

## Article

# Mallard Detection Using Microphone Arrays Combined with Delay-and-Sum Beamforming for Smart and Remote Rice–Duck Farming

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

<sup>1</sup> Faculty of Software and Information Science, Iwate Prefectural University, Takizawa City 020-0693, Japan

<sup>2</sup> Faculty of Systems Science and Technology, Akita Prefectural University, Yurihonjo City 015-0055, Japan; kwatanabe@akita-pu.ac.jp (K.W.); nishiguchi@akita-pu.ac.jp (M.N.); snix@akita-pu.ac.jp (S.N.); ksato@akita-pu.ac.jp (K.S.)

<sup>3</sup> Faculty of Bioresource Sciences, Akita Prefectural University, Akita City 010-0195, Japan; syamamot@akita-pu.ac.jp

<sup>4</sup> Institute of Engineering Innovation, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan; woo@robot.t.u-tokyo.ac.jp

\* Correspondence: hirokazu\_m@iwate-pu.ac.jp; Tel.: +81-019-694-2500

**Abstract:** This paper presents an estimation method for a sound source of pre-recorded mallard calls from acoustic information using two microphone arrays combined with delay-and-sum beamforming. Rice farming using mallards saves labor because mallards work instead of farmers. Nevertheless, the number of mallards declines when they are preyed upon by natural enemies such as crows, kites, and weasels. We consider that efficient management can be achieved by locating and identifying the locations of mallards and their natural enemies using acoustic information that can be widely sensed in a paddy field. For this study, we developed a prototype system that comprises two sets of microphone arrays. We used 64 microphones in all installed on our originally designed and assembled sensor mounts. We obtained three acoustic datasets in an outdoor environment for our benchmark evaluation. The experimentally obtained results demonstrated that the proposed system provides adequate accuracy for application to rice–duck farming.

**Keywords:** delay-and-sum beamforming; microphone array; position estimation; rice–duck farming; remote farming



**Citation:** Madokoro, H.; Yamamoto, S.; Watanabe, K.; Nishiguchi, M.; Nix, S.; Woo, H.; Sato, K. Mallard Detection Using Microphone Arrays Combined with Delay-and-Sum Beamforming for Smart and Remote Rice–Duck Farming. *Appl. Sci.* **2022**, *12*, 108. <https://doi.org/10.3390/app12010108>

Academic Editor: Pasquale Catalano

Received: 25 November 2021

Accepted: 22 December 2021

Published: 23 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Rice–duck farming, a traditional organic farming method, uses hybrid ducks released in a paddy field for weed and pest control [1]. Farmers in northern Japan use mallards because of their utility value as a livestock product [2]. Moreover, mallard farming saves labor because mallards work instead of farmers. Mallard farming is a particularly attractive approach for farmers, especially in a regional society that faces severe difficulties posed by the population decrease and rapid progress of aging [3]. As mallard farming has attracted attention, it has become popular not only in Japan, but also in many Asian countries.

One difficulty posed by mallard farming is that a sord of mallards must be gathered to a specific area in a paddy field. They trample down rice, which produces stepping ponds in which it is difficult to grow rice. Another shortcoming is that weed control effects are not obtained in areas outside the range of mallard activities. For this case, accurate position estimation of mallards in a paddy field is necessary. However, as depicted in the right photograph of Figure 1, detecting mallards among grown rice plants involves difficulties because the targets are not visible. Moreover, the number of mallards decreases because of predation by their natural enemies such as crows, kites, and weasels. to protect mallards from their natural enemies, specifying and managing mallard positions in real time are crucially important tasks.



**Figure 1.** Mallards for rice–duck farming (left). After some growth of rice plants (right), mallards are hidden by the plants. Detecting mallards visually would be extremely difficult.

To mitigate or resolve these difficulties, we evaluated an efficient management system created by locating and identifying mallards and their natural enemies using acoustic information that can be sensed widely in a paddy field. Furthermore, one can provide effective control to elucidate exactly when, where, and what kinds of natural enemies are approaching. This study was conducted to develop a position estimation system for mallards in a paddy field. Developing stable production technology is strongly demanded because rice produced by mallard farming trades at a high price on the market. For this system, we expect technological development and its transfer to remote farming [2], which is our conceptual model for actualizing smart farming.

This paper presents a direction and position estimation method for a sound source of pre-recorded mallard calls using acoustic information with arrayed microphones combined with delay-and-sum (DAS) beamforming. Using acoustic information, the approach can detect mallards that are occluded by stalks or grass. Based on the results, we inferred that an efficient management system can be actualized from locating and identifying mallards and their natural enemies using acoustic information that can widely sense a paddy field. We developed a prototype system with a 64-microphone array installed on our originally designed and assembled mount. To reproduce the sounds obtained using a microphone in advance, we conducted a simulated experiment in an actual environment to evaluate the estimation accuracy of the method.

This paper is structured as follows. Section 2 briefly reviews state-of-the-art acoustic and multimodal methods of automatic bird detection from vision and audio modalities. Subsequently, Sections 3 and 4 present the proposed localization method using an originally developed microphone system based on DAS beamforming. The experimental results obtained using our original acoustic benchmark datasets obtained using two drones are presented in Section 5. Finally, Section 6 presents the conclusions and highlights future work.

## 2. Related Studies

Detecting birds that fly across the sky in three-dimensional space is a challenging task for researchers and developers. Over a half-century ago, bird detection using radar [4] was studied to prevent bird strikes by aircraft at airports. By virtue of advances in software technology and improvements in the performance and affordability of sensors and computers, bird detection methods have diversified, especially in terms of offering improved accuracy and cost-effective approaches. Recent bird-detection methods can be categorized into two major modalities based on survey articles: visual [5–8] and audio [9–11]. Numerous outstanding methods [12–49] have been proposed from both modalities.

Benchmark datasets and challenges for the verification and comparison of detection performance have been expanded in both vision [50–53] and acoustic [54–65] modalities. During the first half of the 2010s, probability models and conventional machine-learning (ML) models combined with part-based features obtained using local descriptors were used widely and popularly. These methods comprise  $k$ -means [66], morphological filtering (MF) [67], principal component analysis (PCA), Gaussian mixture models (GMMs),

expectation–maximization (EM) algorithms [68], boosting [69], scale-invariant feature transform (SIFT) [70], random forests (RFs) [71], bag-of-features (BoF) [72], histograms of oriented gradients (HOGs) [73], support vector machines (SVMs) [74], local binary patterns (LBPs) [75], background subtraction (BS) [76], and multi-instance, multi-label (MIML) [77]. These methods require preprocessing of input signals for the enhancement of the features. Moreover, algorithm selection and parameter optimizations conducted in advance for preprocessing have often relied on the subjectivity and experience of the developers. If data characteristics differ even slightly, then performance and accuracy drop drastically. Therefore, parameter calibration is necessary and represents a great deal of work. In the latter half of the 2010s, deep-learning (DL) algorithms have been predominant, especially after the implementation of the error back-propagation learning algorithm [78] in convolutional neural networks (CNNs) [79].

Tables 1 and 2 respectively present representative studies of vision-based and sound-based bird detection methods reported during the last decade. As presented in the fourth column, the representative networks and backbones used in bird detection methods are the following: regions with CNN (RCNN) [80], VGGNet [81], Inception [82], ResNet [83], XNOR-Net [84], densely connected convolutional neural networks (DC-CNNs) [85], fast RCNN [86], faster RCNN [87], you only look once (YOLO) [88], and weakly supervised data augmentation network (WS-DAN) [89]. End-to-end DL models require no pre-processing for feature extraction of the input signals [90]. Moreover, one-dimensional acoustic signals can be input to the DL model as two-dimensional images [91].

**Table 1.** Representative studies of vision-based bird detection methods reported during the last decade.

Year	Authors	Method	Dataset
2011	Qing et al. [12]	Boosted HOG-LBP + SVM	original
2011	Descamps et al. [13]	Image energy	original
2011	Farrell et al. [14]	HOG + SVM + PNAD	original
2012	Mihreteab et al. [15]	HOG + CS-LBP + linear SVM	original
2013	Liu et al. [16]	HOG + linear SVM	[52]
2014	Xu et al. [17]	linear SVM classifier	original
2015	Yoshihashi et al. [18]	BS + CNN	[18]
2016	T’Jampens et al. [19]	SBGS + SIFT + BoW + SVM	original
2016	Takeki et al. [20]	CNN (ResNet) + FCNs + DeepLab	original
2016	Takeki et al. [21]	CNN + FCN + SP	original
2017	Yoshihashi et al. [22]	CNN (ResNet)	[53]
2017	Tian et al. [23]	Faster RCNN	original
2018	Wu et al. [24]	Skeleton-based MPSC	[50,52]
2019	Lee et al. [25]	CNN	original
2019	Vishnuvardhan et al. [26]	Faster RCNN	original
2019	Hong et al. [27]	Faster RCNN	original
2019	Boudaoud et al. [28]	CNN	original
2019	Jo et al. [29]	CNN (Inception-v3)	original
2020	Fan et al. [30]	RCNN	original
2020	Akcay et al. [31]	CNN + RPN + Fast-RCNN	original
2021	Mao et al. [32]	Faster RCNN (ResNet-50)	original
2021	Marcoñ et al. [33]	RNN + CNN	original

**Table 2.** Representative studies of sound-based bird detection methods reported during the last decade.

Year	Authors	Method	Dataset
2011	Jančovič et al. [34]	GMM	original
2012	Briggs et al. [35]	MIML	original
2014	Stowell et al. [36]	PCA + <i>k</i> -means + RF	[55]
2015	Papadopoulos et al. [37]	GMM	[60]
2015	Oliveira et al. [38]	DFT + MF	original
2017	Adavanne et al. [39]	CBRNN	[62]
2017	Pellegrini et al. [40]	DenseNet	[62]
2017	Cakir et al. [41]	CRNN	[62]
2017	Kong et al. [42]	JDC-CNN (VGG)	[56,62]
2017	Grill et al. [43]	CNN	[56]
2018	Lassecket al. [44]	CNN	[65]
2020	Liang et al. [45]	BNN (XNOR-Net)	[58,61]
2020	Solomes et al. [46]	CNN	[65]
2021	Hong et al. [47]	2D-CNN	[58,59]
2021	Kahl et al. [48]	CNN (ResNet-157)	original
2021	Zhong et al. [49]	CNN (ResNet-50) + GAN	original

Recently, vision-based detection methods have targeted not only birds, but also drones [92]. In 2019, the Drone-vs-Bird Detection Challenge (DBDC) was held at the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) [93]. The goal of this challenge was the detection of drones that appear at several image coordinates in videos without detecting birds that are present in the same frames. The developed algorithms must alert detected drones only, without alerting birds. Moreover, it is necessary to provide estimates of drone locations on the image coordinates. As a representative challenge for acoustic bird detection, bird acoustic detection [65] was conducted at the Detection and Classification of Acoustic Scenes and Events (DCASE) in 2018. Among the 33 submissions evaluated for this challenge, the team of Lasseck et al. [44] was awarded the champion title. For this detection task, the organizer provided 10 h of sound clips, including negative samples with human imitations of bird sounds. The algorithms produced a binary output of whether or not the sounds were birds, based on randomly extracted sound clips of 10 s. There are differences between our study and the extraction challenge of the coordinates of birds inferred from sound clips of the same period. To improve the accuracy for wide applications, the algorithms developed for the challenge incorporated domain adaptation, transfer learning [94], and generative adversarial networks (GANs) [95]. By contrast, our study is specialized for application to mallard farming in remote farming [2].

In comparison to a large number of previous studies mentioned above, the novelty and contributions of this study are as follows:

- The proposed method can detect the position of a sound source of pre-recorded mallard calls output from a speaker using two parameters obtained from our originally developed microphone arrays;
- Compared with existing sound-based methods, our study results provide a detailed evaluation that comprises 57 positions in total through three evaluation experiments;
- To the best of our knowledge, this is the first study to demonstrate and evaluate mallard detection based on DAS beamforming in the wild.

Based on the underlying method used in drone detection [96] and our originally developed single-sensor platform [97], this paper presents a novel sensing system and shows experimentally obtained results obtained from outdoor experiments.

### 3. Proposed Method

#### 3.1. Position Estimation Principle

Figure 2 depicts the assignment of two microphone arrays for position estimation. The direction perpendicular to the array was set to 0° of the beam. Let  $(x, y)$  be the coordinate of an estimation target position based on origin  $O$ . The two microphone arrays, designated as M1 and M2, are installed at distance  $p_x$  horizontally and at distance  $p_y$  vertically.

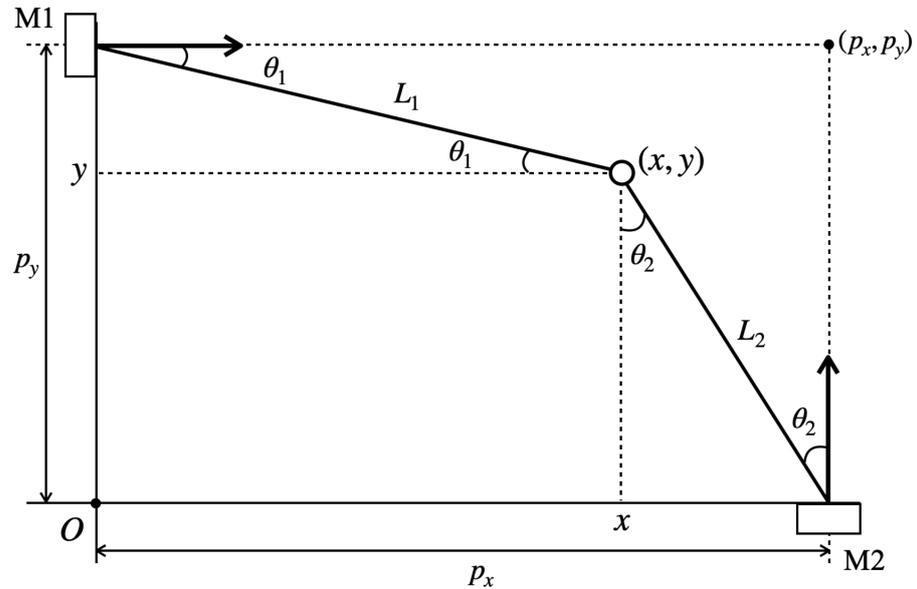


Figure 2. Arrangement of two microphone arrays M1 and M2 for position estimation.

Assuming  $\theta_1$  and  $\theta_2$  as the angles between the lines perpendicular to  $p_x$  and  $p_y$  and the straight line to  $(x, y)$ , and letting  $L_1$  and  $L_2$  represent the straight distances between the respective microphone array systems and  $(x, y)$ , then using trigonometric functions,  $p_x$  and  $p_y$  are given as presented below [98].

$$p_x = L_1 \cos \theta_1 + L_2 \sin \theta_2, \tag{1}$$

$$L_1 = \frac{p_x - L_2 \sin \theta_2}{\cos \theta_1}, \tag{2}$$

$$p_y = L_1 \sin \theta_1 + L_2 \cos \theta_2 \tag{3}$$

$$= \frac{p_x - L_2 \sin \theta_2}{\cos \theta_1} \sin \theta_1 + L_2 \frac{\cos \theta_1 \cos \theta_2}{\cos \theta_1} \tag{4}$$

$$= L_2 \frac{\cos \theta_1 \cos \theta_2}{\cos \theta_1} - L_2 \frac{\sin \theta_1 \sin \theta_2}{\cos \theta_1} + p_x \frac{\sin \theta_1}{\cos \theta_1}. \tag{5}$$

The following expanded equation is obtained by solving for  $L_2$ .

$$L_2 \frac{\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2}{\cos \theta_1} = p_y - p_x \frac{\sin \theta_1}{\cos \theta_1} \tag{6}$$

$$L_2 \frac{-\cos(\theta_1 + \theta_2)}{\cos \theta_1} = \frac{p_y \cos \theta_1 - p_x \sin \theta_1}{\cos \theta_1} \tag{7}$$

$$L_2 = \frac{p_x \sin \theta_1 - p_y \cos \theta_1}{\cos(\theta_1 + \theta_2)}. \tag{8}$$

One can obtain  $(x, y)$  as follows:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p_x - L_2 \cos \theta_2 \\ p_y - L_2 \sin \theta_2 \end{bmatrix}. \tag{9}$$

Alternatively, using  $L_2$ , one can obtain  $p_y$  as follows:

$$L_2 = \frac{p_x - L_1 \cos \theta_1}{\sin \theta_2}, \tag{10}$$

$$p_y = L_1 \sin \theta_1 + \frac{p_x - L_1 \cos \theta_1}{\sin \theta_2} \cos \theta_2 \tag{11}$$

$$= L_1 \frac{\sin \theta_1 \sin \theta_2}{\sin \theta_2} + p_x \frac{\cos \theta_2}{\sin \theta_2} - L_1 \frac{\cos \theta_1 \cos \theta_2}{\sin \theta_2}. \tag{12}$$

The following expanded equation is obtained by solving for  $L_1$ .

$$L_1 \frac{\sin \theta_1 \sin \theta_2 - \cos \theta_1 \cos \theta_2}{\sin \theta_2} = p_y - p_x \frac{\cos \theta_2}{\sin \theta_2} \tag{13}$$

$$L_1 \frac{-\cos(\theta_1 + \theta_2)}{\sin \theta_2} = \frac{p_y \sin \theta_2 - p_x \cos \theta_2}{\sin \theta_2} \tag{14}$$

$$L_1 = \frac{p_x \sin \theta_2 - p_y \sin \theta_2}{\cos(\theta_1 + \theta_2)}. \tag{15}$$

Using  $L_1$  and  $\theta_1$ , one can obtain  $(x, y)$  as follows:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p_x - L_1 \cos \theta_1 \\ p_y - L_1 \sin \theta_1 \end{bmatrix}. \tag{16}$$

Angles  $\theta_1$  and  $\theta_2$  are calculated using the DAS beamforming method, as shown in Figure 3.

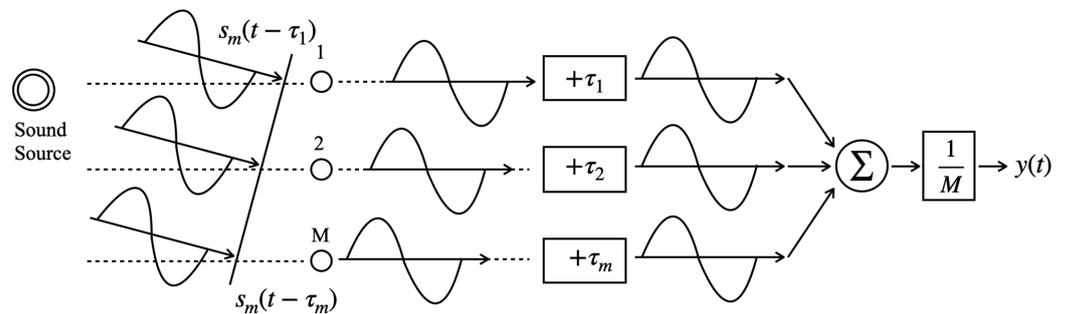


Figure 3. Delay-and-sum (DAS) beamforming in the time domain [96].

### 3.2. DAS Beamforming Algorithm

Beamforming is a versatile technology used for directional signal enhancement with sensor arrays [99]. 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:

$$y(t) = \sum_{m=1}^M w_m(t) \otimes z_m(t), \tag{17}$$

where symbol  $\otimes$  represents convolution.

Based on our previous study [96], we considered DAS beamforming in the temporal domain. Assuming that single plane waves exist and letting  $s_m(t)$  be a set of acoustic signals, then delay  $\tau_m$  as expressed for the formula below occurs for incident waves observed for the  $m$ -th sensor.

$$z_m(t) = s_m(t - \tau_m), \tag{18}$$

where  $M$  represents the total number of microphones.

The delayed  $-\tau_m$  of incident waves was offset by the advanced  $+\tau_m$  of incident waves for a filter. Signals from the direction of  $\theta$  are enhanced because of the gathered phases of signal  $s_m(t)$  in all channels. The temporal compensated filter  $w_m(t)$  is defined as:

$$w_m(t) = \frac{1}{M} \delta(t + \tau_m), \quad (19)$$

where  $\delta$  is Dirac's delta function.

Letting  $\theta$  be an angle obtained using beamforming, then, for the comparison of acoustic directions, the relative mean power level  $G(\theta)$  of  $y(t)$  is defined as:

$$G(\theta) = \frac{1}{T} \sum_{t=0}^T y^2(t), \quad (20)$$

where  $T$  represents the interval time length.

We changed  $\theta$  from  $-90^\circ$  to  $90^\circ$  with  $1^\circ$  intervals. Let  $P_1(\theta)$  and  $P_2(\theta)$  respectively denote  $G(\theta)$  obtained from M1 and M2. Using  $P_1(\theta)$  and  $P_2(\theta)$ , one can obtain  $\theta_1$  and  $\theta_2$  as:

$$\theta_n = \arg \max_{-90 \leq \theta \leq 90} P_n(\theta), \quad n = \{1, 2\} \quad (21)$$

## 4. Measurement System

### 4.1. Mount Design

For this study, we designed an original sensor mount with installed microphones and electric devices: an amplifier, an analog–digital converter, a battery, an inverter, and a laptop computer. Figure 4 depicts the design and assembled mount. We used 20 mm square aluminum pipes for the main frame. The terminals were joined with L-shaped connectors, bolts, and nuts. To set the electrical devices, acrylic boards were used for the plate located at a 575 mm height. Microphones were installed on a bracket arm as an array for a straight 1200 mm long line with a 30 mm gap.

As for the installation of microphones, a spatial foldback occurs in high-frequency signals if the gap is wide. In general [98], the distance  $d$  between microphones and the upper frequency  $f$  where no foldback occurs is defined as:

$$d \leq \frac{c}{2f} \quad (22)$$

where the constant  $c$  is the speed of sound. We confirmed from our preliminary experimental results that the power of mallard calls was concentrated in the band below 5 kHz. Therefore, we installed the microphones in 30 mm intervals because the upper limit is approximately 5.6 kHz for  $d = 0.030$ . In addition, the directivity is reduced for low frequencies, if  $d$  is narrow.

The height from the ground to the microphones was 1150 mm. We used cable ties to tighten the microphones to the bracket arm. We also labeled the respective microphones with numbers that made it easy to check the connector cables to the converter terminals.



**Figure 4.** Originally designed and assembled sensor mount.

#### 4.2. Microphone Array

Table 3 presents the components and their quantities for the respective microphone arrays. We introduced different devices for M1 and M2 because of the various circumstances such as an introduced period, research budget preparation, and coordination with our related research projects. The microphones used for M1 and M2 were, respectively, DBX RTA-M (Harman International Inds., Inc., Stamford, CT, USA) and ECM-8000 (Behringer Music Tribe Makati, Metro Manila, Philippines). Table 4 presents detailed specifications for both microphones. The common specification parameters are an omnidirectional polar pattern and the 20–20,000 Hz frequency range. For the amplifier and analog-to-digital (AD) converter, we used separate devices for M1 and an integrated device for M2. Here, both converters include a built-in anti-aliasing filter. Under the condition of the arrayed microphone interval,  $d = 0.030$  m, grating lobes may appear for the upper limit of 5.6 kHz and above. However, we assumed that the effect is relatively small because the target power of mallard calls is concentrated in the band below 5 kHz. For outdoor data acquisition experiments, we used portable lithium-ion batteries of two types. Moreover, to supply measurement devices with 100 V electrical power, the commercial power supply voltage in Japan, we used an inverter to convert direct current into alternating current.

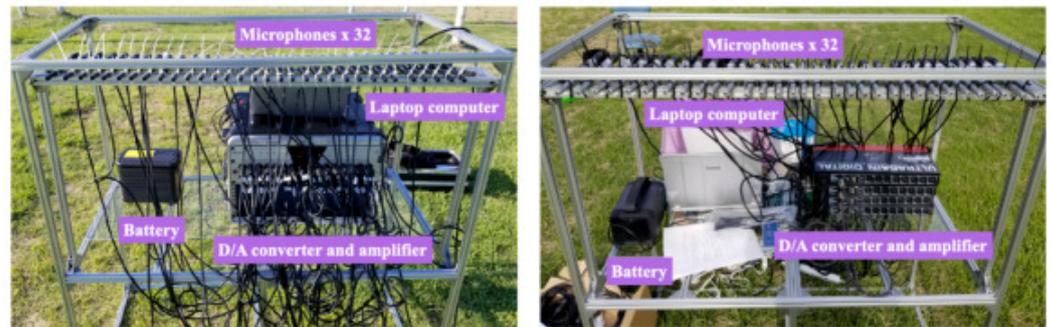
**Table 3.** System components and quantities for the respective microphone arrays: M1 and M2.

Item	M1	Amount	M2	Amount
Microphone	DBX RTA-M	×32	Behringer ECM8000	×32
Amplifier	MP32	×1	ADA8200	×4
AD converter	Orion32	×1	(included amplifier)	
Battery	Jackery 240	×1	Anker 200	×1

**Table 4.** Detailed specifications of microphones of two types.

Parameter	DBX RTA-M	Behringer ECM8000
Polar pattern	Omnidirectional	Omnidirectional
Frequency range	20–20,000 Hz	20–20,000 Hz
Impedance	$259 \Omega \pm 30\%$ (@1 kHz)	600 $\Omega$
Sensitivity	$-63 \pm 3$ dB	$-60$ dB
Mic head diameter	10 mm	(no data)
Length	145 mm	193 mm
Weight	(no data)	120 g

Figure 5 portrays photographs of all the components in Table 3 installed on the mount depicted in Figure 4. In all, the 64 microphones were connected to their respective amplifiers with 64 cables in parallel.

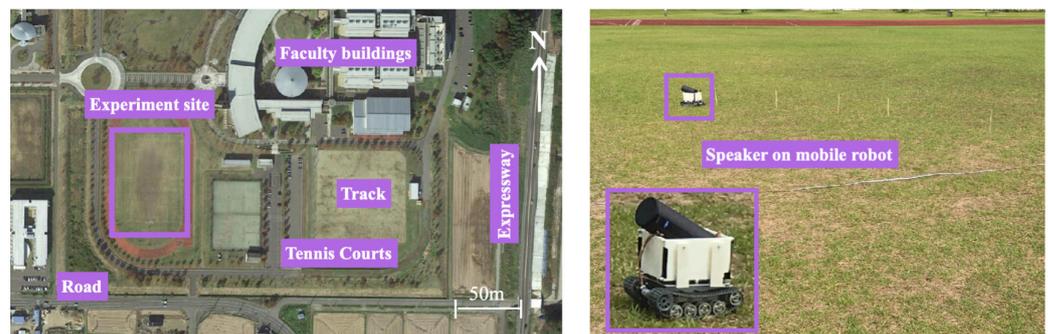


**Figure 5.** Photographs showing the appearance of M1 (left) and M2 (right) after the setup.

## 5. Evaluation Experiment

### 5.1. Experiment Setup

Acoustic data acquisition experiments were conducted at the Honjo Campus of Akita Prefectural University (Yurihonjo City, Akita Prefecture, Japan). This campus, located at  $39^{\circ}23'37''$  north latitude and  $140^{\circ}4'23''$  east longitude, has a size of  $204,379 \text{ m}^2$ . The left panel of Figure 6 portrays an aerial photograph of the campus. The surroundings comprise an expressway to the east and a public road to the south. The experiment was conducted at the athletic field on the west side of the campus, as depicted in the right panel of Figure 6.



**Figure 6.** Aerial photograph showing the experimental environment (left) and UGV used to move the sound source positions (right).

We used a pre-recorded sound of a mallard as the sound source. We played this sound on a Bluetooth wireless speaker that had been loaded on a small unmanned ground vehicle (UGV) to move to the respective measurement positions. In our earlier study, we developed this UGV as a prototype for mallard navigation, particularly for remote farming [2]. We moved it remotely via a wireless network. Our evaluation target was the detection of stationary sound sources. While the UGV was stopped at an arbitrary position, we played a mallard call sound file for 10 s. The sound pressure level produced from the speaker was  $75 \pm 10 \text{ dB}$ .

We conducted experiments separately on 19, 26, and 28 August 2020. Table 5 shows the meteorological conditions on the respective experimental days. All days were clear and sunny under a migratory anticyclone. Experiments were conducted during 14:00 to 15:00 Japan Standard Time (JST), which is 9 h ahead of the Coordinated Universal Time (UTC). The temperature was approximately  $30 \text{ }^{\circ}\text{C}$ . The humidity was in the typical mean range for Japan. The wind speed was less than  $5 \text{ m/s}$ .

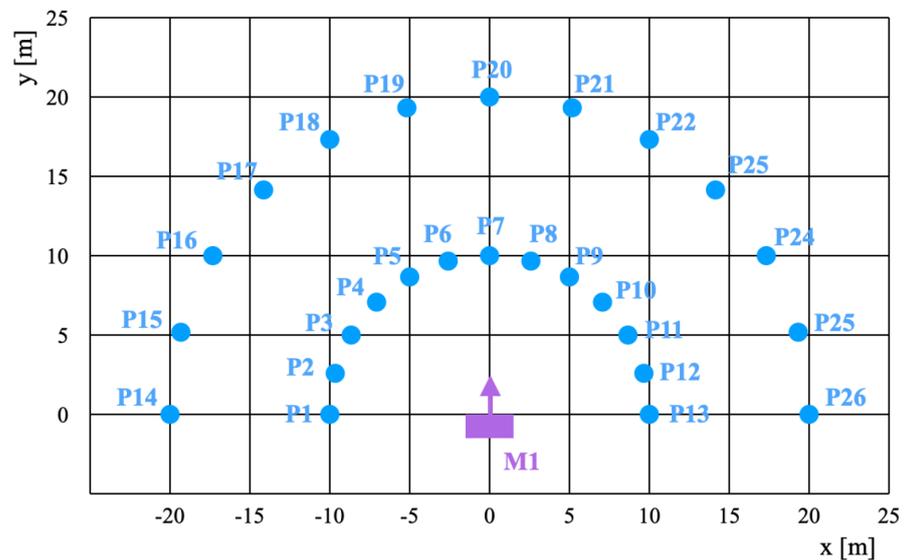
**Table 5.** Meteorological conditions on the respective experiment days.

Parameter	Experiment A	Experiment B	Experiment C
Date	19 August 2020	26 August 2020	28 August 2020
Time (JST)	14:00–15:00	14:00–15:00	14:00–15:00
Weather	Sunny	Sunny	Sunny
Air pressure (hPa)	1008.1	1008.8	1007.2
Temperature (°C)	28.6	31.7	33.4
Humidity (%)	65	53	58
Wind speed (m/s)	4.3	3.8	4.9
Wind direction	WNW	WNW	W

We conducted three evaluation experiments, designated as Experiments A–C. Experiment A comprised a sound source orientation estimation experiment using a single microphone array. The objective of this experiment was to verify the angular resolutions. Subsequently, Experiments B and C provided position estimation experimental results obtained using two microphone arrays. The difference between the two experimental setups was the field size and the orientations of the respective microphone arrays.

5.2. Experiment A

Figure 7 depicts the setup of the sound source positions and the microphone array for Experiment A. We used M1 alone to evaluate the angular resolution as a preliminary experiment. The sound source comprises 26 positions: P1–P26. We divided these positions into two groups as P1–P13 and P14–P26 which were located, respectively, on the circumferences of half circles with radii of 10 m and 20 m. The angle interval of the respective positions from the origin was 15°. Here, P1–P13 existed on the straight lines of P14–P26 from the origin.



**Figure 7.** Setup of sound source positions and the microphone array M1 for Experiment A.

Let  $\theta_g$  denote the ground-truth (GT) angle of  $\theta$ . Figure 8 depicts the angle estimation results for 16 positions at  $\theta_g = \pm 15^\circ, \pm 30^\circ, \pm 60^\circ, \text{ and } \pm 90^\circ$ . The experiment results demonstrated a unimodal output distribution of  $G(\theta)$ . Moreover,  $G(\theta)$  at positions with the radius of 10 m were found to be greater than those of 20 m. The purple-filled circles denote the vertices of the respective output waves. We let  $\theta_m$  represent the maximum angle of  $G(\theta)$  in  $-90^\circ \leq \theta \leq 90^\circ$  calculated from the Formula (21).

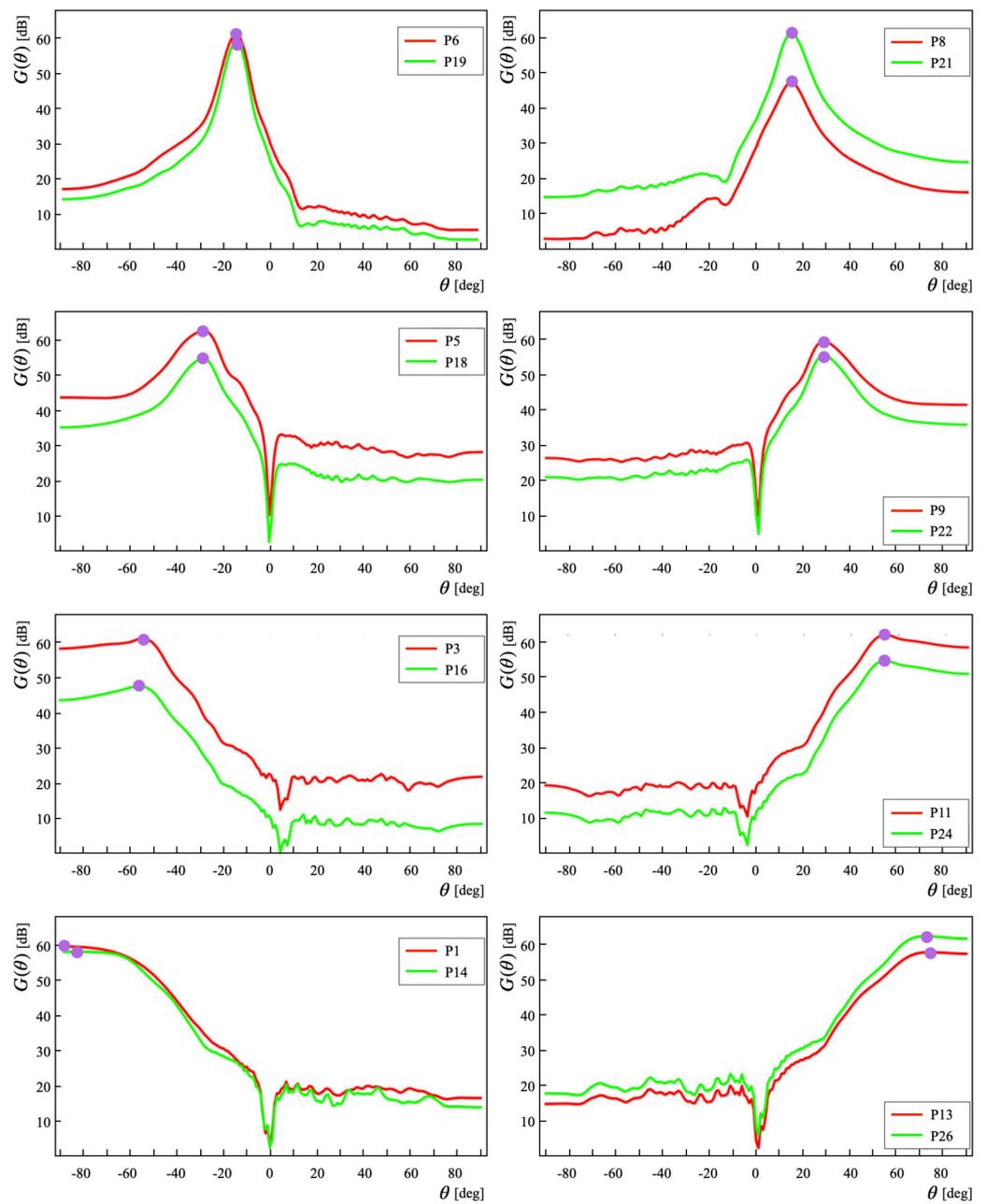


Figure 8. Angle estimation results at 16 positions ( $\theta_g = \pm 15^\circ, \pm 30^\circ, \pm 60^\circ, \text{ and } \pm 90^\circ$ ).

Table 6 presents  $\theta_m$  at the respective positions. The error  $E$  between  $\theta_g$  and  $\theta_m$  is calculated as:

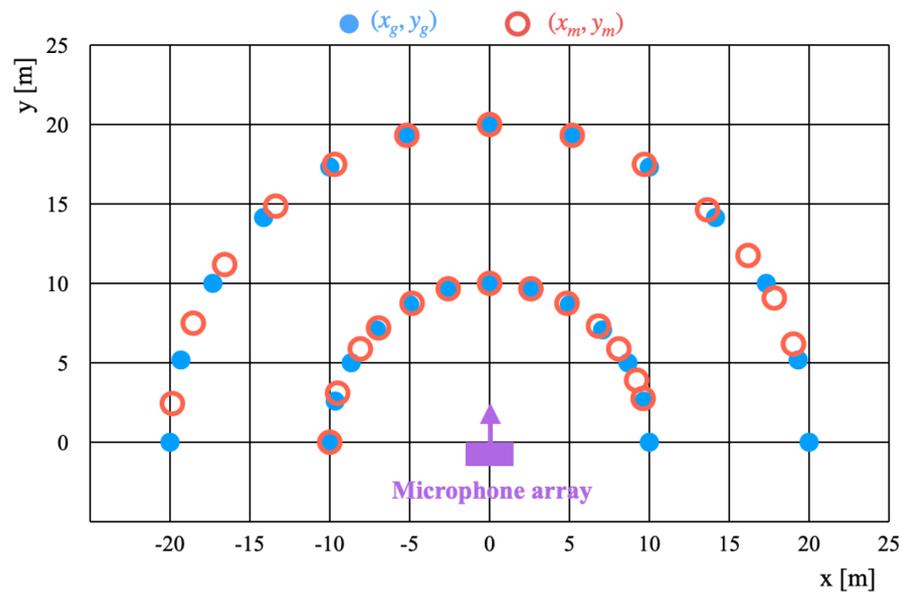
$$E = \sqrt{(\theta_g - \theta_m)^2}. \tag{23}$$

**Table 6.** Angle estimation results and error values found from Experiment A (°).

Position	$\theta_g$	$\theta_m$	$E$	Position	$\theta_g$	$\theta_m$	$E$
P1	-90	-90	0	P14	-90	-83	7
P2	-75	-72	3	P15	-75	-68	7
P3	-60	-54	6	P16	-60	-56	4
P4	-45	-44	1	P17	-45	-42	3
P5	-30	-29	1	P18	-30	-29	1
P6	-15	-15	0	P19	-15	-15	0
P7	0	0	0	P20	0	0	0
P8	15	15	0	P21	15	15	0
P9	30	29	1	P22	30	29	1
P10	45	43	2	P23	45	43	2
P11	60	54	6	P24	60	54	6
P12	75	67	8	P25	75	63	12
P13	90	74	16	P26	90	72	18

The experimentally obtained results demonstrated that  $E = 0$  in  $-15^\circ \leq \theta_m \leq 15^\circ$ . As an overall tendency, the increase of the angles and  $E$  demonstrated a positive correlation apart from the negative angles of the 10 m radius.

Figure 9 presents the two-dimensional distributions of  $(x_m, y_m)$  for comparison with  $(x_g, y_g)$ , which represents the GT coordinates of the respective positions. The theoretical distance between M1 and the respective sound source positions was unobtainable for this experimental setup. Therefore, provisional positions were calculated by substituting  $L_1$  as 10 m or 20 m.



**Figure 9.** Comparative positional distributions between  $(x_g, y_g)$  and  $(x_m, y_m)$  found from Experiment A.

Figure 10 presents scatter plots of  $\theta_g$  and  $\theta_m$ . The distribution results for  $L_1 = 10$  m and  $L_1 = 20$  m are presented, respectively, in the left and right panels. On the plus side,  $\theta_m$  for  $\theta_g$  exhibited smaller values as the angle increased. By contrast, on the minus side, it exhibited greater values as the angle increased. As the angle increased, the absolute  $\theta_m$  for absolute  $\theta_g$  tended to decrease.

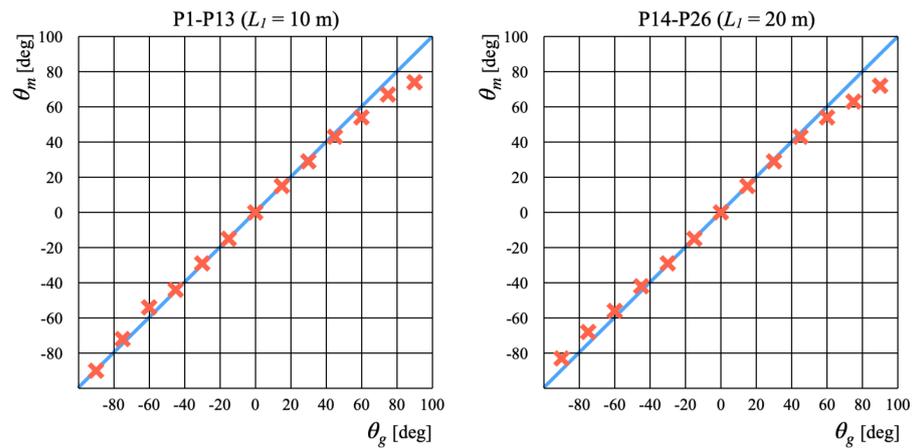


Figure 10. Scatter plots of  $\theta_g$  and  $\theta_m$  for P1–13 (left) and P14–26 (right).

5.3. Experiment B

Experiment B was performed to evaluate the position estimation using two microphone arrays: M1 and M2. Figure 11 depicts the setup of the sound source positions and the microphone arrays for Experiment B. The experiment field was a square area of 30 m in length and width. The coordinates of M1 and M2 were, respectively, (0, 30) in the upper left corner and (30, 0) in the lower right corner. The microphone frontal orientations of M1 and M2 were, respectively, horizontal and vertical. For this arrangement, the positions on the diagonal between M1 and M2 were undetermined because the intersection of  $L_1$  and  $L_2$  was indefinite in the condition of  $\theta_1 = -\theta_2$ . Therefore, the estimation target comprised 12 positions at 10 m intervals along the vertical and horizontal axes, excluding coordinates on the diagonal.

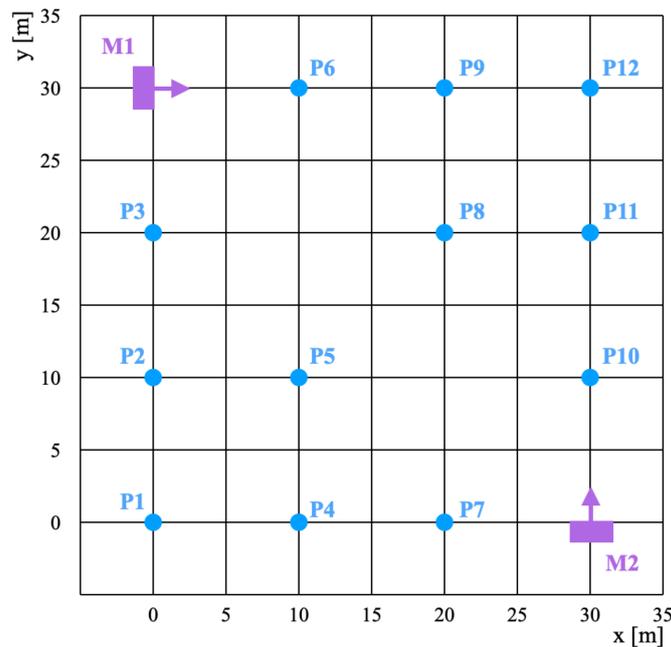
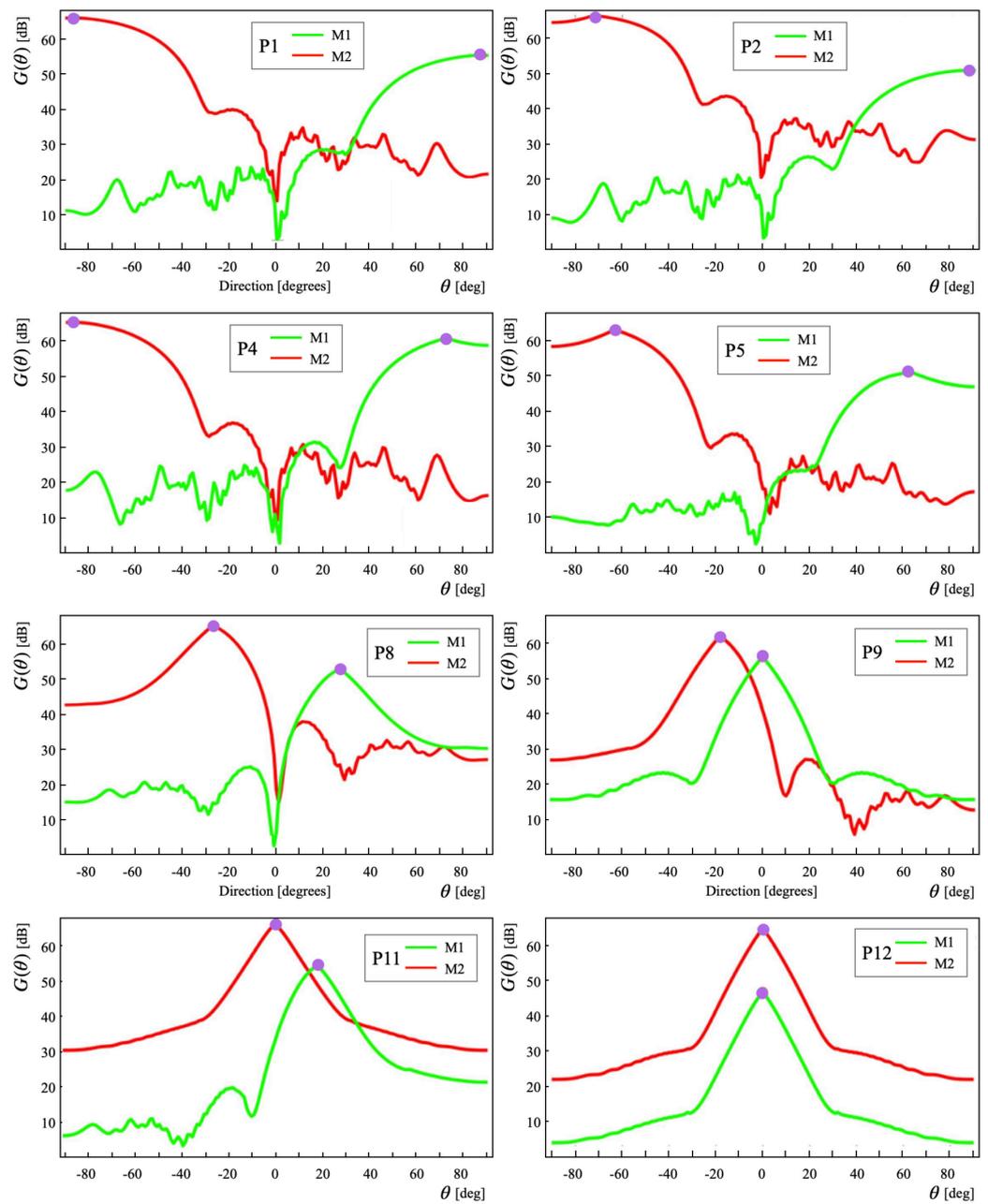


Figure 11. Setup of sound source positions and microphone arrays for Experiment B.

Figure 12 depicts the angle estimation results for 8 of 12 positions. The respective output waves exhibited a distinct peak, especially for shallow angles.



**Figure 12.** Angle estimation results obtained at eight positions (P1, P2, P4, P5, P8, P9, P11, and P12).

Table 7 presents the experiment results:  $\theta_{g1}$ ,  $\theta_{g2}$ ,  $\theta_{m1}$ ,  $\theta_{m2}$ ,  $E_1$ , and  $E_2$ . The mean error values of M1 and M2 were, respectively,  $1.25^\circ$  and  $0.92^\circ$ . The highest error of M1 was  $7^\circ$  at P7. This error increased the mean M1 error.

**Table 7.** Angle estimation results and error values found from Experiment B (°).

Position	$\theta_{g1}$	$\theta_{g2}$	$\theta_{m1}$	$\theta_{m2}$	$E_1$	$E_2$
P1	−90	90	−86	87	4	3
P2	−72	90	−71	87	1	3
P3	−56	90	−56	87	0	3
P4	−90	72	−87	72	3	0
P5	−63	63	−63	62	0	1
P6	−34	0	−34	0	0	0
P7	−90	56	−83	55	7	1
P8	−27	27	−27	27	0	0
P9	−18	0	−18	0	0	0
P10	0	34	0	34	0	0
P11	0	18	0	18	0	0
P12	0	0	0	0	0	0

Table 8 presents the position estimation results. Coordinates  $(x_m, y_m)$  were calculated from  $\theta_{m1}$  and  $\theta_{m2}$  based on the proposed method (21). The error values  $E_x$  and  $E_y$  were calculated from the following.

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} \sqrt{(x_g - x_m)^2} \\ \sqrt{(y_g - y_m)^2} \end{bmatrix}. \tag{24}$$

**Table 8.** Position estimation results obtained from Experiment B (°).

Position	$x_g$	$y_g$	$x_m$	$y_m$	$E_x$	$E_y$
P1	0	0	1.47	2.00	1.47	2.00
P2	0	10	1.05	9.97	1.05	0.03
P3	0	20	0.53	19.88	0.53	0.12
P4	10	0	9.40	1.08	0.60	1.08
P5	10	10	10.73	9.82	0.73	0.18
P6	10	30	9.76	30.00	0.24	0.00
P7	20	0	20.16	1.21	0.16	1.21
P8	20	20	19.87	19.87	0.13	0.13
P9	20	30	20.25	30.00	0.25	0.00
P10	30	10	30.00	9.76	0.00	0.24
P11	30	20	30.00	20.25	0.00	0.00
P12	30	30	30.00	30.00	0.00	0.00

The mean error values of  $E_x$  and  $E_y$  are, respectively, 0.43 m and 0.42 m. The details showed that P11 and P12 had no error. The error at P1 was the highest. Figure 13 presents the distributions of the GT coordinates  $(x_g, y_g)$  and estimated coordinates  $(x_m, y_m)$ . The six positions on the upper right indicate that the error values were small. This resulting tendency was attributed to the measurement angles of M1 and M2 being within 45°. The six positions on the bottom left indicate that the error values were large. This resultant tendency was attributed to the measurement angles of M1 and M2, which were greater than 45°.

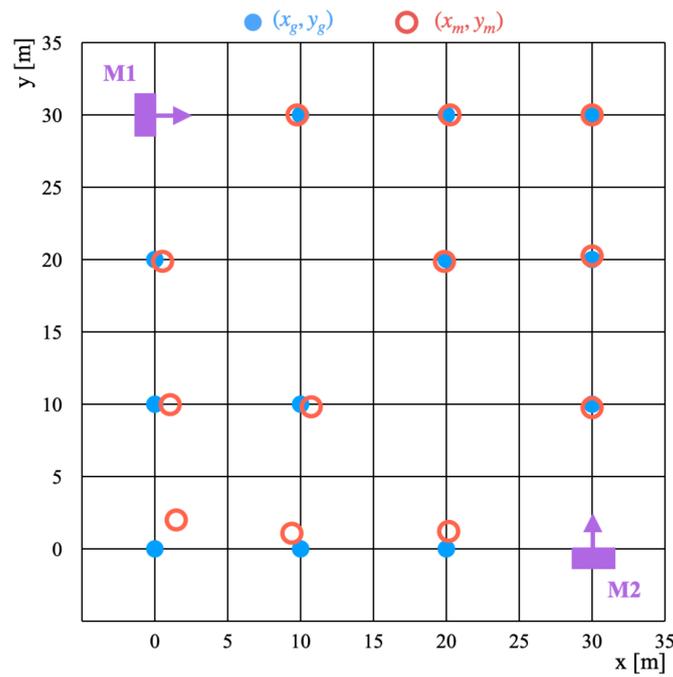


Figure 13. Positional distributions for the GT positions and estimated positions for Experiment A.

Figure 14 presents scatter plots of the GT coordinates and estimated coordinates. The distribution results for  $x$  and  $y$  are presented, respectively, in the left and right panels. Compared to Experiment A, the arrangement of the microphone array for Experiment B used  $90^\circ$ , which is half of the effective measurement range. However, the estimated angle error increased as it approached  $90^\circ$ . The error distribution trend demonstrated that the estimated position coordinates were calculated as a larger value than the GT position coordinates.

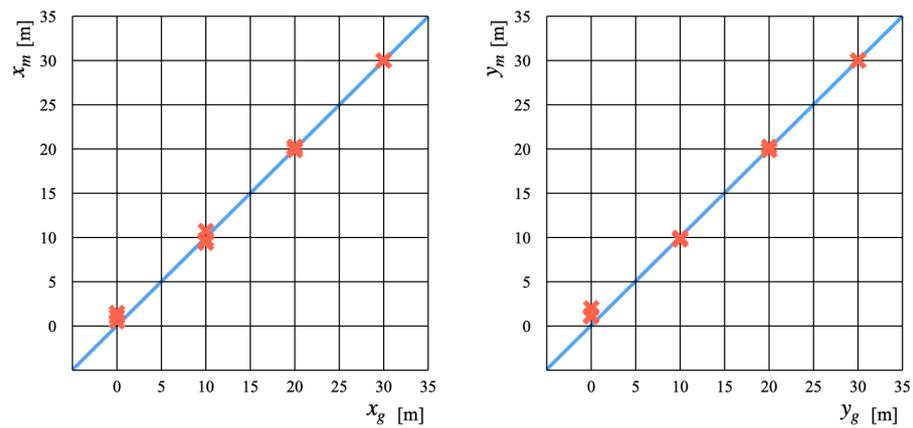


Figure 14. Scatter plots showing the GT coordinates and estimated coordinates for Experiment B.

#### 5.4. Experiment C

Figure 15 shows the experimental setup used for Experiment C. Position estimation was conducted in a rectangular area of 30 m in width and 20 m in length. The two microphone arrays, M1 and M2, were installed, respectively, at coordinates (0, 20) and (30, 0). The microphone frontal orientation of M1 and M2 was set to the equivalent of their diagonals to use the small error range around  $0^\circ$  effectively. The position detection interval was 5 m horizontally and vertically. For Experiment C, the position that could not be calculated on the diagonal was merely (15, 10). This field consists of a rectangular

shape, unlike the square field for Experiment A. The estimation target for Experiment C had 32 positions: P1–P32.

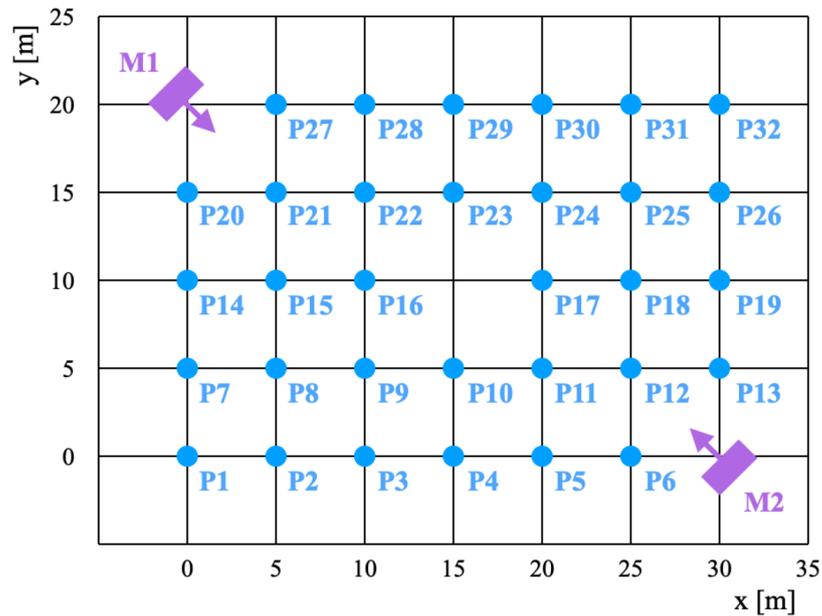


Figure 15. Setup of sound source positions and microphone arrays for Experiment C.

Figure 16 depicts the angle estimation results obtained for each group with similar vertical positions to those shown in Figure 15. The respective results demonstrated that distinct peaks shown as filled purple circles from the output waves were obtained at the source direction angles.

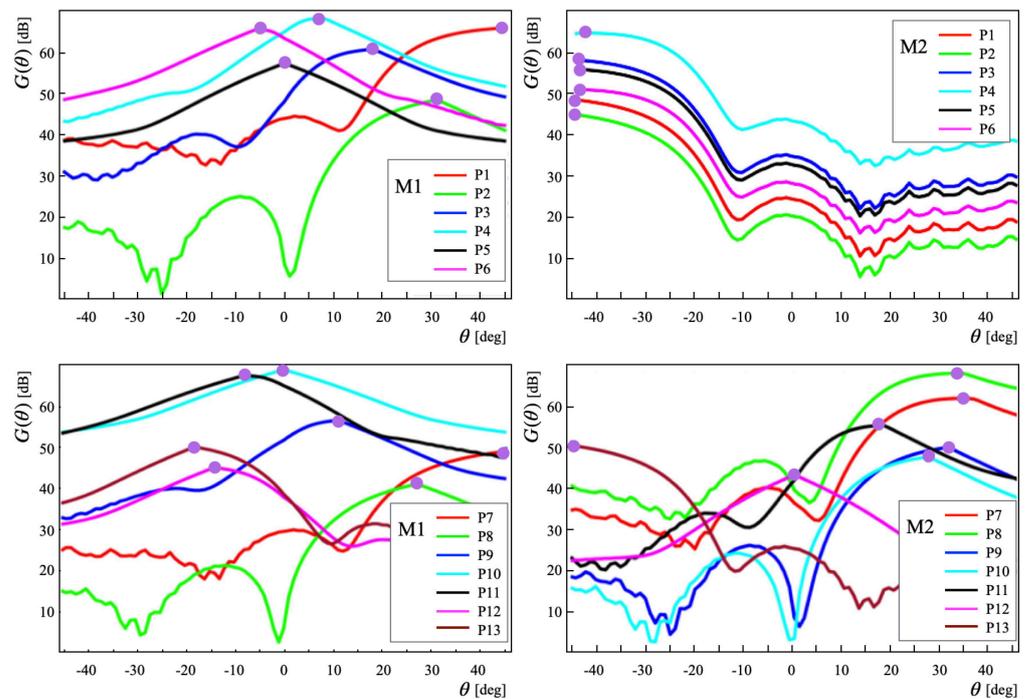


Figure 16. Cont.

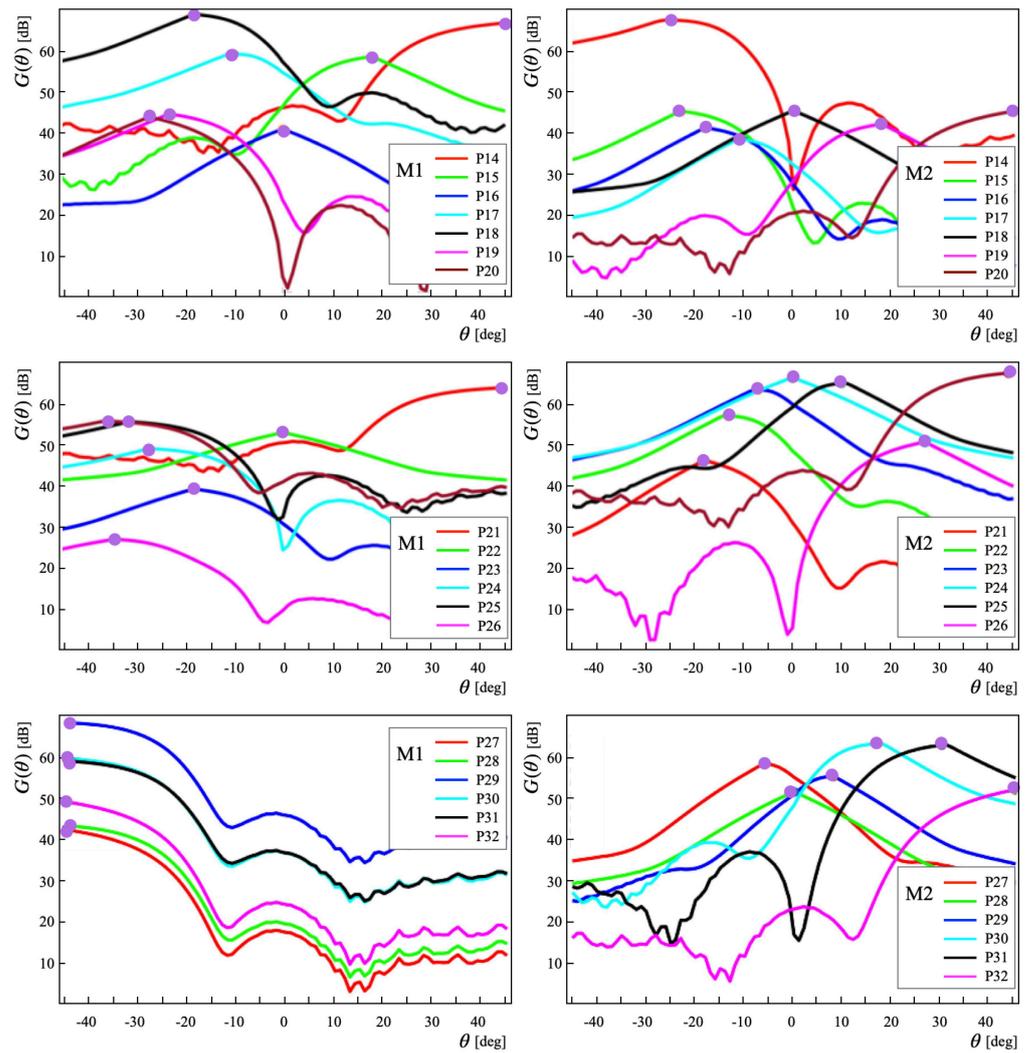


Figure 16. Angle estimation results obtained for P1–P32 ( $y = 0, 5, 10, 15,$  and  $20$ ).

Table 9 presents the estimated angles and error values. The mean error values of M1 and M2 were, respectively,  $0.28^\circ$  and  $0.53^\circ$ . The error values were smaller than  $2^\circ$  because an effective angular range smaller than  $\pm 45^\circ$  was used.

Table 9. Angle estimation results and errors obtained from Experiment C ( $^\circ$ ).

Position	$\theta_{g1}$	$\theta_{g2}$	$\theta_{m1}$	$\theta_{m2}$	$E_1$	$E_2$
P1	45	−45	44	−45	1	0
P2	31	−45	31	−45	0	0
P3	18	−45	18	−44	0	1
P4	8	−45	7	−43	1	2
P5	0	−45	0	−44	0	1
P6	−6	−45	−5	−44	1	1
P7	45	−36	44	−34	1	2
P8	27	−34	27	−33	0	1
P9	11	−31	11	−31	0	0
P10	0	−27	0	−27	0	0
P11	−8	−18	−7	−17	1	1
P12	−14	0	−14	0	0	0
P13	−18	45	−18	45	0	0
P14	45	−27	45	−25	0	2

**Table 9.** *Cont.*

Position	$\theta_{g1}$	$\theta_{g2}$	$\theta_{m1}$	$\theta_{m2}$	$E_1$	$E_2$
P15	18	-23	18	-23	0	0
P16	0	-18	0	-18	0	0
P17	-18	0	-18	0	0	0
P18	-23	18	-23	18	0	0
P19	-27	45	-27	45	0	0
P20	45	-18	44	-18	1	0
P21	0	-14	0	-13	0	1
P22	-18	-8	-18	-7	0	1
P23	-27	0	-27	0	0	0
P24	-31	11	-31	10	0	1
P25	-34	27	-34	27	0	0
P26	-36	45	-35	44	0	1
P27	-45	-6	-45	-6	0	0
P28	-45	0	-44	0	1	0
P29	-45	8	-44	8	1	0
P30	-45	18	-45	17	0	1
P31	-45	31	-44	30	1	1
P32	-45	45	-45	45	0	0

Table 10 presents the position estimation results. The mean error values of  $E_x$  and  $E_y$  were, respectively, 0.46 m and 0.38 m. The error value at P11 was the highest. Moreover, the error values at P6, P21, and P22 were high. These positions were present in the shallow angles. For this setup of the microphone frontal directions, slight angular errors induced greater coordinate errors.

**Table 10.** Position estimation results obtained from Experiment C (m).

Position	$x_g$	$y_g$	$x_m$	$y_m$	$E_x$	$E_y$
P1	0	0	0.35	0.00	0.35	0.00
P2	5	0	4.99	0.00	0.01	0.00
P3	10	0	10.01	0.35	0.01	0.35
P4	15	0	15.22	0.52	0.22	0.52
P5	20	0	19.82	0.18	0.18	0.18
P6	25	0	23.70	0.11	1.30	0.11
P7	0	5	0.25	5.78	0.25	0.78
P8	5	5	4.75	5.37	0.25	0.37
P9	10	5	10.15	4.95	0.15	0.05
P10	15	5	15.19	4.81	0.19	0.19
P11	20	5	16.22	7.33	3.78	2.33
P12	25	5	25.05	4.95	0.05	0.05
P13	30	5	30.00	4.71	0.00	0.29
P14	0	10	0.00	10.92	0.00	0.92
P15	5	10	5.06	10.08	0.06	0.08
P16	10	10	9.61	10.39	0.39	0.39
P17	20	10	20.39	9.61	0.39	0.39
P18	25	10	24.94	9.92	0.06	0.08
P19	30	10	30.00	10.25	0.00	0.25
P20	0	15	0.08	15.24	0.08	0.24
P21	5	15	3.34	16.66	1.66	1.66
P22	10	15	12.65	13.55	2.65	1.45
P23	15	15	14.81	15.19	0.19	0.19
P24	20	15	19.38	15.17	0.62	0.17
P25	25	15	25.09	15.12	0.09	0.12
P26	30	15	29.74	14.76	0.26	0.24
P27	5	20	5.30	20.00	0.30	0.00
P28	10	20	10.18	19.82	0.18	0.18
P29	15	20	15.13	19.74	0.13	0.26
P30	20	20	19.37	20.00	0.63	0.00
P31	25	20	24.76	19.57	0.24	0.43
P32	30	20	30.00	20.00	0.00	0.00

Figure 17 portrays the distributions for the GT positions and estimated positions. This tendency indicated that the error values were significant at positions near the diagonal between M1 and M2.

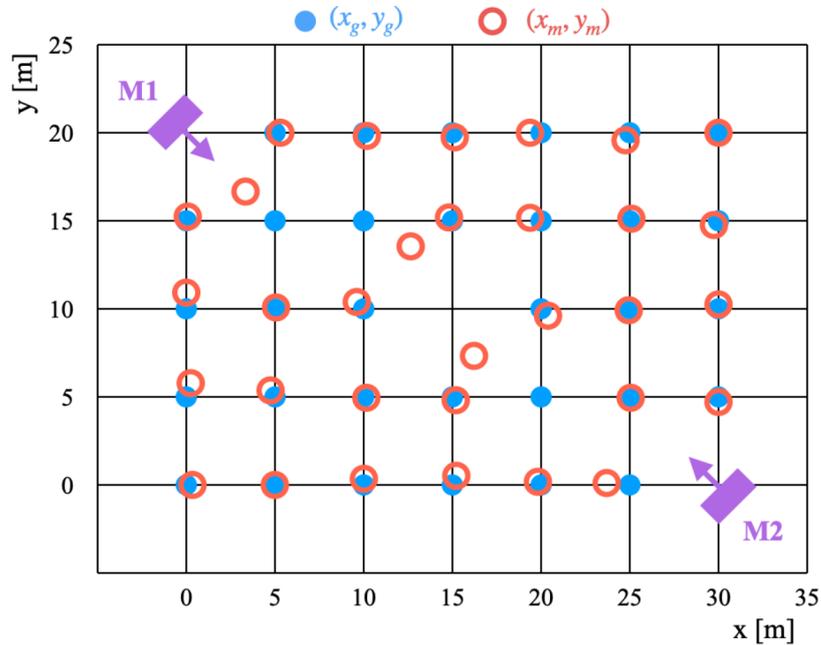


Figure 17. Positional distributions between the GT positions and estimated positions for Experiment C.

Figure 18 portrays the scatter plots between the GT coordinates and estimated coordinates for the respective axes. The distribution variation shown for the results of Experiment C was higher than that of the scatter plots for Experiment B. By contrast, the distribution variation was low in the lower left and upper right regions, which were more distant from both microphone arrays.

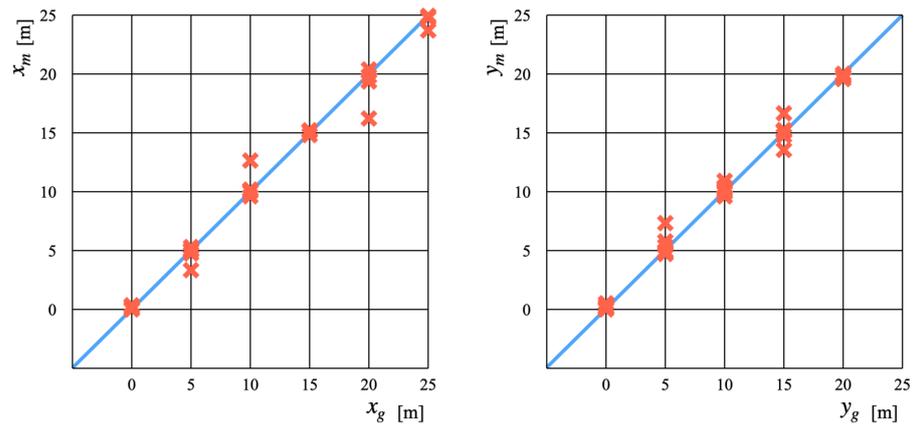


Figure 18. Scatter plots showing the GT coordinates and estimated coordinates for Experiment C.

### 5.5. Discussion

In Japanese rice cultivation, paddy rice seedlings are planted with approximately a 0.3 m separation. Therefore, mallards move in a grid up to 0.3 m. Table 11 presents the simulated evaluation results of the estimation accuracy for mallard detection within grids up to the third neighborhood in each coordinate for Experiments B and C. The experimentally obtained results revealed that the proposed method can detect mallards with 67.7% accuracy in a similar grid. Moreover, accuracies of 78.9%, 83.3%, and 88.3% were obtained, respectively, in the first, second, and third neighborhood regions. We considered

that the allowable errors can be set widely for various mallard group sizes. Although this study was conducted to assess the application of the method in a rice paddy field, the accuracy can be expected to vary greatly depending on other applications.

**Table 11.** Estimation accuracies within grids up to the third neighborhood.

Experiment	Axis	0.3 m	0.6 m	0.9 m	1.2 m
B	<i>x</i>	58.3%	75.0%	83.3%	91.7%
B	<i>y</i>	75.0%	75.0%	75.0%	83.3%
C	<i>x</i>	71.9%	81.3%	87.5%	87.5%
C	<i>y</i>	65.6%	84.4%	87.5%	90.6%
Mean		67.7%	78.9%	83.3%	88.3%

## 6. Conclusions

This study was undertaken to detect mallards based on acoustic information. We developed a prototype system comprising two sets of microphone arrays. In all, we used 64 microphones installed on our originally designed and assembled sensor mounts. For the benchmark evaluation, we obtained three acoustic datasets in an outdoor environment. For the first experiment, the angular resolutions were evaluated using a single microphone array. An error accumulation tendency was demonstrated as the angle increased from the front side of the microphones. For the second experiment, the sound sources were estimated at 12 positions with 10 m grids in a square area. The accumulated position errors were affected by the frontal orientation settings of the two microphone arrays. For the third experiment, sound sources were at 23 positions with 5 m grids in a rectangular area. To minimize blind positions, the two microphone arrays were installed with their front sides facing diagonally. Although the error increased in the diagonal direction because of the effect of the angular resolution of shallow angles, the positional errors were reduced compared to the second experiment. These experimentally obtained results revealed that the proposed system demonstrated adequate accuracy for application to rice–duck farming.

As future work aimed at practical use, we expect to improve our microphone array system in terms of miniaturization and waterproofing to facilitate its practical use for remote farming. For the problem of missing sound sources on the diagonal line between the two microphone arrays, we must consider superposition approaches using probability maps. We would like to distinguish mallard calls from noises such as the calls of larks, herons, and frogs in paddy fields. Moreover, for the protection of the ducks, we expect to combine acoustic information and visual information to detect and recognize the natural enemies of mallards. For this task, we plan on developing a robot that imitates natural enemies such as crows, kites, and weasels. We expect to improve the resolution accuracy of multiple mallard sound sources. Furthermore, we would like to discriminate mallards from crows and other birds and predict the time series mobility paths of mallards using a DL-based approach combine with state-of-the-art backbones. Finally, we would like to actualize the efficient protection of mallards by developing a system that notifies farmers of predator intrusion based on escape behavior patterns. This is because physical protection of mallards, such as electric fences and nylon lines around over paddy fields, requires a significant effort from farmers.

**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 the 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 the results of this study are available upon request to the corresponding author.

**Acknowledgments:** We would like to express our appreciation to Masumi Hashimoto and Tsubasa Ebe, who are graduates of Akita Prefectural University, for their great cooperation with 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, nor in the decision to publish the results.

## Abbreviations

The following abbreviations and acronyms are used for this report.

AD	analog-to-digital
AVSS	advanced video and signal-based surveillance
BoF	bag-of-features
BS	background subtraction
CNN	convolutional neural network
DAS	delay-and-sum
DBDC	Drone-vs-Bird Detection Challenge
DCASE	detection and classification of acoustic scenes and events
DC-CNN	densely connected convolutional neural network
DL	deep learning
EM	expectation–maximization
GAN	generative adversarial network
GMM	Gaussian mixture models
HOG	histogram of oriented gradients
JST	Japan Standard Time
LBP	local binary pattern
MF	morphological filtering
MIML	multi-instance, multi-label
ML	machine learning
PCA	principal component analysis
RCNN	regions with convolutional neural network
RF	random forest
SIFT	scale-invariant features transform
SVM	support vector machine
UGV	unmanned ground vehicle
UTC	Coordinated Universal Time
WS-DAN	weakly supervised data augmentation network
YOLO	you only look once

## References

- Hossain, S.; Sugimoto, H.; Ahmed, G.; Islam, M. Effect of Integrated Rice-Duck Farming on Rice Yield, Farm Productivity, and Rice-Provisioning Ability of Farmers. *Asian J. Agric. Dev.* **2005**, *2*, 79–86.
- Madokoro, H.; Yamamoto, S.; Nishimura, Y.; Nix, S.; Woo, H.; Sato, K. Prototype Development of Small Mobile Robots for Mallard Navigation in Paddy Fields: Toward Realizing Remote Farming. *Robotics* **2021**, *10*, 63. [\[CrossRef\]](#)
- Reiher, C.; Yamaguchi, T. Food, agriculture and risk in contemporary Japan. *Contemp. Jpn.* **2017**, *29*, 2–13. [\[CrossRef\]](#)
- Lack, D.; Varley, G. Detection of Birds by Radar. *Nature* **1945**, *156*, 446. [\[CrossRef\]](#)
- Chabot, D.; Francis, C.M. Computer-automated bird detection and counts in high-resolution aerial images: A review. *J. Field Ornithol.* **2016**, *87*, 343–359. [\[CrossRef\]](#)
- Goel, S.; Bhusal, S.; Taylor, M.E.; Karkee, M. Detection and Localization of Birds for Bird Deterrence Using UAS. In Proceedings of the 2017 ASABE Annual International Meeting, Spokane, WA, USA, 16–19 July 2017.
- Siahaan, Y.; Wardijono, B.A.; Mukhlis, Y. Design of Birds Detector and Repellent Using Frequency Based Arduino Uno with Android System. In Proceedings of the 2017 2nd International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 1–3 November 2017; pp. 239–243.

8. Aishwarya, K.; Kathryn, J.C.; Lakshmi, R.B. A Survey on Bird Activity Monitoring and Collision Avoidance Techniques in Windmill Turbines. In Proceedings of the 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development, Chennai, India, 15–16 July 2016; pp. 188–193.
9. Bas, Y.; Bas, D.; Julien, J.F. Tadarida: A Toolbox for Animal Detection on Acoustic Recordings. *J. Open Res. Softw.* **2017**, *5*, 6. [[CrossRef](#)]
10. Dong, X.; Jia, J. Advances in Automatic Bird Species Recognition from Environmental Audio. *J. Phys. Conf. Ser.* **2020**, *1544*, 012110. [[CrossRef](#)]
11. Kahl, S.; Clapp, M.; Hopping, W.; Goëau, H.; Glotin, H.; Planqué, R.; Vellinga, W.P.; Joly, A. Overview of BirdCLEF 2020: Bird Sound Recognition in Complex Acoustic Environments. In Proceedings of the 11th International Conference of the Cross-Language Evaluation Forum for European Languages, Thessaloniki, Greece, 20–25 September 2020.
12. Qing, C.; Dickinson, P.; Lawson, S.; Freeman, R. Automatic nesting seabird detection based on boosted HOG-LBP descriptors. In Proceedings of the 18th IEEE International Conference on Image, Brussels, Belgium, 11–14 September 2011; pp. 3577–3580.
13. Descamps, S.; B'échet, A.; Descombes, X.; Arnaud, A.; Zerubia, J. An Automatic Counter for Aerial Images of Aggregations of Large Birds. *Bird Study* **2011**, *58*, 302–308. [[CrossRef](#)]
14. Farrell, R.; Oza, O.; Zhang, N.; Morariu, V.I.; Darrell, T.; Davis, L.S. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 161–168.
15. Mihreteab, K.; Iwahashi, M.; Yamamoto, M. Crow birds detection using HOG and CS-LBP. In Proceedings of the International Symposium on Intelligent Signal Processing and Communications Systems, New Taipei City, Taiwan, 4–7 November 2012.
16. Liu, J.; Belhumeur, P.N. Bird Part Localization Using Exemplar-Based Models with Enforced Pose and Subcategory Consistency. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 3–6 December 2013; pp. 2520–2527.
17. Xu, Q.; Shi, X. A simplified bird skeleton based flying bird detection. In Proceedings of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 27–30 June 2014; pp. 1075–1078.
18. Yoshihashi, R.; Kawakami, R.; Iida, M.; Naemura, T. Evaluation of Bird Detection Using Time-Lapse Images around a Wind Farm. In Proceedings of the European Wind Energy Association Conference, Paris, France, 17–20 November 2015.
19. T'Jampens, R.; Hernandez, F.; Vandecasteele, F.; Verstockt, S. Automatic detection, tracking and counting of birds in marine video content. In Proceedings of the Sixth International Conference on Image Processing Theory, Tools and Applications, Oulu, Finland, 12–15 December 2016; pp. 1–6.
20. Takeki, A.; Trinh, T.T.; Yoshihashi, R.; Kawakami, R.; Iida, M.; Naemura, T. Combining Deep Features for Object Detection at Various Scales: Finding Small Birds in Landscape Images. *IPSP Trans. Comput. Vis. Appl.* **2016**, *8*, 5. [[CrossRef](#)]
21. Takeki, A.; Trinh, T.T.; Yoshihashi, R.; Kawakami, R.; Iida, M.; Naemura, T. Detection of Small Birds in Large Images by Combining a Deep Detector with Semantic Segmentation. In Proceedings of the 2016 IEEE International Conference on Image Processing, Phoenix, AR, USA, 25–28 September 2016; pp. 3977–3981.
22. Yoshihashi, R.; Kawakami, R.; Iida, M.; Naemura, T. Bird Detection and Species Classification with Time-Lapse Images around a Wind Farm: Dataset Construction and Evaluation. *Wind Energy* **2017**, *20*, 1983–1995. [[CrossRef](#)]
23. Tian, S.; Cao, X.; Zhang, B.; Ding, Y. Learning the State Space Based on Flying Pattern for Bird Detection. In Proceedings of the 2017 Integrated Communications, Navigation and Surveillance Conference, Herndon, VA, USA, 18–20 April 2017; pp. 5B3-1–5B3-9.
24. Wu, T.; Luo, X.; Xu, Q. A new skeleton based flying bird detection method for low-altitude air traffic management. *Chin. J. Aeronaut.* **2018**, *31*, 2149–2164. [[CrossRef](#)]
25. Lee, S.; Lee, M.; Jeon, H.; Smith, A. Bird Detection in Agriculture Environment using Image Processing and Neural Network. In Proceedings of the 6th International Conference on Control, Decision and Information Technologies, Paris, France, 23–26 April 2019; pp. 1658–1663.
26. Vishnuvardhan, R.; Deenadayalan, G.; Vijaya Gopala Rao, M.V.; Jadhav, S.P.; Balachandran, A. Automatic Detection of Flying Bird Species Using Computer Vision Techniques. In Proceedings of the International Conference on Physics and Photonics Processes in Nano Sciences, Eluru, India, 20–22 June 2019.
27. Hong, S.J.; Han, Y.; Kim, S.Y.; Lee, A.Y.; Kim, G. Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery. *Sensors* **2019**, *19*, 1651. [[CrossRef](#)] [[PubMed](#)]
28. Boudaoud, L.B.; Maussang, F.; Garello, R.; Chevallier, A. Marine Bird Detection Based on Deep Learning using High-Resolution Aerial Images. In Proceedings of the OCEANS 2019—Marseille, Marseille, France, 17–20 June 2019; pp. 1–7.
29. Jo, J.; Park, J.; Han, J.; Lee, M.; Smith, A.H. Dynamic Bird Detection Using Image Processing and Neural Network. In Proceedings of the 7th International Conference on Robot Intelligence Technology and Applications, Daejeon, Korea, 1–3 November 2019; pp. 210–214.
30. Fan, J.; Liu, X.; Wang, X.; Wang, D.; Han, M. Multi-Background Island Bird Detection Based on Faster R-CNN. *Cybern. Syst.* **2020**, *52*, 26–35. [[CrossRef](#)]
31. Akcay, H.G.; Kabasakal, B.; Aksu, D.; Demir, N.; Öz, M.; Erdoğan, A. Automated Bird Counting with Deep Learning for Regional Bird Distribution Mapping. *Animals* **2020**, *10*, 1207. [[CrossRef](#)]
32. Mao, X.; Chow, J.K.; Tan, P.S.; Liu, K.; Wu, J.; Su, Z.; Cheong, Y.H.; Ooi, G.L.; Pang, C.C.; Wang, Y. Domain Randomization-Enhanced Deep Learning Models for Bird Detection. *Sci. Rep.* **2021**, *11*, 639. [[CrossRef](#)]

33. Marcoň, P.; Janoušek, J.; Pokorný, J.; Novotný, J.; Hutová, E.V.; Širůčková, A.; Čáp, M.; Lázníčková, J.; Kadlec, R.; Raichl, P.; et al. A System Using Artificial Intelligence to Detect and Scare Bird Flocks in the Protection of Ripening Fruit. *Sensors* **2021**, *21*, 4244. [CrossRef] [PubMed]
34. Jančovič, P.; Köküer, M. Automatic Detection and Recognition of Tonal Bird Sounds in Noisy Environments. *EURASIP J. Adv. Signal Process.* **2011**, *2011*, 982936. [CrossRef]
35. Briggs, F.; Lakshminarayanan, B.; Neal, L.; Fern, X.Z.; Raich, R.; Hadley, S.J.K.; Hadley, A.S.; Betts, M.G. Acoustic Classification of Multiple Simultaneous Bird Species: A Multi-Instance Multi-Label Approach. *J. Acoust. Soc. Am.* **2012**, *131*, 4640–4650. [CrossRef]
36. Stowell, D.; Plumbley, M.D. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ Life Environ.* **2014**, *2*, E488. [CrossRef]
37. Papadopoulos, T.; Roberts, S.; Willis, K. Detecting bird sound in unknown acoustic background using crowdsourced training data. *arXiv* **2015**, arXiv:1505.06443v1.
38. de Oliveira, A.G.; Ventura, T.M.; Ganchev, T.D.; de Figueiredo, J.M.; Jahn, O.; Marques, M.I.; Schuchmann, K.-L. Bird acoustic activity detection based on morphological filtering of the spectrogram. *Appl. Acoust.* **2015**, *98*, 34–42. [CrossRef]
39. Adavanne, S.; Drossos, K.; Cakir, E.; Virtanen, T. Stacked Convolutional and Recurrent Neural Networks for Bird Audio Detection. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 28 August–2 September 2017; pp. 1729–1733.
40. Pellegrini, T. Densely connected CNNs for bird audio detection. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 28 August–2 September 2017; pp. 1734–1738.
41. Cakir, E.; Adavanne, S.; Parascandolo, G.; Drossos, K.; Virtanen, T. Convolutional Recurrent Neural Networks for Bird Audio Detection. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 28 August–2 September 2017; pp. 1744–1748.
42. Kong, Q.; Xu, Y.; Plumbley, M.D. Joint detection and classification convolutional neural network on weakly labelled bird audio detection. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 28 August–2 September 2017; pp. 1749–1753.
43. Grill, T.; Schlüter, J. Two Convolutional Neural Networks for Bird Detection in Audio Signals. In Proceedings of the 25th European Signal Processing Conference, Kos, Greece, 28 August–2 September 2017; pp. 1764–1768.
44. Lasseck, M. Acoustic Bird Detection with Deep Convolutional Neural Networks. In Proceedings of the IEEE AASP Challenges on Detection and Classification of Acoustic Scenes and Events, Online, 30 March–31 July 2018.
45. Liang, W.K.; Zabidi, M.M.A. Bird Acoustic Event Detection with Binarized Neural Networks. *Preprint* **2020**. [CrossRef]
46. Solomes, A.M.; Stowell, D. Efficient Bird Sound Detection on the Bela Embedded System. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal, Barcelona, Spain, 4–8 May 2020; pp. 746–750.
47. Hong, T.Y.; Zabidi, M.M.A. Bird Sound Detection with Convolutional Neural Networks using Raw Waveforms and Spectrograms. In Proceedings of the International Symposium on Applied Science and Engineering, Erzurum, Turkey, 7–9 April 2021.
48. Kahl, S.; Wood, C.M.; Eibl, M.; Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.* **2021**, *61*, 101236. [CrossRef]
49. Zhong, M.; Taylor, R.; Bates, N.; Christey, D.; Basnet, B.; Flippin, J.; Palkovitz, S.; Dodhia, R.; Ferres, J.L. Acoustic detection of regionally rare bird species through deep convolutional neural networks. *Ecol. Inform.* **2021**, *64*, 101333. [CrossRef]
50. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. Imagenet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
51. Anil, R. CSE 252: Bird's Eye View: Detecting and Recognizing Birds Using the BIRDS 200 Dataset. 2011. Available online: [https://cseweb.ucsd.edu/classes/sp11/cse252c/projects/2011/ranil\\_final.pdf](https://cseweb.ucsd.edu/classes/sp11/cse252c/projects/2011/ranil_final.pdf) (accessed on 21 December 2021).
52. Wah, C.; Branson, S.; Welinder, P.; Perona, P.; Belongie, S. *The Caltech-UCSD Birds-2000-2011 Dataset*; Computation & Neural Systems Technical Report, CNS-TR-2011-001; 2011. Available online: [http://www.vision.caltech.edu/visipedia/papers/CUB\\_200\\_2011.pdf](http://www.vision.caltech.edu/visipedia/papers/CUB_200_2011.pdf) (accessed on 21 December 2021).
53. Yoshihashi, R.; Kawakami, R.; Iida, M.; Naemura, T. Construction of a bird image dataset for ecological investigations. In Proceedings of the IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 4248–4252.
54. Buxton, R.T.; Jones, I.L. Measuring nocturnal seabird activity and status using acoustic recording devices: Applications for island restoration. *J. Field Ornithol.* **2012**, *83*, 47–60. [CrossRef]
55. Glotin, H.; LeCun, Y.; Artiefes, T.; Mallat, S.; Tchernichovski, O.; Halkias, X. *Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data*; Neural Information Processing Systems Foundation: San Diego, CA, USA, 2013.
56. Stowell, D.; Plumbley, M.D. An open dataset for research on audio field recording archives: Freefield1010. *arXiv* **2013**, arXiv:1309.5275v2.
57. Goëau, H.; Glotin, H.; Vellinga, W.-P.; Rauber, A. LifeCLEF bird identification task 2014. In Proceedings of the 5th International Conference and Labs of the Evaluation Forum, Sheffield, UK, 15–18 September 2014.
58. Salamon, J.; Jacoby, C.; Bello, J.P. A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 1041–1044.
59. Vellinga, W.P.; Planqué, R. The Xeno-Canto collection and its relation to sound recognition and classification. In Proceedings of the 6th International Conference and Labs of the Evaluation Forum, Toulouse, France, 8–11 September 2015.

60. Stowell, D.; Giannoulis, D.; Benetos, E.; Lagrange, M.; Plumbley, M.D. Detection and Classification of Audio Scenes and Events. *IEEE Trans. Multimed.* **2015**, *17*, 1733–1746. [[CrossRef](#)]
61. Salamon, J.; Bello, J.P.; Farnsworth, A.; Robbins, M.; Keen, S.; Klinck, H.; Kelling, S. Towards the Automatic Classification of Avian Flight Calls for Bioacoustic Monitoring. *PLoS ONE* **2016**, *11*, e0166866. [[CrossRef](#)]
62. Stowell, D.; Wood, M.; Stylianou, Y.; Glotin, H. Bird Detection in Audio: A Survey and a Challenge. In Proceedings of the IEEE 26th International Workshop on Machine Learning for Signal Processing, Salerno, Italy, 13–16 September 2016; pp. 1–6.
63. Darras, K.; Pütz, P.; Fahrurrozi; Rembold, K.; Tschardt, T. Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biol. Conserv.* **2016**, *201*, 29–37. [[CrossRef](#)]
64. Hervás, M.; Alsina-Pagés, R.M.; Alias, F.; Salvador, M. An FPGA-Based WASN for Remote Real-Time Monitoring of Endangered Species: A Case Study on the Birdsong Recognition of *Botaurus stellaris*. *Sensors* **2017**, *17*, 1331. [[CrossRef](#)] [[PubMed](#)]
65. Stowell, D.; Stylianou, Y.; Wood, M.; Pamula, H.; Glotin, H. Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *Methods Ecol. Evol.* **2018**, *10*, 368–380. [[CrossRef](#)]
66. MacQueen, J.B. Some Methods for classification and Analysis of Multivariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; pp. 281–297.
67. Serra, J.; Vincent, L. An overview of morphological filtering. *Circuits Syst. Signal Process.* **1992**, *11*, 47–108. [[CrossRef](#)]
68. Moon, T.K. The expectation-maximization algorithm. *IEEE Signal Process. Mag.* **1996**, *13*, 47–60. [[CrossRef](#)]
69. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
70. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the International Conference on Computer Vision, Corfu, Greece, 20–25 September 1999.
71. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
72. Csurka, G.; Dance, C.R.; Fan, L.; Willamowski, J.; Bray, C. Visual categorization with bags of keypoints. In Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 1–22.
73. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 886–893.
74. Boser, B.; Guyon, I.; Vapnik, V. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
75. Heikkilä, M.; Schmid, C. Description of interest regions with local binary patterns. *Pattern Recognit.* **2009**, *42*, 425–436. [[CrossRef](#)]
76. Cristani, M.; Farenzena, M.; Bloisi, D.; Murino, V. Background Subtraction for Automated Multisensor Surveillance: A Comprehensive Review. *EURASIP J. Adv. Signal Process.* **2010**, *1*, 24. [[CrossRef](#)]
77. Zhou, Z.H.; Zhang, M.L.; Huang, S.J.; Li, Y.F. Multi-Instance Multi-Label Learning. *Artif. Intell.* **2012**, *176*, 2291–2320. [[CrossRef](#)]
78. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
79. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
80. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
81. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556v6.
82. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
83. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
84. Rastegari, M.; Ordonez, V.; Redmon, J.; Farhadi, A. XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks. In Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 525–542.
85. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 4700–4708.
86. Wang, X.; Shrivastava, A.; Gupta, A. A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 2606–2615.
87. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
88. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
89. Hu, T.; Qi, H.; Huang, Q.; Lu, Y. See Better before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification. *arXiv* **2019**, arXiv:1901.09891.

90. Zinemanas, P.; Cancela, P.; Rocamora, M. End-to-end Convolutional Neural Networks for Sound Event Detection in Urban Environments. In Proceedings of the 24th Conference of Open Innovations Association, Moscow, Russia, 8–12 April 2019; pp. 533–539.
91. Purwins, H.; Li, B.; Virtanen, T.; Schlüter, J.; Chang, S.; Sainath, T. Deep Learning for Audio Signal Processing. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 206–219. [[CrossRef](#)]
92. Seidailyeva, U.; Akhmetov, D.; Ilipbayeva, L.; Matson, E.T. Real-Time and Accurate Drone Detection in a Video with a Static Background. *Sensors* **2020**, *20*, 3856. [[CrossRef](#)]
93. Coluccia, A.; Ghenescu, M.; Piatrik, T.; De Cubber, G.; Schumann, A.; Sommer, L.; Klatte, J.; Schuchert, T.; Beyerer, J.; Farhadi, M.; et al. Drone-vs-Bird Detection Challenge at IEEE AVSS2019. In Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, Taipei, Taiwan, 18–21 September 2019; pp. 1–7.
94. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
95. Pan, Z.; Yu, W.; Yi, X.; Khan, A.; Yuan, F.; Zheng, Y. Recent Progress on Generative Adversarial Networks (GANs): A Survey. *IEEE Access* **2019**, *7*, 36322–36333. [[CrossRef](#)]
96. Madokoro, H.; Yamamoto, S.; Watanabe, K.; Nishiguchi, M.; Nix, S.; Woo, H.; Sato, K. Prototype Development of Cross-Shaped Microphone Array System for Drone Localization Based on Delay-and-Sum Beamforming in GNSS-Denied Areas. *Drones* **2021**, *5*, 123. [[CrossRef](#)]
97. Hashimoto, M.; Madokoro, H.; Watanabe, K.; Nishiguchi, M.; Yamamoto, S.; Woo, H.; Sato, K. Mallard Detection using Microphone Array and Delay-and-Sum Beamforming. In Proceedings of the 19th International Conference on Control, Automation and Systems, Jeju, Korea, 15–18 October 2019; pp. 1566–1571.
98. Van Trees, H.L. *Optimum Array Processing*; Wiley: New York, NY, USA, 2002.
99. Veen, B.; Buckley, K. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Mag.* **1988**, *5*, 4–24. [[CrossRef](#)]