Towards Drone Recognition and Localization from Flying UAVs through Processing of Multi-Channel Acoustic and Radio Frequency Signals: a Deep Learning Approach

Andrea Toma, Niccolò Cecchinato, Carlo Drioli, Gian Luca Foresti, Giovanni Ferrin

University of Udine - Dept. of Mathematics, Computer Science and Physics

ITALY

andrea.toma@uniud.it cecchinato.niccolo@spes.uniud.it carlo.drioli@uniud.it

gianluca.foresti@uniud.it giovanni.ferrin@uniud.it

ABSTRACT

A novel radio-frequency (RF)-assisted algorithm for acoustic recognition and localization of unmanned aerial vehicles (UAVs) in a scenario with small size microphone array sensors is investigated where the multi-channel processing of acoustic signals is assisted by RF power patterns analysis. The propellers of the unidentified drone generate noise that can be used to obtain a number of cues on it, as drones with different size, weight, or mechanical characteristics produce different acoustic signals. Specifically, in this work the spectral signature of the acoustic signal detected by a multi-channel microphone array is used to recognize the drone. Furthermore, RF signals are emitted by Wi-Fi antennas and received signal strength (RSS) is measured to assist the acoustic localization. Both direction of arrival (DOA) and distance from the acoustic source can be predicted. A solution is proposed where a four-stage convolutional neural network (CNN) performs drone recognition through its acoustic spectral signature and produces the RF-assisted acoustic localization through intrinsic feature extraction, fusion of the RF and acoustic features, and regression. Applications are anti-UAV monitoring strategies from flying UAVs against illegal use of UAVs and external UAV attacks. A centralized architecture is proposed for data acquisition and streaming from multiple aerial nodes. A 19 channels spherical microphone array named Zylia is employed. To analyse the current state of this research, the experiments are presented with a description of the results.

1.0 INTRODUCTION

We address the problem of detecting the presence of a unidentified drone acting as an acoustic source, recognizing it among different drones by processing the acoustic signals generated by the propellers' noise, and localizing the drone by estimating both the direction of arrival (DOA) of the acoustic signal and the distance from the drone. A solution is proposed where the acoustic processing is assisted by a radio-frequency (RF) transmission pattern analysis. In this way, when the acoustic localization front-end detects an acoustic activity originating from the direction estimated by the RF antenna components, the acoustic source localization can be refined and the recorded signal enhanced through beamforming. This is motivated by the fact that, when the acoustic recording is performed using small-size microphone arrays installed on multirotor unmanned aerial vehicles (UAVs) as in [1,2,3], the processing and signal enhancement of acoustic sources of interest becomes especially challenging due to the constraints on the size of the microphone array that may lead to poor signal-to-noise enhancement, poor spatial resolution, and incomplete spatial information issues. To tackle such limitations, a novel RF-based processing method for an acoustic source localization has been introduced recently in [4,5], that also enables distance estimation but no recognition capability was introduced. Consequently, we now investigate the performance of the RF-assisted algorithm that also recognizes the unidentified acoustic aerial source. Our algorithm can be applied to anti-UAV monitoring strategies against illegal use of UAVs and external UAV attacks [6,7] even in hostile environments.

The recent computational and performance advances brought by the developments in the research fields of deep learning (DL) and deep neural networks (DNNs) have contributed to the increase in drone recognition algorithms in the literature as in [8,9,10]. In particular, it has been demonstrated that the combined acoustic signal primarily generated by the propellers, motors, and the mechanical vibrations of the body has a sufficiently unique signature and can be used to identify the drone type among a number of drone classes in realistic open world conditions. DL and DNNs are also being investigated in a variety of applications involving multichannel acoustic processing like in [11,12] and [13] where the multichannel spectral phase information is used as input of a convolutional neural network (CNN) for the DOA estimation. In our study, the performance of a CNN-based algorithm with a four-stage network is introduced for the recognition and localization tasks. Two parallel stages process intrinsic features of the RF data and the acoustic data. The third stage performs the acoustic source recognition and the fourth stage performs the regression. This algorithm produces both drone recognition and joint predictions of both DOA and distance from the source. The current state of this research is discussed in this paper.

To investigate our method, we created a semi-simulated scenario with experimental acoustic data generated by two different drones and synthetized RF data from a distributed antenna array. The microphone array is a 19 channels spherical array capable of performing a 3D acoustic scene analysis. An experimental sensor data streaming architecture is also presented where only small-size and low-cost hardware are used for both the acquisition system and the on-board processing units, named single board computers (SBCs), that stream the data towards a ground station (GS) where the CNN-based localization processing can be performed with high computational power.

2.0 THE ACOUSTIC AND RF MODELS

In this section, both the acoustic model and the RF model are briefly shown, whereas a detailed description can be found in [5].

2.1 Acoustic Model

Considering a spherical array with *M* omnidirectional microphones, the far-field model for the acoustic source wave propagation is based on the covariance matrix of the signal array $\mathbf{\Phi}(k,f) = E\{\mathbf{x}(k,f)\mathbf{x}^H(k,f)\}$ computed by averaging the array signal blocks over a number of snapshots *B*, $\hat{\mathbf{\Phi}}(k,f) = \frac{1}{B}\sum_{k_b=0}^{B-1}\mathbf{x}(k-k_b,f)\ \mathbf{x}^H(k-k_b,f)$. The acoustic wave relative to the propellers' noise of a drone at distance d from the microphone array impinges upon the array with a direction θ.

2.2 RF Model

The path loss model (PLM) can accurately describe the received signal strength (RSS) data in light of sight (LOS) conditions as $PL_{dB}^{i,j}(d_{i,j}) = PL_{dB}^{i}(d_0) + 10\alpha\log_{10}(d_{i,j}/d_0)$ and represents the ratio of the transmitted power to the received power of a communication channel. The formulation is based on statistical analysis, where the received power can be expressed as $P_{r,dBm}^{i,j}(d_{i,j}) = P_{r,dBm}^{i}(d_0) + 10\alpha\log_{10}(d_{i,j}/d_0) + \omega^{i,j}$. This RSS-based model follows a log-decreasing law over the distance with the slope determined by α.

3.0 PROPOSED CNN MODEL WITH ACOUSTIC-RF FUSION

The architecture of the proposed multi-stage CNN for drone recognition and localization with Acoustic-RF fusion consists of four networks with convolutional and fully connected layers, as in Figure 1. Two parallel CNNs, the RF-CNN and the Acoustic-CNN, are employed to extract and process intrinsic features from the multi-channel acoustic signal and the RF signal. These features discriminate the input data according to the source drone's propellers noise, the incident angle, and the distance from the acoustic source. The third network performs the drone recognition on the basis of their propellers' noise and outputs a binary value $\hat{C}$.

The forth network performs Acoustic-RF fusion and regression. It outputs not only DOA ($\hat{\theta}$) but also distance ($\hat{d}$) predictions simultaneously.

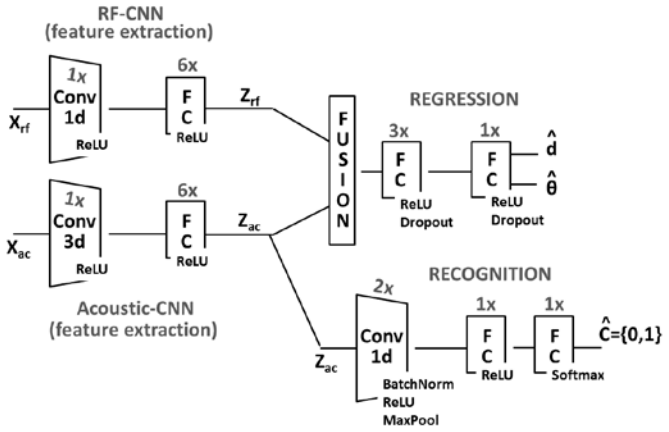

Figure 1. Four-stage CNN-based network architecture for recognition and localization of unidentified drones with RF and Acoustic data processing. The structure is composed of an RF-CNN, an Acoustic-CNN, a regression network, and a binary recognition network

The Acoustic-CNN processes 2-dimensional matrices consisting of phase values computed from the elements in the complex-valued estimated covariance matrix of the multichannel acoustic signal in Sec. 2.1. This can be represented by $\angle\tilde{\mathbf{\Phi}}(k,f)$ where $\angle$ denotes the element-wise phase of the matrix $\tilde{\mathbf{\Phi}}(k,f)$. These matrices consist of $M \cdot M$ elements, with M being the number of acoustic channels. By grouping S frames of such matrices, a 3-dimensional $M \cdot M \cdot S$ array denoted with $\mathbf{X}_{ac}$ is obtained and forms 1 input channel of the Conv3d layer. The RF-CNN processes RSS values related to the received power of RF signals, as in Sec. 2.2, from a transmitter to multiple receivers, namely $P_{r,dBm}^{i,j}(d_{i,j})$ with $i = 1$ and $j = 1, \ldots, N_{rx}$. The total number of elements is $N = N_{rx}$ with N_{rx} the number of receivers. They are processed by N input channels in the convolutional layer. S frames are processed at time so that each of the N input channels of the Conv1d consists of 1-dimensional array with S elements. This input data, N channels with S elements each, is denoted as $\mathbf{X}_{rf}$. In both the CNNs, a convolutional layer kernel operates filtering and activation on the input data. The trained kernel is computed through an optimization method to minimize the loss function intended as a measure of distance between the CNN prediction and the target. The activation function generates the output of the convolutional layer. In our network in Figure 1, the convolution layers with 64 convolutional filters learn intrinsic features in the corresponding RF and acoustic input data. The kernel size is n in Conv1d and (n_1, n_2, n_3) in Conv3d, and the activation is ReLU in both of the convolutions. This is followed by 6 fully connected linear layers with ReLU activation in both of the branches. They output a feature sample with 64 elements, namely $\mathbf{Z}_{rf}$ for the RF branch and $\mathbf{Z}_{ac}$ for the acoustic branch.

3.1 Recognition

The recognizer in Figure 1 provides estimations about the drone according to the noise generated by its propellers. In our case, the feature vector from the Acoustic-CNN, $\mathbf{Z}_{ac}$, is considered since it contains information about the acoustic characteristics of the propellers' noise emitted by the drone. When the recognizer has to differentiate between two different drones, a binary recognizer is employed. Our recognizer network consists of two consecutive convolutional networks with BatchNorm, ReLU, and MaxPool layers, a fully connected linear layer with ReLU activation function, and a fully connected linear layer with a Softmax activation layer. The kernel size of the CNNs is n_c. This network produces the probability for each of the drones that could have generated that noise. From another point view, it contains information about the probability of correct decision and the corresponding probability of false alarm. According to these

expectation values, the recognizer decides about the drone with the highest probability $\hat{C}$. In the case of binary recognition, the decision is either 0 or 1, and it is performed by the Softmax activation function.

3.2 Regression

After fusion of the two feature vectors from the RF-CNN, $\boldsymbol{Z}_{rf}$, and Acoustic-CNN, $\boldsymbol{Z}_{ac}$, regression is performed for drone localization in the prediction network in Figure 1 by the cascade of four fully connected linear layers with 64-elements input feature vector from the fusion. ReLU activations and Dropout layers are employed. In the regression process, linear activations are used to predict values in the continuous range. In particular, the last linear layer produces two outputs for DOA ($\hat{\theta}$) and distance ($\hat{d}$) predictions. In this work, the estimated DOA represents the azimuth angle between the microphone array and the acoustic source.

4.0 EXPERIMENTAL SENSOR AND DATA STREAMING ARCHITECTURE

The experimental architecture that we created for centralized acquisition and processing is shown in Figure 2 and presented in this section as part of the practical implementation of this study. In particular, an encrypted Wi-Fi streaming network infrastructure has been developed in order to send all data captured by on-board drones' acquisition sensors to the GS for a delocalized flows processing of high complexity algorithms as in the previous sections. It operates on the C band (5250MHz -5850MHz), with channels up to 80MHz, and is managed by a TP-Link access-point (AP) EAP-225-outdoor with a MikroTik 15dBi 2x2 multiple-in and multiple-out (MIMO) sectorial antenna MTAS-5G-15D120. The task of the SBC on the acquisition UAV is to capture data from the on-board sensors and send it to the GS via the wireless link. The GS is connected to the AP through a Gigabit Ethernet connection. The network sessions and communications are managed by a data algorithm via Python sockets for both the transmitting and receiving nodes, that assigns a dedicated port number for each flow streamed. The sockets' role is to route the audio, RF, and control signals between nodes on the network properly. Regarding the communication and streaming protocols, a transmission control protocol (TCP) connection with a standard Frame Type can be used. This allows to obtain a reliable connection with the possibility of recovering lost or degraded datagrams.

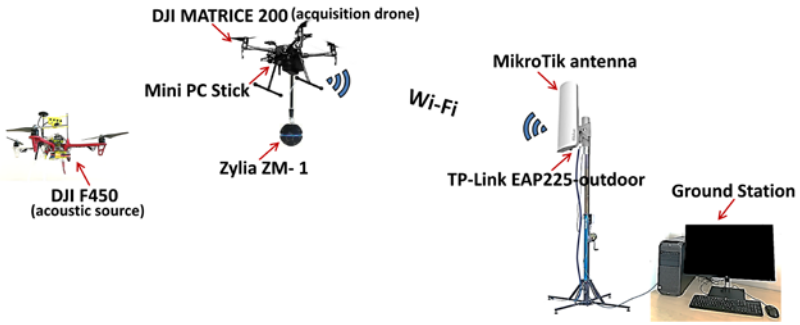

Figure 2. Illustrative scheme of the proposed experimental sensor and data streaming architecture

In our experiments, the Zylia ZM-1 [14] is used as multi-channel acoustic device for acquisition. It is a compact USB spherical array composed of 19 microphone omnidirectional digital MEMS capsules distributed on a sphere with a radius of 4.5 cm. With a 69dB signal-to-noise ratio, the capsules deliver 105dB dynamic range and high output linearity up to 130dB sound pressure level. It is able to capture the whole surrounding acoustic field in 3D with 48 kHz sample rate and resolution of 24 bit. Its gain can be set in the range 0-70dB. It is connected via USB to a Mini PC Stick NV41S computing device SBC with Intel Celeron J4105 (up to 2.5GHz) and Windows 10 Pro. The connection to the AP and the GS, where the acoustic data collected by the Zylia is sent, is enabled by its external dual band Wi-Fi gain antenna. For the experiments in this work, the microphone array is placed on a fixed support on the ground. Indeed, although we are on the

way to mounting it on a drone, the landing phase should be further tested for safety reason. However, this scenario allows for effective investigation towards an hovering situation when the Zylia will eventually be mounted on a flying drone.

Two drones have been used separately as unidentified acoustic sources for this experimental application, a custom DJI F450 quadcopter and a DJI Matrice 200 multirotor with different noise generated by the propellers of the two UAVs, as mentioned in Sec. 5.0 where the experiments are described in detail.

To assist the spherical microphone array, a receiving RF antenna linear array can be built by connecting USB Wi-Fi antenna elements to the Mini PC Stick. The frontal axis of the antenna array can be rotated with respect to the frontal direction of the drone, in order to avoid ill-posed situations from an RF point of view. Each acoustic source is equipped with a transmitting antenna.

5.0 EXPERIMENTS AND RESULTS

We consider the scenario in Figure 3(a) where the acoustic signals, emitted sequentially by acoustic sources from 6 different coordinate points, are measured by the microphone array Zylia. The height of the spherical microphone array with respect to the ground is 1.60 m. Two different UAVs have been employed, one at a time, as unidentified acoustic sources. First, the Matrice 200, and then the DJI F450. They were in hovering during the acquisition time corresponding to each of the 6 coordinate points according to the configurations in Table 1. We chose the relative North direction of the Zylia as the reference direction. In the real experimental setup of Figure 3(b), the Zylia is placed in the centre of a ground field with grass in the university campus in order to reduce the reverberation that could be caused by the nearby building and trees.

To form the acoustic dataset consisting of propellers' noise from the two UAVs in the 6 configurations, the acoustic signals gathered by the Zylia are recorded on the GS as mp4 file. The acquisition time for each configuration is 19 seconds, the sample-rate is 48 MHz, data format is 32-bit float, FFT length is 2048, and the gain of the Zylia is 0dB. The registration script is based on the sounddevice Python module. Concerning the simulated RF data, the channel attenuation factor α is set to 2.3, the noise standard deviation σ_w to 3 dB, and the received power at the reference distance $P^i_{r,dBm}(d_0)$ to -23 dBm. The receiving antenna array consists of $N_{rx} = 2$ elements with spacing 0.5m and frontal axis with respect to the frontal direction of the drone rotated by 60 degrees. The mean value over 20 samples of RSS measurements is taken.

The 4-stages CNN is implemented by using Pytorch. In the regression task, the optimizer is Adamax with learning rate 0.1 and the loss function is MSELoss. In the classification task, the optimizer is Stochastic Gradient Descent (SGD) with learning rate 0.01 and the loss function is CrossEntropyLoss. The number of total data samples is 1491150 divided into 67% training samples (off-line) and 33% testing samples (real-time). The number of epochs for training is 75. The number of frames is $S = 50$ for both the training and testing samples. The acoustic covariance matrix (and its phase values) consists of 19 × 19 values from the estimated covariance matrix, while the RSS array consists of $N = 2$ values. Batch size is 64, for both acoustic and RF inputs. The kernel size of the convolutional layers in the Acoustic-CNN and RF-CNN is $(n_1, n_2, n_3) = (1,1,3)$ and $n = 3$, respectively, and $n_c = 3$ in the recognition network.

(a) (b)

Figure 3. a) Illustrative scheme of the setup where the unidentified acoustic source hovers at 6 different coordinate points (defined by distance and angle) with respect to the Zylia and b) outdoor setup with the Zylia as microphone array and the Matrice 200 as acoustic source

Table 1. Setup values relative to six different configurations according to the acoustic source (DJI Matrice 200 or DJI F450) coordinate points in the experiments

	Configuration #1	Configuration #2	Configuration #3	Configuration #4	Configuration #5	Configuration #6
Distance	5.016 [m]	5.016 [m]	5.016 [m]	10.179 [m]	10.179 [m]	10.179 [m]
Altitude	2.0 [m]	2.0 [m]	2.0 [m]	3.5 [m]	3.5 [m]	3.5 [m]
Azimuth	+45 [deg]	0 [deg]	-45 [deg]	-45 [deg]	0 [deg]	+45 [deg]
Elevation	4.574 [deg]	4.574 [deg]	4.574 [deg]	10.758 [deg]	10.758 [deg]	10.758 [deg]

The validation of the proposed method, over a number of repeated experiments, resulted in a recognition accuracy after the testing phase of 0.957 as in Table 2(a). This means that the recognizer provides satisfying performance and this result could be further improved by acting on the deep model and the training data. Concerning the localization, the results are shown in Table 2(b) for the training and in Table 2(c) for the testing. These preliminary results show that the root mean square error is 41.160 for $\hat{\theta}$ and 2.566 for $\hat{d}$ after training and 42.247 for $\hat{\theta}$ and 3.076 for $\hat{d}$ after testing. Further investigation is required to make the prediction process more effective, by extending the training dataset and introducing learning of temporal dependencies in both the phase covariance matrix of the acoustic signal and the RSS power pattern.

Table 2. (a) Recognition accuracy after testing, (b) localization performance for training, and (c) localization performance for testing

recognition accuracy
0.957

(a)

	minimum absolute error	mean absolute error	RMSE loss
DOA $\hat{\theta}$	0.164	31.602	41.160
distance $\hat{d}$	2.407	2.560	2.566

(b)

	RMSE
DOA $\hat{\theta}$	42.247
distance $\hat{d}$	3.076

(c)

6.0 CONCLUSIONS AND FUTURE WORK

A novel RF-assisted recognition and localization of unidentified aerial acoustic source method is proposed for applications like anti-UAV monitoring strategies from flying UAVs against illegal use of UAVs and external UAV attacks. It is based on a four-stage CNN network consisting of two CNNs for feature extraction, the RF-CNN and the Acoustic-CNN, a recognition network for drone classification, and a regression network with a fusion layer that produces not only DOA predictions but also predictions of the distance between the acquisition drone and the acoustic source. The motivation of considering RF data comes from the issues related to small-size microphone arrays that make acoustic signal processing especially challenging. A semi-simulated scenario has been created where the RF data is synthetized by generating RSS measurements, while the acoustic data is obtained through phase values of the elements in the average covariance matrix of the real propellers' noise generated by two drones in hovering positions. The acoustic signals are collected by a 19 channels spherical microphone array capable of performing a 3D acoustic scene analysis. The current state of this research is analysed in this paper and the results after validation of the four-stage CNN have shown promising performance for the recognition network, while further investigation is required to make the localization more robust and effective.

A dedicated experimental sensor and data streaming architecture is also proposed for a delocalized processing of high complexity algorithms. It enables Wi-Fi streaming towards a GS of data collected by the on-board RF and acoustic sensors. In future work the Zylia will be mounted on a drone along with an antenna array and an extended and fully-real experimental dataset will be collected. Regarding the algorithm, additional study for the localization part will be conducted.

ACKNOWLEDGEMENTS

This research was partially supported by Italian MoD project a2018-045 "A proactive counter-UAV system to protect army tanks and patrols in urban areas" (Proactive Counter UAV), and by the ONRG project N62909-20-1-2075 "Target Re-Association for Autonomous Agents" (TRAAA).

REFERENCES

[1] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," in Wireless Communications and Mobile Computing, vol. 2017, pp. 1–24, 2017.

[2] S. Oh, Y.-J. Go, J. Lee, and J.-S. Choi, "Sound source positioning using microphone array installed on a flying drone," in The Journal of the Acoustical Society of America, vol. 140, no. 4, pp. 3422–3422, 2016.

[3] D. Salvati, C. Drioli, G. Ferrin, and G. L. Foresti, "Beamforming-Based Acoustic Source Localization and Enhancement for Multirotor UAVs," in 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018, pp. 987-991.

[4] A. Toma, N. Cecchinato, C. Drioli, G. L. Foresti and G. Ferrin, "CNN-based processing of radio frequency signals for augmenting acoustic source localization and enhancement in UAV security applications," *2021 International Conference on Military Communication and Information Systems (ICMCIS)*, 4-5 May 2021, pp. 1-5.

[5] A. Toma, D. Salvati, C. Drioli, and G. L. Foresti, "CNN-based processing of acoustic and radio frequency signals for speaker localization from MAVs," in *Proceedings of the 22nd Conference of the International Speech Communication Association (INTERSPEECH)*, 30 Aug-3 Sept 2021.

[6] Y. Kim, Y.G. Min, P.S. Hee, J. Wun-Cheol, S. Soonyong, and H. Tae-Wook, "The analysis of image acquisition method for anti-uav surveillance using cameras image," in 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 549–554.

[7] H. Liu, Z. Wei, Y. Chen, J. Pan, L. Lin, and Y. Ren, "Drone detection based on an audio-assisted camera array," in 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), 2017, pp. 402–406.

[8] H. Kolamunna, T. Dahanayaka, J. Li, S. Seneviratne, K. Thilakaratne, A. Y. Zomaya, and A. Seneviratne. 2021. "DronePrint: Acoustic Signatures for Open-set Drone Detection and Identification with Online Data," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 20 (March 2021), 31 pages.

[9] S. Al-Emadi, A. Al-Ali, A. Al-Ali, "Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks," *Sensors*. 2021; 21(15):4953.

[10] H. Kolamunna, J. Li, T. Dahanayaka, S. Seneviratne, K. Thilakaratne, A. Y. Zomaya, and A. Seneviratne. 2020, "AcousticPrint: acoustic signature based open set drone identification," In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks* (*WiSec '20*), Association for Computing Machinery, New York, NY, USA, 349–350.

[11] J. Choi and J. Chang, "Convolutional neural network-based direction-of-arrival estimation using stereo microphones for drone," in 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1–5.

[12] W. He, P. Motlicek, and J. M. Odobez, "Neural network adaptation and data augmentation for multi-speaker direction-of-arrival estimation," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, pp. 1303–1317, 2021.

[13] D. Salvati, C. Drioli, and G. L. Foresti, "Exploiting CNNs for Improving Acoustic Source Localization in Noisy and Reverberant Conditions," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 2, pp. 103-116, April 2018.

[14] Website (ZYLIA ZM-1 microphone): https://www.zylia.co/zylia-zm-1-microphone.html.