DOI: 10.61782/fa.2025.0269

DESIGN AND ANALYSIS OF MOBILE MICROPHONE ARRAYS FOR ACOUSTIC DRONE TRACKING

Martin Blass* **Franz Graf**
JOANNEUM RESEARCH Forschungsgesellschaft mbH
DIGITAL – Institute for Digital Technologies

ABSTRACT

The growing security threat posed by hostile drones (UAVs) highlights the urgent need for effective detection systems. Besides optical, radar or radio frequency systems, the acoustic domain offers significant but underutilized capabilities. Since human hearing is inherently limited in detecting and localizing weak or high-frequency sounds in noisy or dynamic environments, microphone arrays represent a significant improvement in tracking drone noise. In the literature, most of the existing acoustic systems are stationary and focus on the detection or localization of individual drones. The integration of mobile microphone arrays on vehicle roofs or soldier helmets can enhance military capabilities by enabling real-time tracking of drones in diverse and challenging scenarios. In this paper, we present the design criteria and development of a practical microphone array system capable of detecting and tracking multiple drones. First, we describe the implementation of a hemispherical array of 32 microphones. Then, we consider possible adaptations, e.g. in terms of reducing the number of microphones by selecting subarrays to address low-power consumption considerations and real-time processing for sound source localization. Finally, we show experimental results of a dataset recorded in a fixed and mobile scenario where the microphone array was mounted on a vehicle roof.

Keywords: *microphone, array, tracking, mobile, drone*

**Corresponding author*: *martin.blass@joanneum.at.*

1. INTRODUCTION

Within the recent years, drones and other UAVs have become an established technology in the fields of industrial surveillance, disaster management, public safety, logistics, agriculture and smart cities, to name a few. As commercial drones are available at low cost and do not require explicit expertise, they can also easily be used to intentionally or accidentally violate privacy, cause public nuisance through noise, or even pose a threat to society or public institutions when used as a means of attack. Recent military conflicts, such as those in Ukraine and the Middle East, along with increasing drone-related incidents, highlight the need for effective countermeasures against UAV threats. To enhance civilian security and military defense, numerous research projects and counter-UAV technologies have emerged, employing multi-sensor approaches that integrate radar, computer vision, radio frequency, and acoustic detection [1,2]. While often overlooked due to range limitations, acoustic sensing provides a cost-effective and essential tool for UAV detection, especially in scenarios where other modalities face challenges. Microphone arrays, in particular, offer a practical and low-cost solution for determining the incident angles of sound sources, e.g. drones, within a limited range and hemispheric coverage [3].

In order to localize drone sounds in terms of their incident direction, microphone arrays enable the use of direction-of-arrival estimation (DOAE) and beamforming (BF) methods [4]. In addition, sound source tracking algorithms may be applied to process frame-based DOA estimates and to reconstruct movements of multiple sound sources. In [5] a large microphone array with 64 microphones and 4 m diameter was placed on the ground and used to compare three DOAE algorithms: GCC-PHAT, conventional delay-and-sum BF and differential evolution. Successful localization results were obtained within 80 m. In [6] a 16-channel L-shaped array was chosen together with MUSIC as DOAE method for narrowband, broadband and proposed harmonic focus (2-6 kHz) scenarios. The authors of [7] investigated DOAE using incoherent and proposed coherent BF performed with a 64-channel MEMS microphone array in simulations and outdoor experiments within a range of 200 m. Similar contributions were published in [8, 9].

There is only a limited amount of publications combining both detection and localization to address the sound source association problem to successfully track UAVs while rejecting other sound sources. One promising example is described [10], where UAV noise recorded with a spherical microphone array is used to train a so called *BeamLearning-ID* network to mimic the BF process. Another joint detection and localization approach is described in [11]. In our previous works [12, 13] we presented a framework that enables the joint detection and localization of sound emitted by multiple UAVs in real-time. Similar to [11] we proposed a track-before-detect approach, where multiple hypotheses are generated using a GMM-based sound source tracking algorithm together with a classical ML model relying on drone specific audio features. To the best of our knowledge, our latest paper [13] is one of the first to address mobile UAV tracking, which is of great importance for military applications in view of the global geopolitical situation.

Wearable microphone arrays have been explored for various applications, including sniper detection, situational awareness, and sound event localization. [14] developed a helmet-mounted array for sniper detection, capable of real-time localization by analyzing Mach and muzzle waves from rifle shots. Their system demonstrated robust performance in urban environments with low false alarm rates. [15] presents a similar system with focus on determining the optimum number and position of microphones on a helmet by taking into account diffraction paths and intensity differences. [16] introduced a conformal helmet-mounted array designed to enhance auditory situational awareness while providing hearing protection. Their system integrates BF with head-related transfer functions to maintain spatial hearing and natural localization of sound sources despite protective headgear. [17] focused on a wearable sound event localization and detection (SELD) dataset, using 24 microphones placed on head-worn accessories like glasses and earphones. They evaluated deep learning-based SELD methods, showing that microphone positioning significantly impacts localization accuracy.

In this paper, we want to explore ways to enhance the efficiency of acoustic drone tracking for mobile applications in terms of reducing energy and computational requirements by tuning the number of sensors and the geometry of microphone arrays. First, we present the design and development of *JR-IcoDome32*, a hemispherical 32-channel microphone array that can be deployed in fixed or mobile scenarios, e.g. on vehicle roofs. Second, we analyze a scaled version that could be suitable as a human-worn sensor, e.g. integrated into a soldier's helmet. Then, we derive sub-arrays from the *JR-IcoDome32* to compare their DOAE performance. Finally, we evaluate these array configurations using a dataset collected during measurement campaigns for a fixed and a mobile scenario where the microphone array was placed on the ground and mounted on the roof of a vehicle.

The remainder of this paper is organized as follows: Section 2 explains the microphone array design process including sub-array configurations and the methods used for DOAE and sound source tracking, Section 3 describes the sensor system development and hardware integration, Section 4 presents the experimental setup and results, and Section 5 concludes the work with a future outlook.

2. METHODOLOGY

2.1 Microphone Array Design

Based on the existing knowledge of typical drone sounds and the requirements for mobile application [12, 13], we designed a microphone array suitable for the task of UAV localization: The array should have a rotationally symmetrical directional characteristic, be rigid, lightweight and weatherproof. In addition, a high directivity in the frequency range below 3 kHz and an attenuation of noise below the array in mobile operation (e.g. car engine, rolling noise) should be achieved, which led to a 3D geometry in the design process. The result is a geodesic dome of the type *Icosahedron V2* with a diameter d of 1 m and 26 microphones. By adding 6 more sensors in the base plane, 5 as a pentagon and one in the center, we obtain a 32-channel configuration, of which 16 sensors are placed in 2D and 3D, respectively. In the following we refer to this microphone array as *JR-IcoDome32*. Figure 1 shows the proposed array geometry with microphone positions as connected icosahedron edges and Figure 2 the resulting simulated spatial response given as point spread functions (PSF) for selected octave-band center frequencies and summed over the total frequency range up to 16 kHz.

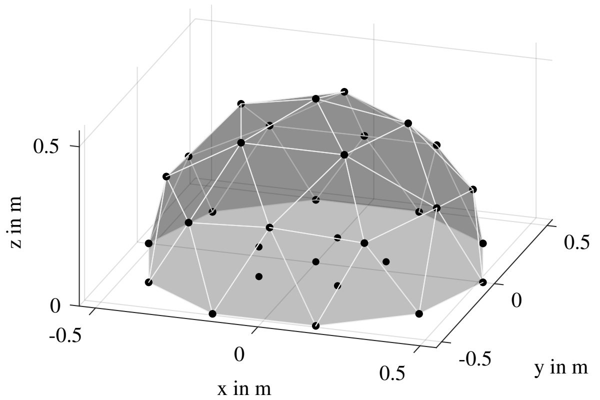


Figure 1: Array geometry of *JR-IcoDome32*.

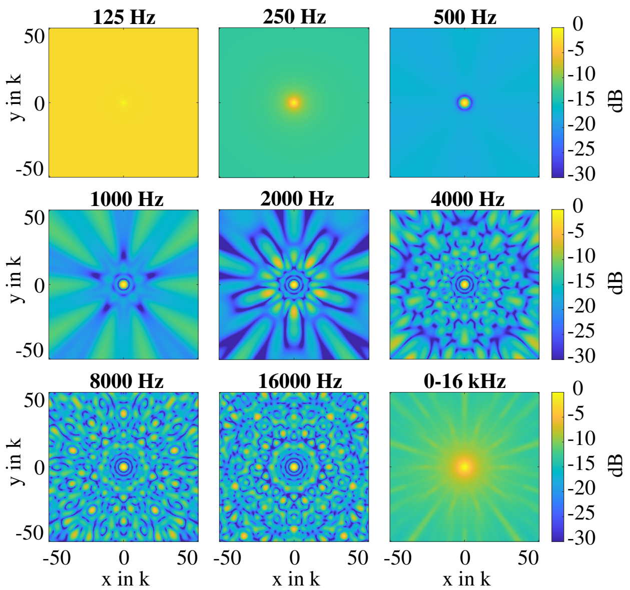


Figure 2: Spatial response as point spread function (PSF) of *JR-IcoDome32* for selected frequencies.

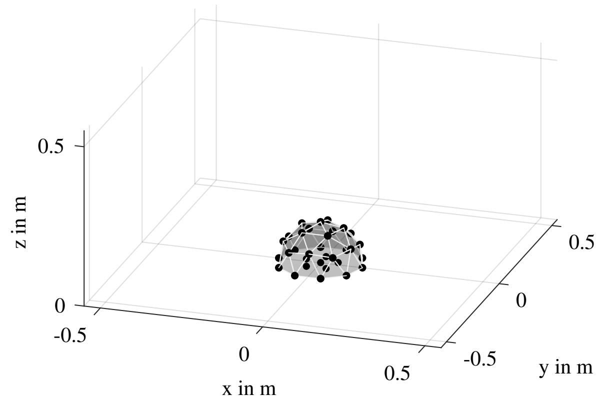


Figure 3: Array geometry of *JR-IcoDome32S*.

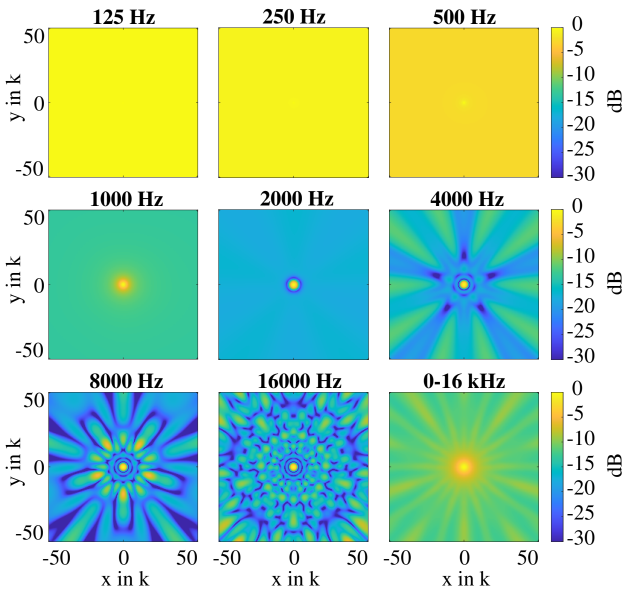


Figure 4: Spatial response of *JR-IcoDome32S* as a scaled version of *JR-IcoDome32* by a factor of 1/4.

In order to visually focus on the inspection of sidelobe levels while maintaining a constant mainlobe width, spatial grids of the simulated beam patterns are normalized by the wavenumber k, defined as $k = 2\pi/\lambda$, where λ denotes the wavelength. We choose a grid size of $k = 100$ in the scanned xy-plane at $z = d/2$, i.e. 0.5 m, over the array for evaluated frequencies. This leads to effective elevation angles of $1°$ and $71°$ for 125 Hz and 16 kHz, respectively.

Since we are not only interested in mobile applications for vehicles but also in portable and wearable sensor solutions, e.g. helmet arrays, we explore the spatial characteristics of a smaller scaled version of the *JR-IcoDome32*. For this purpose, we scale the array by a factor of 1/4 to obtain an aperture of $d = 0.25$ m, *JR-IcoDome32S*, which fulfills the requirements of being easily portable and fitting on a helmet. To make a fair comparison, we leave the geometry as it is and do not remove the six sensors inside the dome structure, which would be necessary when mounting the array on a helmet. Figures 3 and 4 show the geometry and spatial response results of *JR-IcoDome32S* for the same frequencies as shown for the original array configuration in Figure 2. Due to the scaling process, we observe that the same PSF patterns reappear at four times the frequency compared to *JR-IcoDome32*.

The rotational symmetry of the array allows us to easily derive sub-arrays from the original *JR-IcoDome32* geometry, called *dome32* for short: *base16* (2D) and *dome06*, *dome11*, *dome16* (3D). These configurations are shown in Figure 6 and are evaluated in terms of common array performance metrics in Table 1: f_{min} and f_{max} denote the theoretical, aliasing-free frequency range based on $\lambda/2$, HPBW is the half-power beamwidth, MSL and ASL are the maximum and average sidelobe levels, resp., and DI stands for the directivity index [4]. For this evaluation we use 201x201 points in k and average metrics over third-octave band center frequencies in 100-20000 Hz.

Table 1: Array metrics of *JR-IcoDome32* sub-arrays averaged over third-octave band center frequencies.

Metric	*dome06*	*dome11*	*dome16*	*base16*	*dome32*
f_{min} (Hz)	333	206	191	167	167
f_{max} (kHz)	2.80	3.68	5.28	7.21	19.73
HPBW (°)	25.94	21.49	22.89	21.13	21.80
MSL (dB)	-3.02	-3.95	-5.54	-7.95	-7.86
ASL (dB)	-5.53	-7.94	-9.64	-11.96	-13.02
DI (dB)	6.24	8.56	9.88	10.04	12.25

2.2 Direction-of-Arrival Estimation

As a DOAE algorithm we adapted a Gaussian Mixture Model (GMM)-based approach for clustering steered response power (SRP) estimates of a delay-and-sum beamformer with custom FFT frequency and spatial scanning grid settings as proposed in our previous work [12]. This enables probabilistic modeling, tracking of an unknown number of sound sources and source extraction using masked BF signal outputs. The algorithm provides good localization accuracy even with coarse scanning grids and the computational load enables real-time application. Details about the original algorithm can be found in [18].

2.3 Sound Source Tracking

To track potential sound source candidates based on DOA estimates, we use a simple but effective tracking approach to propagate GMM parameters over time as described in [18]: Since the number of active sources may change from frame to frame, each source candidate is assigned a time-to-live (TTL) as soon as it appears. The tracking algorithm then handles the three cases of birth, update and death of up to Q active sources. The birth of a source is controlled by a predefined source confidence threshold γ derived from the GMM parameters $\mathbf{\Sigma}$ and P, as well as by the shrink threshold Υ, i.e. the minimum angular separation between two sources. By using an age-token-based temporal smoothing method, the source locations $\hat{\boldsymbol{\theta}}$, the covariances $\mathbf{\Sigma}$ and the weights P are updated over time.

3. SENSOR SYSTEM

3.1 Microphone Array and Audio Hardware

The implementation of the array design *JR-IcoDome32* as described in Section 2 uses a rigid construction with a 6 mm solid plastic base plate and aluminum struts with 3D-printed corner connectors. ICP electret microphones with suitable signal conditioning and A/D converters were preferred over MEMS microphones due to better signal quality and less development effort in the hardware design. All components provide an IP67-rating to protect the sensors from water and dust. Figure 1 shows the final implementation of *JR-IcoDome32*.

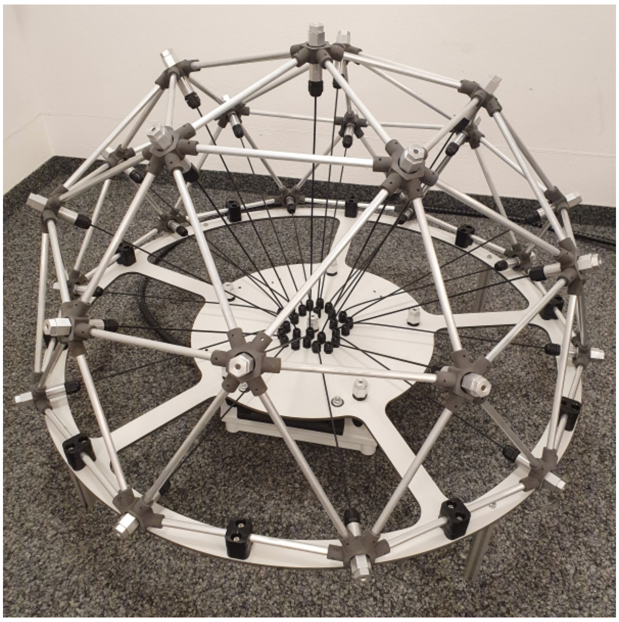
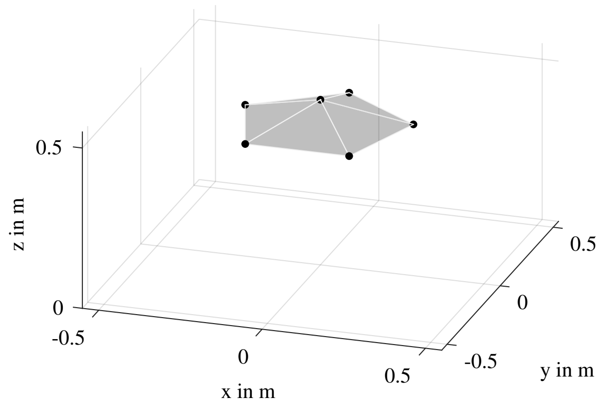
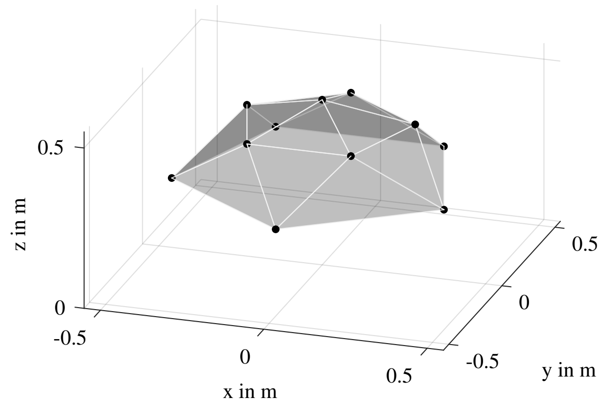


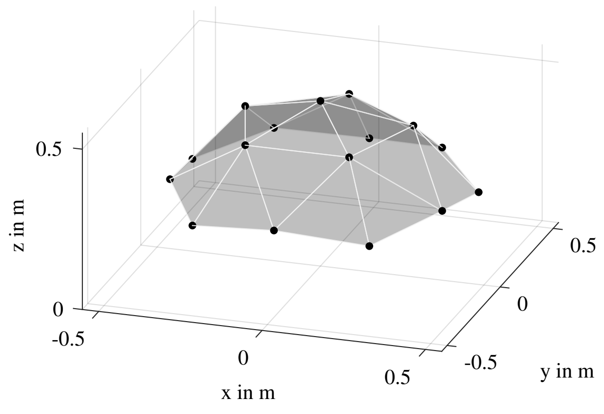
Figure 5: 3D microphone array *JR-IcoDome32*.



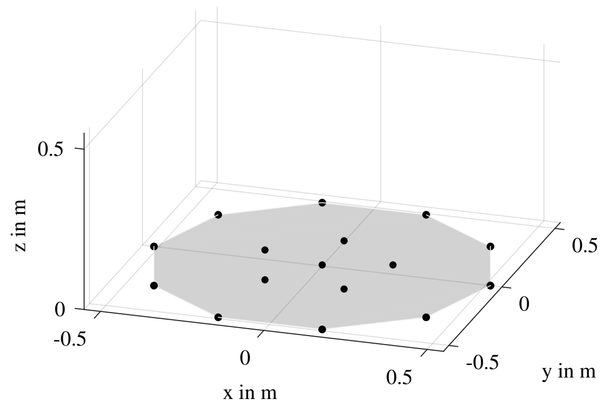
(a) Sub-array geometry of *dome06*.



(b) Sub-array geometry of *dome11*.



(c) Sub-array geometry of *dome16*.



(d) Sub-array geometry of *base16*.

Figure 6: Sub-array configurations derived from *JR-IcoDome32*.

3.2 Position and Orientation Sensors

The mobile use of a microphone array makes it necessary to determine the sensor position (GNSS) and orientation (IMU). This information is needed to transform DOA estimates from a relative to an absolute coordinate system in order to evaluate DOAE w.r.t. the position reference of a UAV. The following commercial sensors were procured:

- GNSS: *u-blox NEO-6M* with integrated antenna ,
- IMU: *Adafruit ADA2472* with *Bosch BNO055*.

The fusion and integration of these sensors was carried out using an *ESP32* micro-controller and a Python module as a serial interface with real-time data processing.

3.3 Vehicle Integration and Processing Hardware

The complete system consists of the *JR-IcoDome32* with spherical windshields, a suspension system of wire rope springs for a roof rack mount to dampen vibrations, the GNSS-IMU module mounted on the array base plate and an in-vehicle hardware rack. 32 BNC cables are routed as three DB-25 cables via a cable pass-through into the vehicle interior where they are connected to a rack containing USB audio interfaces synchronized via an AVB network and an *Intel NUC7i5DNK* processing unit. The system is powered using a portable power station and consumes approximately 120 W in operation. The integration into a vehicle is shown in Figure 7.

Figure 7: *JR-IcoDome32* with GNSS-IMU module mounted onto the rack roof of a car for mobile use.

Figure 8: Multicopter UAVs used for measurements.

Dataset	UAVs	Runs	Patterns	Duration
Fixed	3	11	3	1.5 h
Mobile	3	31	4	1.0 h

Table 2: Recordings of UAV flights conducted with *JR-IcoDome32* and resulting datasets for evaluation.

4. RESULTS

4.1 Experimental Setup

Audio data was recorded at 48 kHz sampling rate and 24 bit together with GNSS-IMU data logs for fixed and mobile setups at *UMFC Stocking* airfield in 2023. Three UAVs (*DJI Mini 2 SE*, *DJI Mavic 2 Pro*, *twinFOLD KAT*) were chosen for both scenarios, representing small, medium and large multicopters as shown in Figure 8. For the fixed setup, the *JR-IcoDome32* was placed on the ground, while the UAVs flew programmed and manual patterns (*hover*, *circle*, *free*) in distances up to 500 m from the array. In the mobile setup, the array was mounted on the car (Figure 7) and the UAVs performed manually controlled patterns (*hover*, *approach*, *follow*, *circle*) while the car drove at different speeds (10, 20, 30 km/h) in fixed gear on a dirt road. Each UAV was flown separately. Table 3 summarizes the recordings of fixed and mobile datasets.

4.2 System Parameters

For STFT processing, we resample the recorded audio signals to 24 kHz and set the window (Hann) and the FFT size to 2048, with a frame overlap of 50%. Considering the predominant characteristics of drone sound and the array metrics, we choose a frequency range of [300 2300] Hz resulting in 171 processed FFT bins. For the DOA scanning grid we choose a unit hemisphere of 200 quasi-uniformly distributed points [19]. The number of initial sound sources in each frame is 3. For source tracking we limit the maximum number of sources to 16 and set an initial and a maximum TTL for each source of 1.5 s, the minimum angular separation $\Upsilon = 30°$, the confidence threshold $\gamma = 0.3$ and specify an exponential smoothing constant of 0.9 corresponding to a time constant of 0.4 s.

4.3 Performance Evaluation

We present the DOAE results in terms of an exemplary run and a statistical evaluation of the DOAE error for all sub-array configurations. For each run and dataset, we compare the DOA estimates of three sound source tracks (SRC1-3) against the position reference (TRUE) obtained from GPS as azimuth and elevation angles. The opacity of each track represents the confidence output of the algorithm. Figures 9 and 10 show the results of a flight of *DJI Mavic Pro 2* in the fixed scenario for *dome06* and *dome32*, resp. In this run, the UAV took off next to the array and

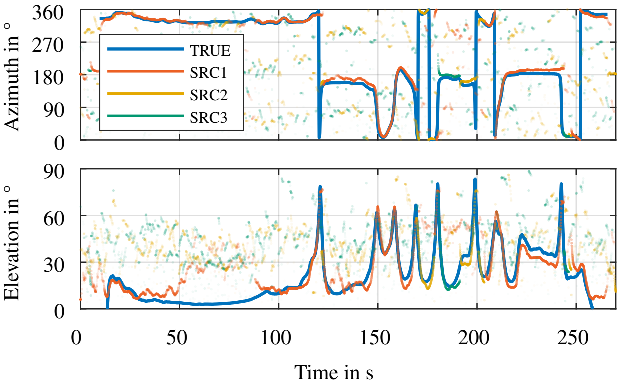


Figure 9: DOAE example with sub-array *dome06*.

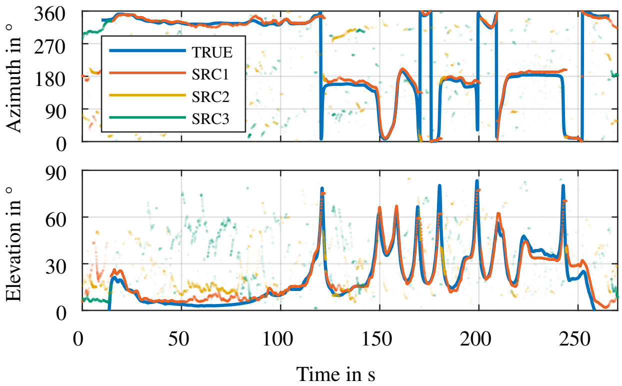


Figure 10: DOAE example with baseline *dome32*.

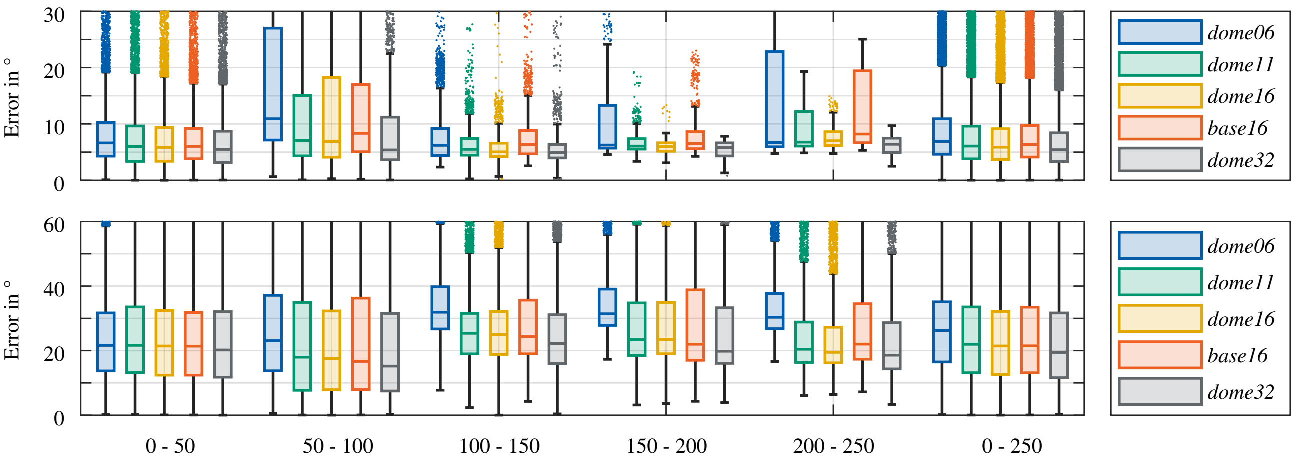


Figure 11: Absolute DOAE error over distance ranges for sub-arrays in fixed (top) and mobile (bottom) setup.

reached a distance of 100 m after 45 s and 260 m after 55 s. It then approached the array and passed 100 m at 90 s and 40 m at 100 s. For the rest of the run, the UAV flew within 80 m and performed overflights. It can be seen that *dome06* lost the UAV from 55-90 s, while *dome32* was able to keep the track (SRC1). In addition, this track is more stable, i.e. fewer changes between track hypotheses.

For each sub-array configuration, we assess the DOAE performance in terms of the absolute angle error for distance ranges in blocks of 50 m and up to 250 m in total. To provide a meaningful summary, we use boxplots to examine the median and interquartile ranges for each sub-array and dataset. Figure 11 presents median errors lower than 10° and 30°, except for *dome06*, for fixed and mobile datasets, respectively. It must be noted that GPS/IMU data may be inaccurate, as we use commercial systems without any enhancement, e.g. DGPS or RTK.

5. CONCLUSION

In this paper, we have presented a 3D microphone array design with possible sub-array configurations and experimental results for acoustic UAV tracking in a fixed and a mobile scenario. Our evaluations show similar results for sub-arrays with 11 or more sensors compared to the 32-channel baseline with 5° and 20° median DOAE error, resp. We argue that *JR-IcoDome32* is a suitable geometry for UAV tracking that also provides good accuracy in elevation. With sufficient drone noise content in the 1-4 kHz range, the proposed scaled version *JR-IcoDome32S* could provide similar performance. Possible future work lies in improving the DOAE and tracking algorithms, e.g. considering SRP amplitudes for weighted GMMs as proposed in [20], JIPDAF with Kalman filters [21], together with extensive evaluations using larger datasets.

6. REFERENCES

[1] M. A. Khan, H. Menouar, A. Eldeeb, A. Abu-Dayya, and F. D. Salim, "On the detection of unauthorized drones — techniques and future perspectives: A review," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11439–11455, 2022.

[2] A. Famili, A. Stavrou, H. Wang, J.-M. Park, and R. Gerdes, "Securing your airspace: Detection of drones trespassing protected areas," *Sensors*, vol. 24, no. 7, p. 2028, 2024.

[3] X. Yan, T. Fu, H. Lin, F. Xuan, Y. Huang, Y. Cao, H. Hu, and P. Liu, "Uav detection and tracking in urban environments using passive sensors: A survey," *Applied Sciences*, vol. 13, no. 20, p. 11320, 2023.

[4] H. L. Van Trees, *Optimum Array Processing*. John Wiley & Sons, 2002.

[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 *Proc. of the 29th Int. Congress on Sound and Vibration*, 2023.

[6] Y. Chen, Y. Chen, P. Chen, and W. Wang, "Acoustic localization using an l-shape microphone array for multirotor drones," in *IEEE 6th Int. Conference on Information Communication and Signal Processing*, pp. 1168–1172, 2023.

[7] M. Varela, W.-D. Wirth, and M. Oispuu, "Wideband direction-of-arrival estimation using microphone arrays," in *Symp. Sensor Data Fusion & Int. Conference on Multisensor Fusion Integration*, pp. 1–6, 2023.

[8] N. Itare, J.-H. Thomas, K. Raoof, and T. Blanchard, "Acoustic estimation of the direction of arrival of an unmanned aerial vehicle based on frequency tracking in the time-frequency plane," *Sensors*, vol. 22, no. 11, p. 4021, 2022.

[9] S. Wu, Y. Zheng, K. Ye, H. Cao, X. Zhang, and H. Sun, "Sound source localization for unmanned aerial vehicles in low signal-to-noise ratio environments," *Remote Sensing*, vol. 16, no. 11, p. 1847, 2024.

[10] 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*, 2022.

[11] Y. Sun, J. Li, L. Wang, J. Xv, and Y. Liu, "Deep learning-based drone acoustic event detection system for microphone arrays," *Multimedia Tools and Applications*, pp. 1–23, 2023.

[12] M. Blass and F. Graf, "A Real-Time System for Joint Acoustic Detection and Localization of UAVs," in *Quiet Drones*, 2020.

[13] M. Blass, S. Grebien, and F. Graf, "Experimental Evaluation of Acoustic Drone Tracking using Mobile Microphone Arrays," in *Quiet Drones*, 2024.

[14] S. Hengy, S. DeMezzo, and P. Hamery, "Sniper detection using a helmet array: first tests in urban environment," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, vol. 6562, pp. 269–278, SPIE, 2007.

[15] H. O. Park, A. A. Dibazar, and T. W. Berger, "Design of a helmet-mounted microphone array for sound localization," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing*, pp. 1478–1481, 2010.

[16] P. Calamia, S. Davis, C. Smalt, and C. Weston, "A conformal, helmet-mounted microphone array for auditory situational awareness and hearing protection," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 96–100, 2017.

[17] K. Nagatomo, M. Yasuda, K. Yatabe, S. Saito, and Y. Oikawa, "Wearable seld dataset: Dataset for sound event localization and detection using wearable devices around head," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing*, pp. 156–160, 2022.

[18] N. Madhu and R. Martin, "A Scalable Framework For Multiple Speaker Localization And Tracking," *Proc. of the International Workshop for Acoustic Echo Cancellation and Noise Control*, 2008.

[19] M. Deserno, "How to generate equidistributed points on the surface of a sphere," tech. rep., Max-Planck-Institut fuer Polymerforschung, Mainz, 2004.

[20] D. Frisch and U. D. Hanebeck, "Gaussian mixture estimation from weighted samples," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 1–5, 2021.

[21] D. Musicki and R. Evans, "Joint integrated probabilistic data association: Jipda," *IEEE transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 1093–1099, 2004.