Neural Drone Localization Exploiting Signal Synthesis of Real-World Audio Data

Ximei Yang*, Patrick A. Naylor†, Simon Doclo*‡, Joerg Bitzer*
* *Fraunhofer Institute for Digital Media Technology IDMT, Oldenburg Branch for Hearing, Speech and Audio Technology HSA*, Germany
† *Dept. of Electrical and Electronic Engineering, Imperial College London*, UK
‡ *Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg*, Germany

***Abstract*—As unmanned aerial vehicles (UAVs) become increasingly common, concerns about security, privacy, and noise pollution have intensified. As a result, the need for accurate and efficient UAV localization and tracking has become critical for security operations and timely intervention, yet algorithmic audio-based localization methods have limitations in complex outdoor environments. This study presents an approach to synthesizing drone acoustic signals and generating training datasets designed for deep neural network (DNN)-based localization. Using these simulated signals, two neural networks, SELDnet ACCDOA and Neural SRP, were trained and evaluated for accurate direction-of-arrival (DOA) estimation, addressing challenges specific to outdoor acoustic conditions. Their performance was benchmarked against the steered response power with phase transform (SRP-PHAT) methods. To further validate the models' effectiveness, real-world drone data were collected and used for testing. Experimental results indicate that neural networks trained on synthesized data achieve effectiveness comparable to SRP-PHAT, validating the reliability of the simulation approach, with Neural SRP even outperforming SRP-PHAT-based algorithms in DOA estimation accuracy.**

***Index Terms*—UAVs, localization, sound synthesis, direction of arrival estimation, deep neural network.**

I. Introduction

As drones have become more affordable and advanced, their widespread use raises concerns about security, privacy, and noise [1]. Automated drone localization and tracking are essential for security services and timely intervention. Recent research on audio-based drone localization has explored methodologies based on signal processing as well as techniques leveraging deep learning. Algorithmic methods, such as cross-correlation methods like generalized cross-correlation with phase transform (GCC-PHAT) [2] and SRP-PHAT [3], or subspace methods like multiple signal classification (MUSIC) [4], exhibit limitations in handling complex outdoor environments characterized by echoes, diverse noise sources, and distant or multiple UAVs [5]. Consequently, deep learning techniques have emerged as a promising alternative, demonstrating superior potential to address the challenges of complex acoustic environments [6] [7].

Grumiaux et al. [8] extensively reviewed various neural network approaches for sound source localization (SSL) in indoor environments, covering aspects such as network architecture, input features, output strategies, dataset types, and model training strategies. However, two key distinctions exist between conventional SSL research and drone localization: the acoustic environment and the characteristics of the acoustic signals. While SSL studies predominantly focus on indoor settings, drone operations typically occur in outdoor environments, including suburban and wilderness areas. Moreover, SSL research primarily focuses on human speech, which differs significantly from the mechanical noise produced by drones. Furthermore, the movement patterns of humans and drones vary considerably, necessitating further investigation into the unique acoustic and spatial properties of drone signals.

Another challenge in neural network-based drone localization is the need for large-scale datasets to ensure reliable training, which is both laborious and resource-intensive. To address this, various studies have analyzed drone acoustics and developed sound synthesis models. Acoustic analysis of drone noise has been conducted in [9] and [10], while an auralization model incorporating oscillators and autoregressive noise modeling was proposed in [11]. Heutschi et al. [12] synthesized real-world drone signals based on laboratory recordings, considering factors such as drone type, flight path, and wind conditions. These studies demonstrate the feasibility of drone sound modeling, providing a foundation for dataset generation in localization research.

This paper is organized as follows. Section II details the synthesis of drone signals and dataset generation we employed, while Section III presents the implemented localization algorithms, encompassing two neural networks alongside SRP-PHAT and its modified versions. Section IV describes datasets, evaluation metrics, and the experimental setup, with results analyzed in Section V. Finally, Section VI concludes the paper and explores future research directions.

II. Synthesis of Drone signals and Dataset generation

The construction of a drone auralization model is essential to simulate the spectro-temporal and spatial characteristics of a drone during flight. Its noise signal, primarily generated by rotating propellers, is synthesized using a procedural audio model that integrates oscillators and digital filters. This approach captures tonal components with modulations in

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369 'SOUNDS'.

amplitude and frequency while incorporating white noise to represent stochastic elements, resulting in a realistic emission signal.

A. Drone tonal sound emission

In emission models, the tonal characteristics of a drone are determined primarily by the rotational speed of its rotors, with the revolutions per minute (RPM) being a key factor. Rotor speeds vary dynamically based on drone type and maneuver types, such as hovering, climbing, descending, or forward flight. These variations in RPM can be referenced in [12], which discusses the relationship between different maneuver types, drone models, and corresponding rotor speeds, with the fundamental frequency determined by multiplying the RPM by the number of blades per motor, effectively doubling for typical two-blade configurations. To reproduce the tonal component, an oscillator is used to model the amplitudes of purely mono-frequent oscillators. This can be achieved using sinusoidal waves, or alternatively, pulse wave oscillators or bipolar pulse wave. This process simulates the tonal characteristics of the drone in flight, thus generating the characteristic "bee" sound of a drone in auralization.

B. Drone stochastic sound emission

In realistic scenarios, the audible rotor speed of a drone also varies due to propeller imperfections, transitions between different flight maneuvers, and external environmental influences such as wind. To achieve perceptually plausible auralizations, it is essential to account for fluctuations in the sound field.

Since random fluctuations caused by varying wind conditions lead to frequency variance, it exhibits a linear relationship with the normalized standard deviation of rotational speed σ_n and the average wind speed ($\overline{v}_{\text{w}}$, e.g., 2.5 m/s for a specific flight) [12]. This relationship can be expressed as $\sigma_n = a + b \cdot |\overline{v}_{\text{w}}|$ where parameters a and b are available in [12].

To ensure smooth transitions in rotational speed during maneuver changes, a sigmoid function is employed to avoid abrupt shifts. Furthermore, due to propeller imperfections, the fundamental frequency derived from the rotor speed is inherently variable. Therefore, the mean of the fundamental frequency is slightly modulated over time. For example, small random shifts within a range of -5 Hz to +5 Hz are periodically introduced to reflect natural variability.

Drones typically have multiple rotors, most commonly four or six, whose signals are combined with added white noise to represent stochastic components.

C. Amplitude and frequency modulations

Since the dominant harmonics of rotor-generated tones generally fall between 100 Hz and 2000 Hz [5], amplitude modulation is applied to these harmonics to enhance realism, followed by a low-pass filter, such as a second-order Butterworth filter with a cutoff frequency of 2400 Hz, to smooth the signal and remove high-frequency noise.

D. Movement and spreading

To generate realistic drone signals at the positions of the simulated microphone array, several acoustic propagation effects must be accounted for, including radiation directivity, geometric spreading, reflections, and background noise. The received signal $m_i(t)$ at the i^{th} microphone consists of a delayed and attenuated version of the drone's emitted sound and background noise, modeled as $m_i(t) = L_{\text{prop}} \cdot L_{\text{rad}} \cdot s(t - \tau_i) + v_i(t)$, where $s(t)$ is the emitted drone signal, and τ_i is the propagation delay determined by the geometric distance between the drone and the microphone. A Doppler shift further affects the perceived frequency when the drone is in motion. The term L_{rad} represents radiation directivity, which depends on the emission angle but remains independent of rotor speed and flight procedure [11] [12], while L_{prop} accounts for distance-based propagation attenuation. The background noise $v_i(t)$ originates from various environments (urban, suburban, or rural) and can be obtained from real-world recordings captured with an actual sensor array or simulated under spatial constraints, such as a diffuse noise model [13].

Additionally, ground reflections influence the received signal with the reflection coefficient which varies with ground material properties (e.g., asphalt, soil, or grass), affecting how much sound is absorbed or reflected. By incorporating these factors, we ensure a realistic synthesis of drone signals, enabling accurate dataset generation for localization tasks. Code for the synthesis of drone signals and dataset generation in this paper is available on GitHub[1].

III. Localization Algorithms

A. SRP-PHAT-based algorithms

SRP-PHAT is chosen as the baseline method due to its widespread adoption and effectiveness in estimating the DOA of a sound source through beamforming and phase-based processing. This paper employs the conventional SRP-PHAT approach [3], along with its modified variants, SRP-PHAT-β [14] and SRP-PHAT-mask, for drone localization.

The conventional SRP-PHAT computes the correlation between the signals received by the microphones and the cumulative steered response at each candidate location. SRP-PHAT-β enhances it by incorporating a weighting factor β ($0 \leq \beta \leq 1$) into the GCC-PHAT computation, preserving partial magnitude information and enhancing performance in low signal-to-noise ratio (SNR) environments [14] [15]. Since drone signals are primarily concentrated within a specific frequency range [5], we propose SRP-PHAT-mask, which applies a binary mask in the frequency domain during GCC-PHAT computation to preserve relevant components while suppressing noise.

B. Data-driven methods

The primary focus of this paper is a data-driven approach for drone DOA estimation, utilizing neural networks trained on simulated datasets. Two models, SELDnet ACCDOA and

[1]https://github.com/SOUNDS-RESEARCH/DroneSynthesis2DOA

Neural SRP, are selected for reproduction and evaluation to assess their effectiveness.

1) SELDnet ACCDOA: SELDnet [Sound Event Localization and Detection Network] is a convolutional recurrent neural network (CRNN) originally designed for joint sound event detection (SED) and three-dimensional (3D) DOA estimation [16]. It employs three convolutional layers for feature extraction, two gated recurrent units (GRUs) for temporal dependencies, and dual output branches for DOA regression and sound event classification. The output is a tensor that allows for real-time, frame-by-frame prediction of the source's angular position, with each frame providing an updated estimate of the direction. This direction is represented by a unit vector with Cartesian coordinates $(x, y, z) \in [-1, 1]$.

For drone localization, we adapted SELDnet into ACCDOA [Activity-Coupled Cartesian DOA] [17] by removing the SED branch and simplifying the loss to mean square error (MSE). Additionally, the original input features, based on the short-time Fourier transform (STFT), were replaced with GCC-PHAT to enhance spatial information [18], and the model was simplified to single-event regression to reduce complexity.

2) Neural SRP: The Neural SRP model extends sound event localization and detection (SELD) to arbitrary microphone array geometries by incorporating microphone coordinates into its feature representation, enabling it to process signals from different arrays without being constrained to the specific array used during training [19]. Its architecture consists of a pairwise network P and a global decoder D. The pairwise network processes the GCC-PHAT between two channels and corresponding microphone coordinates through three convolutional layers, two GRUs, and a two-layer multilayer perceptron (MLP). For M microphones, $M(M-1)/2$ such pairs are processed in parallel, with their outputs summed and fed into the global decoder D. For single-drone localization, the SED branch is removed as well, retaining only the DOA estimation branch with a two-layer MLP. This design integrates the coordinate information of the array, enabling robust generalization across diverse array configurations.

IV. EVALUATION

A. Datasets

Data-driven localization algorithms are trained using synthetic data, whereas all algorithms are evaluated on both synthetic and real-world datasets. The synthetic dataset is generated using real-world drone parameters from [12], including DJI Mavic 2 Pro, Inspire 2, S-900, and F-450, all featuring four or six rotors with two-blade propellers, ensuring the realism of the simulations. Both datasets employ an 8-microphone array arranged in a 15 cm cubic configuration, with its coordinate system and corresponding image presented in Figure 1.

1) Synthetic drone dataset: In the simulations, a far-field sound source is assumed with two types of ground reflections based on surface materials. The microphone array is fixed at the 3D coordinate center, while the drone follows a 3D flight path within a ±200 m range in the x-y plane and a maximum altitude of 47 m, always flying above the array with only one drone per flight.

Mic	X (m)	Y (m)	Z (m)
1	0.05625	0.0375	0.15
2	0	0.09	0.15
3	0.01875	0.15	0.1145
4	0.05625	0.15	0.009
5	0.15	0.15	0.075
6	0.13125	0.075	0.009
7	0.15	0.01875	0.15
8	0	0	0

(a) Microphone array coordinates

(b) Real-world microphone array

Fig. 1: Microphone array configuration

The synthetic dataset models realistic drone dynamics with random transitions between hovering, ascending, descending, and forward flight. Each flight includes four state transitions, totaling 1400 flights lasting between 10 and 121 seconds, resulting in approximately 1226 minutes of data. Half of the samples contain no background noise, while the other half incorporate six distinct environmental noise recordings (e.g., urban, suburban, rural), simulated via the diffuse model [13]. For the samples with environmental noise, the SNR, defined as the ratio of the received drone signal power to background noise power, is dynamically set, ranging approximately from -5 dB to 15 dB at 3 m and from -30 dB to -10 dB at 50 m, making it distance-dependent.

2) Real-world drone dataset (ground truth localization): The real-world drone dataset was collected by Fraunhofer IDMT during an experiment in August 2022 in Berne, Lower Saxony, Germany. The dataset includes recordings of drones (DJI Phantom, DJI Mavic, and Align models) performing static, linear, and random flights. A total of eight reliable flight recordings were chosen, with durations ranging from 4 to 23 minutes, summing to approximately 106 minutes. Most flights were within 200 m of the array, with a maximum distance of 430 m.

Each drone and array was equipped with a GPS-Logger 3 (SM Modellbau), providing high-precision location data with ±2.5 m accuracy and a 10 Hz update rate. This real-time latitude and longitude data allowed calculation of the drone's distance and azimuth relative to the microphone array. The coordinate system was defined with the x-axis northward, the z-axis upward, and the y-axis following the right-hand rule, enabling azimuth angle computation.

The Haversine formula [20] was used to calculate the shortest distance between two points on the Earth's surface, assuming equal altitude. The azimuth was computed based on the dihedral angle in space. These calculations provided the drone's direction and corresponding azimuth angles, critical for validating the trained models.

B. Evaluation metrics

1) DOA error: The DOA error is computed as the angular difference between the predicted and true DOA in 3D space.

TABLE I: Comparison of localization algorithms using different GCC-PHAT inputs on synthetic and real-world datasets.

		Synthetic Drone Test Set				**Real-World Drone Dataset**			
Algorithms	GCC-PHAT	-	with β	with mask	β+mask	-	with β	with mask	β+mask
SELDnet ACCDOA	DOA error	9.7	11.1	**9.5**	11.1	37.3	42.2	39.2	44.1
	ER_{20°	11.9%	12.1%	11.6%	13.0%	48.3%	52.1%	52.5%	49.5%
Neural SRP	DOA error	9.6	11.1	10.9	10.6	20.4	**15.9**	23.3	19.1
	ER_{20°	**10.4%**	11.4%	13.1%	12.3%	28.5%	**23.4%**	33.6 %	30.3%
SRP-PHAT	DOA error	27.8	25.4	24.7	25.1	29.4	24.1	23.0	21.6
	ER_{20°	28.5%	24.3%	24.9%	24.6%	29.4%	26.0%	25.6%	24.2%

2) Error rate ER_{20°: The ER_{20° is calculated by considering frame-wise predictions as true positives only when the angular difference from the reference is less than 20°.

Although the flight durations vary, the overall DOA error and ER_{20° are calculated as the unweighted average of the individual errors across all flights, regardless of their respective durations.

C. Experiments

1) Model training and parameter settings: Different DOA estimation algorithms require specific parameter configurations. For SRP-PHAT and its modified versions, 1000 candidate directions are uniformly distributed using spherical Fibonacci mapping [21], ensuring an even distribution across the surface. β is chosen to be 0.7 for SRP-PHAT-β algorithm [15]. For SRP-PHAT-mask, a mask spanning 250 Hz to 7000 Hz is strategically designed to preserve relevant spectral components of the drone signal while effectively mitigating noise interference.

To ensure real-time applicability, both SELDnet ACCDOA and Neural SRP employ unidirectional GRU layers. The model sizes for SELDnet ACCDOA and Neural SRP are 0.80M and 0.79M parameters, respectively.

All algorithms make frame-wise predictions with a 32 ms frame length and no overlap. To handle the varying sequence lengths in dynamic drone flight durations, Bucket Batching is used to group sequences of similar lengths, minimizing padding. Shorter sequences are padded within each batch, with the padded output for DOA coordinates set to zero (x, y, z = 0), and a masking mechanism is applied to exclude the padded regions from training.

The synthetic drone dataset follows a standardized data split of 70% for training, 15% for validation, and 15% for testing. Model training is conducted over 300 epochs with early stopping, utilizing a dynamic learning rate initialized at 0.001 and a batch size of 32.

2) Selecting model input format: The original SELDnet model used STFT-based features, but GCC-PHAT showed improved performance [18]. However, GCC-PHAT alone may be insufficient for drone signals due to their narrowband nature and noise sensitivity. To better capture drone characteristics, modified formats, GCC-PHAT-β and GCC-PHAT-mask, were introduced, both derived from SRP-PHAT cross-correlation computations (Section III-A) and using the same parameters to ensure comparability. Furthermore, considering the inter-microphone time lag, 48 central GCC bins were used for each frame.

V. RESULTS AND DISCUSSION

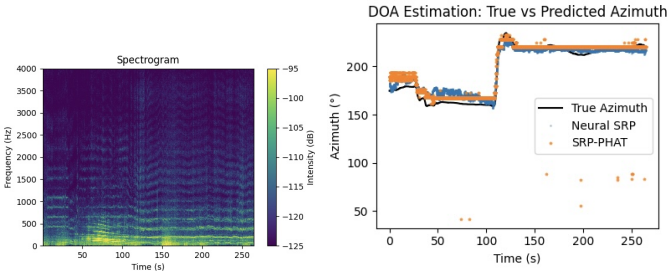
Table I presents the DOA prediction results for both the synthetic test set and the real-world dataset, with neural networks trained solely on the synthetic train set. On the synthetic test set, SRP-PHAT performs worse than neural network-based methods, with DOA errors ranging from 25° to 28°, primarily due to its poor performance on noisy samples. In contrast, SELDnet ACCDOA and Neural SRP achieve lower DOA errors, demonstrating greater robustness in both clean and noisy conditions. SELDnet ACCDOA with GCC-PHAT-mask achieves the lowest DOA error of 9.5°, while Neural SRP with GCC-PHAT yields the lowest ER_{20° at 10.4%. In other cases, performance differences remain minor.

The trained models are also evaluated on the real-world drone dataset, with results presented in the right half of Table I alongside comparisons to SRP-PHAT-based algorithms. Since the real-world dataset provides only azimuth ground truth, DOA error is computed as the angular difference between the predicted and true azimuth values, without considering elevation. The results demonstrate that even with unseen and slightly mismatched data, the models effectively predict DOA angles, confirming the simulated dataset's validity and the models' generalization ability. The best performance is achieved by Neural SRP with GCC-PHAT-β, yielding a DOA error of 15.9° and an ER_{20° of 23.4%, outperforming the best SRP-PHAT-based result (21.6° DOA error, 24.2% ER_{20°) obtained with β and mask. In general, the DOA error remains high due to outdoor environmental noise and the typically large drone-to-microphone distances, reaching over 400 m. Neural SRP surpasses SRP-PHAT-based algorithms, indicating that the simulated dataset closely approximates real-world drone conditions, including dynamic flights with varying background noise. SELDnet ACCDOA underperforms compared to SRP-PHAT, potentially due to its increased sensitivity to noise variations leading to weaker generalization.

Notably, for SRP-PHAT-based methods, incorporating β or a mask consistently reduces errors in both datasets. However, for neural networks, these modifications do not always improve performance and may even degrade accuracy, particularly on the test set. In the real-world dataset, using β, which means adding some amplitude information, improves accuracy for Neural SRP. In contrast, applying a mask worsens

performance, likely due to reduced input information or the networks' inherent noise suppression capability, indicating that neural networks already learn to handle noise without additional masking.

To further evaluate the efficacy of the trained models, a sample from the real-world drone dataset is selected for visualization. Figure 2 presents the drone spectrogram from one array channel alongside azimuth angle estimations obtained from Neural SRP (using GCC-PHAT-β as input) and SRP-PHAT-β-mask, based on one real drone flight. While the errors of both methods are comparable, SRP-PHAT exhibits significant outliers, whereas Neural SRP demonstrates robust performance without such anomalies. This indicates that, compared to SRP-PHAT, the recurrent structures in neural networks provide some degree of tracking capability, further highlighting the importance of dynamic data.



(a) Drone spectrogram (b) Azimuth estimation in DOA

Fig. 2: Visualization of a real DJI Mavic flight

VI. Conclusions and Future Work

This study synthesized a drone audio dataset and evaluated various SSL methods, including SRP-PHAT-based algorithms, SELDnet ACCDOA, and Neural SRP. It highlights the feasibility of training localization models using synthetic drone acoustic data based on drone tonal characteristics, thereby reducing reliance on real-world data collection. Experimental findings further validate this approach for DOA estimation, showing that trained neural networks achieve performance comparable to or even exceeding that of SRP-PHAT. Future work includes integrating motor sounds into the synthesis process, exploring Kalman filters for DOA refinement, and developing specialized neural network architectures for UAV localization. Challenges also remain in locating multiple drones and using multi-array systems for precise spatial coordinates.

References

[1] Y. Zhi, Z. Fu, X. Sun, and J. Yu, "Security and privacy issues of uav: A survey," *Mobile Networks and Applications*, vol. 25, no. 1, pp. 95–101, 2020.

[2] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[3] J. H. DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*. Brown University, 2000.

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[5] A. Altena, S. Luesutthiviboon, G. de Croon, M. Snellen, and M. Voskuijl, "Comparison of acoustic localisation techniques for drone position estimation using real-world experimental data," in *Proceedings of the 29th International Congress on Sound and Vibration, ICSV 2023*. Society of Acoustics, 2023.

[6] É. Bavu, H. Pujol, A. Garcia, C. Langrenne, S. Hengy, O. Rassy, N. Thome, Y. Karmim, S. Schertzer, and A. Matwyschuk, "Deeplomatics: A deep-learning based multimodal approach for aerial drone detection and localization," in *QUIET DRONES Second International e-Symposium on UAV/UAS Noise*, 2022.

[7] Z. Xiao, H. Hu, G. Xu, and J. He, "Tame: Temporal audio-based mamba for enhanced drone trajectory estimation and classification," 2025. [Online]. Available: https://arxiv.org/abs/2412.13037

[8] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, "A survey of sound source localization with deep learning methods," *The Journal of the Acoustical Society of America*, vol. 152, no. 1, p. 107–151, Jul. 2022. [Online]. Available: http://dx.doi.org/10.1121/10.0011809

[9] M. Podsedkowski, R. Konopiński, and M. Lipian, "Sound noise properties of variable pitch propeller for small uav," in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2022, pp. 1025–1029.

[10] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "Dregon: Dataset and methods for uav-embedded sound source localization," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1–8.

[11] Dreier, Christian and Vorländer, Michael, "Drone auralization model with statistical synthesis of amplitude and frequency modulations," *Acta Acust.*, vol. 8, p. 35, 2024. [Online]. Available: https://doi.org/10.1051/aacus/2024026

[12] K. Heutschi, B. Ott, T. Nussbaumer, and P. Wellig, "Synthesis of real world drone signals based on lab recordings," *Acta Acustica*, vol. 4, no. 6, p. 24, 2020.

[13] E. A. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, 2008.

[14] A. Ramamurthy, H. Unnikrishnan, and K. D. Donohue, "Experimental performance analysis of sound source detection with srp phat-β," in *IEEE Southeastcon 2009*, 2009, pp. 422–427.

[15] K. D. Donohue, J. Hannemann, and H. G. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments," *Signal Processing*, vol. 87, no. 7, pp. 1677–1691, 2007.

[16] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, p. 34–48, Mar. 2019. [Online]. Available: http://dx.doi.org/10.1109/JSTSP.2018.2885636

[17] K. Shimada, Y. Koyama, N. Takahashi, S. Takahashi, and Y. Mitsufuji, "Accdoa: Activity-coupled cartesian direction of arrival representation for sound event localization and detection," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 915–919.

[18] A. Politis, A. Mesaros, S. Adavanne, T. Heittola, and T. Virtanen, "Overview and evaluation of sound event localization and detection in dcase 2019," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, p. 684–698, 2021. [Online]. Available: http://dx.doi.org/10.1109/TASLP.2020.3047233

[19] E. Grinstein, C. M. Hicks, T. van Waterschoot, M. Brookes, and P. A. Naylor, "The neural-srp method for universal robust multi-source tracking," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 19–28, 2024.

[20] A. Khamis, *Optimization Algorithms: AI techniques for design, planning, and control problems*, 2024, p. 4-22. [Online]. Available: https://books.google.de/books?id=jI8eEQAAQBAJ

[21] B. Keinert, M. Innmann, M. Sänger, and M. Stamminger, "Spherical fibonacci mapping," *ACM Trans. Graph.*, vol. 34, no. 6, Nov. 2015. [Online]. Available: https://doi.org/10.1145/2816795.2818131