Drone audition: dataset and methods for ground surface material classification using drone noise in outdoor environment

Tsubasa Yano*, Benjamin Yen*† and Kazuhiro Nakadai*
* Tokyo Institute of Technology
E-mail: {yano, benjamin, nakadai}@ra.sc.e.titech.ac.jp

Abstract—**This paper considers a technique to estimate the material of ground surfaces directly below the drone based on the rotor noise generated by the drone. It is anticipated that this technique will enable us to assess the damage in areas that are not easily accessible by humans. In this study, drone noise was recorded while flying over multiple materials, such as asphalt, soil, and water, in a real environment with external noise, such as wind. Following, a machine learning-based model was trained using this dataset to classify the various materials using the recorded audio. Results indicate that while the classification was successful, there is a need for a classification method that is more robust to external noise.**

I. Introduction

In recent years, natural disasters such as earthquakes and landslides caused by torrential rains have become more frequent and often cause significant damage. In Japan, Typhoon No. 19 in 2019 caused more than 950 landslides in 20 prefectures, mainly in eastern Japan [1], and torrential rain that hit Kumamoto Prefecture in July 2020 caused several national roads to be closed due to overflowing rivers and landslides. Furthermore, the 2024 Noto Peninsula earthquake that occurred in January of this year caused more than 90 sections of prefectural roads to be closed to traffic [2]. Natural disasters caused by extreme weather and other factors are on the rise, not only in Japan but around the world [3], and countermeasures and responses are required.

One solution to these problems is the use of highly manoeuvrable drones. Drones can be used to quickly and inexpensively survey areas that are difficult to access directly by humans.

In this paper, we make use of the fact that the acoustic signal recorded by the microphone mounted on the drone includes not only noise directly from the drone's rotors, but also rotor noise that is reflected once on the ground surface. As such, we can address the problem of ground surface material estimation using this reflected sound with deep learning. Since there is no dataset for estimating the ground surface material outdoors, the creation of such a dataset will be conducted as well. In addition, since there is concern about the effects of external noise caused by strong winds in an outdoor environment, we also propose a custom-made windshield for the microphone array used in this study and investigate its effect against various forms of windshield materials.

II. Related work

The majority of studies conducted to assess the disaster situation by installing sensors on drones utilize cameras and microphones as sensors [4]. Although cameras are effective sensors and have been employed in disaster-stricken areas, their use is limited under conditions of poor visibility, such as at night. Microphones, however, are immune to such limitations and are currently being investigated as a potential solution, often referred to as ”drone audition” [5]–[10].

Most drone audition approaches involve suppressing drone rotor noise in order to detect voices coming from victims. In other words, research has been conducted with the objective of suppressing drone rotor noise. On the other hand, this paper takes a different approach to drone audition research in that it actively utilizes drone rotor noise not as an object to be suppressed but as useful information for understanding the material properties of the ground surface. If ground surface material estimation through reflected drone rotor noise is feasible, as illustrated in Fig. 1, it opens the possibility to detect landslides or road collapses by identifying changes in ground surface material or features. This is expected to be beneficial in comprehending the extent of the damage. To the best of our knowledge, few such studies have been conducted, with the exception of ours. Additionally, while our previous study [11] estimates the ground surface material using noise recorded while changing the ground surface material, it has yet to show that it is effective in a real environment. This is due to the data being recorded in an anechoic chamber, and thus noise other than drone rotor noise (e.g. wind noise) was not considered. Furthermore, the amount of data measured was extremely small (about one minute for each condition).

A potential application that utilizes research on the estimation of material properties from acoustic signals includes the sound impact inspection of bridge structures. Generally, in such inspections, skilled workers estimate the characteristics and degree of deterioration of bridge structures from the sound generated when the bridge is struck. However, recent years have seen a growing number of studies conducted to learn the skills of skilled inspectors using machine learning methods

[†] 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.

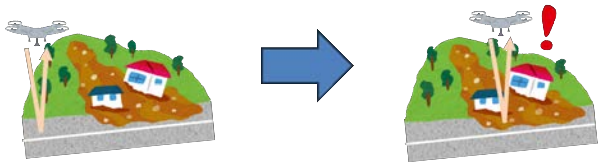


Fig. 1: Necessity of ground surface material estimation.

to obtain equivalent performance [12]. Some of these studies have reported the use of drones equipped with a percussion inspection function [13]. However, these studies were conducted on concrete, and no research has been conducted from the viewpoint of discriminating various materials from the reflected sound. Furthermore, drone rotor noise is treated as unnecessary background noise even when a drone with a percussion inspection function is used, and there is a lack of perspective on actively using drone noise, as in this paper.

III. Dataset construction

A. Guidelines for dataset design

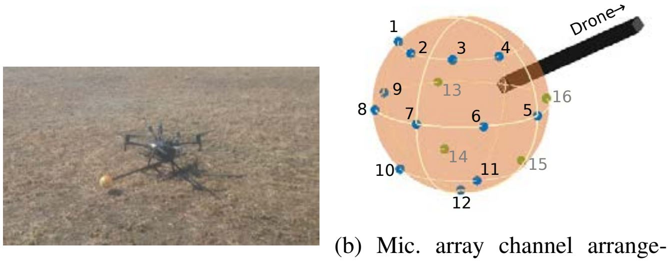
The factors that affect the recorded drone rotor noise include not only the material of the ground surface but also the drone's altitude, the rotational speed of each motor, and the geometry of the ground surface. In this paper, as a preliminary step in the application to outdoor environments, we decided to create a dataset based on the following guidelines.

1) All parameters are fixed except for the ground surface material and the drone's altitude.
2) The ground surface material is selected based on the actual disaster situation.
3) For the drone altitude, two types of altitude are set: one is the same height as that often used in actual disaster rescue scenes, and the other is lower than that but still high enough to reach above the human head. (This is because it is anticipated that the attenuation level of reflected drone noise, when flown too high, will make material estimation difficult.)
4) To secure the amount of data, the recording time for each material is as long as possible within the limits of the drone's battery.

B. Dataset creation method

As shown in Fig. 2, a 16-channel spherical microphone array is installed a distance forward of the drone, and a Raspberry Pi for recording and storage is installed at the rear. The installation position is adjusted with consideration of the center of gravity of the drone. In this state, when the drone is in operation, the radiated noise from the rotors is recorded by all channels of the microphone array for various types of ground surface materials.

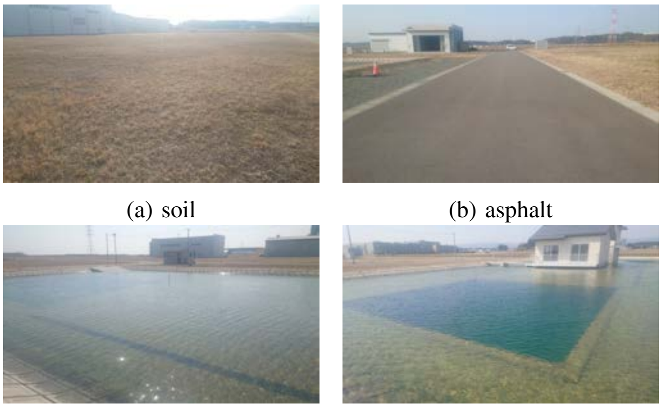
The materials selected for the ground surface were soil, asphalt, and water. Two types of water surface were selected: 0.7 m deep (hereafter, "shallow water") and 5 m deep (hereafter, "deep water"), for a total of four materials. All surfaces were nearly horizontal, without any slopes or large irregularities. The surface of each material is shown in the Fig. 3. Two drone altitudes, 5 m and 10 m, were selected in accordance with the previously established guidelines.



(a) Drone with equipments.

(b) Mic. array channel arrangement.

Fig. 2: Drone and microphone array for recording.



(a) soil

(b) asphalt

(c) shallow water (depth: 0.7 m)

(d) deep water (depth: 5 m)

Fig. 3: Ground surface materials.

Audio recording was first conducted at an altitude of 5 m above each material for 8 minutes while hovering. Upon completion, the drone ascended to an altitude of 10 m and recorded for another 8 minutes while hovering. For safety reasons, if the drone's low battery alarm sounded during the recording, the recording process was terminated and the drone was immediately returned for landing. In fact, during data collection for audio at an altitude of 10 m in shallow water, the recording was stopped after 6 minutes and 20 seconds due to the drone's low battery alarm sounding midway through the recording. This was probably because the drone's battery consumption was higher than expected due to the strong winds. Eventually, we were able to record approximately 1 h of data. The other conditions are shown in TABLE I.

C. Analysis of recorded data

To ascertain whether the recorded drone noise exhibits any anomalies, we conducted a comprehensive examination of

TABLE I: Dataset preparation specifications.

Aircraft in use	DJI Matrice 210
Number of array channels	16 (Sphereical)
Sampling rate	16 kHz

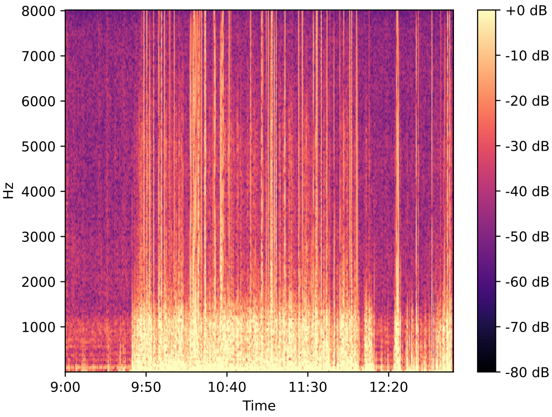


Fig. 4: Occurrences of sound clipping.

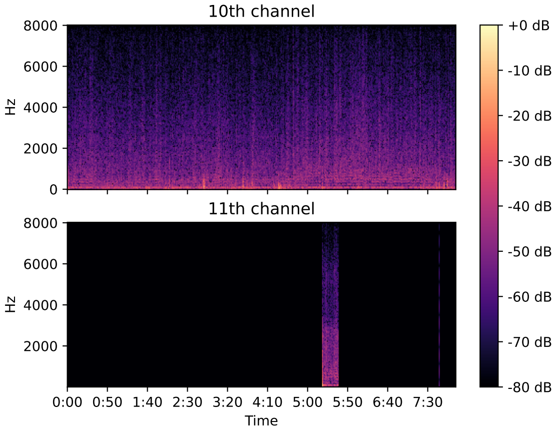


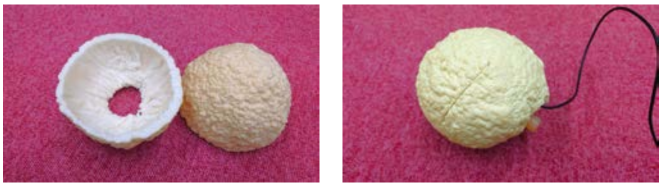
Fig. 5: Interval in which data value is zero.

the spectrogram and waveforms, complemented by a detailed auditory analysis. This process led to the identification of two defects.

- Sound clipping is frequently observed in the recorded data.
- There are intervals in which the recorded data value is zero or very close to zero.

Fig. 4 shows a spectrogram of the recording of the 14th channel (located at the rear side, see Fig. 2(b)) recorded over shallow water, with brighter colors indicating stronger intensity. It can be seen that impulsive signals are frequently recorded in the low and high frequency bands. The sound clipping caused by such impulsive noise occurs at a common time in many channels. This indicates that the cause of the clipping was strong winds during the recording. When such clipping occurs, the reflected sound signal cannot be picked up properly, which may result in degraded estimation performance. In contrast, a spectrogram from recordings of the 10th and 11th channels recorded above the soil in Fig. 5 illustrates a phenomenon in which the value of the recorded data becomes zero. As shown, especially in the 11th channel, nearly 90% of the recordings have a value of zero. The length of time and frequency of this phenomenon differ from channel to channel and recording to recording. After consulting with the microphone manufacturer, it was found that strong winds can cause the diaphragm of the MEMS microphones used to be stuck to the inner wall of the microphone, resulting in extremely low sensitivity.

As mentioned in Sec. III-B, it was windy at the time of recording, with an average wind speed of 4˜7m/s and a maximum of 17 m/s observed with the anemometer on hand. As a consequence of the aforementioned issues, a significant proportion of the dataset is unusable for ground surface material estimation. Consequently, only the usable parts of the recorded noise was extracted for the formation of the dataset. The specific method is described in Sec. V-A.



(a) Windshield components. (b) Attached to microphone array.

Fig. 6: Proposed windshield.

IV. Windshield Installation on Microphone Array

As a consequense of the findings from Sec. III-C, in order to robustly perform ground surface material estimation in an outdoor environment, it is critical to address the presence of external noise, including strong winds. One effective approach to achieve this is by attaching a windshield to the microphone array. Materials with porous structures are renowned for their exceptional sound absorption capabilities, with polyurethane being a particularly prevalent choice in civil construction and transportation applications [14]. We propose the use of a windshield made of porous polyurethane foam, as illustrated in Fig. 6. This windshield is capable of enclosing the entire microphone array through the combination of two hemispherical parts. The material utilized for this windshield contains numerous air bubbles within the windshield due to its inherent ability to expand 100 times in volume when combined with water. Consequently, it is remarkably lightweight at 21 grams in total, making it an ideal choice for use in drones with tight payload restrictions, and it is also anticipated to exhibit high sound absorption properties.

In this paper, we verify the performance of the windshield through experiments to determine whether ground surface material estimation is possible using drone noise recored with the proposed windshield attached to the microphone array. Details regarding the methodology and results of the experiment can be found in Sec. V-B.

V. Evaluation

A. Ground surface material classification

The feasibility of the created dataset was evaluated by proposing a machine learning-based drone noise to ground surface material classifier.

1) Data to be used: It is necessary to eliminate from the recorded data any intervals that are unsuitable for the purposes of training and estimation. Specifically, the following steps were taken:

- The material was limited to three types: soil, asphalt, and shallow water.
- Only data collected at an altitude of 5 meters were used.
- For the shallow water recordings, the sections with severe clipping (the initial one minute and a half and the last 10 seconds) were removed and 6 minutes 17 seconds of the eight minutes were extracted.
- Data with intervals which data value is zero among 16 channels were removed (found in channels 1, 4, 9, 10, 12, 13, 14, and 15).

Hereafter, shallow water is simply referred to as “water”.

2) Networks and experimental conditions: In conducting deep learning-based classification, two networks were employed as classifiers: a four-layer convolutional neural network (CNN) depicted in Fig. 7 and ResNet18 [15] utilized in previous study [11]. To create the input data for the networks, each recording was divided into three different sections (training, validation, and test), and converted into the time-frequency domain using short-time Fourier transform (STFT) under the conditions outlined in TABLE II. Subsequently, the spectrograms were extracted with a data length of 256 frames and a shift length of 128 frames, and the real and imaginary components of each extracted spectrogram were utilized as input. In this experiment, eight channels were utilized, resulting in the acquisition of 1,094 input features with a length of 16×256×256 for training, and 106 for both the validation and testing. The labels of the training data were three-dimensional one-hot vectors corresponding to the ground surface materials.

TABLE III illustrates the length of each section in each recording, and TABLE IV illustrates the training conditions for the CNN and ResNet18. The epoch with the optimal performance was selected for evaluation within the range where overlearning does not occur based on the learning curve. The evaluation was conducted using the following accuracy measure.

$$\text{Accuracy}(\%) = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}} \times 100$$

Additionally, the confusion matrices were computed and analyzed for trends by material.

3) Results: TABLE V illustrates the appropriate response rate for each data section. The accuracy of the test data was 70.8% for the CNN using the model at the conclusion of 24 epochs and 82.1% for ResNet18 using the model at the conclusion of 38 epochs. The confusion matrix heatmaps in these cases are shown in Fig. 8.

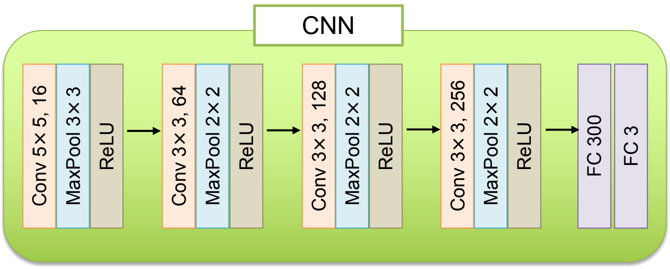


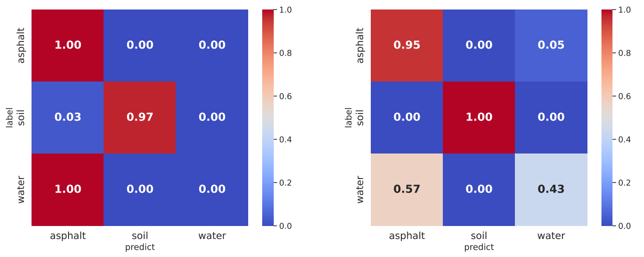
Fig. 7: CNN architecture used in this experiment.

TABLE II: STFT parameters

STFT frame length	512
STFT hop length	128
STFT window function	Hanning window

4) Discussions: As shown in TABLE V, the accuracy for the CNN on the test data is as high as 70%, while the accuracy on the validation data is only 38%. In light of the fact that the task in this paper is 3-class classification task and the accuracy is only 33%, it would be premature to accept these results at face value. An analysis of the cause is therefore necessary. The next step is to shuffle the data and perform cross-validation in order to ascertain the source of the bias. Conversely, ResNet18 yielded an accuracy exceeding 80% for all data sections, suggesting that the learning process for ground surface material estimation was effective. It is postulated that the superior learning capabilities of ResNet18 relative to CNN are attributable to its greater number of layers, which facilitate the suppression of noise other than reflected sound.

The heatmaps of the confusion matrices indicate that while both networks demonstrate high reproducibility for asphalt and soil, the reproducibility for water is relatively low. In particular, the CNN did not have any data point that could be estimated as water, indicating that it was not possible to distinguish between water and asphalt. This is likely due to the dynamic nature of the wind generated by the drone’s propeller, which causes the water surface to be in a state of flux, and the irregularity of the



(a) CNN

(b) ResNet18

Fig. 8: Confusion matrices for ground surface material classification.

TABLE III: Time length of each section (seconds)

Section	Training	Validation	Test
Asphalt	405	40	40
Soil	407	40	40
Water	313	32	32

TABLE IV: Training parameters

Loss function	Cross entropy
Optimizer	Adam
Learning rate	10^{-4}
Epochs	100

characteristics of the reflected sound. However, this remains a hypothesis at this point, and future analysis will be conducted in conjunction with the performance differences among the networks.

B. Windshield effects

1) Experimental conditions: A running car was used to record the sound in order to reproduce a high wind environment. Specifically, a jig with a microphone array attached was held out of the window from a moving car to record the target sounds (whistle sound and voice). The microphone array was kept at approximately 1 m from the car window while controlling it so that it did not rotate. To assess the efficacy of the proposed windshield, the following four types of microphone arrays were utilized.

1) No windshield was attached.
2) The windshield proposed in this paper was attached.
3) A 10 mm thick polyurethane (PU) foam material was affixed to cover the few channels located on the windward side.
4) A 10 mm thick polyethylene (PE) foam material was affixed to cover the few channels located on the windward side.

The recordings were made on a circuit of approximately 350 m per lap, with three to six laps run for each condition. To measure the wind speed, the anemometer was placed outside the window along with the jig. The wind speed during the run ranged from 12 to 20m/s, with no significant differences between each condition. The performance of the windshield is evaluated by comparing the spectrum and the rate of clipping occurrence per unit time for 13th channel, located on the windward side, of each recording.

2) Results and Discussions: Fig. 10 shows a spectrum of the recordings from each windshield configuration. It can be observed that the proposed windshield exhibits a sound insulation effect of approximately 10 to 20 dB across a wide range of frequencies, with the highest level of sound insulation overall compared to other foams. In particular, the sound insulation performance is markedly elevated in the high-frequency range above 3000 Hz. Given that the target sound in this experiment is situated between 100 and 3000 Hz and that wind noise is replete with high-frequency components, it is anticipated that the proposed windshield will be effective in preventing sound clipping due to strong winds.

TABLE V: Accuracy (in % of correct material estimation) for each method and dataset.

Network	Training	Validation	Test
CNN	68.1%	36.8%	70.8%
ResNet18	100%	70.8%	82.1%

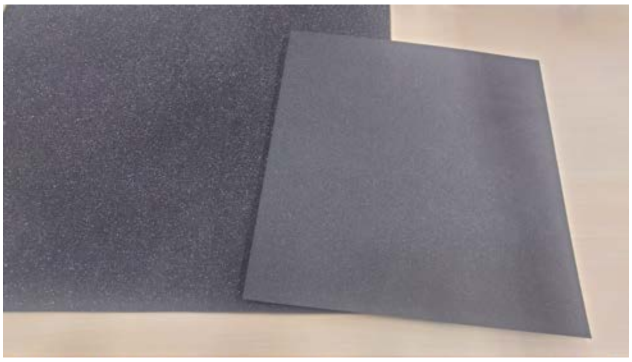


Fig. 9: PU foam (left in the figure) and PE foam (right in the figure)

TABLE VI depicts the number of observation points per second, wherein the absolute value of the recorded sound observation exceeded a specified threshold (th). While other foam materials did not prevent sound clipping, the proposed windshield does not have a single data point that suggests sound clipping, and the absolute value of the maximum observed value is approximately 0.67, indicating that the windshield can withstand even stronger winds. Therefore, together with the spectrum shown earlier, it can be shown that the proposed windshield is extremely effective against strong winds.

VI. Conclusion

In this paper, we presented a robust ground surface estimation method from drone noise reflections. To this end, we

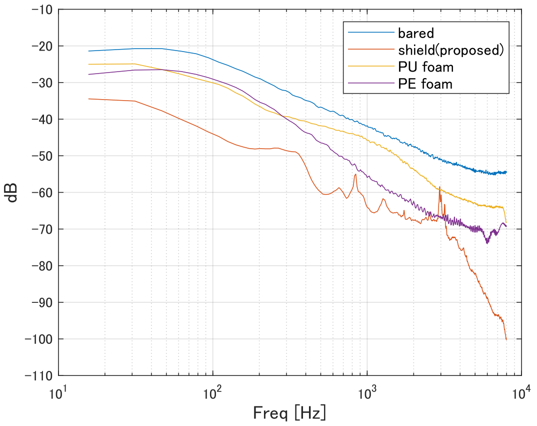


Fig. 10: Spectrum of recordings from each windshield configuration.

TABLE VI: the observation points where the absolute value $\geq th$ per second

threshold	bared	shield(proposed)	PU foam	PE foam
$th = 0.99$	16.59	0	0.358	0.682
$th = 0.9$	41.73	0	1.433	2.317

created a dataset of rotor noise and four types of ground surface materials (soil, asphalt, and two types of water at different depths) in an outdoor environment. Based on this dataset, we developed a convolutional neural network (CNN) and ResNet18-based ground surface material estimation method for the three types of materials except for deep water. The efficacy of the developed method was evaluated using the generated datasets, and a 70.8% correct response rate was observed on ResNet18. Moreover, we proposed a novel windshield for the microphone array as a means of mitigating external noise, and demonstrated its efficacy in preventing sound clipping caused by strong winds. The forthcoming stages of this project will involve the expansion of the dataset, data recording using the proposed windshield, an analysis of the ease of identification due to differences in networks, and real-time processing.

Acknowledgment

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

References

[1] Ministry of Land, Infrastructure, Transport and Tourism, *Overview of landslides caused by typhoon no. 19 of 2019 ver2.1 (in Japanese)*, https://www.mlit.go.jp/river/sabo//jirei/r1dosha/r1typhoon19_gaiyou191224r.pdf, Dec. 2019.

[2] I. prefecture, *Information on the 2024 noto peninsula earthquake, information on damage, etc. (12th report) (in Japanese)*, https://www.pref.ishikawa.lg.jp/saigai/documents/202401041500higaihou.pdf, Jan. 2024.

[3] W. M. Organization, *Wmo atlas of mortality and economic losses from weather, climate and water extremes (1970–2019)*, https://library.wmo.int/idurl/4/57564, 2021.

[4] K. Matsui, K. Hasegawa, J. Ohya, Y. Kato, and M. Yokozawa, "Study on estimation method of landslide location and scale by segmentation of rgbd images acquired by the camera attached to a drone," *Research Report: Computer Vision and Image Media (CVIM)*, no. 16, pp. 1–6, 2023.

[5] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3288–3293, 2012. DOI: 10.1109/IROS.2012.6385994.

[6] K. Hoshiba, K. Washizaki, M. Wakabayashi, *et al.*, "Design of uav-embedded microphone array system for sound source localization in outdoor environments," *Sensors*, vol. 17, no. 11, 2017, 2535.

[7] 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. DOI: 10.3390/drones7070405.

[8] M. Kumon, H. G. Okuno, and S. Tajima, "Alternating drive-and-glide flight navigation of a kiteplane for sound source position estimation," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2114–2120, 2021.

[9] B. Yen, T. Yamada, K. Itoyama, and K. Nakadai, "A performance assessment on rotor noise-informed active multidrone sound source tracking methods," *Drones*, vol. 8, no. 6, 2024.

[10] M. C. L. Wang and A. G. Rossberg, "Drone audition for bioacoustic monitoring," *Methods in Ecology and Evolution*, vol. 14, no. 12, pp. 3068–3082, 2023.

[11] T. Yano, K. Nishida, K. Itoyama, and K. Nakadai, "Ground surface material estimation by using drone rotor noise," *SICE SI*, 2023.

[12] H. Emoto, Y. Baba, H. Asano, and Y. Nagase, "Comparision of ai method on hammering sounds at concrete bridge," *Artificial Intelligence and Data Science*, vol. 1, no. J1, pp. 514–521, 2020. DOI: 10.11532/jsceiii.1.J1_514.

[13] F. Uehan, "Inspection techniques of concrete bridge members using uav," *Concrete Journal*, vol. 57, no. 9, pp. 699–704, 2019. DOI: 10.3151/coj.57.9_699.

[14] S. Dong, Y. Duan, X. Chen, *et al.*, "Recent advances in preparation and structure of polyurethane porous materials for sound absorbing application," *Macromolecular Rapid Communications*, p. 2 400 108, 2024.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.