Prototype Development of Cross-Shaped Microphone Array System for Drone Localization Based on Delay-and-Sum Beamforming in GNSS-Denied Areas

Hirokazu Madokoro 1,2,*, Satoshi Yamamoto 3, Kanji Watanabe 2, Masayuki Nishiguchi 2, Stephanie Nix 2, Hanwool Woo 4 and Kazuhito Sato 2

Abstract: Drones equipped with a global navigation satellite system (GNSS) receiver for absolute localization provide high-precision autonomous flight and hovering. However, the GNSS signal reception sensitivity is considerably lower in areas such as those between high-rise buildings, under bridges, and in tunnels. This paper presents a drone localization method based on acoustic information using a microphone array in GNSS-denied areas. Our originally developed microphone array system comprised 32 microphones installed in a cross-shaped configuration. Using drones of two different sizes and weights, we obtained an original acoustic outdoor benchmark dataset at 24 points. The experimentally obtained results revealed that the localization error values were lower for 0° and ±45° than for ±90°. Moreover, we demonstrated the relative accuracy for acceptable ranges of tolerance for the obtained localization error values.

Keywords: acoustic information; delay-and-sum beamforming; drone localization; global navigation satellite system-denied; microphone array; world coordinates

1. Introduction

Drones have virtually unlimited potential [1], not only for hobby use but also as innovative sensing platforms for numerous industrial applications. The global navigation satellite system (GNSS), which comprises the global positioning system (GPS), global navigation satellite system (GLONASS), BeiDou navigation satellite system (BDS), and Galileo, provides drones with autonomous flight and hovering with high accuracy based on absolute location information. Using high-precision positioning combined with real-time kinematic (RTK) technology [2], drone applications have been expanding rapidly in terms of precision agriculture [3] and smart farming [4], infrastructure inspection [5], traffic monitoring [6], object recognition [7], underground mining [8], construction management [9], parcel and passenger transportation [10], public safety and security [11], and rescue operations at disaster sites [12]. Particularly, improved efficiency, cost reduction, and labor savings have been achieved through drone applications in numerous operations and missions of various types. Nevertheless, the GNSS signal reception sensitivity is significantly lower in areas between high-rise buildings, under bridges, and in tunnels [13]. As localization techniques, location estimation and identification for drones in such GNSS-denied areas are challenging research tasks [14]. Moreover, for security and safety, various studies have been conducted to detect unknown drones [15–17].
In indoor environments that are frequently GNSS unavailable, simultaneous localization and mapping (SLAM) [18] are used widely for localization and tracking of drones [19,20]. In recent years, environmental mapping and location identification have been actualized from visual SLAM [21–24] using depth cameras and light detection and ranging (LiDAR) sensors [25]. Numerous visual SLAM-based methods have been proposed [26], not only for research purposes but also for practical applications such as floor-cleaning robots [27]. Another strategy that can be adopted without creating or using a map is an approach using motion capture sensors [28]. The use of motion capture is mainly confined to indoor areas because of their limited sensing range. The sensing range in outdoor environments is markedly greater than that in indoor environments. Moreover, in outdoor environments, sensors are affected by sunlight, wind, rain, snow, and fog. Therefore, localization and identification in GNSS-denied environments are tremendously challenging tasks. Taha et al. [15] categorized drone detection technologies into four categories of radar, acoustic, visual, and radio-frequency (RF)-based technologies. They presented both benefits and shortcomings of the respective technologies based on various evaluation criteria. In addition, they inferred the radar-based approach as the most realistic in terms of accuracy and detection range. Moreover, they surveyed that vision-based methods using static or dynamic images are consistent with deep learning [29], which has undergone rapid evolution and development in recent years. However, vision-based methods merely detected relative positions in image coordinates. Absolute positions as world coordinates remained unconsidered for detection results.

This study specifically examines acoustic-based techniques to acquire absolute positions as world coordinates. Drones with multiple rotors that revolve at high speed constantly make noise. This feature differs from fixed-wing gliders in the universal category of unmanned aerial vehicles (UAVs). Originally, the drone designation signified continuous low-frequency humming sounds, similar to that of honeybees. A salient benefit of using acoustic information is that it is available for a wider sensing range than when using vision information. Moreover, acoustic information performs robustly in environments with limited visibility and occlusion between objects.

The objective of this study is to develop a prototype microphone array system that identifies the drone position in world coordinates and to evaluate its quantitative accuracy. Our microphone array system comprises 32 omnidirectional microphones installed on an original cross-shaped mount. We conducted a flight evaluation experiment in an outdoor environment using two drones of different body sizes and payloads. Experimentally obtained results achieved using originally collected benchmark datasets revealed that our proposed system identified drone locations in the world coordinates within the accuracy range of GPS. Moreover, we demonstrated properties and profiles of position identification error values according to distances and angles.

This paper is structured as follows. Section 2 briefly reviews related studies of state-of-the-art acoustic and multimodal methods of drone detection and localization. Subsequently, Section 3 presents our proposed localization method using an originally developed microphone system based on delay-and-sum (DAS) beamforming. Experiment results obtained using our original acoustic benchmark datasets obtained using two drones are presented in Section 4. Finally, Section 5 concludes this explanation and highlights future work.

2. Related Work

As comprehensive research, Taha et al. [15] surveyed state-of-the-art acoustic methods for drone detection [30–36]. They asserted that far fewer research efforts have been based on acoustic methods compared with other modalities. Moreover, they indicated a shortage of acoustic benchmark datasets because of difficulties related to annotation compared with other modalities, especially in visual benchmark datasets. In our brief literature survey here, four additional references below are reviewed as reports of related work.

Sedunov et al. [37] proposed an acoustic system for detecting, tracking, and classifying UAVs according to their propeller noise patterns. Their system included three sensor
array nodes: each node comprised 15 microphones with 100 ± 20 m intervals. To detect sound source directions, they used the steered-response phase transform (SRP-PHAT) [38] based on the computation and addition of generalized cross-correlation phase transform (GCC-PHAT) [39]. The flight evaluation experiment results conducted outdoors using five drones of different sizes and payloads revealed that their proposed method obtained notably good performance compared to the existing commercial acoustic-based systems. Specifically, their system detected a drone at up to 350 m with average precision of 4°. Moreover, they demonstrated real-time drone tracking within a predicted coverage distance of approximately 250 m.

Chang et al. [40] proposed a systematic method using two acoustic arrays for drone localization and tracking. Both arrays were composed of four microphone sensors separated by 14 m gap distance. They developed a time difference of arrival (TDOA) estimation algorithm using the Gaussian prior probability density function to overcome multipath effects and a low signal-to-noise ratio (SNR). The analysis of data obtained during field experiments leads to satisfactory performance for real-time drone localization and tracking. However, they merely demonstrated a single experimental result using a drone in the limited range of 50 × 50 m.

Dumitrescu et al. [41] proposed an acoustic system for drone detection using 30 spiral-shaped digital microphones in a micro-electro-mechanical system (MEMS) fabrication. For drone identification and classification, they proposed a concurrent neural network (CoNN) that included supervised and unsupervised learning paradigms simultaneously. They achieved simultaneous detection using their originally developed benchmark datasets using six drones: three commercial drones and three originally developed drones.

Blanchard et al. [42] proposed an arrayed microphone system for 3D localization and tracking of a drone. Their system comprises ten microphones arranged in three orthogonal axes. They employed time-frequency DAS beamforming and a Kalman filter for 3D localization and tracking. They conducted an outdoor experiment to evaluate localization and tracking errors. However, the flight range was limited to a relatively small area of approximately 10 × 10 m horizontally and 5 m vertically.

As a multimodal device with new possibilities, audio images obtained using an acoustic camera have come to be used for drone detection. Zunino et al. [43] developed an optical video camera prototype with 128 digital MEMS microphones. Using a turntable that mechanically enhances the range of camera view, the maximum field of view is 90° in elevation and 360° in azimuth were achieved. Experimental results obtained using their original datasets demonstrated real-time detection and tracking of drones and people indoors, and motorbikes outdoors. The relation between the sound source positions and the rotary table angles strongly affected the detection accuracy.

Liu et al. [44] proposed a modular camera array system combined with multiple microphones. They specifically examined a camera array system that was able to use large-scale airspace observation outdoors. Their proposed system integrated vision and audio information from multiple sensors in surroundings of various directions. They identified the characteristics of various drones to achieve higher monitoring efficiency. However, their proposed system did not support localization in world coordinates.

Svanström et al. [45] developed an automatic multisensor drone detection system for achieving multiple modality sensing. Their proposed system comprised a normal video camera, a fisheye lens camera, a thermal infrared camera, and a microphone. They obtained original datasets containing 650 annotated infrared and visible video files of drones, birds, airplanes, and helicopters combined with audio files of drones, helicopters, and background noise. Their comparison results obtained using a you only look once (YOLO) v2 detector [46] revealed improved accuracy with multiple modalities over a single modality.

Izquierdo et al. [47] proposed a combined method for detecting drones and persons simultaneously. Their acoustic camera system comprised a monocular camera with 64 MEMS microphones. They used a 2D beamforming algorithm for target detection. The
experimental results revealed that their proposed system identified rotor noise not only by the arrival angles of direct sound signals but also by the first echo reflected by the reflective surface. However, mere evaluation experiments were conducted under limited conditions in a small room as a feasibility study. The novelty and contributions of this study are the following.

- Our proposed method can detect the horizontal position of a drone using three parameters obtained from the drone and our cross-shaped microphone array system. The challenging detection range is greater than the previous approach [42].
- Compared with previous studies [44–47] based on multimodal devices combining a camera and arrayed microphones, our method can detect a drone using only acoustic information similar to previous studies [37,40–42].
- Compared with previous studies [37,40–42] based on a single modality of arrayed microphones, we provide a detailed evaluation that comprises 24 positions and two different drones in an actual outdoor environment.
- To the best of our knowledge, this is the first study to demonstrate and evaluate drone localization based on DAS beamforming used in GNSS-denied areas.

By contrast, as a limitation of our method, the target drone must accurately transmit its flight altitude information to our proposed system. Another limitation is that we do not currently consider applications in environments where sound reflection occurs, such as in tunnels or under bridges.

3. Proposed Method

3.1. Acoustic Localization Method

Figure 1 depicts the principle used to calculate position coordinates. We let O represent the origin for the center position of the cross-shaped microphone array, and (x, y) be the coordinates of the drone flight position. In addition, we let $\theta_h$ and $\theta_e$ respectively represent the azimuthal and elevation angles between the microphone array and the drone. Furthermore, $h$ denotes the drone flight altitude; $h_m$ represents the height from the ground to O. Vertical length $h_d$ from O to the drone is calculated as shown below.

$$h_d = h - h_m. \quad (1)$$

![Figure 1. Drone position coordinate calculation.](image-url)
Using $h_d$ and $\theta_v$, length $r$ from $O$ to $(x, y)$ is obtained as

$$r = \frac{h_d}{\tan \theta_v}. \quad (2)$$

Using $r$ and $\theta_h$, one can calculate length $x$ as

$$x = r \cos \theta_h. \quad (3)$$

Subsequently, using $x$ and $\theta_h$, the length $y$ is calculated as

$$y = x \tan \theta_h. \quad (4)$$

For this study, we set $h_m$ to 900 mm when developing our original sensor mount. Parameter $h$ is obtained from the flight controller via a drone controller. Although GNSS provides high vertical ranging resolutions, it drops significantly in horizontal ranging resolutions. Therefore, drones use not only IMU values but also barometric pressure and temperature values to calculate the altitude. Parameters $\theta_h$ and $\theta_v$ are calculated using the beamforming method.

### 3.2. Delay-and-Sum Beamforming

As depicted in Figure 2, beamforming is a versatile technology used for directional signal enhancement with sensor arrays [48]. Letting $y(t)$ be a beamformer output signal at time $t$ and letting $M$ be the number of microphones, then for $z_m(t)$ and $w_m(t)$, which respectively denote a measured signal and a filter of the $m$-th sensor, $y(t)$ is calculated as

$$y(t) = \sum_{m=1}^{M} w_m(t) \otimes z_m(t). \quad (5)$$

![Figure 2. Delay-and-sum (DAS) beamforming in a time domain.](image)

For this study, we regard DAS beamforming as a temporal domain. Assuming that single plane waves exist and letting $s_m(t)$ be a set of sound signals, then delay $\tau_m$ as expressed for the formula below occurs for incident waves observed for the $m$-th sensor as

$$z_m(t) = s_m(t - \tau_m), \quad (6)$$

where $M$ represents the total number of microphones.

By using the filter, the delayed $-\tau_m$ of incident waves and the advancing $+\tau_m$ of incident waves were offset. Signals from the direction of $\theta$ are enhanced because of gathered phases of signal $s(t)$ in all channels. For this study, the temporal compensated filter $w_m(t)$ is defined as

$$w_m(t) = \frac{1}{M} \delta(t + \tau_m), \quad (7)$$

where $\delta$ is Dirac’s delta function.
Letting $\theta$ be an angle as a variable parameter, then for the comparison of sound directions, the relative mean power level $G(\theta)$ of $y(t)$ is defined as

$$G(\theta) = \frac{1}{T} \sum_{t=0}^{T} y(t)^2,$$

where $T$ represents the length of the interval time.

We changed $\theta$ from $-90^\circ$ to $90^\circ$ with $1^\circ$ intervals. We assumed $P_h(\theta)$ and $P_e(\theta)$ as respectively denoting $G(\theta)$ obtained from the horizontal and vertical microphone arrays. Using $P_h(\theta)$ and $P_e(\theta)$, one can obtain $\theta_h$ and $\theta_e$ as

$$\{\theta_h, \theta_e\} = \arg \max_{-90^\circ \leq \theta \leq 90^\circ} P\{h,e\}(\theta).$$

### 3.3. Devices for Experimentation

Figure 3 depicts the design output of a cross-shaped sensor mount for installing microphones as an array system. The mount dimensions are 1200 mm length, width, and height. The horizontal arm is located 900 mm above the ground. The cross-attached mount legs were designed to be wide to prevent falling. Two aluminum frames were placed in parallel to fasten microphones at their front and back. We installed 16 microphones at constant intervals at a 600 mm width. Similarly, for the vertical side, we mounted 16 microphones in a height range of the positions from 600 mm to 1200 mm. In all, we used 32 microphones as our first prototype.

For assembly, aluminum straight pipes with a square-cross sectional of $20 \times 20$ mm were used for the mount framework structure. To maintain greater strength, L-shaped brackets were used for all joints. After assembling the mount, the 32 microphones were installed using plastic for tightening up. Figure 4 depicts photographs of the completed microphone array system from the front, right, and rear side views.

Figure 5 depicts an 8-channel audio interface unit (Behringer ADA8200; Music Tribe; Makati, Metro Manila, Philippines) that includes an amplifier and an analog-to-digital (A/D) converter. According to the specifications of ADA8200, the sampling rate is 44.1/48 kHz with synchronization across the channels. Our system comprises four interface units because it uses 32 microphones. The mean power consumption of each interface unit is 15 Wh. Digital signals from them are integrated with a hub module. We used a laptop computer to capture and store digital signals. The mean power consumption of the laptop computer is 9 Wh. Therefore, the total system power consumption is approximately 69 Wh. We used a portable battery (PowerHouse 200; Anker Innovations Technology Limited, Shenzhen, China) with 213 Wh battery capacity. Using this battery, the operating time of our proposed system is approximately 3 h.
Drones 2021, 5, 123

Figure 4. Originally developed microphone array system from front, right, and rear side views. Several microphones are installed with slightly unequal intervals. Omnidirectional microphones allow this installation.

Figure 5. Appearance of the combined four audio interface units and our measurement system.

Figure 6 portrays the omnidirectional microphones (Behringer ECM8000; Music Tribe; Makati, Metro Manila, Philippines) used for this study. Their input frequency range is 15–20,000 Hz. The microphone tip diameter is 12 mm. We installed microphones on the mount at 40 mm intervals horizontally and vertically.

Figure 6. Installation of omnidirectional microphones to our originally developed sensor mount prototype.

Figure 7 presents two drones (Matrice 200 and Matrice 600 Pro; SZ DJI Technology Co., Ltd.; Shenzhen, China) used as detection targets for our experiments. Both drones are widely applied in various industrial fields for aerial measurement and observation. The sound pressure level range from the respective drones was 50–80 dB depending on the payload and flight conditions. Table 1 presents the respective specifications.
Table 1. Major specifications of two industrial drones.

| Items                        | Matrice 200       | Matrice 600 Pro   |
|------------------------------|-------------------|-------------------|
| Diagonal wheelbase           | 643 mm            | 1133 mm           |
| Dimensions (L × W × H)       | 887 × 880 × 378 mm| 1668 × 1518 × 727 mm|
| Rotor quantity               | 4                 | 6                 |
| Weight (including standard batteries) | 6.2 kg          | 9.5 kg            |
| Payload                      | 2.3 kg            | 6.0 kg            |
| Maximum ascent speed         | 5 m/s             |                   |
| Maximum descent speed        |                   | 3 m/s             |
| Maximum wind resistance      | 12 m/s            | 8 m/s             |
| Maximum flight altitude      | 3000 m            | 2500 m            |
| Operating temperature        | −20 °C to 45 °C   | −10 °C to 40 °C   |
| GNSS                         |                   | GPS+GLONASS       |

4. Position Estimation Experiment

4.1. Benchmark Datasets

We conducted acoustic data collection experiments at a track of the Honjo campus (39°39'35"N, 140°7'33"E), Akita Prefectural University, Yurihonjo city, Japan. The left panel of Figure 8 exhibits results of some aerial photography conducted at the experimental site. This campus is surrounded by paddy fields. An expressway and an ordinary road are located, respectively, east and south of the campus. We conducted this experiment during the daytime, as depicted on the right panel of Figure 8. Although the traffic was scarce, noise from passing cars was included in the obtained datasets. There was no noise output from the faculty building located to the north.

Table 2 presents details of ground truth (GT) positions and angles at 24 points of drone flights for obtaining sound datasets. Here, $x_g$ and $y_g$ correspond to GT coordinates in the evaluation process of estimation results. We used a tape measure (USR-100; TJM Design Crop.; Tokyo, Japan) to obtain GT positions. The minimum range of the tape measure is 2 mm. Additionally, $\theta_{gh}$ and $\theta_{gv}$ respectively signify the azimuthal angle and elevation.
angles calculated from \( x_g, y_g \), and \( h \) using our proposed method. The respective positions were assigned independent labels as P1–P24.

Table 2. Positions and angles at 24 points of drone flights for obtaining sound datasets.

| Position | Group | \( x_g \) [m] | \( y_g \) [m] | \( h \) [m] | \( \theta_{gh} \) [°] | \( \theta_{gv} \) [°] |
|----------|-------|--------------|--------------|-------------|--------------------|--------------------|
| P1       | 4     | −50          | 0            | 50          | −90                | 45                 |
| P2       | 3     | −40          | 0            | 40          | −90                | 45                 |
| P3       | 2     | −30          | 0            | 30          | −90                | 45                 |
| P4       | 1     | −20          | 0            | 20          | −90                | 45                 |
| P5       | 1     | 20           | 0            | 20          | 90                 | 45                 |
| P6       | 2     | 30           | 0            | 30          | 90                 | 45                 |
| P7       | 3     | 40           | 0            | 40          | 90                 | 45                 |
| P8       | 4     | 50           | 0            | 50          | 90                 | 45                 |
| P9       | 1     | −20          | 20           | 28          | −45                | 45                 |
| P10      | 1     | 0            | 20           | 20          | 0                  | 45                 |
| P11      | 1     | 20           | 20           | 28          | 45                 | 45                 |
| P12      | 2     | −30          | 30           | 42          | −45                | 45                 |
| P13      | 2     | 0            | 30           | 30          | 45                 | 45                 |
| P14      | 2     | 30           | 30           | 42          | 45                 | 45                 |
| P15      | 3     | −40          | 40           | 57          | −45                | 45                 |
| P16      | 3     | 0            | 40           | 40          | 45                 | 45                 |
| P17      | 3     | 40           | 40           | 57          | 45                 | 45                 |
| P18      | 4     | −50          | 50           | 71          | −45                | 45                 |
| P19      | 4     | 0            | 50           | 50          | 45                 | 45                 |
| P20      | 4     | 50           | 50           | 71          | 45                 | 45                 |
| P21.1    | –     | 20           | 20           | 5           | 45                 | 8                  |
| P21.2    | –     | 20           | 20           | 50          | 45                 | 60                 |
| P21.3    | –     | 20           | 20           | 100         | 45                 | 74                 |
| P21.4    | –     | 20           | 20           | 150         | 45                 | 79                 |
| P22.1    | –     | 50           | 50           | 5           | 45                 | 3                  |
| P22.2    | –     | 50           | 50           | 50          | 45                 | 35                 |
| P22.3    | –     | 50           | 50           | 100         | 45                 | 54                 |
| P22.4    | –     | 50           | 50           | 150         | 45                 | 65                 |
| P23      | –     | 70           | 70           | 100         | 45                 | 45                 |
| P24      | –     | 355          | 355          | 150         | 45                 | 17                 |

Here, the following three steps were used to determine the drone’s flight position.

1. A 2D position on the ground was measured using a tape measure and marked.
2. After placing a drone at the mark, it was flown to an arbitrary height in the vertical direction.
3. The 3D flight position was confirmed by visual observation from the ground and field-of-view (FOV) images transmitted from an onboard camera of the drone.

We used the Matrice 200 at P1–P20 and the Matrice 600 Pro at P21–P24. Datasets at P1–P20, P21–P22, P23–P24 were obtained on different days. Table 3 denotes the meteorological conditions on the respective experimental days. On all three days, the meteorological conditions were found to be suitable for drone flights.

The coordinates of flight positions P1–P20 are depicted in Figure 9. The microphone array was placed at the coordinate (0, 0). For the \( x \)-axis direction, the flight positions were assigned from \( ±20 \text{ m} \) to \( ±50 \text{ m} \) at \( ±10 \text{ m} \) intervals. For the \( y \)-axis direction, the flight positions were assigned from \( 20 \text{ m} \) to \( 50 \text{ m} \) at \( 10 \text{ m} \) intervals. Based on the \( x \) axis positive direction, angles were assigned from \( 0° \) to \( 180° \) at \( 45° \) intervals. The flight altitudes were assigned from \( 20 \text{ m} \) to \( 70 \text{ m} \) at \( 10 \text{ m} \) intervals.
Table 3. Meteorological conditions on respective experiment days.

| Parameter        | P1–P20       | P21–P22       | P23–P24       |
|------------------|--------------|--------------|--------------|
| Date             | 17 July 2020 | 27 August 2020 | 16 October 2020 |
| Time (JST)       | 14:00–15:00 | 14:00–15:00 | 14:00–15:00  |
| Weather          | Sunny        | Sunny        | Cloudy       |
| Air pressure [hPa] | 1006.8       | 1007.7       | 1019.4       |
| Temperature [°C] | 28.0         | 33.1         | 14.8         |
| Humidity [%]     | 60           | 50           | 48           |
| Wind speed [m/s] | 1.8          | 5.3          | 1.1          |
| Wind direction   | W            | W            | ENE          |

Figure 9. Flight positions (P1–P20).

Similarly to P11, the P21 flight coordinate was set to 20 m on the x and y axes. Moreover, similarly to P20, the P22 flight coordinate was set to 50 m on the x and y axes. For both positions, the flight altitudes were set to four steps: 5 m, 50 m, 100 m, and 150 m. The Japanese Civil Aeronautics Law regulates the maximum flight altitude from the ground up to 150 m.

For P23 and P24, the microphone array system locations were changed, but the flight position was fixed. Figure 10 depicts the positional relations between the drone flight point and measurement position at P23 and P24, which correspond to the installation positions of the microphone array system. The distances from P23 and P24 to the drone in flight were, respectively, 70 m and 355 m on the x and y axes. The flight altitudes at P23 and P24 were, respectively, 100 m and 150 m. At all locations except for P23 and P24, we were able to hear the drone propeller rotation sound.

For the respective positions, we recorded 10 s sound data while the drone was hovering. We extracted 1 s sound data randomly at time t for input to DAS beamforming as s(t).

4.2. Experiment Results

The experimentally obtained results are presented in groups according to the distances between the microphone array system and flight positions. Herein, Group 1 includes P4, P5, P9, P10, and P11, which are located ±20 m in the x axis and 20 m in the y axis from the microphone array system. Groups 2–4 comprise the positions that are equally distant at 10 m intervals.
Figure 10. Relations among flight points and measurement positions at P23 and P24.

Figure 11 depicts DAS beamforming output distributions and estimated angles for Group 1. The left and right panels respectively correspond to azimuthal estimation angle $\theta_{hm}$ and elevation estimation angle $\theta_{vm}$. The curves show the mean power changes on the vertical axis. The angle on the horizontal axis is shifted from $-90^\circ$ to $90^\circ$ step by $1^\circ$. The angles with maximum values, which are marked with purple circles, are extracted as $\theta_{hm}$ and $\theta_{vm}$. For $\theta_{hm}$, from left to right on the graph, the maximum values of P10, P9, P11, P5, and P4 are located respectively at approximately $-90^\circ$, $-45^\circ$, $0^\circ$, $45^\circ$, and $90^\circ$. For $\theta_{vm}$, the maximum values are located at approximately $45^\circ$ at the respective locations because the flight altitude was 20 m or 30 m.

Figure 11. DAS beamforming output distributions and estimated angles for Group 1: P10, P9, P11, P5, and P4.

Angle estimation results for Groups 2, 3, and 4 are presented respectively in Figures 12–14. The estimated angles for $\theta_{hm}$ are located at approximately $45^\circ$ intervals. The estimated angles for $\theta_{vm}$ are located at approximately $45^\circ$ in all cases. The experimentally obtained angle estimation results revealed that each output curve exhibited a unimodal distribution with a distinct peak.

Figures 15 and 16 respectively depict angle estimation results obtained at P21 and P22. Both results include four flight altitudes of 5 m, 50 m, 100 m, and 150 m. Regarding horizontal results, $\theta_{hm}$ was extracted at around $45^\circ$ in all altitude patterns. As a comprehensive trend, the flight altitudes and the mean power level exhibit an inversely proportional trend except at 5 m altitude. We consider that the mean power level at the 5 m altitude was low because of the effects sound waves reflected from the ground. For $\theta_{hm}$, the flight altitudes and estimated angles were found to have positive correlation. The respective curves exhibit a unimodal distribution with a distinct peak.
Figure 12. DAS beamforming output distributions and estimated angles for Group 2: P6, P12, P3, P13, and P14.

Figure 13. DAS beamforming output distributions and estimated angles for Group 3: P7, P15, P2, P16, and P17.

Figure 14. DAS beamforming output distributions and estimated angles for Group 4: P8, P18, P1, P19, and P20.

Figure 15. DAS beamforming output distributions and estimated angles at P21 for four flight altitudes.
Figure 16. DAS beamforming output distributions and estimated angles at P22 for four flight altitudes.

Figures 17 and 18 respectively depict angle estimation results obtained at P23 and P24. For these experiments, P23 and P24 respectively represent 70 m and 355 m distance in both $x$ and $y$ axis directions. These are equivalent to approximately 100 m at P23 and 500 m at P24 in straight-line distances. Horizontally, $\theta_{hm}$ was obtained at $45^\circ$ for both positions. Vertically, $\theta_{vm}$ was obtained at $45^\circ$ for P23 and at $17^\circ$ for P24. Although sound volumes that were heard subjectively were low, the respective curves were presented as a unimodal distribution with a distinct peak.

All of the results described above are presented in Table 4. Here, positional error values $x_e$, $y_e$, and $E$ were calculated as presented below.

$$ E = \sqrt{x_e^2 + y_e^2}, $$

$$ \begin{bmatrix} x_e \\ y_e \end{bmatrix} = \begin{bmatrix} x_m - x_g \\ y_m - y_g \end{bmatrix}. $$
Table 4. Localization results obtained for angles and position coordinates.

| Position | θhm [°] | θvm [°] | x[m] | y[m] | xe [m] | ye [m] | E |
|----------|---------|---------|------|------|--------|--------|---|
| P1       | -86     | 44      | -50.7| 3.5  | 3.5    | 0.7    | 3.6 |
| P2       | -84     | 44      | -40.3| 4.2  | 4.2    | 0.3    | 4.2 |
| P3       | -84     | 44      | -30.0| 3.1  | 3.1    | 0.0    | 3.1 |
| P4       | -87     | 45      | -19.1| 1.0  | 1.0    | -0.9   | 1.3 |
| P5       | 85      | 45      | 19.0 | 1.7  | 1.7    | 1.0    | 2.0 |
| P6       | 85      | 45      | 29.0 | 2.5  | 2.5    | 1.0    | 2.7 |
| P7       | 85      | 45      | 39.0 | 3.4  | 3.4    | 1.0    | 3.5 |
| P8       | 85      | 45      | 48.9 | 4.3  | 4.3    | 1.1    | 4.4 |
| P9       | -45     | 45      | -19.2| 19.2 | -0.8   | -0.8   | 1.1 |
| P10      | 0       | 45      | 0.0  | 19.1 | -0.9   | 0.0    | 0.9 |
| P11      | 45      | 45      | 19.2 | 19.2 | -0.8   | 0.8    | 1.1 |
| P12      | -45     | 45      | -29.1| 29.1 | -0.9   | -0.9   | 1.3 |
| P13      | 0       | 45      | 0.0  | 29.1 | -0.9   | 0.0    | 0.9 |
| P14      | 45      | 45      | 29.1 | 29.1 | 0.9    | 0.9    | 1.3 |
| P15      | -45     | 45      | -39.7| 39.7 | -0.3   | -0.3   | 0.4 |
| P16      | 0       | 45      | 0.0  | 39.1 | -0.9   | 0.0    | 0.9 |
| P17      | 45      | 45      | 39.7 | 39.7 | -0.3   | 0.3    | 0.4 |
| P18      | -45     | 45      | -49.6| 49.6 | -0.4   | -0.4   | 0.6 |
| P19      | 0       | 45      | 0.0  | 49.1 | -0.9   | 0.0    | 0.9 |
| P20      | 44      | 45      | 48.7 | 50.4 | 0.4    | 1.3    | 1.4 |
| P21.1    | 45      | 8       | 20.6 | 20.6 | 0.6    | -0.6   | 0.8 |
| P21.2    | 45      | 60      | 20.0 | 20.0 | 0.0    | 0.0    | 0.0 |
| P21.3    | 44      | 73      | 21.0 | 21.8 | 1.8    | -1.0   | 2.1 |
| P21.4    | 45      | 79      | 20.5 | 20.5 | 0.5    | -0.5   | 0.7 |
| P22.1    | 45      | 3       | 55.3 | 55.3 | 5.3    | -5.3   | 7.5 |
| P22.2    | 45      | 35      | 49.6 | 49.6 | -0.4   | 0.4    | 0.6 |
| P22.3    | 45      | 54      | 50.9 | 50.9 | 0.9    | -0.9   | 1.3 |
| P22.4    | 45      | 65      | 49.2 | 49.2 | -0.8   | 0.8    | 1.1 |
| P23      | 45      | 45      | 70.1 | 70.1 | -0.6   | 0.6    | 0.8 |
| P24      | 45      | 17      | 344.8| 344.8| -9.2   | 9.2    | 13.0|

Figure 19 presents localization results obtained at P1–P20. Filled circles and rings respectively correspond to GT positions and estimation positions. As an overall trend, no estimation position differs to a considerable degree from the GT positions. With respect to the angular characteristics, the error values at ±90° tend to be larger than those at 0° or ±45°. The cumulative error increased because of the higher delay ratio, defined theoretically as $\tau_m$.

Figure 20 portrays scatter plots of correlation between GT and estimation positions at P1–P20 for each coordinate. These results indicate that the variance on the y axis of the right panel is greater than that at the x axis of the left panel.

Figure 21 depicts localization results obtained at P21 and P22. In comparison to the GT position, the P21 estimation results are located farther away from the microphone array system. Similarly, in comparison to the GT position, the P22 estimation results are located closer to and farther away from the microphone array system. Large error values were found at altitudes of 100 m for P21 and 5 m for P22. The localization results at altitudes of 50 m and 150 m were stable.
Figure 19. Localization results obtained at P1–P20.

Figure 20. Scatter plots of correlation between GT and estimation positions for $x$ axis (left) and $y$ axis (right).

Figure 21. Localization results obtained at P21 (left) and P22 (right).

Figure 22 depicts distributions of estimated locations at P23 and P24. The localization error of P23 was 0.8, which was smaller than those of P1–P22. Although the localization
error of P24 was 13.0, which was higher than those of P1–P23, it corresponds to 3.4%. We regard these results as reference values because they were conducted only in a single trial.

Figure 22. Localization results obtained at P23 (left) and P24 (right).

4.3. Discussion

Table 5 presents $E$ of each group with respect to the distance between the microphone array system and the drone. The total $E$ represents the accumulated values of five positions in each group. The mean $E$ shows a slight trend for the error to increase slightly according to the growing distance.

Table 5. Total and mean $E$ of each group.

|          | Group 1 | Group 2 | Group 3 | Group 4 |
|----------|---------|---------|---------|---------|
| Total [m]| 6.48    | 9.24    | 9.50    | 10.83   |
| Mean [m] | 1.30    | 1.85    | 1.90    | 2.17    |

Generally, GPS localization error values are approximated at 2 m, which is dependent on receivers, antennas, and obstacles in surrounding environments. Modsching et al. [49] reported that their experimentally obtained GPS error was 2.52 m in a medium-size city with few obstructions. We evaluated the simulated localization accuracy for 28 results at P1–P22 for changing tolerance ranges from 3.0 m to 0.5 m step by 0.5 m, as presented in Table 6. The localization accuracy of our proposed method is 71.4% with 2.0 m tolerance. The localization accuracy dropped to 67.9% with 1.5 m tolerance. However, it dropped to 39.3% with 1.0 m tolerance.

Table 6. Simulated localization accuracy for 28 results at P1–P22 for changing tolerance ranges.

| Tolerance | 3.0 m | 2.5 m | 2.0 m | 1.5 m | 1.0 m |
|-----------|-------|-------|-------|-------|-------|
| Accuracy [%]| 78.6  | 75.0  | 71.4  | 67.9  | 39.3  |

5. Conclusions

This paper presented a drone localization method based on acoustic information using a microphone array in GNSS-denied areas. Our originally developed microphone array system comprised 32 microphones installed in a cross-shaped configuration. Using two drones of different sizes and weights, we obtained an original acoustic outdoor benchmark dataset at 24 points. The experimentally obtained results revealed that the localization error values were lower at 0° and ±45° than at ±90°. Moreover, we evaluated the imaginary localization accuracy for 28 results at 22 points for changing tolerance ranges from 3.0 m to 0.5 m step by 0.5 m. The localization accuracy of our proposed method is 71.4% with 2.0 m tolerance.
Future work includes improving the angular resolution of elevation angles, comparative evaluation of sound source localization with other methods, further investigation of the magnitude and effects of noise, simultaneous localization of multiple drones, real-time drone tracking, and expansion of benchmark datasets. Moreover, we must evaluate altitude measurement errors using a total station theodolite, which is an optical instrument device used in surveying and building construction.

**Author Contributions:** Conceptualization, H.M.; methodology, K.W. and M.N.; software, S.N.; validation, K.W. and M.N.; formal analysis, S.Y.; investigation, S.Y.; resources, K.W. and M.N.; data curation, S.N.; writing—original draft preparation, H.M.; writing—review and editing, H.M.; visualization, H.W.; supervision, H.M.; project administration, K.S.; funding acquisition, H.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 17K00384 and 21H02321.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** Datasets described as a result of this study are available on request to the corresponding author.

**Acknowledgments:** We would like to express our appreciation to Tsubasa Ebe and Masumi Hashimoto, who are graduates of Akita Prefectural University, for their great cooperation in the experiments.

**Conflicts of Interest:** The authors declare that they have no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or in interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

**Abbreviations**

The following abbreviations are used in this manuscript:

- A/D: analog-to-digital
- BDS: BeiDou navigation satellite system
- CoNN: concurrent neural network
- DAS: delay-and-sum
- FOV: field of view
- GCC-PHAT: cross-correlation phase transform
- GLONASS: global navigation satellite system
- GNSS: global navigation satellite systems
- GPS: global positioning system
- GT: ground truth
- LiDAR: light detection and ranging
- RF: radio-frequency
- RTK: real-time kinematic
- SRP-PHAT: steered-response phase transform
- SNR: signal-to-noise ratio
- SLAM: simultaneous localization and mapping
- TDOA: time difference of arrival
- UAV: unmanned aerial vehicle
- YOLO: you only look once

**References**

1. Floreano, D.; Wood, R. Technology and the Future of Small Autonomous Drones. *Nature* **2015**, *521*, 460–466.
2. Henkel, P.; Sperl, A. Real-Time Kinematic Positioning for Unmanned Air Vehicles. In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA, 5–12 March 2016; pp. 1–7.
3. Rao, U.M.; Deepak, V. Review on Application of Drone Systems in Precision Agriculture. *Procedia Comput. Sci.* **2018**, *133*, 502–509.
4. Bacco, M.; Berton, A.; Ferro, E.; Gennaro, C.; Gotta, A.; Matteoli, S.; Paonessa, F.; Ruggeri, M.; Virone, G.; Zanella A. Smart Farming: Opportunities, Challenges and Technology Enablers. In Proceedings of the IoT Vertical and Topical Summit on Agriculture, Tuscany, Italy, 8–9 May 2018; pp. 1–6.
5. Shakhatreh, H.; Sawalmeh, A.H.; Fuqaha, A.A.; Dou, Z.; Almaita, E.; Khalil, I.; Othman, N.S.; Khreishah, A.; Guizani, M. Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. *IEEE Access* 2019, 7, 48572–48634.

6. Na, W.S.; Baek, J. Impedance-Based Non-Destructive Testing Method Combined with Unmanned Aerial Vehicle for Structural Health Monitoring of Civil Infrastructures. *Appl. Sci.* 2017, 7, 15.

7. Madokoro, H.; Sato, K.; Shimoi, N. Vision-Based Indoor Scene Recognition from Time-Series Aerial Images Obtained Using a MAV Mounted Monocular Camera. *Drones* 2019, 3, 22.

8. Shahmoradi, J.; Talebi, E.; Roghanchi, P.; Hassanalian, M. A Comprehensive Review of Applications of Drone Technology in the Mining Industry. *Drones* 2020, 4, 34.

9. Li, Y.; Liu, C. Applications of Multirotor Drone Technologies in Construction Management. *Int. J. Constr. Manag.* 2019, 19, 401–412.

10. Kellermann, R.; Biehle, T.; Fischer, L. Drones for parcel and passenger transportation: A literature review of Transportation Research. *Interdiscip. Perspect.* 2020, 4, 100088.

11. He, D.; Chan, S.; Guizani, M. Drone-Assisted Public Safety Networks: The Security Aspect. *IEEE Commun. Mag.* 2017, 55, 218–223.

12. Alotaibi, E.T.; Alqefari, S.S.; Koubaa, A. LSAR: Multi-UAV Collaboration for Search and Rescue Missions. *IEEE Access* 2019, 7, 55817–55832.

13. Shen, N.; Chen, L.; Liu, J.; Wang, L.; Tao, T.; Wu, D.; Chen, R. A Review of Global Navigation Satellite System (GNSS)-Based Dynamic Monitoring Technologies for Structural Health Monitoring. *Remote Sens.* 2019, 11, 1001.

14. Chen, C.; Tian, Y.; Lin, L.; Chen, S.; Li, H.; Wang, Y.; Su, K. Obtaining World Coordinate Information of UAV in GNSS Denied Environments. *Sensors* 2020, 20, 2241.

15. Taha, B.; Shoufan, A. Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research. *IEEE Access* 2019, 7, 138669–138682.

16. Lykou, G.; Moustakas, D.; Gritzalis, D. Defending Airports from UAVs: A Survey on Cyber-Attacks and Counter-Drone Sensing Technologies. *Sensors* 2020, 20, 3537.

17. S. Park, H.T. Kim, S. Lee, H. Joo and H. Kim, Survey on Anti-Drone Systems: Components, Designs, and Challenges. *IEEE Access* 2021, 9, 42635–42659.

18. Leonard, J.J.; Durrant-Whyte, H.F. Simultaneous Map Building and Localization for an Autonomous Mobile Robot. In Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems, Osaka, Japan, 3–5 November 1991; pp. 1442–1447.

19. Perez-Grau, F.J.; Ragel, R.; Caballero, F.; Viguria, A.; Ollero, A. An architecture for robust UAV navigation in GPS-denied areas. *J. Field Robot.* 2018, 35, 121–145.

20. López, E.; García, S.; Barea, R.; Bergasa, L.M.; Molinos, E.J.; Arroyo, R.; Romera, E.; Pardo, S.A. Multi-Sensorial Simultaneous Localization and Mapping (SLAM) System for Low-Cost Micro Aerial Vehicles in GPS-Denied Environments. *Sensors* 2017, 17, 802.

21. Krul, S.; Pantos, C.; Frangulea, M.; Valente, J. Visual SLAM for Indoor Livestock and Farming Using a Small Drone with a Monocular Camera: A Feasibility Study. *Drones* 2021, 5, 41.

22. Bloesch, M.; Czarnecki, J.; Clark, R.; Leutenegger S.; Davison, A.J. CodeSLAM—Learning a Compact, Optimisable Representation for Dense Visual SLAM. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2560–2568.

23. Motlagh, H.D.K.; Lotfi, F.; Taghirad, H.D.; Germi, S.B. Position Estimation for Drones based on Visual SLAM and IMU in GPS-denied Environment. In Proceedings of the 7th International Conference on Robotics and Mechatronics, Tehran, Iran, 20–21 November 2019; pp. 120–124.

24. Karimi, M.; Oelsch, M.; Stengel, O.; Babaians, E.; Steinbach, E. LoLa-SLAM: Low-Latency LiDAR SLAM Using Continuous Scan Slicing. *IEEE Robot. Autom. Lett.* 2021, 6, 2, 2248–2255.

25. Horaud, R.; Hansard, M.; Evangelidis, G.; Ménier, C. An overview of depth cameras and range scanners based on time-of-flight technologies. *Mach. Vis. Appl.* 2016, 27, 1005–1020.

26. Taketomi, T.; Uchiyama, H.; Ikeda, S. Visual SLAM algorithms: a survey from 2010 to 2016. *IPSJ Trans. Comput. Vis. Appl.* 2017, 9, 16.

27. Tribelhorn, B.; Dodds, Z. Evaluating the Roomba: A low-cost, ubiquitous platform for robotics research and education. In Proceedings of the IEEE International Conference on Robotics and Automation, Rome, Italy, 10–14 April 2007; pp. 1393–1399.

28. Mashood, A.; Dirir, A.; Hussein, M.; Noura, H.; Awwad, F. Quadrotor Object Tracking Using Real-Time Motion Sensing. In Proceedings of the 5th International Conference on Electronic Devices, Systems and Applications, Ras Al Khaimah, United Arab Emirates, 6–8 December 2016; pp. 1–4.

29. LeCun, Y.; Bengio, Y. Hinton, G. Deep Learning. *Nature* 2015, 521, 436–444.

30. Nijim, M.; Mantrawadi, N. Drone Classification and Identification System by Phenome Analysis Using Data Mining Techniques. In Proceedings of the IEEE Symposium on Technologies for Homeland Security Waltham, MA, USA, 10–11 May 2016; pp. 1–5.

31. Jeon, S.; Shin, J.W.; Lee, Y.J.; Kim, W.H.; Kwon, Y.; Yang, H.Y. Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 28 August–2 September 2017; pp. 1858–1862.
34. Yue, X.; Liu, Y.; Wang, J.; Song, H.; Cao, H. Software Defined Radio and Wireless Acoustic Networking for Amateur Drone Surveillance. *IEEE Commun. Mag.* 2018, 56, 4, 90–97.

35. Seo, Y.; Jang, B.; Im, S. Drone Detection Using Convolutional Neural Networks with Acoustic STFT Features. In Proceedings of the 15th IEEE International Conference on Advanced Video Signal Based Surveillance, Auckland, New Zealand, 27–30 November 2018; pp. 1–6.

36. Matson, E.; Yang, B.; Smith, A.; Dietz, E.; Gallagher, J. UAV Detection System with Multiple Acoustic Nodes Using Machine Learning Models. In Proceedings of the third IEEE International Conference on Robotic Computing, Naples, Italy, 25–27 February 2019; pp. 493–498.

37. Sedunov, A.; Haddad, D.; Salloum, H.; Sutin, A.; Sedunov N.; Yakubovskiy, A. Stevens Drone Detection Acoustic System and Experiments in Acoustics UAV Tracking. In Proceedings of the IEEE International Symposium on Technologies for Homeland Security, Woburn, MA, USA, 5–6 November 2019; pp. 1–7.

38. Cobos, M.; Marti, A.; Lopez, J.J. A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization With Scalable Spatial Sampling. *IEEE Signal Process. Lett.* 2011, 18, 1, 71–74.

39. Knapp, C.; Carter, G. The Generalized Correlation Method for Estimation of Time Delay. *IEEE Trans. Acoust. Speech Signal Process.* 1976, 24, 4, 320–327.

40. Chang, X.; Yang, C.; Wu, J.; Shi, X.; Shi, Z. A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays. In Proceedings of the IEEE 10th Sensor Array and Multichannel Signal Processing Workshop, Sheffield, UK, 8–11 July 2018; pp. 573–577.

41. Dumitrescu, C.; Minea, M.; Costea, I.M.; Cosmin Chiva, I; SemenesCU, A. Development of an Acoustic System for UAV Detection. *Sensors* 2020, 20, 4870.

42. Blanchard, T.; Thomas, J.H.; Raoof, K. Acoustic Localization and Tracking of a Multi-Rotor Unmanned Aerial Vehicle Using an Array with Few Microphones. *J. Acoust. Soc. Am.* 2020, 148, 1456.

43. Zunino, A.; Crocco, M.; Martelli, S.; Trucco, A.; Bue, A.D.; Murino, V. Seeing the Sound: A New Multimodal Imaging Device for Computer Vision. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 6–14.

44. Liu, H.; Wei, Z.; Chen, Y.; Pan, J.; Lin, L.; Ren, Y. Drone Detection Based on an Audio-Assisted Camera Array. In Proceedings of the IEEE Third International Conference on Multimedia Big Data, Laguna Hills, CA, USA, 19–21 April 2017; pp. 402–406.

45. Svanström, F.; Englund, C.; Alonso-Fernandez, F. Real-Time Drone Detection and Tracking with Visible, Thermal and Acoustic Sensors. In Proceedings of the 25th International Conference on Pattern Recognition Milan, Italy, 10–15 January 2021; pp. 7265–7272.

46. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July, 2017; pp. 7263–7271.

47. Izquierdo, A.; del Val, L.; Villacorta, J.J.; Zhen, W.; Scherer, S.; Fang, Z. Feasibility of Discriminating UAV Propellers Noise from Distress Signals to Locate People in Enclosed Environments Using MEMS Microphone Arrays. *Sensors* 2020, 20, 597.

48. Van Veen, B.D.; Buckley, K.M. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Mag.* 1988, 5, 2, 4–24.

49. Modsching, M.; Kramer, R.; ten Hagen, K. Field Trial on GPS Accuracy in a Medium Size City: The Influence of Built-up. In Proceedings of the Third Workshop on Positioning, Navigation and Communication, Hannover, Germany, 16 March 2006.