2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

2025 IEEE/SICE International Symposium on System Integration (SII) | 979-8-3315-3161-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/SII59315.2025.10871001

Locating Survivors' Voices in Disaster Sites Using Quadcopters Based on Modeling Complicated Environments by PyRoomAcoustics and SSL by MUSIC-based Algorithms

Masachika Kamada[1], Junji Yamato[2], Yasuhiro Oikawa[1], Hiroshi G. Okuno[1], and Jun Ohya[1]

***Abstract*— To enable the practical use of quadcopters equipped with microphone arrays in disaster sites for locating survivors' voices, this paper proposes a comprehensive method for modeling and simulating complex acoustic environments using PyRoomAcoustics, and for locating sound sources using variants of the MUSIC algorithms. By comparing impulse responses in PyRoomAcoustics simulations with those in real environments, we observed a high degree of correlation, indicating the simulations' suitability for real-world applications. Utilizing these simulations, we identified the optimal microphone array configuration for sound source localization (SSL) and examined the relationship between flight altitude and SSL performance. Key insights include minimizing ground reflection impacts at higher altitudes and enhancing SSL performance at lower altitudes with reduced ground reflectivity. Additionally, we found that power variations among multiple sound sources significantly affect the SSL performance of weaker sources. Among the MUSIC algorithm variants, iGEVD-MUSIC achieved the highest SSL performance, successfully locating multiple sound sources, including human voices. These findings demonstrate that the proposed simulation method is a valuable tool for developing and optimizing SSL techniques and parameters for real-world quadcopter applications. Furthermore, the insights gained from these simulations can be directly applied to the practical deployment of microphone array-equipped quadcopters in disaster response scenarios, aiding in the precise localization of survivors. This research significantly advances SSL methods and the practical realization of quadcopters for disaster site applications, ultimately enhancing the effectiveness and reliability of search and rescue operations.**

I. INTRODUCTION

Many regions on Earth face the threat of natural disasters such as earthquakes. Japan's 2024 Noto Peninsula earthquake has painfully reminded us of this fact. This earthquake caused devastating damage, resulting in 241 fatalities, over 1,000 injuries, and approximately 70,000 homes damaged. However, this issue is not limited to Japan. Historically, major earthquakes have occurred globally, such as the 2005 Kashmir earthquake and the 2023 Turkey-Syria earthquake, each causing over 70,000 deaths and significant damage. Additionally, the 1994 Northridge earthquake in California, USA, led to substantial economic losses and significant casualties due to the collapse of buildings.

Not only earthquakes, but also landslides caused by heavy rains, which are considered to be a consequence of recent global warming, underline the importance of rescuing people buried under debris. Moreover, it's not just natural disasters that cause the collapse of buildings; wars in regions such as the Ukraine and the Gaza Strip have led to similar destruction, making rescue efforts equally crucial.

Given these circumstances, establishing an effective human rescue process during disasters is urgent. Specifically, there is significant demand for technologies capable of quickly and accurately identifying hidden survivors beneath collapsed buildings, debris, and other obstacles.

This study focuses on developing a solution that utilizes quadcopters equipped with microphone arrays, as illustrated in Fig. 1, to locate survivors by their voices. This approach can provide the rescue teams that operate the quadcopters with the information to conduct rapid and efficient rescue operations at remote disaster sites, where the rescue team members do not need to enter narrow gaps or spaces in collapsed buildings or debris to search for survivors, thus avoiding secondary injuries for the members.

Despite ongoing research into sound source localization (SSL) quadcopters, practical application has yet to be achieved. The reasons for this include the lack of comprehensive testing under real-world disaster conditions and the prohibitive costs associated with such testing. Therefore, simulation-based approaches have become very important. The requirements for such a simulator include:

- Reproduce highly complicated acoustic environments accurately.
- Obtain the optimal configuration for the microphone array attached to the quadcopter.
- Clarify factors and conditions that affect SSL performance.

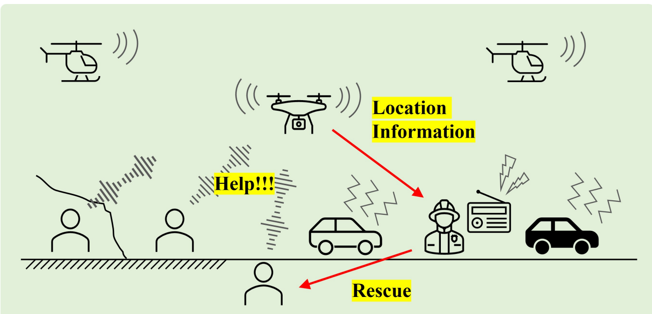

Fig. 1. The envisioned rescue process using SSL quadcopters.

[1]Waseda University, Tokyo, Japan. Masachika Kamada kmd@ruri.waseda.jp; Yasuhiro Oikawa yoikawa@waseda.jp; Hiroshi G. Okuno okuno@ieee.org; Jun Ohya ohya@waseda.jp.
[2]Kogakuin University, Tokyo, Japan. yamato@ieee.org

- Clarify which variant of the MUSIC algorithm can robustly achieve SSL in environments with ego-noise caused by the quadcopter.
- Locate multiple sound sources, including human voices, accurately.

To achieve the above-mentioned requirements, this paper presents the development of an acoustic simulation designed to assist rescue teams in conducting more efficient operations. Our approach is to develop a PyRoomAcoustics-based simulator that models real acoustic environments, including the above-mentioned complexities, and tests various SSL algorithms. Such a simulator is expected to evaluate the performance of unmanned aerial vehicles (UAVs)-based SSL for disaster applications.

II. RELATED WORK

In the quest to realize the rescue process depicted in Fig. 1, the necessity for real-time SSL technology becomes apparent, as it addresses the weaknesses of image-based survivor detection, such as occlusions and limitations caused by imaging conditions. Accurate SSL results can significantly enhance the capabilities of sound source separation and speech enhancement. However, multimodal SSL methods utilizing audio and visual information [1], [2] face challenges in obstructed or excessively dark environments. Our study focuses on survivor detection solely through audio means.

In the realm of SSL, the enhanced versions of the Multiple Signal Classification (MUSIC) algorithm [3], known as SEVD-MUSIC (Standard EigenValue Decomposition MUSIC), have been widely applied. The MUSIC algorithm employs eigenvalue decomposition, represented by:

$$R\mathbf{e}_i = \lambda_i \mathbf{e}_i, \tag{1}$$

where R denotes the spatial correlation matrix, $\mathbf{e}_i$ the eigenvector, and λ_i the eigenvalue. SEVD-MUSIC requires the number of sound sources to be known in advance and identifies larger eigenvalues as sound signals, while treating the smaller ones as directional noise. The MUSIC spectrum, indicating the probability of sound source presence, is calculated by:

$$P(\theta) = \frac{|\mathbf{v}^H(\theta) \cdot \mathbf{v}(\theta)|}{\sum_{i \in noise} |\mathbf{v}^H(\theta)\mathbf{E}_i|^2}. \tag{2}$$

Here, $\mathbf{v}(\theta)$ is the steering vector in the direction of θ, and $\mathbf{E}_n$ represents the array of noise eigenvectors $\{\mathbf{e}_{N+1}, \ldots, \mathbf{e}_M\}$, where $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ are signal eigenvectors each of which corresponds to each of the N sound sources. The steering vector element $\mathbf{v}_{m,k}$ directed towards the mth microphone is expressed as:

$$\mathbf{v}_{m,k} = r_m e^{-j2\pi f_k \tau_m}, \tag{3}$$

where r_m is the attenuation rate of sound from the source to the mth microphone, τ_m the duration for sound propagation, f_k the frequency, and j the imaginary unit.

GEVD-MUSIC [4] employs a noise correlation matrix (NCM), denoted as K, to mitigate directional noise interferences and alleviate the specified requirement. This technique enhances signal clarity by leveraging generalized eigenvalue decomposition (GEVD) and noise whitening, expressed mathematically as follows:

$$R\mathbf{e}_i = \lambda_i K \mathbf{e}_i. \tag{4}$$

In contrast, GSVD-MUSIC [5] utilizes generalized singular value decomposition (GSVD), offering an expedited approach compared to GEVD. The iterative versions, iGEVD-MUSIC [6] and iGSVD-MUSIC [7], introduce an advanced level of noise reduction, extending their capabilities to non-stationary noise through periodic updates to the NCM. The continuous advancements in MUSIC technology, underscored by the strategic manipulation of the NCM, emphasize the critical role of refining NCM identification as a focal point of ongoing research.

Ince et al. advanced the idea of creating a comprehensive database for NCMs specific to robot motor noise, i.e., *ego-noise*, utilizing monitoring data from robot motors as a key to query and retrieve suitable matrices [8]. However, this concept faces implementation challenges with quadcopters, given their complex control and posture parameters, which result in the need for a vast database to accommodate the varied conditions.

Furukawa et al. proposed a novel approach to estimate the NCM directly from quadcopter monitoring data using Gaussian process regression [9]. This method aims to suppress the drone's ego-noise effectively. The testing of this approach was carried out in an anechoic chamber, focusing on direct sounds. Nonetheless, its application in real-world scenarios is limited by environmental factors such as sound reflections and reverberations that were not considered in the controlled setting.

Yen et al. introduced an alternative strategy to obtain the NCM by employing directional microphones to capture the motor noise directly [10]. While this method promises direct noise measurement in a static condition where the quadcopter is supported, it has the drawback of increasing the overall device size, posing a conflict with the design goal of maintaining compact and user-friendly systems.

Strauss et al. contributed to the field by developing the DREGON dataset, designed explicitly for SSL research involving UAVs with integrated microphone arrays [11]. This dataset, consisting of indoor recordings, is instrumental in evaluating SSL methods, particularly for tasks like quadcopter-based search and rescue operations. However, its utility is somewhat limited by the specific microphone array configuration, which might not easily adapt to new or varying experimental setups.

Current research on employing quadcopters for SSL in disaster search and rescue scenarios has identified significant challenges. These include managing multiple sound sources, such as the voices of survivors amidst environmental noise, and addressing the intricacies of navigating complex terrains, which are highlighted concerns. A practical solution to these issues is vital for applying quadcopter-based SSL technologies in disaster response efforts, emphasizing the need for robust advancements suitable for dynamic and unpredictable disaster environments.

III. ENHANCED ACOUSTIC SIMULATION FOR SSL QUADCOPTERS IN DISASTER SITES

In this study, we introduce a methodology that leverages PyRoomAcoustics [12] to simulate the acoustic dynamics of disaster environments, facilitating the recreation of soundscapes that can be detected by microphone arrays mounted on quadcopters. Our simulations, based on real-world acoustic characteristics, allow for preliminary SSL performance estimation and operational strategy assessment before deployment. This framework replicates the acoustic challenges quadcopters face in disaster settings, permitting the assessment of SSL performance within these contexts. The MUSIC algorithm underscores the viability of both 2D and 3D SSL by constructing steering vectors, as delineated in Equation (2). For clarity and focus, this study concentrates on 2D SSL within the horizontal plane, simplifying the environmental model and narrowing the scope to evaluate factors influencing localization accuracy.

In the simulation of disaster environments, a pivotal differentiation is established between the following two scenarios: one where humans are located on the ground surface, termed "*Surface Scenario Simulation*," enabling direct transmission of their sounds to the microphones, and the other where survivors are trapped beneath the rubble or confined, referred to as "*Beneath-Debris Scenario Simulation*," which obstructs the direct sound propagation. This distinction is essential for precisely emulating and assessing the SSL capabilities of quadcopters in a range of intricate and challenging disaster contexts. Note that, in the case of "*Beneath-Debris Scenario Simulation*," survivors in real space are trapped beneath debris, etc.; however, due to the complexities discussed later, simulations are conducted under the conditions of the "*Surface Scenario Simulation*."

A. Surface Scenario Simulation

In this simulation, the primary challenges related to SSL are depicted in Fig. 2. Figure 2 (left) shows the paths of ego-noise toward the microphone array, while Fig. 2 (right) depicts the paths of a survivor's voice reaching the microphone array. In addition to the direct sound path (red), multiple indirect paths (blue) contribute to the sound reflections received by the microphone array. Our simulation replicates these effects, which are typically not considered in anechoic chamber settings.

In disaster-impacted areas, terrain complexities often arise from collapses and ground movements. To simulate this complexity accurately, we randomly assign coordinates to points and connect them with lines to form barriers within PyRoomAcoustics, as shown in Fig. 3. The diagram uses brown lines to depict the ground, and by specifying different materials, we can adjust the sound reflection characteristics. Coordinates can also be freely assigned, for example, to form shapes of buildings or houses. By using materials defined in PyRoomAcoustics for the ground, it is possible to simulate different absorption rates at different sound frequencies.

In PyRoomAcoustics simulations, the number of walls is related to a significant increase in computational load.

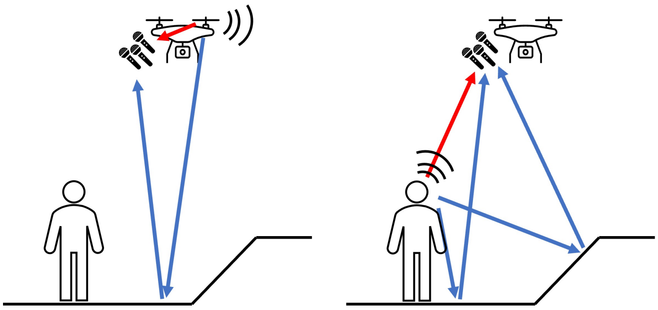

Fig. 2. Illustration of ego-noise and a human's voice propagation paths towards the microphone array, emphasizing direct sound paths in red and reflected sound paths in blue.

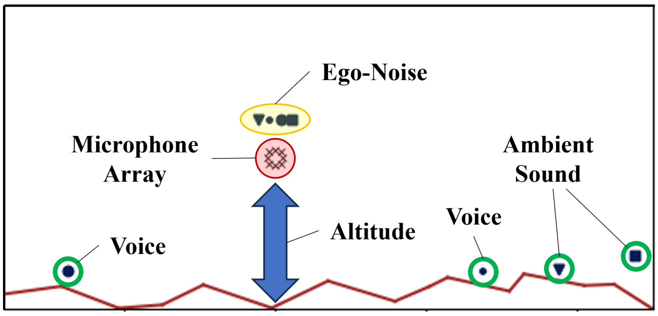

Fig. 3. A disaster site modeled using PyRoomAcoustics, featuring a microphone array and sound sources, including ambient sounds.

Therefore, by simplifying the shape of the ground, it is possible to simulate dominant sound events with a lower computational load. By setting the energy absorption rate of walls other than the ground to 1, it is possible to recreate an open space where sound does not reflect off surfaces other than the ground.

We also place sources of ambient sounds, such as survivors' voices and engine noises, alongside the quadcopter's ego-noise, including rotor and propeller noises and the microphone array, to create a comprehensive model of the disaster site environment, assuming the quadcopter in a hovering state. These sound sources can be freely positioned using pre-recorded audio files.

A key parameter in our simulation is the number of sound reflections considered. Opting for a low number of reflections may oversimplify the model, detracting from its realism, whereas a high number of reflections significantly increases computational costs. Notably, the computational burden escalates exponentially with each additional reflection, necessitating careful selection of a minimal yet realistic reflection count. We found a balance between SSL performance and processing time by adjusting the number of reflections, ultimately fixing it at four. This decision stems from the observation that, in outdoor settings, reflections typically dissipate upward after a few bounces and cease to contribute meaningfully to further sound reflections. Our simulations confirmed that increasing the number of reflections beyond four does not significantly enhance SSL results, thus validating our choice.

B. Beneath-Debris Scenario Simulation

This subsection delves into scenarios where direct sound fails to reach the microphone array, such as when individuals are trapped beneath debris. Survival rates drop to nearly zero for those fully buried without gaps, but they can be higher for those in spaces under collapsed structures, where their voices might escape through gaps, as illustrated in Fig. 4 (a). In PyRoomAcoustics, acoustic simulations are calculated using the Image Source Model, which places virtual sound sources on the opposite side of walls matching the actual sound source to compute sound reflections. Therefore, it cannot handle the spatial diffusion of sound to locations that cannot be reached by reflections.

To overcome this challenge and accurately simulate the acoustic environment beneath the debris, we divide the problem into two steps. First, within the spaces under the debris, sound attenuates, and the sound leaking to the surface weakens. Taking this into account, we determine the appropriate signal-to-noise ratio (SNR) at the openings post-attenuation. Second, we assume that the sound is emitted from the positions of the openings and reaches the microphone. Using the determined SNR settings, we can reproduce the behavior of the sound, which is actually emitted from beneath the debris. Therefore, we only need to model the transmission from the openings to the microphone array, which simplifies the situation as depicted in Fig. 4 (b). That is, by redefining the situation in Fig. 4 (a) as shown in Fig. 4 (b), and aligning it with the surface scenario, we make it possible to simulate the beneath-debris scenario, thereby resolving the limitations of PyRoomAcoustics.

C. SSL Quadcopter Strategy and Evaluation

Employing the outlined acoustic simulation approach for SSL on quadcopters at disaster sites facilitates an exhaustive pre-production analysis of SSL performance and the development of operational strategies. This evaluative framework includes a series of essential steps, elaborated upon below, guaranteeing the refined optimization and effectiveness of quadcopter deployment in actual disaster contexts.

1) Optimization of Microphone Array Geometry
2) Development of Operational Strategies
3) Selection of SSL Algorithms

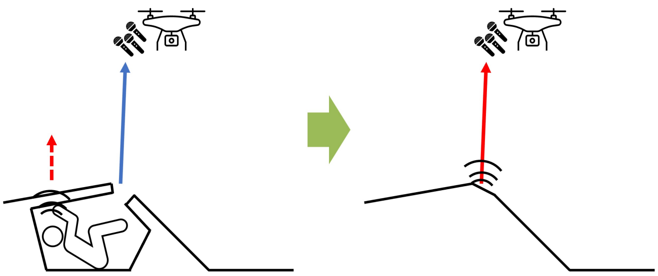

Fig. 4. Acoustic Propagation in Beneath-Debris Scenarios: (a) Illustration of sound paths from beneath the debris to surface gaps; (b) Simplified model for scenarios lacking direct sound paths, excluding openings.

IV. ASSESSING THE REALISM OF PYROOMACOUSTICS SIMULATIONS

This section assesses how well PyRoomAcoustics simulations can mirror real-world acoustic settings. Despite PyRoomAcoustics being a prominent acoustic simulator, there has been a lack of comprehensive validation of its simulation's fidelity to real environments, as outlined in the foundation papers [11]. Establishing the accuracy of PyRoomAcoustics is crucial for demonstrating the validity and reliability of our simulations.

For our methodology, we obtained impulse responses from outdoor settings and matched them against simulations from PyRoomAcoustics. Given that PyRoomAcoustics models acoustics through impulse response calculations, this approach validates the simulation's accuracy in replicating the acoustic features of real-world environments.

Figure 5 illustrates the experimental arrangement. To accurately capture sound emission in outdoor scenarios and to capture both direct sound and ground-reflected echoes, we positioned the loudspeaker facing the ground vertically. A microphone placed 0.5 meters beneath the loudspeaker recorded the sound directly from the source and reflections bouncing off the ground during quiet periods to minimize environmental noise. Table I provides comprehensive details on the experimental tools employed in this research.

Utilizing the Impulse Response Measurer application in MATLAB, we emitted a 10-second Exponential Swept Sine signal four times to measure impulse responses. The spatial

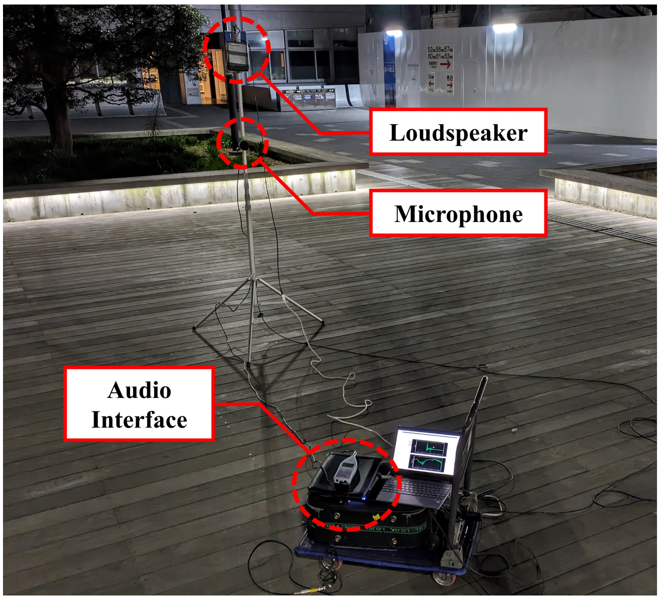

Fig. 5. Experimental setup for measuring impulse responses.

TABLE I
SUMMARY OF EQUIPMENT USED IN THE EXPERIMENT.

Equipment Type	Model
Loudspeaker	YAMAHA MS101III
Microphone	RION NL-32
Audio Interface	MOTU UltraLite AVB

arrangement between the loudspeaker and the microphone remained unchanged, but we altered the microphone's elevation above the ground to 1.5 meters, 2.0 meters, and 2.5 meters to obtain varied measurements. These tests were conducted on wooden, grass, and concrete surfaces.

Figure 6 illustrates the impulse responses derived from our experiments. Specifically, Fig. 6 (a) illustrates the impulse response recorded in the actual environment, whereas Fig. 6 (b) depicts the impulse response generated by PyRoomAcoustics under comparable conditions. We applied normalization to both sets of responses for consistency. Table II indicates a comparison of acoustic reflection times measured in actual environments against those simulated using PyRoomAcoustics, across different ground materials and heights. This comparison shows deviations confined to within 1% for most scenarios, demonstrating PyRoomAcoustics' remarkable accuracy in replicating real-world acoustic behavior.

As evidenced in Fig. 6, the simulation also accurately replicates the amplitudes, which can be fine-tuned by modifying the energy absorption rate in PyRoomAcoustics. While Fig. 6 showcases the results for wood, similar outcomes were observed for other materials. The waveform from 0 to 0.005 seconds in the actual measurements, marked by direct sound followed by noise, as depicted in Fig. 6 (a), is due to the speaker's performance and resonance, ideally aligning more closely with PyRoomAcoustics simulations. Thus, the negligible discrepancy between PyRoomAcoustics simulations and actual measurements validates the effectiveness of our simulation approach in accurately replicating the acoustic properties of genuine environments.

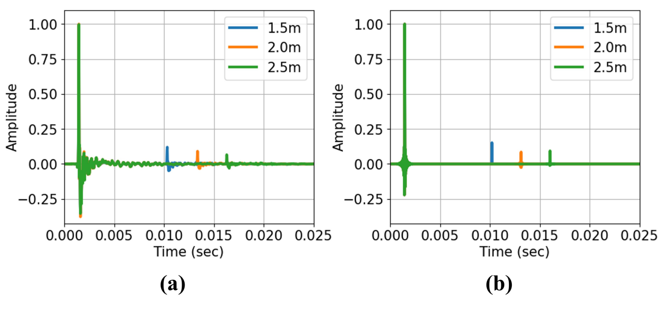

Fig. 6. Comparison of impulse responses on a wooden floor: (a) actual measurement and (b) PyRoomAcoustics simulation.

TABLE II
COMPARISON OF ACOUSTIC REFLECTION TIMES IN ACTUAL AND SIMULATED ENVIRONMENTS BY PYROOMACOUSTICS.

Height (m)	Ground	Actual (s)	Simulated (s)	Ratio
1.5	Wood	0.0103	0.0102	0.9884
1.5	Grass	0.0102	0.0102	0.9972
1.5	Concrete	0.0102	0.0102	1.0038
2.0	Wood	0.0134	0.0131	0.9826
2.0	Grass	0.0132	0.0131	0.9927
2.0	Concrete	0.0132	0.0131	0.9909
2.5	Wood	0.0163	0.0160	0.9843
2.5	Grass	0.0162	0.0160	0.9912
2.5	Concrete	0.0162	0.0160	0.9898

V. VALIDATION OF OPERATIONAL CAPABILITIES IN ACOUSTICS

A. *Optimization of Microphone Array Geometry*

This subsection explores the optimization of microphone array configurations for UAV-based SSL. Building on the findings of Hoshiba et al. [13], who identified spherical arrays as the most consistently performing configuration, this study further refines array placement and dimensions to mitigate rotor noise and adhere to weight restrictions. Unlike their design using counterbalances, our strategy positions the array beneath the quadcopter's chassis, as shown in Fig. 7. This positioning reduces weight without losing SSL performance, accommodating the quadcopter's limited carrying capacity.

Our objective is to pinpoint the ideal radius for the spherical microphone array. The placement of the microphones within the spherical array is fixed based on previous studies. We adjusted the radius in increments of 0.05 m, ranging from 0.05 m to 0.45 m, as detailed in the experimental setup shown in Fig. 8. Analyzing the MUSIC spectrum generated by the SEVD-MUSIC algorithm, we identified the most suitable radius for the spherical microphone array to be mounted on the SSL quadcopter.

The results, illustrated in Fig. 9, indicate that a larger radius improves the spectrum's amplitude at the correct sound source directions, indicating higher SSL performance for locating survivors. However, increasing the microphone array's size adds weight, which is contrary to the requirements of rescue scenarios. Thus, balancing efficiency and

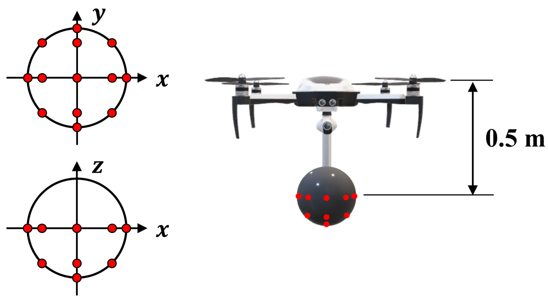

Fig. 7. Microphone array placement. Red dots represent the microphones.

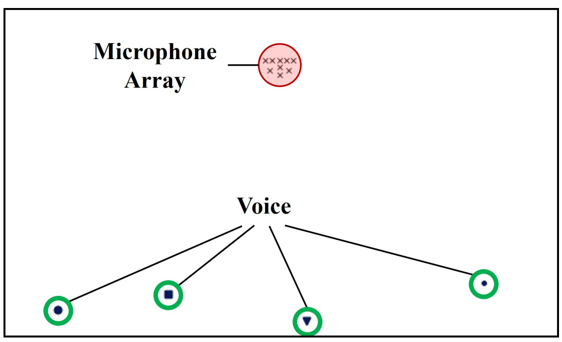

Fig. 8. Experimental setup in PyRoomAcoustics for determining the optimal radius of the spherical microphone array for the SSL quadcopter.

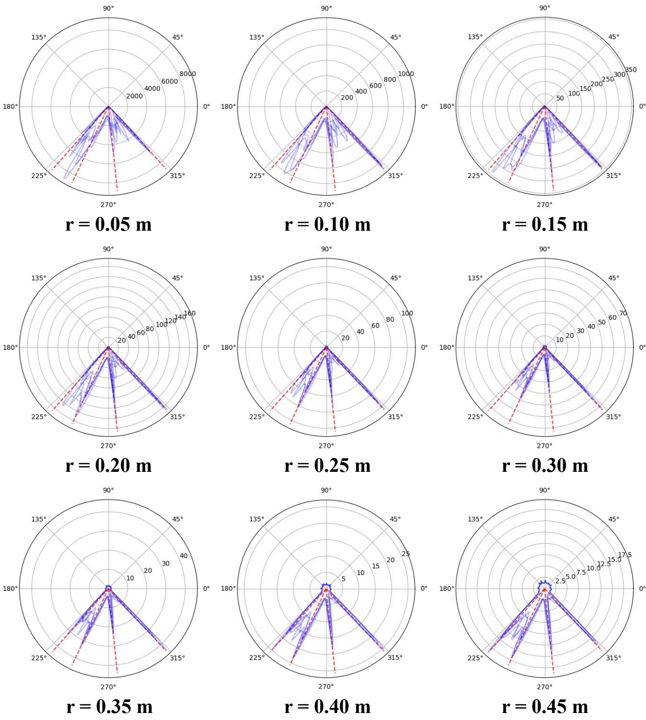

Fig. 9. MUSIC spectrum for different radii in polar coordinates.

practicality, this study identifies a radius of 0.2 m as the maximum feasible size, deeming it the optimal choice.

B. Development of Operational Strategy

This subsection outlines the approach to developing operational strategies through simulations set in disaster environments, with the parameters detailed in Table III. The standard SNR is established under conditions where the number of signal and noise sources is equal and their distance to the microphone array is the same.

For SSL, we applied the methods outlined in Table IV. iGEVD utilizes the NCM from moments just before localization. Based on [6], an NCM was constructed by extracting 140 frames, starting with 90 frames before reaching the optimal value during quadcopter flight. Methods labeled with a "dir" suffix employ ego-noise from reflection-free settings, while those with a "rev" suffix incorporate ground reflections into the NCM construction. The suffixes stable, acc, and diff in the method names represent different approaches to constructing the NCM. "Stable" assumes that ego-noise is stationary and uses all time frames to construct the NCM. "Acc" represents the accurate NCM for the time frame used for SSL. "Diff" adds a random error within ±5% to the accurate NCM defined by "acc", inspired by Furukawa et al. [9], who regressed the elements of the NCM from quadcopter monitor information.

The evaluation metrics utilized were Recall and Precision. The correct direction of the sound source was defined as the direction from the center of the microphone array to the sound source, within a ±3-degree range, considering only results within this criterion as True Positive.

TABLE III

DISASTER ENVIRONMENT SIMULATION EXPERIMENT PARAMETERS.

Parameter	Values
Altitude	{ 2 m, 3 m, 4 m, 5 m }
Ground Roughness	{ 0.1 - 1.0 m, 0.2 - 1.2 m }
Ground Material	{ Hard surface, Plasterboard, Wood }
Number of Voices	{ 1, 2, 3 }
Number of Ambient Sounds	{ 0, 1, 2 }
Standard SNR-ego	{ 8 dB, 11 dB, 14 dB }
Standard SNR-ambient	{ -3 dB, 0 dB, 3 dB }

TABLE IV

SUMMARY OF METHODS AND NOISE CORRELATION MATRIX.

Method Name	Algorithm	NCM	Reverberation
SEVD	SEVD-MUSIC	–	–
iGEVD	GEVD-MUSIC	Incremental	–
GEVD_stable_dir	GEVD-MUSIC	Entire	
GEVD_stable_rev	GEVD-MUSIC	Entire	✓
GEVD_acc_dir	GEVD-MUSIC	No Error	
GEVD_acc_rev	GEVD-MUSIC	No Error	✓
GEVD_diff_dir	GEVD-MUSIC	Error 5%	
GEVD_diff_rev	GEVD-MUSIC	Error 5%	✓

Figure 10 demonstrates how the metrics vary with the altitude of the microphone array, indicating an improvement in recall at higher altitudes. This trend prompted a deeper investigation into the underlying factors. Figure 11 shows the variation in metrics by ground material at different altitudes. At a 2 m altitude, the performance metrics improve in the order of hard surface (HS), plasterboard (PB), and wood (WL), correlating with a decrease in reflectivity. Nevertheless, at a 5 m altitude, this trend diminishes. Therefore, the degradation in metrics at lower altitudes can be attributed mainly to increased reflections from the ground.

Next, we analyzed data incorporating an equal mix of human voices and ambient sounds at altitudes where the impact of reflections is minimized (4 m and 5 m) to illustrate the variation in recall with the Standard SNR-ambient (Fig. 12). It was observed that both human voices and ambient sounds show a decrease in recall when their SNR is lower relative to the other. This finding suggests that SSL performance for the less audible sound sources diminishes when there is a significant power discrepancy among various sounds. Thus, it is inferred that an increase in power variance among multiple sound sources at lower altitudes leads to poorer performance metrics.

Subsequently, we explored the influence of the Standard SNR-ego on localization accuracy (Fig. 13). While metrics improve when the human voice is more prominent than ego-noise, significant changes in methods with already high SSL performance were not observed. This lack of significant change is likely due to the propellers being opposite the SSL direction, resulting in a low correlation with the steering vector.

Based on these observations, it is recommended that operations be conducted at the highest feasible altitude (5 m) to optimize performance. Additionally, minimizing ego-noise is crucial for enhancing SSL performance.

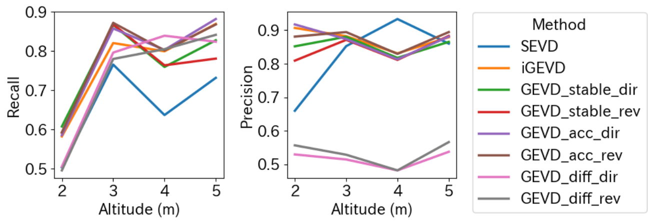

Fig. 10. SSL Results by Altitude.

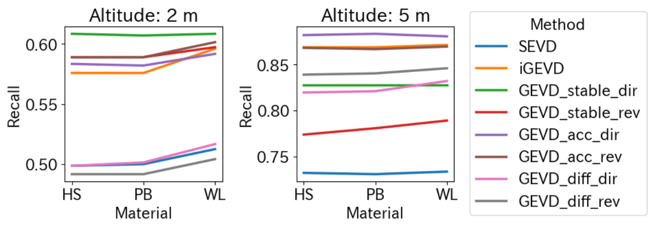

Fig. 11. SSL Results by Ground Material, Left: at 2 m Altitude, Right: at 5 m Altitude.

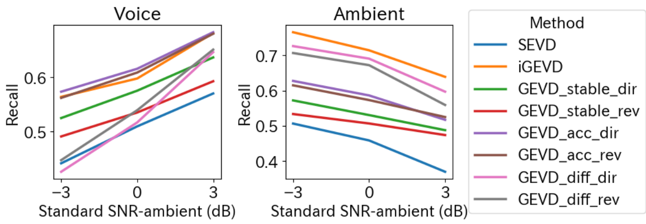

Fig. 12. SSL Results by Standard SNR-ambient, Left: Human Voices, Right: Ambient Sounds.

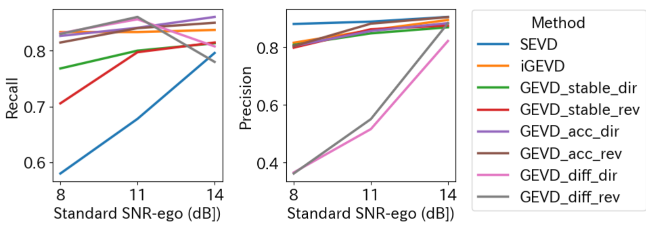

Fig. 13. SSL Results by Standard SNR-ego.

C. *Selection of SSL Algorithm*

At the altitude of 5 m, as depicted in Fig. 10, GEVD_acc, which is theoretical and used only in simulations, and iGEVD stand out for their high accuracy, with GEVD_stable also showing notable performance. On the other hand, SEVD struggles due to its insufficient noise cancellation capabilities, and GEVD_diff falls short by incorporating inaccuracies into the NCM, rendering it less effective than its counterparts. Therefore, among the methods compared in this study, it can be concluded that iGEVD-MUSIC is the most suitable for the practical application of SSL quadcopters.

Although iGEVD-MUSIC is designed to identify directional changes from one moment to the next, which could potentially challenge its ability to identify sustained sounds, the results of this study suggest otherwise. Specifically, conditions that could potentially lead to instability in SSL for sustained sounds include scenarios with prolonged human voice sounds. To address this concern, we included such sounds in our dataset and conducted experiments. The inclusion of prolonged human voice sounds did not impair its performance. Instead, iGEVD-MUSIC demonstrated a competent level of SSL ability, thanks to its straightforward approach in NCM derivation. This simplicity suggests robustness in its methodology, indicating that iGEVD-MUSIC is certainly capable of handling complex acoustic scenarios with efficiency.

VI. CONCLUSIONS

To locate the voices of survivors beneath debris in disaster sites using quadcopters equipped with microphone arrays, this paper proposes a method for modeling and simulating complex acoustic environments using PyRoomAcoustics and for locating sound sources using variants of MUSIC algorithms. Considering the limitations of PyRoomAcoustics in modeling sound transmission from beneath debris, we propose a simplification where the attenuation of sound as it leaks from beneath the debris to the surface is taken into account. By treating the sound source as if it were on the surface, it becomes feasible to simulate it in the same manner as a surface-level sound source.

To verify the conformity of PyRoomAcoustics simulations with reality, we compared impulse responses in PyRoomAcoustics simulation environments and in real space, confirming a 99% match rate, as shown in Table II. This verification, which has not been conducted in past works, underscores the effectiveness of the proposed approach and implies that PyRoomAcoustics simulations fit well with real-world applications.

Utilizing PyRoomAcoustics simulations, (1) we revealed the optimal configuration of microphone arrays for SSL, and (2) we examined the relationship between flight altitude and SSL performance using variants of the MUSIC algorithm.

In (1), we evaluated SSL results by changing the microphone array radius within the simulation environment. As a result, it was found that increasing the radius of the microphone array as much as possible, unless the weight of the microphone array exceeds the upper limit of the payload of the quadcopter, improves SSL performance. Furthermore, the proposed simulation method elucidated the optimal configuration of microphone arrays for SSL quadcopters, explored the relationship between flight altitude and SSL performance using variants of the MUSIC algorithm, and examined comparisons of SSL algorithms.

In (2), simulations that change parameters such as the altitude of the microphone arrays, ground reflectivity, the number of sound sources, and SNR clarified conditions that enhance SSL performance. Important findings include that higher altitudes minimized the impact of ground reflections, leading to improved SSL performance, while at lower altitudes, lower ground reflectivity improved SSL performance, and the reflection of ego-noise significantly affected SSL performance. Additionally, as can be seen in the comparison of Standard SNR-ambient, power variations among multiple sound sources reduced the SSL performance of weaker sources.

Moreover, as a result of comparing MUSIC-based SSL methods, it was confirmed that iGEVD-MUSIC achieves the most accurate SSL, demonstrating robustness to variations in Standard SNR-ego and suitability for SSL quadcopter applications. In addition, iGEVD-MUSIC can successfully locate each of multiple sound sources, including human voices.

Existing research has proposed groundbreaking SSL methods, but a significant challenge has been the inability to adequately consider the acoustic complexity of disaster site environments, resulting in a lack of practical application. This study proposes a method to simulate the complex acoustic environments required for practical use at disaster sites. This enables low-cost testing of SSL systems and methods using quadcopters, as demonstrated by the results and insights obtained from the aforementioned simulations.

Looking forward, we anticipate that advancements in managing sound source dynamics, minimizing computational costs during 3D simulations, and addressing the effect of wind noise will energize the research community. By focusing on improving simulation adaptability and forging stronger links between theoretical models and practical applications, we aim to contribute to the advancement of technologies that support rescue activities at disaster sites. We hope this study will expedite the social adoption of sophisticated search and rescue techniques and bolster disaster response effectiveness, thereby making a significant impact on real-world disaster management and response efforts.

References

[1] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, ”Audio-visual sensing from a quadcopter: dataset and baselines for source localization and sound enhancement,” in Proc. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, Nov. 2019, pp. 5320-5325.

[2] R. Sanchez-Matilla, L. Wang, and A. Cavallaro, ”Multi-Modal Localization and Enhancement of Multiple Sound Sources from a Micro Aerial Vehicle,” in Proc. of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, Oct. 2017, pp. 1591-1599.

[3] R. Schmidt, ”Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas Propagat., vol. 34, no. 3, pp. 276-280, Mar. 1986.

[4] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, ”Intelligent sound source localization for dynamic environments,” in Proc. of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, Dec. 2009, pp. 664-669.

[5] K. Nakamura, K. Nakadai, and G. Ince, ”Real-time super-resolution Sound Source Localization for robots,” in Proc. of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, Dec. 2012, pp. 694-699.

[6] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, “Incremental Noise Estimation in Outdoor Auditory Scene Analysis using a Quadrocopter with a Microphone Array,” JRSJ, vol. 31, no. 7, pp. 676-683, 2013 (In Japanese).

[7] T. Ohata, K. Nakamura, T. Mizumoto, T. Taiki, and K. Nakadai, ”Improvement in outdoor sound source detection using a quadrotor-embedded microphone array,” in Proc. of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, Sept. 2014, pp. 1902-1907.

[8] G. Ince, K. Nakamura, F. Asano, H. Nakajima, and K. Nakadai, ”Assessment of general applicability of ego noise estimation,” in Proc. of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, May 2011, pp. 3517-3522.

[9] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H. G. Okuno, ”Noise correlation matrix estimation for improving sound source localization by multirotor UAV,” in Proc. of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, Nov. 2013, pp. 3943-3948.

[10] B. Yen, Y. Li, and Y. Hioka, ”Rotor Noise-Aware Noise Covariance Matrix Estimation for Unmanned Aerial Vehicle Audition,” IEEE/ACM Trans. Audio, Speech, and Language Processing, vol. 31, pp. 2491-2506, June 2023.

[11] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, ”DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization,” in Proc. of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, Oct. 2018, pp. 1-8.

[12] R. Scheibler, E. Bezzam, and I. Dokmanić, ”Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms,” in Proc. of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, Apr. 2018, pp. 351-355.

[13] K. Hoshiba, K. Washizaki, M. Wakabayashi, T. Ishiki, M. Kumon, Y. Bando, D. Gabriel, K. Nakadai, and H. G. Okuno, ”Design of UAV-embedded microphone array system for sound source localization in outdoor environments,” Sensors, vol. 17, no. 11, Nov. 2017, Article 2535.