BATradar: Acoustic Drone Localization Engine

Loris Fabian Klindworth, Colin Wai-Loen Berendt
School of Computer Science SCS, University of St. Gallen HSG
Rosenbergstrasse 30, 9000, St. Gallen, Switzerland
E-mail: [lorisfabian.klindworth, colinwai-loen.berendt]@student.unisg.ch

***Abstract*—Amidst the rising popularity of unmanned aerial vehicles (UAVs) in various areas such as recreational videography, but also surveillance and warfare, the need for advanced techniques to detect and localize drones has become increasingly crucial. Conventional monitoring methods, such as radio frequency (RF) tracking, often face significant challenges as they rely on active signal transmission, which may be absent in autonomous drone operations. This creates a critical demand for passive, non-line-of-sight identification methods. In this paper, a localization program for drones is introduced that works solely based on acoustic signals. Our system, BATradar, utilizes a specialized three-step pipeline comprising Time Difference of Arrival (TDOA), Direction of Arrival (DOA), and multilateration algorithms. By leveraging Recurrent Neural Networks (RNN) and Mel-frequency Cepstral Coefficients (MFCCs), the engine is designed to reliably identify unique acoustic signatures even in environments with low signal-to-noise ratios. Our experimental evaluation demonstrates that a Recurrent Neural Network (RNN) architecture achieves a detection accuracy of over 99% in quiet environments and remains robust under low signal-to-noise conditions. However, while the system shows high precision for distinct sounds, the results for monotone drone signatures highlight physical limitations of TDOA-based localization with low-budget equipment in narrowband scenarios, suggesting a need for a combination of RF and acoustic localization in future applications.**

I. Introduction

The idea of Unmanned Aerial Vehicles (UAVs) dates back to the early 20th century, with early pioneers such as Nikola Tesla envisioning remotely controlled devices as early as 1898, followed by the development of the "Kettering Bug" during World War I [1]. In recent years, drones have increased in popularity not only in recreational videography but also in areas such as surveillance or warfare. This growth has also impacted the demand for anti-drone systems such as Drone-Detection Systems. As drones become smaller and more agile, traditional monitoring methods often face significant challenges. Radio frequency monitoring relies on active signal transmission, which may not be present in autonomous or "dark" drone operations. This creates a critical need for passive, non-line-of-sight identification methods that can reliably detect targets based on their inherent physical characteristics.

The motivation for this research stems from the increasing security risks posed by unauthorized drone activity in sensitive airspaces, such as airports, industrial complexes, and public events. By focusing on the unique acoustic sound of drones, we investigate the usefulness of a localized sensing engine designed to detect and pinpoint unauthorized aerial activity. Acoustic sensing offers a cost-effective, passive alternative that can operate in diverse environmental conditions where other sensors might fail. This paper explores new technologies and documents the design and implementation of a specialized system utilizing acoustic localization. By bridging the gap between simple detection and precise spatial localization using only microphone arrays, we propose an engine that combines detection and localization solely based on acoustic signals.

The remainder of this paper is structured as follows: Section II reviews related work in acoustic-based drone audition, comparing traditional digital signal processing with modern deep learning architectures as well as other sectors where acoustic localization is utilized. Section III describes the design and implementation in detail, covering the hardware configuration and the three-step pipeline involving Time Difference of Arrival (TDOA), Direction of Arrival (DOA), and multilateration algorithms for the localization step. It also highlights our implementation of a detection system to classify sounds and thus detect drones. Evaluation of the obtained results is provided in Section IV. Finally, Section V concludes the paper with a summary of findings, and conclusions we took from working on this project over the whole semester. It also highlights which limitations or challenges we faced during our implementation of the drone localization engine. Finally, the paper concludes potential future work, such as hybrid RF-acoustic sensing or future interception of unauthorized drones or UAVs.

II. Related Work

Acoustic-based drone audition has, in recent years, gathered significant traction, due to the potential application in critical fields, such as search and rescue operations, surveillance, package delivery, and airspace safety. Current research in this field is mostly divided into two different categories: For one, Drone Detection and Identification, which focuses on recognizing the presence and type of a drone, and sometimes even localizes it with the help of radio frequencies (RF) or other methods. Secondly, Onboard Sound Source Localization, where drones act as mobile sensor platforms to find external sound sources.

Early approaches to drone detection relied on traditional machine learning techniques. Jeon et al. [2] established a foundational baseline in this domain by investigating the performance of Gaussian Mixture Models, Convolutional Neural Networks, and Recurrent Neural Networks and showcased, that deep neural networks outperform GMMs.

Al-Emadi et al. [3], [4] in their 2019 and 2024 studies then highlighted that these earlier approaches relying on Digital Signal Processing (DSP) combined with SVM or KNN required manual feature extraction and were quite sensitive to noise. Then, by utilizing deep learning approaches, specifically RNN and LSTM architectures, they were able to achieve a higher accuracy and better noise immunity as they could capture temporal dependencies of motor sounds. In their 2024 work they classified drones through a fusion of Radio Frequency (RF) signatures and acoustic features building a Fusion RNN architecture that connects RF and acoustic signals which lead to a strong noise immunity.

Additionally, in the domain of onboard sound source localization, several studies have focused on embedding microphone arrays directly onto drones to localize external sound sources, while navigating the challenges of flight-induced noise.

Manamperi et al. [5] investigated Direction of Arrival (DOA) estimation using onboard microphone arrays on drones operating under low signal-to-drone-noise ratio (SdNR) conditions. While their goal originally was to localize other sounds with a drone, their methodology for suppressing ego-noise is very relevant for such kinds of localization. They demonstrate, that the rotor noise dominates the spectrum below 2 kHz and that the microphone array geometry plays an important role in the localization accuracy.

Wang and Cavallaro [6] proposed a two-stage model (TFS-DNN) that integrates neural networks for noise suppression and localization with which they outperformed traditional SRP-DNN methods. However, in extremely low Signal-To-Noise Ratio (SNR) conditions (below –25 db) they noticed a performance degradation.

In a complementary direction, Hoshiba et al. [7] developed a drone-embedded spherical microphone array system (SMAS) specifically designed for outdoor operation. Their compact array was able to achieve a reliable localization even at signal-to-noise ratios as low as –10 dB.

Lastly, Wakabayashi et al. [8] also do localization, but additionally find features specific to each sound source to identify not just single sound sources, but concurrent sources through data association.

For ground based drone localization, Chang et al. [9] proposed a system to localize amateur drones using a Time Difference of Arrival estimation algorithm. While their work shows the potential of TDOA for drone tracking, it relies on highly optimized ground-array geometries and specific statistical priors to handle the narrowband nature of drone signatures.

Overall, the reviewed literature mainly indicates that deep learning-based methods, usually RNNs, achieve the best accuracy and robustness for both detection and localization tasks compared to traditional approaches. Most existing studies, however, address either drone detection or acoustic source localization independently, which is where we identified the gap to first detect but then localize drones solely based on their acoustic signals with accessible hardware while navigating the inherent physical limitations of acoustic localization in dynamic environments.

TABLE I: Comparison of Related Works in Acoustic Drone Audition

Ref.	Focus	Method/Model	Key Contribution
[2]	Detection	GMM, CNN, RNN	Established DL superiority for temporal audio patterns.
[3]	Detection	RNN, LSTM	Captured motor dependencies for higher noise immunity.
[4]	Detection	Fusion RNN	Combined RF and acoustic signals for robust classification.
[9]	Ground Loc.	TDOA (Ground)	Managed multipath effects using statistical priors.
[5]	Onboard Loc.	DOA Estimation	Identified <2kHz rotor noise as primary error source.
[6]	Onboard Loc.	TFS-DNN	Two-stage model for joint denoising and localization.
[7]	Onboard Loc.	Spherical Array	Specialized hardware (SMAS) for low SNR (-10dB).
[8]	Onboard Loc.	MUSIC/Assoc.	Concurrent source identification via feature association.

III. Design and Implementation

This Section describes in detail how the Drone Localization Engine was implemented in our project. While there are existing work experiments with acoustic drone detection and also sound source location estimation, the approach that we used for our Engine was the combination of detection and localization of drones solely based on their acoustic signals. The following subsections detail the hardware configuration, the signal processing pipeline, and the triangulation algorithms used to determine the drone's position.

A. Localization

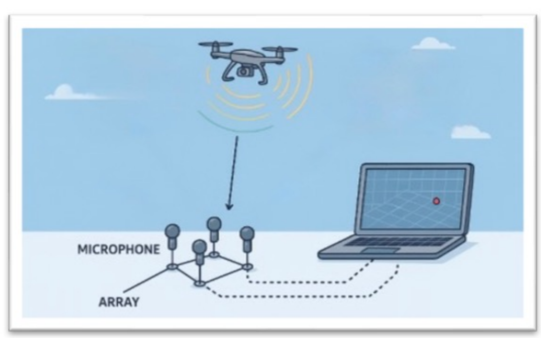

Fig. 1: Architectural overview of the Drone Localization Engine.

To be able to determine drone locations solely based on their acoustic signals, an array of synchronized microphones were required to do the necessary calculations. (Fig.

1) The localization was split into a three-step pipeline for the best results:

1) Time Difference of Arrival Calculation (TDOA): Based on the method from Manamperi et. al [5] the calculation of the location of the sound first requires the estimation of the direction of arrival (DOA). To utilize this estimation technique, the Time Difference of Arrival (TDOA) at different microphone pairs is required. To improve accuracy, we implemented a bandwidth filter to filter out irrelevant noise for our engine such as High Frequencies from 200 - 4000 HZ.

Additionally, we implemented several preprocessing steps specifically tailored for continuous periodic drone signals. A gentle highpass filter at 60 Hz removes low-frequency wind noise and environmental rumble while preserving the drone's fundamental frequency, which typically lies in the 100-300 Hz range. To eliminate power line interference that can cause false correlation peaks, we apply notch filters at 50 Hz and 60 Hz with a Q-factor of 30. A Tukey window with an alpha of 0.1 was applied to each audio segment. This ensures that the signal boundaries are smoothed, reducing unwanted technical noise. These preprocessing steps are particularly critical for drone signals because, unlike distinct sounds such as claps or snaps, continuous periodic signals are more susceptible to interference from environmental noise sources.

Afterward, a signal processing algorithm called Generalized Cross Correlation (GCC-PHAT) is applied to estimate the time delay of the same signal received by two or more microphones. For this step, it is crucial that the recordings of each microphone are synchronized as it would otherwise cause the calculation of the TDOA to be distorted. To create the TDOA microphone pairs, one microphone from the array is designated as the "reference" microphone. This reference is compared against each corresponding microphone signal, yielding three TDOA values in total. Using Fast Fourier Transform (FFT), the audio signals are converted from the time domain to the frequency domain (Fig. 2) for computational efficiency.

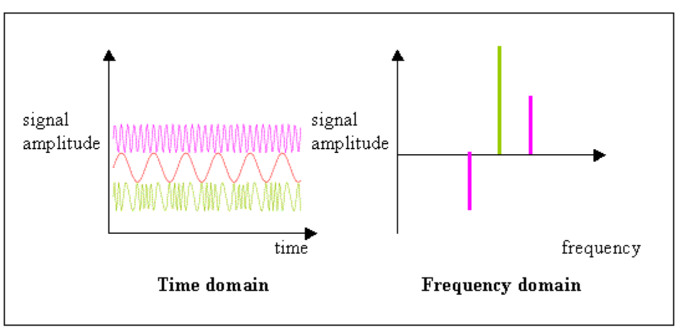

Fig. 2: Time and Frequency Domain

The GCC-PHAT algorithm normalizes the amplitudes to make it more robust against reverberation in enclosed spaces. By ignoring the volume (magnitude) of the sound, the algorithm focuses solely on the phase information. The phases from the two microphone recordings are compared, and the resulting difference is calculated within the cross-spectrum.

$$\text{cross\_spectrum} = FFT1 \cdot FFT2^*$$

$$\text{normalized\_cross\_spectrum} = \frac{\text{cross\_spectrum}}{|\text{cross\_spectrum}| + 1e-10}$$

Finally, to improve accuracy even further, an interpolation technique is applied. As the system records audio at a fixed sample rate, a sound wave may arrive at a microphone at any arbitrary time, often falling between two discrete sampling intervals. By utilizing the magnitude of the detected discrete peak $y_1$and its immediate neighbors y_0 and y_2, a parabolic fit is used to estimate the fractional sample offset δ:

$$\delta = 0.5 \cdot \frac{y_0 - y_2}{y_0 - 2y_1 + y_2}$$

The standard GCC-PHAT approach works well for distinct sounds, however drone acoustic signatures are characterized by clear, repeating sound patterns that occur at regular intervals from propellor rotation, which is why we developed an enhanced TDOA calculator that processes multiple frequency bands separately. The audio spectrum is divided into five bands: 80-200 Hz (fundamental frequency range, weight 1.0), 200-400 Hz (first harmonic, weight 0.9), 400-800 Hz (second harmonic, weight 0.8), 800-1600 Hz (higher harmonics, weight 0.6), and 1600-3200 Hz (upper range, weight 0.4). Each band is processed independently through GCC-PHAT, and then the resulting TDOA estimates are combined using adaptive weights that depend both on the base weights mentioned above and the measured signal-to-noise ratio and correlation quality. Bands with SNR below 3 dB are automatically excluded to prevent noise from corrupting the final TDOA estimate.

2) Direction of Arrival Calculation (DOA): After having calculated TDOA values, we are now able to determine the direction of arrival of the source sound. The direction is defined by the azimuth angle, which describes the horizontal orientation measured counterclockwise from a reference direction. We chose the x-Axis as starting angle (0°). Each TDOA defines a hyperbola of possible source locations which are then converted into range differences (Δd_i).

$$\Delta d_i = \text{TDOA} \cdot v_s$$

A least-squares problem is then solved to identify the point that best fits all TDOA measurements. This result is represented as a 2D position vector, which is used to compute the azimuth angle, providing the horizontal direction of the source.

3) Multilateration Calculation: Multilateration is a positioning technique used to determine the location of an object by measuring the difference in distance to several known receivers which in this case are the four microphones in the array. Unlike basic DOA estimation, which only provides an angle, we used the known fixed coordinates of each microphone to interpolate the source's absolute position. This process effectively transforms the raw acoustic data into comprehensible spatial coordinates. This last step utilizes both previously calculated values for TDOA pairs and DOA. We first align the TDOA values relative to a reference microphone and convert them into range differences, Δd_i, by multiplying the time differences by the speed of sound $v_s = 343 m/s$ exactly as in the DOA calculation.

To ensure the optimizer converges correctly and avoids local minima, we implemented an intelligent search initialization. If a DOA estimate is available, the algorithm projects a vector from the array center along the estimated azimuth to set an initial guess p_0. We then defined a combined residual function that the solver iteratively minimizes. For each microphone i, a TDOA-based error $e_i(p)$ is calculated using the formula:

$$e_i(p) = \sqrt{w_i} \cdot \left((\|p - m_i\| - \|p - m_{ref}\|) - \Delta d_i \right)$$

where $p = (x, y)$ represents the candidate source position, m_i and m_{ref} are the fixed coordinates of the microphones, and w_i represents optional quality weights used to prioritize reliable sensor data. To further enhance robustness against noise, we implemented an additional penalty term, $e_{DOA}(p)$, which is appended to the error vector as:

$$e_{DOA}(p) = \text{rad}(\Delta\theta) \cdot r_{max} \cdot w_{DOA}$$

In this equation, $\Delta\theta$ is the angular deviation between the candidate position and the measured DOA bearing, r_{max} is the maximum array radius used to scale the angular error into a distance-equivalent magnitude, and w_{DOA} is a weighting factor. Finally, the combined vector of residuals $\mathbf{E} = [e_1, \ldots, e_{n-1}, e_{DOA}]$ is passed to a non-linear *least_squares* solver that adjusts p starting from p_0 until it finds the optimal spatial coordinates that best satisfy both the timing and directional data. The optimization employs a trust region reflective algorithm with a soft L1 loss function to provide greater robustness against outliers than standard L2 loss. To avoid convergence to local minima, the initial guess p_0 is determined through an intelligent search strategy: the algorithm projects a unit vector from the array center along the estimated azimuth angle to a distance of 70% of the maximum search radius, ensuring the starting point is both physically plausible and aligned with the directional information.

4) Alternative Localization Approaches: During our development phase, we additionally experimented with several alternative localization methods to identify the most suitable approach, even if the recommendation by our advisors was to use a TDOA-based approach. With Beamforming (SRP-PHAT) we tried to implement a steered response power approach that searches over a grid of candidate source positions, computing the sum of pairwise GCC-PHAT correlations for each position, which is quite expensive computationally. Also, we tried the so called MUSIC-algorithm, which is a subspace algorithm that decomposes the signal covariance matrix into signal and noise subspaces through eigendecomposition. The algorithm then searches for source positions whose steering vectors are orthogonal to the noise subspace, which would have required more high-quality microphones for reliable performances. Lastly, we also tried a Harmonic SRP-approach, which extends beamforming by first detecting the fundamental frequency of the drone through autocorrelation or harmonic product spectrum methods, then creating a frequency mask that focuses correlation on the detected propeller harmonics.

B. Detection

For our drone detection system we experimented with different machine learning techniques to distinguish drone audio from non-drone audio. This section details the data acquisition pipeline, our preprocessing techniques, and the feature extraction methodology used.

1) Data Acquisition and Dataset Composition: The first step necessary for our drone detection was to generate a dataset with both drone and non-drone sounds. For that we used the public dataset by BowonY on GitHub [10], containing 1,332 drone and 10,372 "unknown" audio clips. Recognizing, that this pre-existing dataset lacked specific rotor characteristics and specific background noise profiles, such as talking, we expanded the training set with our own custom recordings of two different additional drones, with various background noises, as well as additional not-drone sounds.

This was important, as in the beginning in the presence of non-stationary background noise, such as human speech, we were not able to successfully detect the drone, which we then mitigated by simply expanding the training dataset with audio files, where we are simply talking over the drone for example.

2) Signal Preprocessing and Standardization: Each audio clip is fixed to a duration of exactly one second. This was already the case with the pre-existing dataset, and we chose to go with it, as this window size is an optimal trade-off between spectral resolution and computational efficiency – there is enough time to capture rotor harmonics while minimizing the latency of real-time inference. Small inaccuracies are handled either per zero-padding or trimming, to ensure a fixed-size input for the neural network architectures.

Furthermore, input signals are resampled to a 16 kHz mono format. Again, this was the case with the pre-existing data and again it makes sense, as according to the Nyquist-Shannon sampling theorem, this rate allows for the accurate representation of frequencies up to 8kHz into which most drones fall, while higher sample rates would mostly add ultrasonic detail that would not significantly improve detection performance.

3) Feature Extraction: In our project we experimented with different machine learning architectures, and used two distinct feature extraction pathways depending on the concrete architecture.

a) MFCC-Features(SVM, DNN, KNN, GMM, RNN): For most of the models we tested, we employ Mel-frequency Cepstral Coefficients (MFCCs), which are effective at capturing the spectral envelope of non-stationary signals. These features capture the phonetic information of audio and are commonly used in speech recognition and audio analysis. They provide a compact and robust representation of acoustic patterns such as motor noise, propeller harmonics, and mechanical vibrations, making them well suited for our drone detection audio analysis.

We extract 40 static MFCC coefficients to represent the instantaneous spectral shape, and also compute their first- and second-order deltas (derivatives) to model velocity and acceleration of spectral changes over time.

By computing the mean and standard deviation, we aggregate these 120 features (40 coefficients x 3), resulting in a 240-dimensional feature vector that we can then use for our detection.

b) Log-Mel Spectrograms (CNN): For the Convolutional Neural Network (CNN) architecture, we instead use Log-Mel-Spectrograms. The audio is converted into a 64x100x1 image-like representation, which preserves the spatial-temporal relationship of the sound, allowing the CNN to learn patterns in the time-frequency similar to how an image recognition model would identify edges and textures.

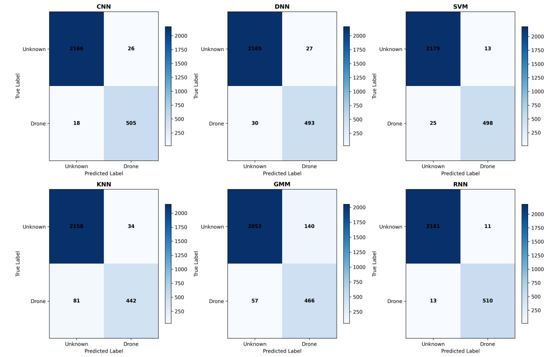

Fig. 3: Confusion Matrices

IV. Experimental Evaluation

A. Detection

In this section, we evaluate the results we were able to achieve and explain why specific results did or did not match our original intention.

Starting with the detection, we achieved astonishing results. To provide a comprehensive assessment of the models, we first focus on Accuracy, Precision, Recall, and the F1-Score 4, which can be calculated by using the confusion matrices 3.

- Accuracy: Measures the overall percentage of correct predictions (both drone and non-drone)
- Precision: Quantifies the reliability of a positive drone detection
- Recall: Measures the ability to detect all drones present in the audio
- F1-Score: The harmonic mean of precision and Recall, serving as a single performance metric for imbalanced data scenarios

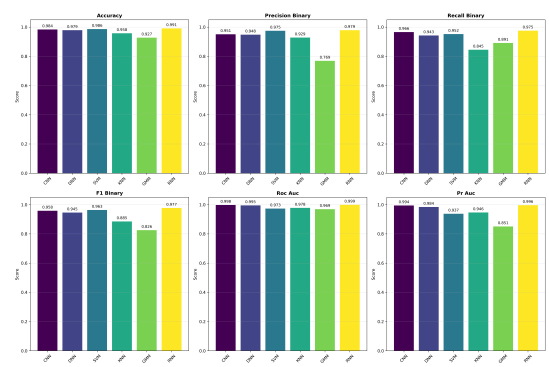

Fig. 4: Metrics Comparison

The RNN architecture using MFCCs achieved the highest performance across all metrics. This confirms that the temporal evolution of audio signals is a the most discriminative feature for identifying drone signatures. This result directly corroborates the findings of Al-Emadi et al. [3], [4], who demonstrated that deep learning approaches significantly outperform traditional DSP methods reliant on manual feature extraction.

Additionally, looking at the ROC curves 5 and the Precision-Recall curves 6 you get the same result.

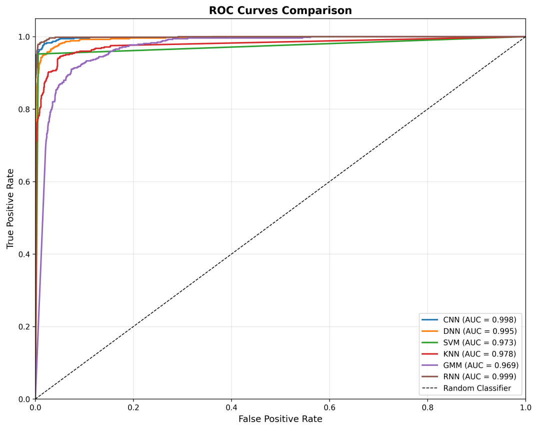

Fig. 5: ROC Curves

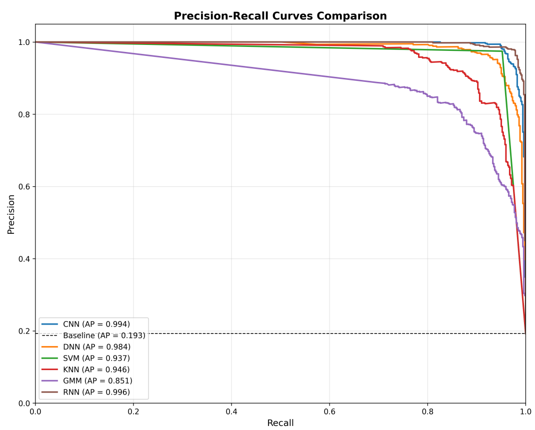

Fig. 6: Precision Recall Curves

The primary advantage of RNNs, and the reason why they perform better than other architectures, lies in their ability to process sequential data which allows them to preserve a temporal structure. The result: we were able to perfectly detect drones solely based on their acoustic signals even with a low signal-to-noise ratio.

B. Localization

In the localization domain, our system demonstrated a high precision when processing sharp, distinct sounds such as hand claps 7. However, it encountered substantial difficulties when attempting to localize continuous drone signatures 8.

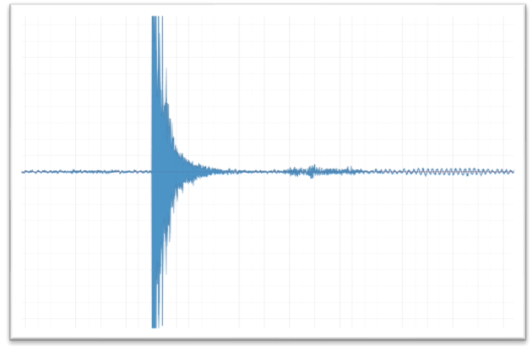

Fig. 7: Clap Audio File

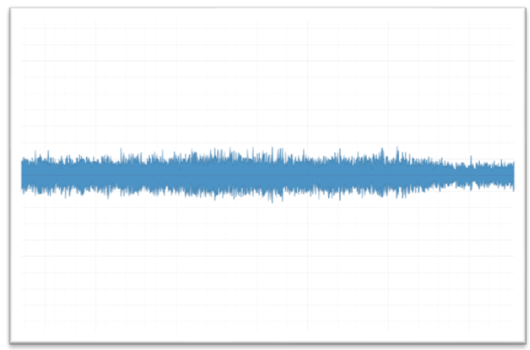

Fig. 8: Drone Audio File

Impulsive signals are broadband transients that produce clear, unambiguous cross-correlation peaks. This is why our TDOA-based algorithm worked well to calculate relative delays between microphones with near-perfect accuracy, resulting in reliable triangulation of the source position with minimal angular error.

In contrast, the localization of drone signals unfortunately yielded arbitrary and inconsistent results. This performance gap is tied to the limitations that a TDOA-based approach brings with it when applied to the periodic, narrowband signatures typical for drones. This aligns with the spectral analysis by Manamperi et al. [5], who noted that rotor noise dominates the frequency spectrum below 2 kHz. This narrowband dominance worsens the "cycle jumping" phenomenon, where multiple correlation peaks appear equally valid due to the periodic oscillation of the propellers, leading to arbitrary coordinate estimation. Furthermore, while Wang and Cavallaro [6] successfully utilized neural networks (TFS-DNN) to suppress noise and improve localization, and Hoshiba et al. [7] achieved reliable results using a specialized spherical microphone array (SMAS), our approach was restricted by standard array geometry and basic TDOA processing. Without the specialized hardware of Hoshiba et al. or the deep learning denoising layers proposed by Wang and Cavallaro, the disturbing Doppler effect and lack of broadband features rendered standard TDOA ineffective for the drone source.

This lead us to consider the alternative high-resolution methods Beamforming, MUSIC, or Harmonic SRP, mentioned in the previous chapter, however several hardware and environmental constraints made them non-viable. For example, spatial sampling requires that microphones be spaced no more than half the wavelength ($\lambda/2$) apart from the highest frequency being tracked.

Our findings define the upper performance limit for accessible, low-cost acoustic drone audition: Our low-quality microphones and the setup in general did not allow us to meaningfully pursue these other approaches.

Consequently, while the project successfully demonstrated robust acoustic identification, the transition to precise acoustic localization remains restricted by the physical and mathematical limitations of the current TDOA-based hardware configuration.

V. Summary, Conclusions, Challenges and Future Work

A. Summary

In this paper, we attempted to develop a drone localization engine that solely works with acoustic signals, that can be built with simple accessible hardware. Our research began with an evaluation of existing literature regarding acoustic sensing in search and rescue operations, as well as the detection and localization of drones via radio frequencies.

Building upon these methodologies, our design focused on the combination of detection and the localization of a drone. To determine a drone's position, we built an array of synchronized microphones and split the calculation into a three-step pipeline. The first stage involved estimating the Direction of Arrival (DOA) of the drone sound, which in turn required calculating the Time Difference of Arrival (TDOA) across various microphone pairs. For the computation of the TDOA, we used an algorithm called GCC-PHAT. These TDOA values were subsequently utilized to determine the DOA by solving a least-squares problem to find the optimal fit for all measurements, defining the direction through the azimuth angle. Finally, by combining the TDOA and DOA data, a positioning technique called multilateration is applied to determine the location of an object by measuring the difference in distance across the four microphones.

For the drone detection, we trained six different models with the RNN Model achieving the highest accuracy. Regarding drone detection, we evaluated six different machine learning architectures, with the Recurrent Neural Network (RNN) model achieving the highest accuracy. To incorporate detection of specific rotor characteristics and background noise, a public dataset was extended with additional recordings. Each audio clip is fixed to a duration of exactly one second as this window size is an optimal trade-off between spectral resolution and computational efficiency.

B. Conclusions

The TDOA-based approach with our hardware constraints proved insufficient for the reliable localization of drones as the signals often looked identical. Drone emissions are classified as narrowband signals, which lack the sharp temporal transients (edges) necessary for high-precision TDOA estimation. However, we can perfectly locate sharp sound sources based on their acoustic features if there is only a single distinct sound and if the sound remains inside the microphone array, as this is the area where TDOA calculation works best. In contrast, we were very successful concerning drone detection solely based on their acoustic signals. We managed to achieve a detection accuracy of 99% and higher in a quiet environment. Even under conditions with a low signal-to-noise ratio, the RNN model maintained a robust performance with an accuracy of at least 80%. While we sometimes had occasional successful trials, the limitations of our localization engine and available hardware resources prevented consistent and reliable feedback. We therefore conclude that additional input, such as radio frequencies or comparable data, are essential to achieve precise drone localization and, more importantly, do it reliably over longer distances and different environments.

C. Challenges

Throughout the project, we faced several challenges that impacted our progress. One of the primary obstacles was the lack of a functional drone until eight weeks after we began developing the localization engine. This delay forced us to seek alternatives for our initial testing, which is why our first localization trials were conducted using hand-claps instead of drone signatures.

We also faced substantial hardware limitations, particularly regarding the synchronization of our microphones. Because they lacked a shared hardware clock, the Time Difference of Arrival (TDOA) calculations were frequently distorted and inaccurate. Further complicating the setup was an issue with the Raspberry Pi's USB architecture. From a software perspective, the USB slots were not consistently mapped; for instance, plugging a microphone into the top-left USB port did not guarantee it would be recognized as the top-left sensor in our 2×2 array. This inconsistent indexing required constant manual verification and troubleshooting to ensure the spatial coordinates in our algorithms matched the physical microphone placement.

In the beginning, the Raspberry PI did not provide enough computational power for our detection models which forced us to invest a substantial amount of time into optimizing the models which diverted our focus from further progressing in other areas of the project. All in all, overcoming these cumulative challenges was a valuable learning experience. It helped us expand our knowledge for the hardware used but it was also a leading factor in us not being able to realize the localization engine that we had envisioned at the start.

D. Future Work

Based on the results and limitations identified in this study, several paths for future research emerge. A primary objective is combining acoustic data with radio frequency (RF) analysis. Since acoustic TDOA calculations struggle with the narrowband characteristics of drone rotors over long distances, integrating RF signals could provide the necessary precision for reliable tracking. Such a hybrid system would leverage the strengths of both methods. Furthermore, the transition from stationary microphone arrays to mobile sensing represents a significant evolutionary step. By mounting synchronized microphone arrays onto drones, the system's operational range could be expanded substantially. These mobile UAV's would not only cover larger geographical areas but could also actively minimize the distance to the target, thereby maintaining a high signal-to-noise ratio and improving the performance of the RNN-based detection models. Ultimately, the refinement of these localization techniques serves as the foundation for moving beyond passive monitoring toward active interception. Once a drone's position can be determined with high reliability in real-time, the system can be integrated with automated response protocols such as signal jamming or physical capture.

Acknowledgments

This work was supported by the School of Computer Science at the University of St. Gallen. We extend our special thanks to our supervisors, Professor Bruno Rodrigues and Karim Khamaisi, for their continual support and guidance throughout this project.

References

[1] J. F. Keane and S. S. Carr, "A brief history of early unmanned aircraft," *Johns Hopkins APL Technical Digest*, Vol. 32, No. 3, pp. 558–571, 2013.

[2] S. Jeon, J.-W. Shin, Y.-J. Lee, W.-H. Kim, Y. Kwon, and H.-Y. Yang, "Empirical study of drone sound detection in real-life environment with deep neural networks," *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1858–1862.

[3] S. A. Al-Emadi, A. K. Al-Ali, A. Al-Ali, and A. Mohamed, "Audio based drone detection and identification using deep learning," *IWCMC 2019 Vehicular Symposium (IWCMC-VehicularCom 2019)*, Tangier, Morocco, jun 2019.

[4] ——, "Fusion of rf and audio for drone classification," *Sensors*, Vol. 24, No. 2, p. 812, 2024.

[5] M. Manamperi, H. Zhang, and D. Florencio, "Doa estimation with onboard microphone arrays on drones," *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 266–270.

[6] H. Wang and A. Cavallaro, "Deep-learning-assisted source localization with drones under ego-noise," *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 426–430.

[7] T. Hoshiba, W. Onishi, and Y. Tokunga, "Uav-embedded spherical microphone array system for outdoor sound source localization," *2017 IEEE Sensors*, 2017, pp. 1–3.

[8] H. Wakabayashi, K. Uehara, S. Kasai, and S. Seto, "Multiple sound source localization and identification for uavs," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 5644–5650.

[9] X. Chang, C. Yang, Z. Wu, and J. Wu, "An acoustic-based surveillance system for amateur drones detection and localization," *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 3, pp. 2731–2739, 2020.

[10] BowonY, "drone-audio-detection: Detecting drone sound in real time," 2025, gitHub repository, accessed January 28, 2026.

Appendix

Code Implementation

The implementations of our proposed framework and experimental evaluations are publicly available on GitHub:

https://github.com/BP-BATradar/BATradar (final version [uses localization for impulse sounds, so reliable results can be reproduced in that domain at least])

https://github.com/BP-BATradar/TDOA-DOA (drone localization experiments)

https://github.com/BP-BATradar/drone-classification (drone detection experiments)

Contribution Statement

Both authors contributed equally to the research, design, execution, and writing of this research.

Aids

Aid	Usage	Affected Part
ChatGPT	Text reduction, Grammar correction, Code explanations, Comments, Frontend Inspiration, Syntax for LaTeXa	Complete Paper / Code
Gemini	Grammar correction, wording, translation	Complete Paper
DeepL	Translation from German to English	Complete Paper

Declaration of Authorship

"We hereby declare,

- that we have written this thesis independently,
- that we have written the thesis using only the aids specified in the index;
- that all parts of the thesis produced with the help of aids have been declared;
- that we have handled both input and output responsibly when using AI. we confirm that we have therefore only read in public data or data released with consent and that we have checked, declared and comprehensibly referenced all results and/or other forms of AI assistance in the required form and that we are aware that we are responsible if incorrect content, violations of data protection law, copyright law or scientific misconduct (e.g. plagiarism) have also occurred unintentionally;
- that we have mentioned all sources used and cited them correctly according to established academic citation rules;
- that we have acquired all immaterial rights to any materials we may have used, such as images or graphics, or that these materials were created by us;
- that the topic, the thesis or parts of it have not already been the object of any work or examination of another course, unless this has been expressly agreed with the

faculty member in advance and is stated as such in the thesis;

- that we are aware of the legal provisions regarding the publication and dissemination of parts or the entire thesis and that we comply with them accordingly;
- that we are aware that my thesis can be electronically checked for plagiarism and for third-party authorship of human or technical origin and that we hereby grant the University of St.Gallen the copyright according to the Examination Regulations as far as it is necessary for the administrative actions;
- that we are aware that the University will prosecute a violation of this Declaration of Authorship and that disciplinary as well as criminal consequences may result, which may lead to expulsion from the University or to the withdrawal of our title."

By submitting this thesis, we confirm through our conclusive action that we are submitting the Declaration of Authorship, that we have read and understood it, and that it is true.