BeBeC-2026-D03

Acoustic tracking of small UAVs: An experimental evaluation using microphone arrays and the Acoular framework

Paweł Baran[1] and Benjamin Janek Wagner[2] and Thijs Hampsink [3]
[1]Warsaw University of Technology
pl. Politechniki 1, 00-661, Warsaw, Poland
[2]Technische Universität Berlin
[3]TU Delft

10th May 2026

Abstract

The increasing proliferation of small, cost-effective Unmanned Aerial Vehicles (UAVs) across commercial, industrial, and military sectors has created a growing need for reliable detection and tracking systems. As these drones become more advanced—often utilizing fiber-optic tethers or flying autonomously, which renders them virtually invisible to conventional radar and radio-frequency (RF) detection - acoustic localization has emerged as a crucial surveillance method.

This project investigates the feasibility of localizing sub-250g multirotor drones solely on the acoustic noise generated by their propulsion systems. The research methodology combined numerical simulations with empirical measurements conducted in an anechoic chamber. Utilizing the Python-based Acoular library, various microphone array geometries and beamforming algorithms were evaluated for their spatial resolution and computational efficiency. The study successfully reconstructed flight trajectories from acoustic data, which were subsequently compared against optically recorded ground truth footage.

1 INTRODUCTION

This study investigates the possibilities of tracking moving objects using microphone arrays, established beamforming methods, and the Acoular Python library. Electric multicopter drones were selected as the specific sound sources for tracking. This decision was driven by their surging civilian popularity, expanding industrial applications, and, crucially, their transformative impact on modern security and defence paradigms.

In recent military conflicts, the tactical deployment of commercial-off-the-shelf (COTS) and custom-built Unmanned Aerial Vehicles (UAVs) has increased exponentially. These small platforms have introduced a highly asymmetric threat to the battlefield; high-value military assets, such as armoured vehicles worth millions of dollars, can be effectively disabled by inexpensive drones carrying small payloads. Furthermore, the operational stealth of these systems is rapidly evolving. The emergence of fiber-optic-tethered drones, along with fully autonomous flight capabilities, essentially eliminates their radio frequency (RF) footprint. Combined with their minimal radar cross-section (RCS), these advancements render them virtually undetectable by traditional electronic warfare (EW) systems or conventional radar. Consequently, the acoustic signature of their rapidly spinning propellers remains one of the few reliable physical phenomena that can be exploited for early detection and localization, which serves as the core motivation for this research.

This project aimed to evaluate the detection and tracking capabilities of two sub-250g quadcopters using well established microphone array layout and beamforming algorithms. The expected outcome was the generation of tracking diagrams mapped onto a specified area. All measurements were conducted in an anechoic chamber to simulate a near-reverberant-free outdoor environment. A key requirement of the study was the implementation of the Acoular library, accounting for both its advantages and technical constraints. Before performing the measurements, some extensive preliminary simulations were done, to test multiple array geometries and various beamforming algorithms on synthesized drone signals in a test environment set up in Acoular library which provides various beamforming algorithms.

1.1 Theoretical background

All of the three used algorithms can be used to track moving objects with unknown trajectories.

The core of the localization framework is based on delay-and-sum (DAS) beamforming, which aligns signals from multiple microphones to steer the array's sensitivity toward specific spatial coordinates. In this study, three distinct implementations within the Acoular library were evaluated:

- Time Domain Beamforming (BeamformerTimeSq) - utilizes a squaring operation on the output signal, which effectively enhances the signal-to-noise ratio (SNR) and results in clearer acoustic maps compared to basic frequency domain methods.
- Frequency Domain Beamforming (BeamformerBase) - provides a standard cross-spectral matrix-based approach for identifying sound pressure levels (SPL) across a defined grid.
- CLEAN-SC Deconvolution (BeamformerCleansc) - an advanced frequency-domain algorithm used to mitigate the effects of side-lobes and spatial artifacts. While it offers

high precision in source localization, it introduces significant computational overhead, reflected in a lower Real-Time Factor (RTF).

As a computationally efficient alternative for real-time applications, the Generalized Cross-Correlation with Phase Transform (GCC-PHAT) was implemented. Unlike the "acoustic camera" grid-based approach, GCC-PHAT estimates the Direction of Arrival (DoA) in terms of azimuth and elevation angles, offering incomparably greater computation speeds, even on low-end hardware.

The spatial performance of the tracking system is fundamentally defined by the physical arrangement of the microphones. The angular resolution $\Delta\theta$ of the array is inversely proportional to its aperture D and the frequency of the sound source f. It can be approximated using the following relationship:

$$\Delta\theta \approx \frac{c}{f \cdot D}$$

where c is the speed of sound. In this study, an array with an aperture of 1.5m was utilized to ensure high resolution even at lower frequencies, which carry the bulk of the drone's acoustic energy. To ensure unambiguous localization, the effect of spatial aliasing must be considered. For a regular grid, the maximum distance between sensors d_{max} to avoid aliasing is governed by the Nyquist criterion in the spatial domain:

$$d_{max} \leq \frac{\lambda}{2} = \frac{c}{2f_{max}}$$

However, to mitigate the side-lobe levels and artifacts inherent in regular geometries like rectangular or concentric grids, a non-uniform, aperiodic distribution was chosen. By employing the Vogel spiral geometry, the array suppresses spatial aliasing even when the inter-element spacing exceeds the theoretical limit for the highest analysed frequencies, effectively broadening the operational bandwidth of the system [5].

1.2 Methodology

For the scope of this project, the Acoular framework [1] was used to implement delay-and-sum beamforming algorithms (BeamformerTimeSq, BeamformerBase, and BeamformerCleansc). The data processing parameters used for beamforming are summarized in Table 1. The focus grid was mapped to the room dimensions with a slight overlap to ensure symmetry with respect to the array position. A standard Hann window with a length of 4096 samples was used for the FFT in order to maintain a quasi-stationary state within each window. An evaluation frequency of 2 kHz was selected, as most of the acoustic energy emitted by the drone lies below 4 kHz. Furthermore, the objective of this study is source localization rather than an accurate determination of the sound pressure level.

The tracking codebase consists of the DBSCAN clustering algorithm and Kalman filtering. The DBSCAN algorithm was selected for clustering, as it has shown good performance in acoustic applications in previous research as a density-based algorithm [6]. In this paper, the clustering algorithm is used to group data points associated with a single drone. A key advantage of DBSCAN is that it does not require the number of clusters as an input parameter, which makes it suitable for tracking multiple moving sources in the presence or absence of background noise.

Table 1: Data processing parameters for beamforming

Sampling rate	51200 Hz
Focus grid [m]($l_x \times l_y$)	9,4 × 14,2
Focus grid resolution	0,25 m
FFT window	4096 samples, Hanning window
Frequency band	2000 Hz, third octave

Prior to clustering, data points with sound pressure levels 10 dB lower than the maximum value per frame were filtered out to reduce input data and improve clustering results. In this project, DBSCAN was implemented using Scikit-learn Python library [4]. The algroithm requires the input parameters ε and *min_samples*, which were optimised in prior simulations. For beamforming algorithms without deconvolution, *min_samples* was set to 5, while for those with deconvolution, it was set to 2 due to the reduced number of remaining data points.

For the ε parameter, an adaptive approach was chosen, as a fixed ε value led to unstable clustering results due to variations in the number of points and the shape of the point clouds over time. The parameter ε was determined by identifying the nearest neighboring points and assigning the 85th percentile of these distances as ε. This method yielded stable results in simulation tests.

The resulting point clouds were converted into single representative points by calculating the energy center of each cluster. To obtain dynamic trajectories, each source was assigned to a path using a confirmation logic that only recognizes new trajectories if they persist for at least five time steps. Trajectories that are not observed for five consecutive time steps are terminated. This procedure ensures that short-lived noise sources are effectively filtered out.

A Kalman filter was implemented to stabilize the estimated trajectory and reduce the impact of measurement noise, using the FilterPy python library [2]. The setup was evaluated during simulations and the parameters were selected to track a fast moving drone while providing satisfactory results. For the Kalman filter execution, three time steps were considered for state estimation. The Parameters P, R and Q were defined as scaled Identity Matrices, with scaling factors chosen as follows: $P = 1000$, $R = 5$ and $Q = 1$.

2 SIMULATION ENVIRONMENT

Before physical testing, extensive numerical simulations were conducted using synthesized drone signals. Utilizing various Acoular processing techniques, we characterized the acoustic footprint of a drone by defining it as four dipole point sources. These sources generated harmonics derived from a broadband noise, correlating to a propeller rotational speed of approximately 8,000–9,000 RPM. The simulation model accounted for ground reflections and incorporated several flight paths, including straight-line passes, loitering patterns, figure-eights, and 'attack-and-egress' manoeuvres. Furthermore, multiple microphone array geometries were evaluated, with a specific focus on spatial aliasing constraints and the analysis of Point Spread Functions (PSF). This allowed us to identify layouts providing optimal angular resolution while minimizing the number of microphones to reduce computational overhead.

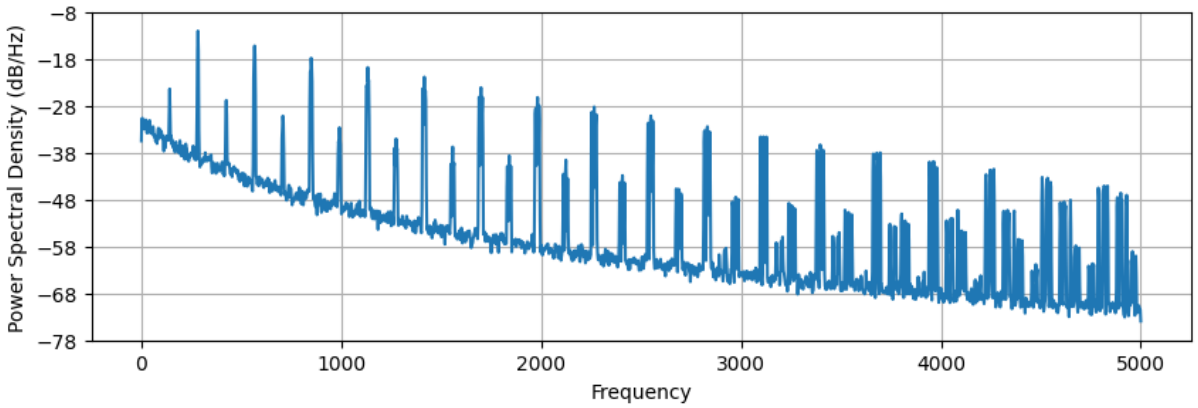

Figure 1: Spectral Density of the sound generated by the modeled drone

To localize the sound sources, several beamforming algorithms implemented in Acoular were benchmarked. The most effective methods selected for further investigation were the standard Delay-and-Sum beamformers in both time and frequency domains (*BeamformerTime* and *BeamformerBase*), as well as the CLEAN-SC deconvolution algorithm, which was employed to enhance map clarity and source definition.

Based on these extensive simulations, several key conclusions were drawn:

- A higher microphone count allows for a larger array aperture, significantly improving angular resolution; however, this substantially increases computational intensity.
- Microphones should be distributed non-uniformly and aperiodically to mitigate spatial aliasing and artifacts. Although regular geometries (such as rectangular grids or concentric circles) are easier to design in CAD, they tend to perform poorly in beamforming applications due to high side-lobe levels.
- While higher sampling rates allow for a broader analysis bandwidth, the majority of acoustic energy, even for small quadcopters, remains well below 8 kHz. Thus, lowering the sampling frequency can optimize computational performance without significant data loss, especially since analysing frequencies beyond the spatial aliasing limit is counterproductive.

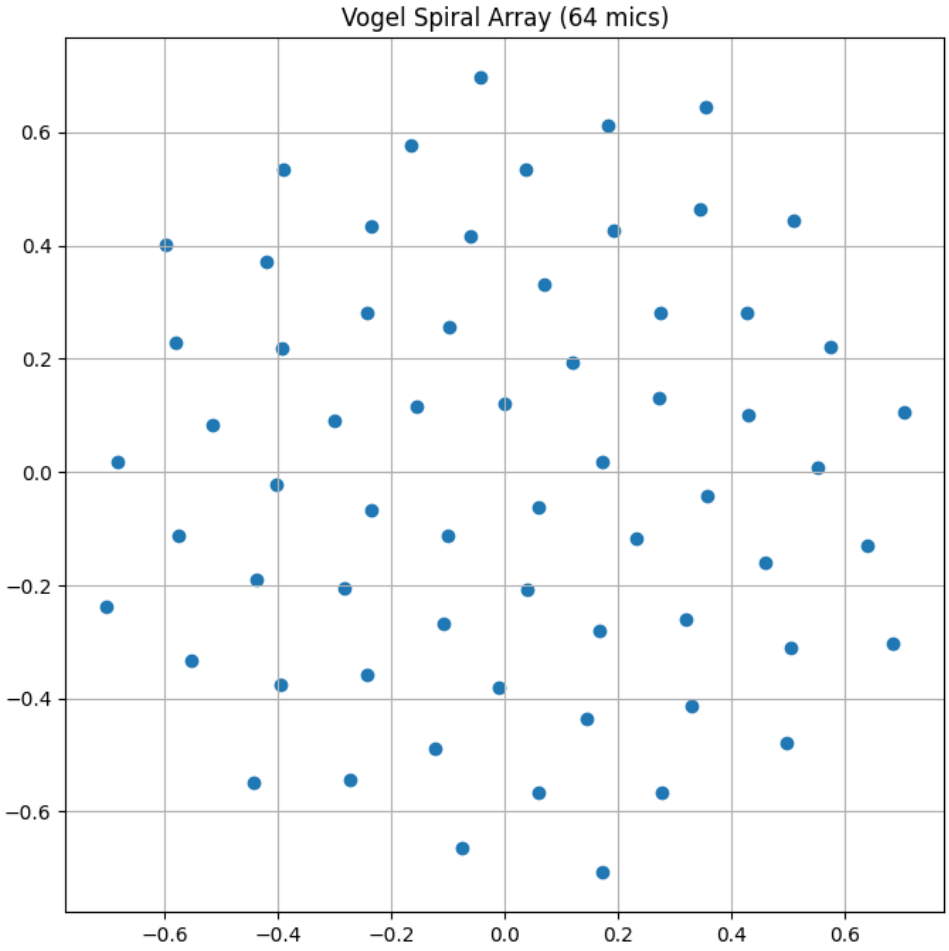

Figure 2: Vogel spiral array geometry used during physical experiments

3 EXPERIMENTAL SETUP

All actual measurements were conducted in the anechoic chamber in the acoustic test facilities of TU Berlin. The room dimensions were measured to be 13,5 m x 8,5 m x 8 m. The array used was a vogel spiral array with 64 microphones and an aperture of 1,5 m. It was used for drone tracking before with satisfactory results for tracking in the azimuth plane et al. [3] and was additionally tested in prior simulations. The aluminium mounting plate with the array was positioned horizontally on four wooden blocks with a layer of insulation to decouple the array from the wooden blocks. Additionally, sheets of absorption foam were positioned around the array to mitigate unwanted reflections. At the centre of the array, a camera was positioned for validation purposes. The whole setup is shown in Figure 3 (left).

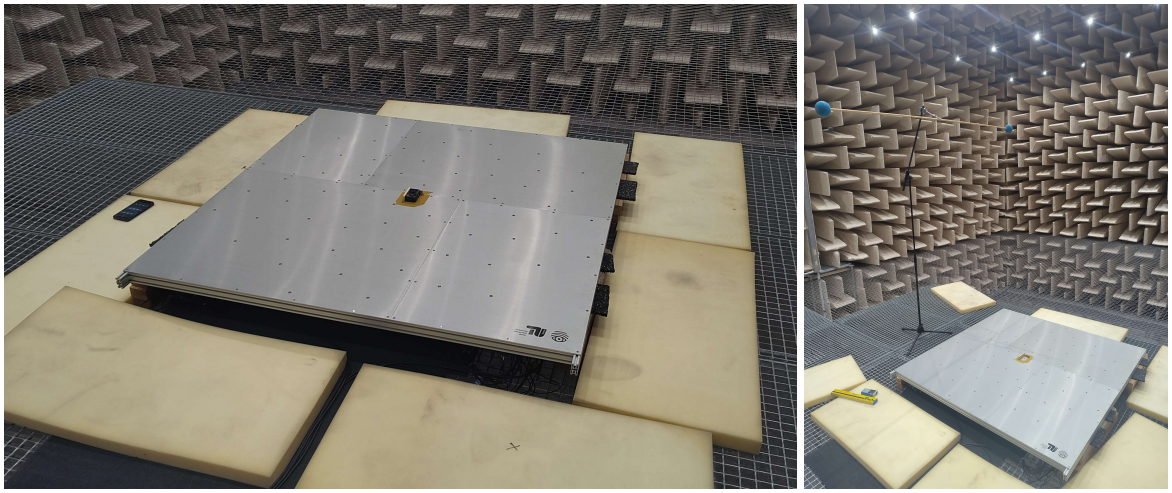

Figure 3: Setup of the microphone array with the camera in the centre (left) and setup for camera calibration (right)

The used drones were a DJI Mavic Mini and a DJI Tello seen in figure 5. Both drones have the same amount of rotors but have different speeds and weights. All the equipment will be listed in the following:

- Microphone Array with mounting plate
- 64 GRAS 40 PK microphones
- An anechoic chamber with dimensions 13,5 m x 8,5 m x 8 m
- DJI Mavic Mini
- DJI Tello
- DJI Osmo Action camera (1st Gen)
- Two blue balls on a stick 2m apart
- Weather station

- Typhoon recording system with 96ch soundcard
- Cables, extension boxes, mounts, tools, and radios
- Calibrator (94 dB at 1 kHz)

The calibration of the camera was performed with two blue balls on a stick. The stick has a defined length of 2 m and two different calibration positions were recorded as can be seen in figure 3 (right). Additionally, an audio source was positioned at the calibration positions to calibrate the acoustic tracking from audio to the camera footage and de-warp the videos. All the 64 microphones were calibrated using the calibration device.

Four different drone trajectories were measured with both drones, executing every measurement twice to cover for potential error. After both the acoustic and visual measurement were started, every run included a clap before the flyby, to synchronise these measurements. The measured trajectories were a straight trajectory, a deploy and depart trajectory, a circling trajectory, and the figure eight as symbolised in Figure 4. All measurements were conducted at roughly the same height.

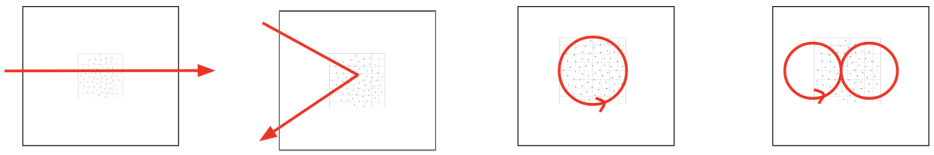

Figure 4: Representations of the measured flight trajectories from left to right: straight, deploy and depart, circling, and figure eight

Additionally, the straight and circling trajectory were measured with a stationary disturbance sound source with the DJI Mavic Mini.

Figure 5: DJI Mavic Mini (left) and DJI Tello (right) drones used for the experiments

4 RESULTS

From the outset of the project, the desired results were clearly defined: the primary objective was to accurately estimate the drone's location and track its trajectory in the time domain. While the obtained results are highly satisfactory, their representation in a static written document is inherently challenging due to the dynamic nature of multidimensional data, which evolves simultaneously in both the spatial and temporal domains. Consequently, it is highly recommended to view the supplementary video material[1], where several flyovers are presented with overlaid beamforming maps to visualize the acquired data in the most effective manner.

Nevertheless, this section provides a detailed presentation of the most significant results obtained, including a comparative analysis of various algorithms and flight trajectories. Figure 6 illustrates a comparison of acoustic maps generated by different beamforming algorithms at the same temporal instance of a single flyover. To ensure a valid comparison, all maps have been normalized to a consistent sound pressure level range, specifically from 0 to 50 dB SPL.

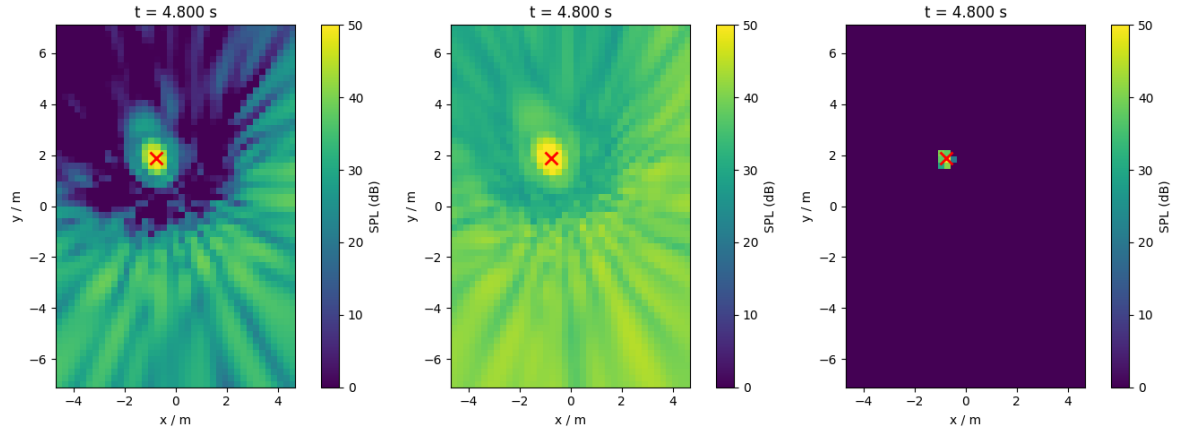

Figure 6: Comparison of maps from 3 different beamforming algorithms inside Acoular. From the left: TimeSQ, Base, CleanSC

As is evident, the squaring operation in the initial algorithms enhances the signal-to-noise ratio (SNR), resulting in clearer acoustic maps while maintaining the same spatial resolution for drone localization. Due to the inherent physical characteristics of the microphone array and the nature of beamforming, prominent artifacts are visible surrounding the identified source. In contrast, by employing the deconvolution techniques available in the CLEAN-SC beamformer, the map is clearer, allowing for a highly precise determination of the object's coordinates. However, this increased precision entails a significant computational cost. Table 2 presents a comparison of the Real Time Factor (RTF) calculated for all three methods as follows:

$$RTF = \frac{t_m}{t_c}$$

where t_m denotes the total measurement duration and t_c represents the time required for computation. It should be noted that these measurements account strictly for the core beamforming

[1] https://youtu.be/okVygPVcjbI

processing time, excluding data acquisition, I/O operations, and the rendering of the acoustic maps.

	TimeSQ	**Base**	**CleanSC**
RTF	0,98	0,53	0,15

Table 2: Real time factors for all 3 beamformers, for 64 microphones and 51200 Hz sampling rate, for 2262 point grid

The table clearly illustrates the disparity in computational workload between the evaluated algorithms, particularly when a domain transformation and deconvolution operations are required. It is important to note that the processing was performed on all 64 channels with a sampling frequency of 51,200 Hz, dictated by the parameters of the recording station employed. As observed in the spectral analysis, the majority of the drone's acoustic energy is concentrated in frequencies well below 4 kHz. Consequently, the sampling rate in a practical, real-time system could be significantly reduced, thereby substantially accelerating the computation process.

The calculations were executed on a modern workstation equipped with an AMD Ryzen 9 9950X3D CPU and 64 GB of DDR5 RAM. Processing a 28-second audio clip required between 30 seconds and 3 minutes, depending on the algorithm. However, it should be emphasized that these computations were performed using a single-threaded Python environment, which is not optimized for maximum execution speed. Performance could undoubtedly be enhanced by optimizing the workflow, utilizing multi-threading, or migrating the core computational kernels to a lower-level language or dedicated hardware.

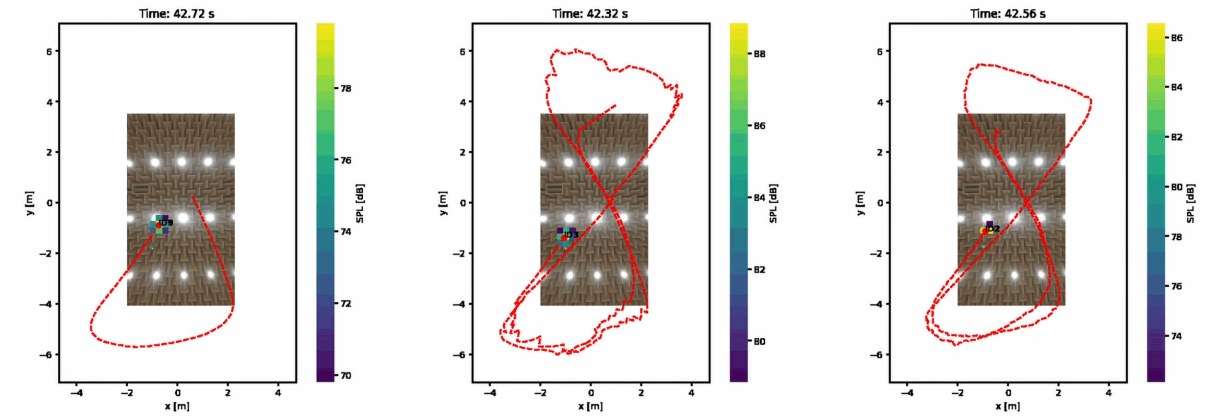

Figure 7: Trajectories traced for every tested algorithm. From the left: TimeSQ, Base, CleanSC

Figure 7 presents the trajectories acquired via beamforming, overlaid on video footage captured during the experiments. Throughout the testing phase, it became evident that fine-tuning the clustering and Kalman filter parameters is highly complex; achieving consistent tracking without signal dropouts or ghosting (sudden multiplication of traces) requires significant calibration. Furthermore, the optimal parameter set varied across different beamforming algorithms. Nevertheless, every evaluated method appears to provide sufficient tracking precision, provided that the parameters are meticulously selected. In Figure 8 the direct comparison of the calculated trajectories from the different beamforming algorithms can be seen. The comparison

indicates that there is no substantial visual difference in tracking performance between the various beamformers towards the center of the microphone array. In this area the difference between the methods is below 0,2 m. In further distance, more than 5 m away from the array, it rises to 0,8 m. Consequently, the high-precision maps generated by CLEAN-SC may not be strictly necessary, as the standard Delay-and-Sum beamformer achieves good precision in close distance to the array and adequate position for higher distances for this application without the associated computational overhead.

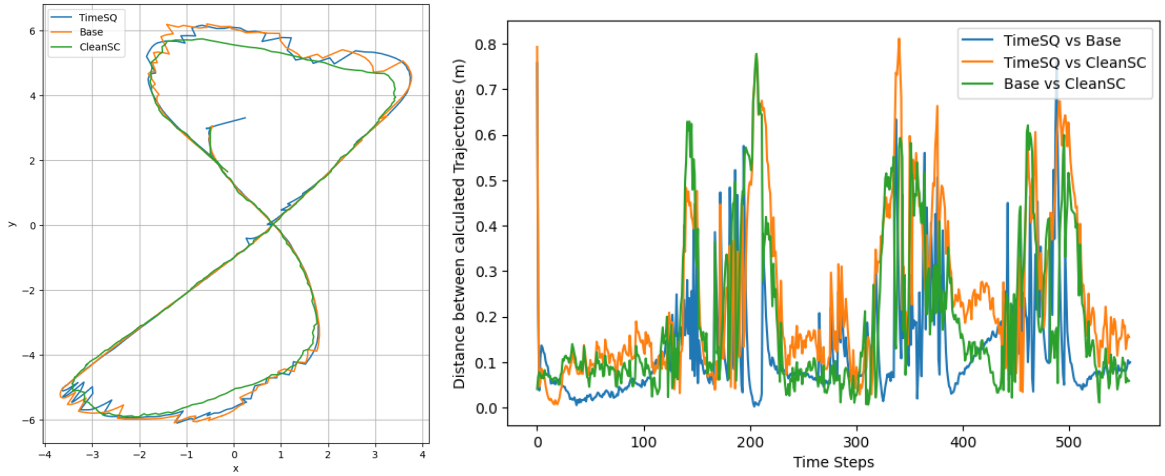

Figure 8: direct comparison of the trajectories obtained with TimeSQ, Base and CleanSC beamforming (left) Distance between calculated trajectories between differnt beamforming algorithms (right)

Regarding the primary use case — a cost-effective, distributed sensor network — generating high-resolution maps from every array may be impractical due to computational constraints and the difficulty of reporting such dense data to a central command. To address this, we experimented with the Generalized Cross-Correlation with Phase Transform (GCC-PHAT), implemented independently of the Acoular library. Using custom functions, we were able to determine the Direction of Arrival (DoA) of the drone in both the azimuth and elevation planes. This approach corresponds more closely to a radar-like representation, facilitating the guidance of countermeasures toward the detected target.

By deploying multiple distributed arrays, it would be possible to perform triangulation to estimate the drone's exact coordinates within a networked environment. The primary advantage of this approach is the significantly reduced computation time. Using this implementation, we achieved an RTF exceeding 10, even with a high microphone count and a sampling frequency far above the minimum requirements. A single frame of this evaluation is shown in Figure 9.

Unfortunately, a rigorous quantitative estimation of the localization error for each algorithm could not be established, primarily due to the lack of precise spatial correlation between the video footage and the acoustic maps. While calibration was performed using four reference points in a plane parallel to the array, this proved insufficient for an accurate de-warping of the footage. Consequently, the perspective remains skewed, and the trajectories on the actual flight plane lack perfect rectilinearity. Furthermore, the reconstruction plane for the beamformers was arbitrarily set at a specific height. However, verification of the actual flight altitude was chal-

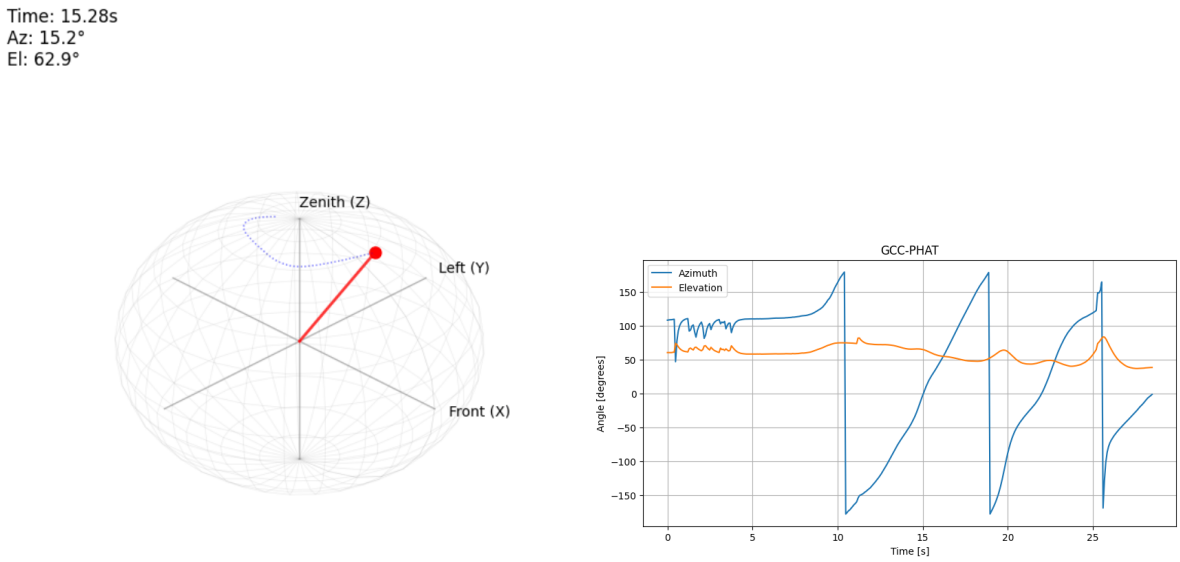

Figure 9: Representation of angle of arrival using GCC-PHAT

lenging, as GNSS signals are unavailable within the anechoic chamber, and the drone exhibited minor altitude fluctuations throughout the duration of the tests, so even laser measurements were burdened by a significant error.

As a result, visible discrepancies in the drone's position appear within the overlaid video frames. While the acoustic and visual localizations align closely near the center of the frame, they diverge as the drone approaches the perimeters of the chamber and the edges of the image. For these reasons, a formal quantitative error analysis was not conducted, as the resulting data would predominantly reflect inaccuracies in the camera-array calibration rather than the inherent performance of the beamforming algorithms themselves.

5 CONCLUSIONS

Both simulations and real-world measurements on acoustic signals using non-trivial microphone array geometries was done. The proposed workflow and measurement methodology turned out to be effective. Consequently, the multiple flight trajectories could be reconstructed using the acoustic measurements and qualitatively compared against visual recordings. All beamforming operations were computed in two dimensions on a plane, roughly corresponding to the estimated flight altitude of the drone. Furthermore, a functional processing pipeline was successfully implemented using the Acoular library and various available beamforming algorithms were benchmarked. Through this process, the software parameters that exert the most significant impact on both computational speed and localization precision were identified.

However, Acoular's inherent design, which essentially enforces an "acoustic camera" paradigm, may not be optimal for applications requiring real-time, wide-area drone tracking. For such scenarios, alternative approaches must be considered. To address this, A separate processing pipeline outside the Acoular environment was experimented with, utilizing the GCC-PHAT algorithm. This alternative approach yields both azimuth and elevation angles toward the

sound source, providing full 3D directional information (Direction of Arrival) without distance estimation, but at an incomparably faster processing speed. For real-world applications, further research into this methodology needs to be done, particularly regarding how the exact 3D coordinates of an aircraft could be determined using multiple distributed arrays and triangulation techniques.

Future development of this research could move toward more realistic experimental settings that extend beyond localization alone. In practical deployments, a monitoring system must answer not only where a drone is, but also what type of drone it is, since the latter is critical for risk assessment and downstream decision-making. Another potential avenue for expansion involves scenarios with multiple drones flying simultaneously. In such cases, spatial responses may interfere, and competing or overlapping peaks can degrade localization performance, especially when sources are close in angle or distance. Moreover, the acoustic signatures of different drones can partially overlap, making separation and attribution more challenging. Further optimization could focus on real-time operation under practical constraints. The trade-off between latency and accuracy should be quantified by evaluating different window sizes, update rates, and beamforming/tracking configurations, as well as by measuring end-to-end delay on target hardware. In parallel, designing computationally efficient processing blocks (e.g., streamlined feature extraction, reduced search spaces, and lightweight tracking updates) and developing robust automatic parameter selection would significantly reduce the need for manual tuning. Ultimately, these theoretical improvements would aim to produce a system capable of running reliably in an online deployment while remaining stable across changing acoustic conditions and operational environments.

REFERENCES

[1] "Acoular - acoustic testing and source mapping software." URL `https://www.acoular.org/`.

[2] "Filterpy." URL `https://filterpy.readthedocs.io/`.

[3] G. Herold, A. Kujawski, C. Strümpfel, S. Huschbeck, M. Ujit de Haag, and E. Sarradj. "Detection and separate tracking of swarm quadcopter drones using microphone array measurements." *BeBeC-2020-D05*, 2020.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, 12, 2825–2830, 2011.

[5] E. Sarradj. "A generic approach to synthesize optimal array microphone arrangements." In *6th Berlin Beamforming Conference*. Berlin, Germany, 2016. BeBeC-2016-S4.

[6] H. Yan, T. Chen, P. Wang, L. Zhang, R. Cheng, and Y. Bai. "A direction-of-arrival estimation algorithm based on compressed sensing and density-based spatial clustering and its application in signal processing of mems vector hydrophone." *Sensors*, 21(6), 2021. ISSN 1424-8220. doi:10.3390/s21062191.