



HAL
open science

Drone intrusion detection using a network of acoustic arrays running a neural network for real-time drone detection, localization and identification

Sébastien Hengy, Sébastien de Mezzo, Marco Sangwa-Simba, Alexis Matwyschuk, Hadrien Pujol, Eric Bavu

► To cite this version:

Sébastien Hengy, Sébastien de Mezzo, Marco Sangwa-Simba, Alexis Matwyschuk, Hadrien Pujol, et al.. Drone intrusion detection using a network of acoustic arrays running a neural network for real-time drone detection, localization and identification. 30th International Congress on Sound and Vibration, Jul 2024, Amsterdam, Netherlands. hal-04789559

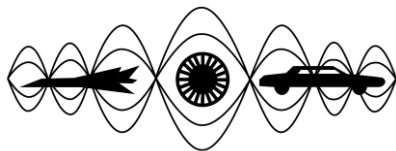
HAL Id: hal-04789559

<https://hal.science/hal-04789559v1>

Submitted on 18 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



30th International Congress on Sound and Vibration



DRONE INTRUSION DETECTION USING A NETWORK OF ACOUSTIC ARRAYS RUNNING A NEURAL NETWORK FOR REAL-TIME DRONE DETECTION, LOCALIZATION AND IDENTIFICATION

Sébastien Hengy, Sébastien De Mezzo, Marco Sangwa-Simba, Alexis Matwyschuk
ISL, French-German research institute of Saint-Louis
email: sebastien.hengy@isl.eu

Hadrien Pujol
ACOEM, Lyon

Eric Bavu
LMSSC laboratory, Le CNAM, Paris

Over the last few years, a data fusion solution has been developed, that integrates information from multi-modality sensors (acoustic, optical) to detect, identify, locate and track a drone entering on an area to be protected. To date, this approach exploits the information transmitted by acoustic sensors, which enable initial detection. Each individual microphone array has an embedded deep neural network that analyses the incoming sound data in real-time and estimates whether a drone is present, its type, and its direction of arrival every 20 milliseconds. The presence of several microphone arrays in the area enables the position of the potential threat to be estimated. The estimated position is then transmitted to a camera mounted on a motorized stand, which then confirms the actual presence of a drone thanks to an on-board A.I. and starts tracking the object.

The present paper describes the results obtained with the acoustic sensors during the final measurement campaign of the presented study. Five acoustic sensors were deployed. The fusion computer estimated the drone's position every 200 milliseconds. The drone's estimated position has been compared with the ground truth data recorded using an RTK-GPS carried on-board the drones during the flight.

Over the 14 flights for which RTK-GPS data were available, for a total duration of 95 minutes and 29 seconds, the median radial error in estimating the UAV's position averaged 10.7 meters, with a mean standard deviation of 15 meters. These data cover the full range of flight scenarios, including long-distance and high-speed UAVs. UAV recognition rates based on acoustic data fusion alone are 99.2% and 95.8% respectively for DJI S1000 and Phantom flights. Future solutions for enhanced detection range and handling of multiple drone intrusions are introduced.

Keywords: drone localization, acoustics, identification, deep learning

1. Introduction

Modern conflicts show that the use of drones for sudden attack is a major concern on the battlefield. The “low-cost” aspect of this type of weapon, its scalability and easy deployment made it a major threat on the battlefield [1]. Drones are also a major concern in civilian environment. Incidents of drones breaching the airspace over prisons [2,3] and nuclear facilities [4] underscore the urgent need for robust detection, localization and identification solutions.

Among these, the localization of drones using multiple modalities of ground sensors, associated to data fusion capabilities, are promising solutions to detect unauthorized incursions. Real-time monitoring and precise localization capabilities can now be achieved with low-cost, passive, omnidirectional technologies, including acoustics arrays. This paper aims to explore the principles, challenges, and advancements in acoustic signal processing for drone detection, localization and identification.

On this basis, this paper presents the latest results obtained using a network of acoustic sensors equipped with an embedded deep neural network that handles every 20 milliseconds a detection, localization and identification task. Metadata integrating the information relative to these tasks are then transmitted to a data fusion computer that enables the estimation of the position of the potential threat. The estimated position is then transmitted to a camera mounted on a motorized stand and using multiple wavelengths, which then confirms the actual presence of a drone thanks to an on-board A.I. and begins tracking the object [5].

This paper presents the results of drone localization during the project's final trials held in May 2022. The drone's position was estimated using data fusion, and compared with the ground truth recorded using a RTK-GPS on-board the drones during each flight. Fourteen flights have been recorded, of a duration of 10 to 25 minutes each, for a total duration of 95 minutes and 29 seconds. In the first section, the individual acoustic arrays and the data fusion process are presented. Then the second section describes the field experiment setup and results obtained for the detection and tracking of single drone flying near to the monitored area. Finally, future developments are discussed, introducing new methodologies that will be used in future iterations for the tracking of multiple drones arriving near to a restricted area.

2. Drone tracking using multiple acoustic arrays

In order to detect and track a drone flying over a restricted area, a network of multiple acoustic arrays is deployed on the field. Each individual acoustic array is running a deep neural network that allows for real-time detection, localization and identification (first subsection). The data fusion is handled on a personal computer every 200 milliseconds and allows for the estimation of the position of the detected threat in a geographical referential (second subsection).

2.1 Individual acoustic array

A set of compact, transportable and independent broadband microphonic antennas are deployed around the area of interest. The overall range of monitoring using the audio modality therefore depends only on the number of microphonic antennas backed up by their individual neural network in the monitoring network. From a technological point of view, the selected acoustic antennas are based on MEMS digital microphones (19 microphones per antenna), which offer decisive advantages in terms of signal-to-noise ratio, miniaturization and densification of portable antennas. This relatively inexpensive technology enables a large number of sensors to be deployed over large areas. Each module is linked (*Fig. 1*) via a USB (Universal Serial Bus) connection to a miniature, low-power Nvidia Jetson Xavier computational unit, which performs real-time local localization and target recognition tasks, and integrates deep

neural networks pre-trained using databases created during the three years project that lead to the final experiment presented in this paper.

The specialized AIs were trained on a multi-channel dataset of acoustic signals recorded from realistic conditions. These data acquisitions were augmented by a 3D spatializer. This augmentation enabled the neural network to respond as efficiently as possible to the localization and source identification tasks that will be performed simultaneously by the AI modules at the microphone arrays, even in noisy environments and in the presence of dominant sources covering the acoustic signature of drones flying over the site to be protected. Each intelligent acoustic module in the network transfers its recognition and localization data at a rate of 40 Hz to the data fusion computer. The orientation of the motorized pan&tilt unit and the field-of-view parameters of the optical system are controlled on the basis of the data fusion localization results.



Figure 1: View of the 19 MEMS microphone antenna without (left) and with (right) environmental protection, and Nvidia Jetson associated with the antenna.

During multiple field experiments involving various types of drones and flight patterns, a multi-channel audio dataset and a realistic in-flight drone dataset were built in order to train the deep neural network for acoustic drone localization and recognition. Acoustic signals recorded by microphone arrays intrinsically contain information about the position and nature of the acoustic source. The aim of the Beam-Learning deep neural network presented in [6] is to recover this information by supervised learning from the raw temporal data, without any pre-processing. Supervised learning requires a priori knowledge of the position and identification information of UAVs in flight which has been possible with the use of an RTK-GPS system mounted on the various types of drones used during the organized field experiments. Hence, the acoustic dataset annotations included the drone position information.

The individual detection, localization and identification performance of the individual acoustic arrays has been evaluated for various drone sizes. Three of the drones available in the evaluation dataset are:

- A DJI S1000 octocopter,
- A DJI Mavic pro quadcopter,
- A DJI spark quadcopter.

The deep neural network associates to each of the analysed input buffers a confidence index. For a confidence index of 85% or more, the median direction of arrival estimation error has been computed for each drone type. The median localization error (azimuth and elevation) for the DJI S1000, Mavic pro and Spark drones is respectively 4.8, 5.2 and 8.2 degrees. The detection range has also been evaluated for each drone model, leading to respectively 180, 120 and 90 meters. As expected, the larger the drone, the better the detection range is. The correct identification rate is of the order of magnitude of 60% for the DJI S1000 drone when it flies at a range of 150 to 180m (above 90% for shorter range <120m).

2.2 Data fusion process

For each acoustic array deployed on the field, the beamlearning deep neural network handles the detection, localization and identification process every 20 milliseconds. The process result is then transmitted to the data fusion computer within a standard TCP-IP network. Each sensor transmits frames based on a proprietary NMEA sentence protocol. The transmitted NMEA message is called a PABET (Proprietary Acoustic Bearing and Elevation to Target) message and is generated by acoustic sensors deployed in the field when an estimate of the direction of arrival of the sound wave generated by the source of interest is available. It contains information on the position of the sensor that transmitted the message (latitude, longitude in degrees/minutes and altitude in meters), the estimated azimuth and elevation of the detected source, the identifier associated with the estimated drone model (integer from 0 to 5 as 5 drones models were available in the learning database) and the associated confidence index, as well as the distance separating the drone and the antenna. Initially, the distance separating the antenna and the drone is set to 1000 meters, and is then updated when the fusion software is able to estimate the geographical position of the detected sound source.

The data fusion process is programmed as a background task, which is triggered at a regular time interval of 200 milliseconds, so that information on the threat's position can be sent to the camera within a reasonable time. Indeed, if the information transmission delay is too slow, there is a risk that, at high speed, the target will no longer be visible in the camera's field of view. Hence, every 200 milliseconds, the data fusion process performs several checks. First, all the messages received from the sensors in the field are analyzed. When several messages have been received from a single sensor, a first filtering stage, based on a particle filter [7], smoothes the received estimated azimuth and elevation values. Furthermore, the information concerning the drone identifier is integrated. The smoothed azimuth and elevation values are then used to calculate the position of the localized source.

When more than two sensors transmit messages, the fusion initially uses the azimuth information estimated by each antenna to determine an initial rough estimate of the source's position in the ENU (East - North - Up) reference frame associated with the reference sensor defined by the fusion program every 200 milliseconds. This reference sensor may be different for each fusion iteration, and is usually the sensor that provided the first received message. Changing the reference sensor to generate the ENU reference frame has no effect, since at the end of the fusion process, the estimated position of the detected object is recalculated and transmitted in the geographic reference frame (latitude, longitude, altitude).

When two or more sensors transmit valid PABET frames, data fusion estimates the position of the barycenter G of the points of intersection between each LOB (Line Of Bearing), starting from the position of each sensor, and from the direction of the estimated azimuth at that sensor. This first calculation is performed in a two-dimensional plane (East - North). In this process, the information transmitted by each antenna is merged to validate the consistency of the azimuth estimates transmitted by each sensor. To do this, the fusion checks that the scalar product between the vectors starting from position M_i of the i -th sensor and the calculated point of intersection G_{ij} and the director vector of the LOB associated with the i -th sensor is positive. The fusion also checks that the distance separating the intersection point and each sensor is less than the individual range limit for each antenna, set at 500 meters.

Once the center of gravity G of the LOB intersection points G_{ij} has been estimated, the altitude h of the UAV is calculated and averaged for all the elevation (φ_i) and altitude h_i information from each sensor that has transmitted valid information.

$$h = \frac{1}{N} \cdot \sum_{i=1}^N (h_i + \|\overrightarrow{M_i G}\| \cdot \tan(\varphi_i)) \quad (1)$$

The drone identification process consists in analysing the N received PABET messages, in which a parameter called target ID (ID_n) is available. The ID_n value associated to the n -th message is an integer value ranging from 0 to 5, representing the set of UAVs that have been included in the deep neural network learning database. Each integer value represents a type of drone:

- 0: unknown drone
- 1: DJI S1000
- DJI Phantom
- DJI Mavic Pro
- DJI Mavic Air
- DJI Spark

Similarly, a confidence index (∂_n) varying between 0 and 1 is associated to each of the N messages. The fusion identifies the detected drone ID_{final} by determining the most represented identifier in the set of PABET messages received and used to determine the position of the detected drone. This is done by summing the confidence index for each of the N ID_n PABET messages with the ID_n value as their identifier.

$$ID_{final} = ID \left(\max_n \left(\sum_{i=1}^N ID_n \partial_n \right) \right) \quad (2)$$

3. Final tests

3.1 Field experiment setup

The data fusion process was tested and validated in real time during field trials organized in May 2022. The fusion server was able to record and analyze in real time the PABET frames transmitted by the five acoustic antennas deployed in the field (Fig. 2), at a rate of one message transmitted every 20 to 30 milliseconds by each antenna.

Data fusion takes place every 200 milliseconds, estimating the position of the detected object in the geographical reference frame from all the messages received during the last 200 milliseconds. Once the position has been estimated, the drone's estimated trajectory is smoothed using particle filtering, which automatically resets itself if too much time has passed between two estimates.

The filtered position is then transmitted to the camera, whose motorized pan&tilt unit is controlled in azimuth and elevation to observe the object on the image. All messages received and transmitted by the fusion server from each sensor are saved in a csv file. The result of each data fusion estimation is compared with information on the drone's position measured with the RTK-GPS in order to evaluate the performance of the complete process. The estimation error is displayed in an ENU frame with the UAV's position at its center. The measured error thus represents the absolute distance separating the drone's position from its estimate.

In the final trials, 14 flights were recorded using RTK-GPS. The total duration of the flight recordings was 105 minutes and 21 seconds. During these recordings, the data fusion transmitted drone position and identification information during 95 minutes and 29 seconds. The data fusion was implemented autonomously, automatically alternating between tracking phases and "radio silence" phases when the fusion considers that no drone is present (for instance when the distance separating the drone from the antennas was greater than the detection distance). The data presented here includes flight data from three UAV models: Phantom, S1000 and Inspire UAVs. The first two were part of the training dataset, the last UAV was never tested before this field experiment.

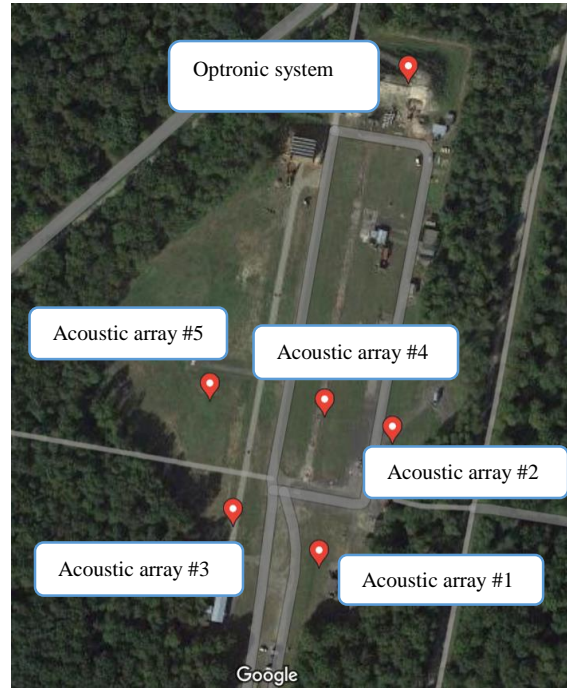


Figure 2: Position of the 5 acoustic antennas and the camera deployed on the field test ground, google maps type of view.

3.2 Results

For all the flights recorded, the errors observed on the X, Y and Z axes (in the East - North - Up reference frame) averaged less than 10 meters, indicating that the drone's location after data fusion is relevant and sufficiently accurate to enable the drone to be observed in the field of view of the optics associated with the optronic system (Fig. 3, Fig. 4). Similarly, the median radial error of localization (distance between the estimated position and the actual position) is equal to 10.7 meters over all 14 trials. The mean standard deviation of the radial error is 15.6 meters, ranging from 4.5 meters in the best case (flight involving the Phantom drone) to 28.8 meters for the worst case scenario (again a flight involving the Phantom drone).

Focusing on the two flights with extreme performance values, the ideal flight (flight #5 for the Phantom drone) involves drone movements strictly within the group of five distributed acoustic systems. This scenario enables us to assess performance when the drone penetrates the safety perimeter defined by the deployment of the acoustic antennas. The high performance observed in this particular case means that when a non-cooperative drone flies directly over the antennas, tracking is excellent.

During flight #7 for the Phantom drone, with the lowest performance levels, it is observed that the greatest localization errors occur during phases when the drone's flying height is very low (less than 5 meters), and therefore in scenarios that are hardly representative of drone use. What's more, these errors are very patchy, appearing only over very short observation periods. By adding a time delay to the data merge to ensure that a sufficient number of coherent merges are performed (e.g. 5 merges in less than 2 seconds), these outliers can be easily removed. What's more, transmitting this information to the optical system also means that no false alarm is raised in this case, since the drone is not present in the image when the error is so large.

The average error is of the same order of magnitude for each drone and flight scenario which shows the relevance of the proposed technology and data fusion process.

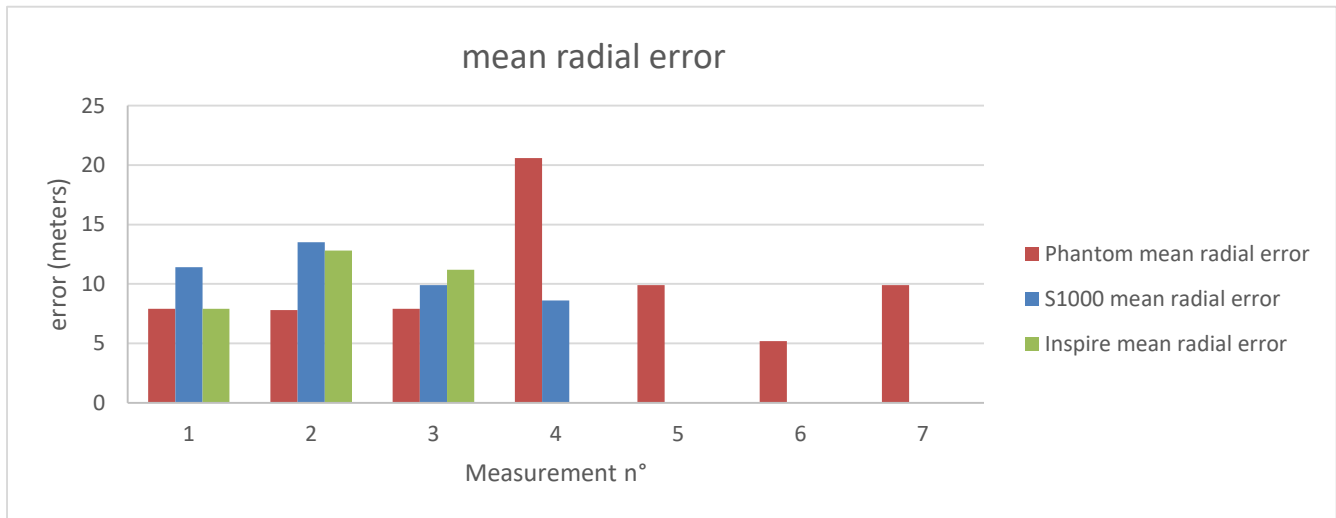


Figure 3: Mean radial error for each model of drone used during the field experiment and for each flight (7 flights with Phantom, 4 flights with S1000 and 3 flights with Inspire).

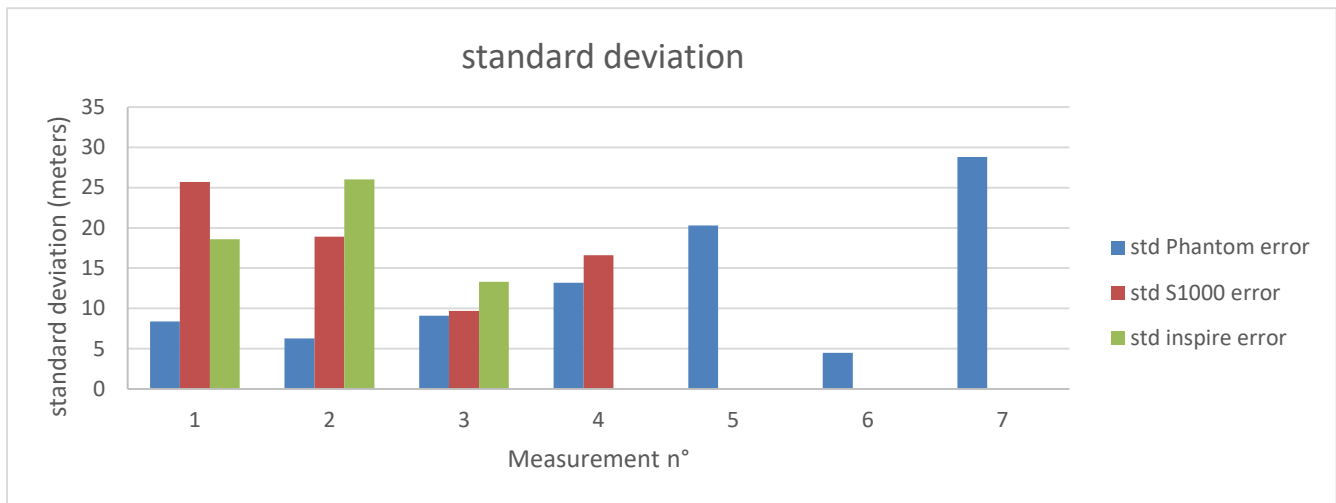


Figure 4: Standard deviation measured for each model of drone used during the field experiment and for each flight.

Acoustic data fusion thus provides a reliable and fast estimate of the position of a drone attempting to intrude on a sensitive site. More, it also makes drone identification using acoustics alone more reliable. Analysis of the data fusion data obtained for the 3 UAVs used in the tests shows that when the UAV is known in the database, the recognition rate is significantly increased. When flying the DJI S1000 and Phantom UAVs, which are included in the database, the identification rate was measured at 99.2% and 95.8% respectively after data fusion. As for the Inspire drone, its non-integration in the learning process meant that it could not be recognized, and the identification phase tended to classify it as a phantom (53%) or Mavic Pro (45%), whose acoustic signatures are closer to those of the Inspire. It's worth noting that, despite the fact that it's impossible to identify the inspire drone, the detection and localization phases don't show any noticeable drop in performance.

4. Discussion

The proposed methodology for drone intrusion detection shows satisfactory performance in terms of tracking of a single drone flying over an area surveyed with a network of acoustic arrays. For future developments, we are studying the integration of two techniques that aim at enhancing the detection range of the acoustic antennas that use the beamlearning algorithms and aim at allowing the tracking of multiple targets entering the surveyed area. A first step will be to integrate some beamforming capabilities to the detection algorithm, by using for instance standard array signal processing techniques like Delay-and-Sum or differential beamformers [8] in order to focus the acoustic array in privileged directions. In initial tests, the integration of Delay&Sum allowed to double the detection range for each of the three drone involved in the field experiment presented in this paper, by the cost of increasing the false alarm rate as no identification process is included in the process. Therefore, a combination of the beamforming and beamlearning algorithms should allow for better detection range performance.

In a second step, we plan to integrate some joint probabilistic data association (JPDA) filter [9] usually used for multiple target tracking. This approach should lead to robust tracking of one or more UAVs entering the monitored area, leading to better comprehension of the flight patterns of each individual UAV while the actual system only allows for the tracking of a single intrusion.

Finally, as is the case for checking the validity of input data, a consistency check of fusion output data will ensure greater localization reliability when fusion data is transmitted only transiently.

REFERENCES

- 1 <https://www.reuters.com/world/europe/three-rescue-workers-killed-russian-drone-attack-ukraines-kharkiv-officials-say-2024-04-03/>
- 2 <https://www.gov.uk/government/news/new-prison-no-fly-zones-for-drug-delivering-drones>
- 3 https://www.lemonde.fr/societe/article/2023/10/09/le-fleau-de-la-livraison-par-drone-dans-les-prisons_6193221_3224.html
- 4 <https://ukdefencejournal.org.uk/drone-threat-concerns-as-25-spotted-over-nuclear-facilities/>
- 5 Hengy, S., et al., DeepDynamics project: Deep-learning for the real-time multimodal localization and identification of small UAVs. NATO SET-312 proceedings Symposium on "Distributed Multi-spectral/statics sensing", Bled, SI, 23-24 May 2022, 2022
- 6 Pujol, H., Bavu, E., & Garcia, A. BeamLearning: An end-to-end deep learning approach for the angular localization of sound sources using raw multichannel acoustic pressure data. *The Journal of the Acoustical Society of America*, **149** (6), 4248-4263, (2021)
- 7 Orlande, H., et al., *Kalman and Particle filters*. METTI V-Thermal Measurements and Inverse Techniques (2011)
- 8 Chen, J., Benesty, J., & Pan, C. On the design and implementation of linear differential microphone arrays. *The Journal of the Acoustical Society of America*, **136**(6), 3097-3113, (2014).
- 9 Svensson, L., Svensson, D., Guerriero, M., & Willett, P. Set JPDA filter for multitarget tracking. *IEEE Transactions on Signal Processing*, **59**(10), 4677-4691, (2011).