on an unmanned aerial vehicle," in *IWAENC*, IEEE, 2016, pp. 1–5.
[16] S. Yoon *et al.*, "Two-stage adaptive noise reduction system for broadcasting multicopters," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016, pp. 219–222.
[17] L. Wang and A. Cavallaro, "Acoustic sensing from a multi-rotor drone," *IEEE Sensors Journal*, vol. 18, no. 11, pp. 4570–4582, 2018.
[18] Z.-W. Tan *et al.*, "An efficient dilated convolutional neural network for UAV noise reduction at low input SNR," in *APSIPA ASC*, 2019, pp. 1885–1892.
[19] B. Yen *et al.*, "Multi-sensory sound source enhancement for unmanned aerial vehicle recordings," *Applied Acoustics*, vol. 189, p. 108 590, 2022.
[20] W. N. Manamperi *et al.*, "Drone audition: Audio signal enhancement from drone embedded microphones using multi-channel Wiener filtering and Gaussian-mixture based post-filtering," *Applied Acoustics*, vol. 216, p. 109 818, 2024.
[21] L. Wang and A. Cavallaro, "Deep learning assisted time-frequency processing for speech enhancement on drones," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 6, pp. 871–881, 2021.
[22] S. Uemura *et al.*, "Outdoor acoustic event identification using sound source separation and deep learning with a quadrotor-embedded microphone array," in *The 6th International Conference on Advanced Mechatronics*, 2015.
[23] T. Morito *et al.*, "Partially shared deep neural network in sound source separation and identification using a UAV-embedded microphone array," in *IROS*, 2016, pp. 1299–1304.
[24] O. Sugiyama *et al.*, "Outdoor acoustic event identification with DNN using a quadrotor-embedded microphone array," *JRM*, vol. 29, no. 1, pp. 188–197, 2017.
[25] T. Fuhrman *et al.*, "An interactive indoor drone assistant," in *IROS*, 2019, pp. 6052–6057.
[26] M. Kumon *et al.*, "Alternating drive-and-glide flight navigation of a kiteplane for sound source position estimation," in *IROS*, 2021, pp. 2114–2120.
[27] T. Yamada *et al.*, "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, 2021, ISSN: 1660-4601.
[28] T. Yamada *et al.*, "Placement planning for sound source tracking in active drone audition," *Drones*, vol. 7, no. 7, 2023.
[29] B. Yen *et al.*, "A performance assessment on rotor noise-informed active multidrone sound source tracking methods," *Drones*, vol. 8, no. 6, 2024.
[30] S. Niijima *et al.*, "Semi-automatic town-scale 3D mapping using building information from publicly available maps," *IEEE Access*, vol. 10, pp. 32 244–32 254, 2022.
[31] Y. Sato *et al.*, "STP4: Spatio temporal path planning based on pedestrian trajectory prediction in dense crowds," in *RiTA 2022: The 10th International Conference on Robot Intelligence Technology and Applications*, 2022.