BeBeC-2020-D05

DETECTION AND SEPARATE TRACKING OF SWARM QUADCOPTER DRONES USING MICROPHONE ARRAY MEASUREMENTS

Gert Herold, Adam Kujawski, Christoph Strümpfel,
Svenja Huschbeck, Maarten Uijt de Haag, Ennes Sarradj
Technische Universität Berlin
Germany

Abstract

The use of unmanned aerial systems (UAS) or drones for a wide variety of applications, such as law enforcement, search and rescue, agriculture, infrastructure inspection, mapping, filming and journalism has become common. As it can be expected that the number of applications will increase in the near future, the impact of these systems on the environment – urban or rural – is a rising concern. Furthermore, airports are faced with an increased security threat of unknown UAS penetrating airport areas and interfering flight operation.

This contribution aims at providing a tool for localizing and tracking simultaneously flying drones via acoustic measurements. Measurements of a swarm of four quadcopter drones using a 64-channel microphone array are evaluated with a multi-step process. First, the position of the drones are determined via short-time frequency domain beamforming. Based on the positions, an α-β filter is used to reconstruct the individual trajectories. The accuracy of the acoustically determined positions is evaluated by comparing it to positional data obtained from video recordings with a dual-camera setup.

1 Introduction

To this day, there are only few scientific studies dealing with noise effects and sound propagation properties of unmanned aerial systems (UAS). The noise emitted by UAS differs in many ways from conventional aircraft or road vehicle noise. A study conducted by the NASA Langley Research Center concluded that UAS noise was perceived by study participants as more annoying than road noise [4]. It can also be expected that UAS will generally operate closer to people on the ground and that their use will not be limited to a specific area (e.g. around

airports). Against this background it must be assumed that previous noise certification procedures – e.g. according to ICAO Annex 16 [7] – are only partially applicable to the new and very specific UAS noise.

A study performed by the German Federal Environmental Agency investigated sound emissions of multiple quadcopter UAS with a maximum take-off mass below 2 kg, including (psycho-)acoustic measurements during both indoor and outdoor flight campaigns as well as the definition of horizontal and vertical noise directivity patterns [14]. The vulnerability of the measuring method for determining the sound power level according to ISO 3744 [8] for slight changes of the 3D position was identified. Furthermore, the use of the aforementioned measurement method to determine a reproducible sound power level is only useful for UAS in hovering, which represents only a very specific operating mode. Typically, such devices are also intended to perform translatory movements.

Performing measurements including flight tests, on the other hand, poses challenges regarding the reproduciblity or tracking of the flight paths, which is necessary for a correct source characterization. Such tests usually have to be performed in open spaces, where additional noise sources might be present. This makes it necessary to additionally filter signals emitted by the measured object. Moreover, with increasing number of applications for UAS, monitoring the noise impact caused by the drones will become important. If several drones are present in the vicinity, it is necessary to acoustically separate the immissions from the individuals, i.e. follow multiple trajectories at once.

In what follows, a method for reconstructing trajectories of multiple simultaneously flying quadcopter drones from microphone array measurements is presented. The method was tested in an anechoic chamber with up to four UAS flying at the same time in five different flight scenarios. The accuracy of the acoustic position detection is quantified by calculating the deviations to positions detected using a dual-camera setup for 3D localization. Finally, coherent trajectories with time-dependent position and velocity information are calculated and attributed to the inividual drones.

2 Materials and methods

2.1 Measurement setup

Measurements were performed in the anechoic chamber at the Department of Engineering Acoustics at TU Berlin. For acoustic recording, a planar array consisting of 64 wall-mounted microphones with an aperture of 1.5 m was oriented towards the ceiling. While not ideal for 3D localization, this setup ensures that the equipment does not interfere with the flight paths. Battista et al. [3] showed that source positions in 3D can be obtained from measurements with a planar array even though the resolution capabilities in z direction is lower than in lateral direction.

Additionally, synchronous video recordings were done with two cameras, with one of the cameras positioned at the array center with the same orientation as the array and the second camera fixed to the chamber wall, with its principle axis pointing towards the focus area above the array. The setup as viewed from above is depicted in Figure 1.

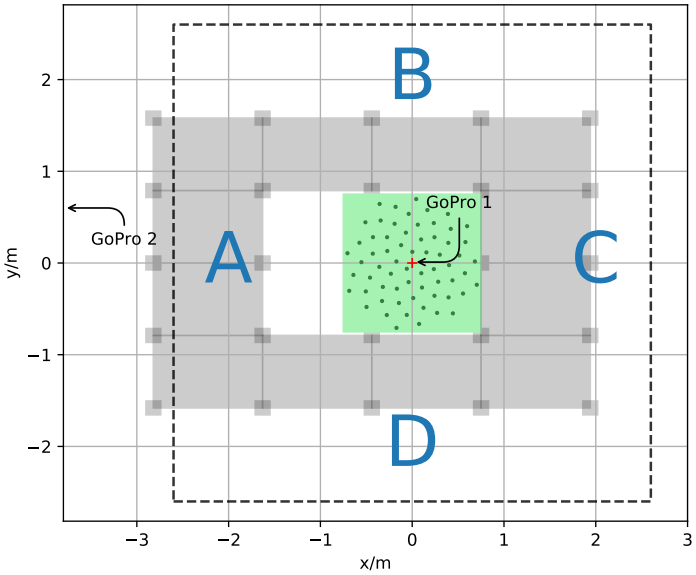

Figure 1: Measurement setup with the 64-channel microphone array and the FATO sites A-D. The dashed line marks the x-y-extent of the area monitored by the array.

2.2 Measurement object

Four small quadcopter UAS (Type: Tello EDU, Manufacturer: Ryze Technology, see Fig. 2) were used for the flight experiments, enabling remote-controlled as well as pre-programmed auto piloted swarm-like flights above the microphone array. The Tello EDU is a small UAS (sUAS), with a weight of 87 g (incl. propellers and batteries), sizing 9.8 cm by 9.2 cm (without propellers and protection guards) and a propeller diameter of 7.6 cm. As the Tello EDU only uses optical flow sensors and an inertial measurement unit for inflight positioning, it is suitable for indoor flights and comparable small flight distances, where global navigation satellite systems like GPS or GALILEO are not available.

All sUAS were integrated into a WiFi network and controlled via a Python-based programming interface. In this experimental setup, the Tello EDU only accepts vector- and speed-based control inputs (up-, down-, for- and backwards, yaw along z axis, speed control), thus an implementation of flight plans consisting of several 3D coordinates and time was not possible.

We designed five flight missions above the array, from single flights to swarm flights using four sUAS. The final setup of final approach and take-off (FATO) sites (A, B, C, D) is shown in Figure 1. Due to suboptimal flight conditions (e.g. low light and uneven floor), the navigation performance in vertical and lateral extension repeatedly deviated from the target flight path.

- Scenario 1: One sUAS is taking off from FATO A upwards to 1.3 m flight height and flies towards FATO C, where it lands.
- Scenario 2: Two sUAS are taking off from FATO A and FATO C respectively, upwards to 1.3 m flight height, and simultaneously fly in opposite directions above the array center towards the opposite FATO.

Figure 2: Tello EDU (left) and preparation of measurement flights with 4 identical Tello EDU sUAS (right)

- Scenario 3: Two sUAS are taking off from FATO A and FATO C respectively, upwards to 1.3 m flight height, and simultaneously fly in opposite directions towards the array center, where a static hover maneuver with a 90° turn towards FATOs B and FATO C is performed.
- Scenario 4: Four sUAS are taking off from FATO A-D and fly at different flight heights to the opposite FATO.
- Scenario 5: Four sUAS fly remote-controlled along arbitrary paths.

Schematics for the first 4 scenarios are given in Fig. 3.

2.3 Flight path reconstruction

Flight records with 4D-trajectory information (3D position + time) of the sUAS are not available directly but could be calculated based on the parsing raw measurements of the IMU (translatory acceleration and rotation rates). However, due to drift effects and non-accurate timestamps this approach is not accurate enough to determine the true position above the array. Hence, flight paths of the drones were only reconstructed by evaluating video and audio recordings respectively.

Optical tracking

To visually monitor the trajectory of the moving drones during the experiment, consumer-grade cameras (GoPro Hero 7 black) were used in a stereo setup. Each of the cameras covered a different view of the volume of interest, however their exact alignment was not known a-priori.GoPro 1 was positioned on the array, with its field of view covering the ceiling of the chamber; GoPro 2 was mounted on the wall in “A”-direction, oriented towards the opposite wall. Figure 4 shows the views of the respective cameras.

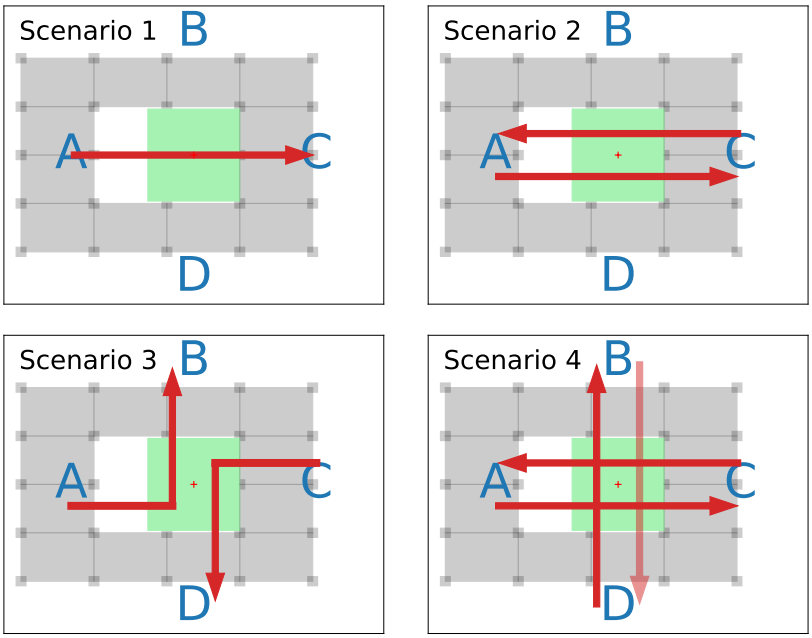

Figure 3: Pre-programmed flight scenarios of the drones.

A shooting mode with an aspect ratio of 4:3 (2704 × 2028), 2.7k resolution, and a frame rate of 50 fps was selected to capture the spatially limited recording area. The wide field of view of the camera's fisheye lens made it possible to observe a large area in the volume of interest. However, the images contained additional radial and tangential distortion.

In order to use the measured 2D images to reconstruct the drones' flight path in the real world 3D coordinate system, the lens distortion and the relative positioning, including rotation and translation of the camera views, must be known. The determination of unknown lens parameters (tangential, radial distortion coefficients, focal length, principal points) and alignment of optical sensors is an often encountered problem in computer vision [6], widely known as camera calibration.

As described by Abdel-Aziz et al. [1], the mapping from measured image points to object space involves a two-step transformation, which they introduced as Direct Linear Transform (DLT). The DLT coefficients can be determined using the Sparse Bundle Adjustment (SBA) method [10], which solves the nonlinear least-squares problem based on multiple measurements of a calibration object with known position in the field of view projected on the image plane. Due to the size of the observation volume, no object that completely fills the measuring field and has known 3D dimensions could be used for calibration. Instead, two blue balls attached to a rod with fixed and known distance, often referred to as a "wand" in the literature [13], were moved in the field of view as shown in Figure 4. A total of 200 frames with paired points was used for optimization.

The edges of several visible absorbing wedges and the lamps were used as additional 64 points that project from the real world coordinate system on the image plane of the two cameras for optimization. For tracking of the wand and calculation of the DFT-coefficients via SBA, the Python-based software *Argus* was used [9]. The world coordinate frame was chosen to

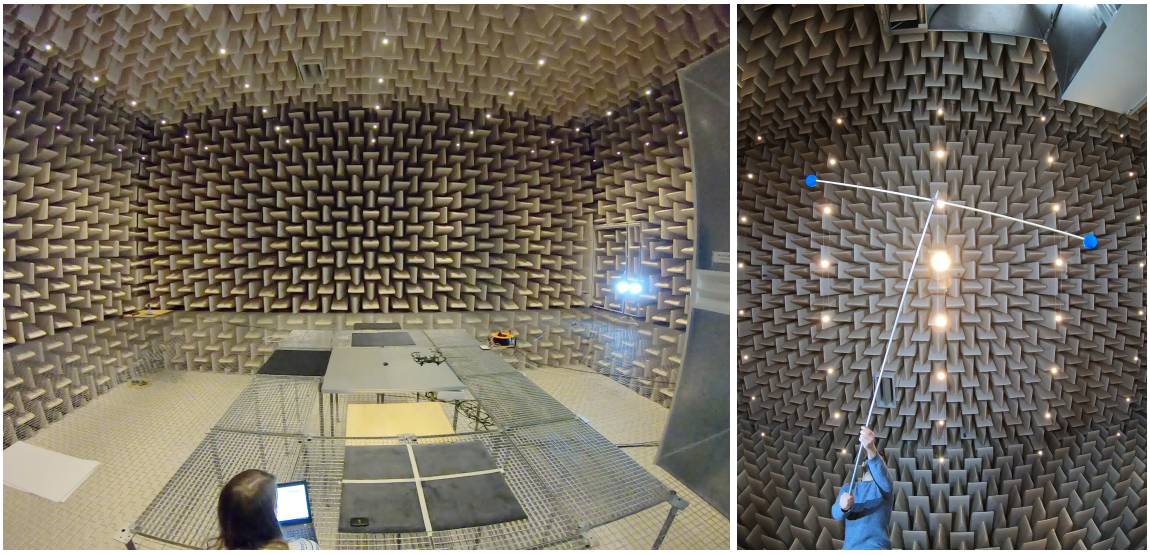

Figure 4: Fields of view of the wall-mounted GoPro 2 (left) and the GoPro 1 positioned at the array center including camera calibration procedure with the "wand" (right).

be aligned with that of the camera positioned on the microphone array. To ensure that the camera coordinate system also matches that of the microphone array, a small loudspeaker was positioned at a distance of 1 m to the center of the array and used as visual an acoustic reference point. The center of the camera coordinate system was then adjusted accordingly in x, y and z direction so that the position of the reference point in the camera coordinate system coincides with that of the microphone array. It should be noted that this step only accounts for a possible translation between the two coordinate systems but did not compensate any rotation. However, based on the comparison of acoustically obtained trajectories and those determined from image data, it is assumed that the rotation is negligible.

Acoustic tracking

The reconstruction of the flight paths from microphone array measurements is done in a multi-step process. In a first step, the recording is divided into short tracks of $\Delta t = 0.1\,\mathrm{s}$ length, with the assumption that within that time, a moving drone has changed its position sufficiently little to still be detectable with stationary frequency domain beamforming.

Using Welch's method [16] with a block length of 512 and an overlap of 50 %, a cross-spectral matrix (CSM) is estimated for each discrete frequency by:

$$\boldsymbol{C} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{p}_k \boldsymbol{p}_k^{\mathrm{H}}\,, \tag{1}$$

with the number of averaged cross-spectra $K = 19$ in this case. The vector $\boldsymbol{p}_k \in \mathbb{C}^M$ contains the complex spectral data for each of the M microphones.

For beamforming in the frequency domain, the sound propagation model (i.e. shifting the phase and correct the amplitude according to distances of microphones to focus points $r_{s,m}$) is

Table 1: Measurement and data processing parameters for the beamforming.

Total number of microphones	64
Array aperture	1.5 m
Sampling rate	51 200 Hz
Focus grid (excl. mirror) ($l_x \times l_y \times l_z$)	5.2 m × 5.2 m × 1.7 m
Focus grid resolution	0.05 m
FFT window	512 samples, von Hann
Averaging time	0.1 s
Beamforming	Functional Beamforming ($\nu = 8$)
	Removed main diagonal of CSM
Reference position	$r_{s,0} = 1\,\mathrm{m}$

described via the steering vector $\boldsymbol{h}$, whose entries are calculated by

$$h_m = \frac{1}{r_{s,m}\sqrt{M\sum_{l=1}^{M} {r_{s,l}}^{-2}}}\,\mathrm{e}^{-\mathrm{j}k(r_{s,m}-r_{s,0})}\,, \qquad m = 1\ldots M\,, \tag{2}$$

with $s = 1\ldots N$ being the focus positions of interest. The reference distance at which the levels are evaluated is set to $r_{s,0} = 1\,\mathrm{m}$ for all focus points. The formulation in (2) ensures the correct detection of a local maximum at the location of a source [11].

For this step, the major objective consists of finding the sources, whereas the exact quantitative information about its strength is not of importance. Therefore, the main diagonal of the CSM is set to zero (as are its resulting negative eigenvalues) and Functional Beamforming [5] can be applied:

$$b(\boldsymbol{x}_s) = \left(\boldsymbol{h}^{\mathrm{H}}(\boldsymbol{x}_s)\,\boldsymbol{C}^{\frac{1}{\nu}}\,\boldsymbol{h}(\boldsymbol{x}_s)\right)^{\nu}\,. \tag{3}$$

The exponent $\nu = 8$ ensures sufficient side lobe suppression. The beamforming is done using the Python package *Acoular* [12] for a three-dimensional, regularly discretized area in which the drones are to be tracked. Acoustic data processing details are summarized in Table 1.

In the next step, local maxima in the focus area are detected using the multi-dimensional maximum filter from the *SciPy* package [15]. To avoid false positives, a hard level threshold below which source candidates are discarded, is set. Furthermore, a slight position resolution enhancement is achieved by calculating the center of mass of the focus grid point containing the maximum and its neighboring points (including up to three neighbors in every direction).

The above steps are repeated for every time step, yielding positions where acoustic sources are detected, which are assumed to be attributed to a drone. A quantitative comparison of the positions calculated from the video recordings to those calculated from the microphone array measurements is done in Section 3.1.

The extracted positional information is, however, still insufficient in two aspects. Firstly, it is not ensured that a movement along the reconstructed points is physical, and secondly, for the case of multiple drones being present in the monitored area, it is necessary to attribute the points to the respective trajectories of the individuals. Both problems are assessed via a two-step process consisting of trajectory prediction combined with solving a linear assignment problem (LAP), based on the method described by Wu et al. [17].

Let the time-dependent state $\boldsymbol{x}(t)$ of a moving object be described by its 3D position and

velocity:

$$\boldsymbol{x}(t) = (x, y, z, u, v, w)^T \tag{4}$$

With the assumption that the current state exclusively depends on the previous state, one can postulate that

$$\boldsymbol{x}(t) = \boldsymbol{F}\,\boldsymbol{x}(t-1) + \sigma_\mathrm{p}(t)\ , \tag{5}$$

with the transition matrix

$$\boldsymbol{F} = \begin{pmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{6}$$

and some process noise σ_p. The observed quantities are only the current positions $\boldsymbol{z}(t)$, which are described by

$$\boldsymbol{z}(t) = \boldsymbol{H}\,\boldsymbol{x}(t) + \sigma_\mathrm{m}(t)\ , \tag{7}$$

with the observation matrix

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{8}$$

and the unknown measurement noise σ_m.

With a known state $\boldsymbol{x}(t-1)$ and following Eq. (5), the current state can predicted via

$$\hat{\boldsymbol{x}}(t) = \boldsymbol{F}\,\boldsymbol{x}(t-1)\ . \tag{9}$$

This prediction may not necessarily agree with the associated observation $\boldsymbol{z}(t)$, however, the prediction can be updated to estimate the true state with

$$\boldsymbol{x}(t) = \hat{\boldsymbol{x}}(t) + \boldsymbol{G}(\boldsymbol{z}(t) - \boldsymbol{H}\,\hat{\boldsymbol{x}}(t))\ . \tag{10}$$

Equations (9) to (10) describe a Kalman filter [2]. While it is possible to determine an optimal Kalman gain $\boldsymbol{G}$, an efficient simplification is given by the α-β filter with fixed coefficients:

$$\boldsymbol{G} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \\ \frac{\beta}{\Delta t} & 0 & 0 \\ 0 & \frac{\beta}{\Delta t} & 0 \\ 0 & 0 & \frac{\beta}{\Delta t} \end{pmatrix}\ , \tag{11}$$

where α and β are weight factors between measurement and prediction. In this case, values of $\alpha = \beta = 0.2$ yield satisfactory results.

Before applying Eq. (10), the observation $\boldsymbol{z}(t)$ associated to the prediction $\hat{\boldsymbol{x}}(t)$ has to be identified. This is trivial in case of a single drone, but can be challenging in case of multiple drones flying in close vicinity of each other. Therefore, a linear assignment problem (LAP) is

Table 2: Parameters for the trajectory generation.

Source level threshold	0 dB
Max distance of trajectory points	0.83 m
Max. number of frames w/o source	5
Min. number of valid trajectory points	5
Smoothing filter	$\alpha = \beta = 0.2$

formulated with a cost matrix $\boldsymbol{D}$:

$$D_{ij} = \|\boldsymbol{z}_i(t) - \boldsymbol{H}\,\hat{\boldsymbol{x}}_j(t)\|\ , \tag{12}$$

where i and j denote the specific observations and predictions respectively. The LAP is solved using *SciPy*'s linear sum assignment implementation [15].

This process is repeated for every time step, including all detected maxima. If a maximum can not be associated with an existing trajectory, a new trajectory state is initialized. Furthermore, if no associated points can be found for an active trajectory, the drone is assumed to remain at constant speed for a fixed number of time steps. If, after that time, no further valid points are detected, the trajectory is set to inactive and the presumed trajectory states are discarded. Only trajectories with a minimum number of points are regarded as valid. Table 2 summarizes important parameter for the trajectory generation.

3 Results

The measured drones emit a very characteristic humming noise while flying. Figure 5 shows an averaged narrow-band spectrum ($\Delta f = 12.5\,\mathrm{Hz}$) and spectrogram for the single drone of Scenario 1, measured at the center-most microphone of the array. The spectrum features strong tonal components, with multiples of the 558 Hz blade passing frequency (BPF). With two rotor blades on each rotor, this amounts to an average rotor speed of 16 740 rpm . Around second 13, the drone passes the microphone, inducing flow noise on the microphone. This is visible in the spectrogram through an increase of the level at the lowest frequencies.

For the acoustic detection of the position, small frequency bands of 558 Hz bandwidth are used. With the block size of 512, this amounts to an evaluation of 6 discrete frequencies per band. The center frequencies for the evaluations in the following section are multiples of 2, 5, 8, 10, 12, and 15 times the BPF.

3.1 Positional accuracy

For the verification of the accuracy of the drone tracking by the array, the drone positions obtained from short-time beamforming of scenario 1 were compared to the positions yielded by the dual-camera system. It should be mentioned that the camera data can not serve as an exact ground truth of the trajectory, since the accuracy of the camera system itself highly depends on the quality of the calibration, which was not validated in the case and can also vary spatially. However, since the two methods for path detection work completely independent of each other – one relying purely on video while the other only evaluates acoustic data – both results can be compared against each other to quantify the uncertainty.

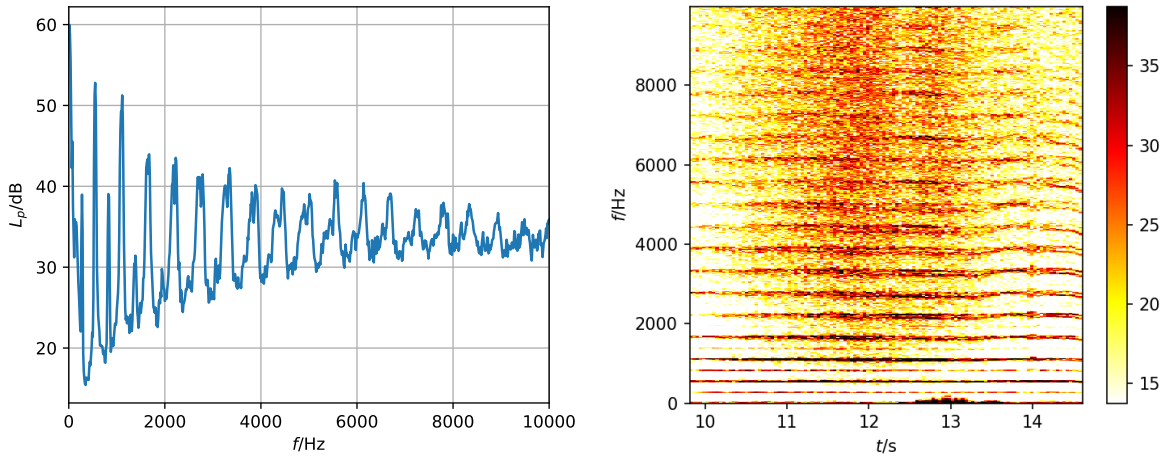

Figure 5: Narrow band spectrum (left) and spectrogram (right) of a single-drone flight.

Acoustic positions were calculated with 10 frames per seconds, optical positions with 50 frames per seconds. Frame recording was not synchronized initially. Therefore, the position determined with the microphone array at one respective time step was compared with the position from image data at the closest time step.

Table 3 shows the average deviation of the acoustically obtained positions along x, y and z direction over a number of snapshots for different frequencies.

Table 3: Mean value and standard deviation of drone x, y, z detected via camera vs. array setup.

f/Hz	μ_x/m	μ_y/m	μ_z/m	σ_x/m	σ_y/m	σ_z/m
1116	−0.085	0.021	−0.024	0.126	0.029	0.077
2790	−0.059	0.015	−0.048	0.077	0.015	0.027
4464	−0.072	0.018	−0.026	0.078	0.018	0.033
5580	−0.069	0.016	−0.034	0.072	0.016	0.030
6696	−0.086	0.016	−0.044	0.088	0.015	0.035
8370	−0.075	0.016	−0.035	0.077	0.016	0.030

Along the x direction, slightly higher deviations can be observed on average than along the other directions. However, the mean spatial difference is always below 9 cm. A very good fit with deviations of about 2 cm can be observed along the y dimension. In general, the accuracy of the tracking does not vary a lot for the tested frequencies. At the lowest evaluated frequency of 1116 Hz, the standard deviation increases in all directions.

Figure 6 shows detected points for both methods. As image processing was done at a higher frame rate than acoustic processing, more positions could be determined with the cameras. While the results are in very good agreement towards the C side, on the A side the positions appear to deviate more. This could be attributed to the camera calibration being mostly done with video recordings where the wand was on this side. The comparison proves that for this setup, acoustic tracking can be used as an alternative with similar accuracy as camera tracking.

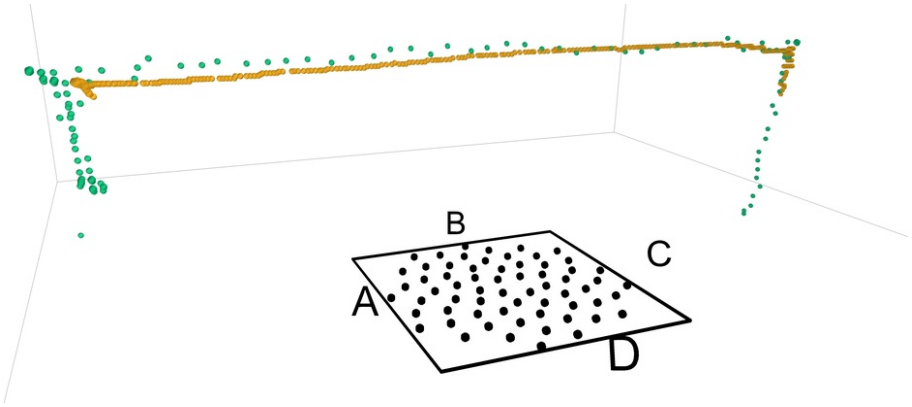

Figure 6: Detected positions of the drone flying Scenario 1. Orange: from video recordings, green: from array measurements (evaluated at 5580 Hz).

3.2 Reconstruction of drone trajectories

For the trajectory detection of the different scenarios, all evaluations were done with the array following the multi-step data processing described in Section 2.3 at a frequency of 5580 Hz.

The calculated trajectory plots are shown in Figures 7 to 11. The lines indicate the calculated flight paths, with their color (blue to yellow) matching the velocity of the drone at that position. The corresponding scale is given on the vertical colorbar. Each trajectory line begins and ends with a colored sphere, whose color (from white to red) indicates the instance in time at this position. The corresponding horizontal colorbar shows the respective times, which do not necessarily start at zero, since the evaluated acoustic measurement may have started some time before the event of interest happened.

Trajectories beginning after or ending before the start or end of the recording time may indicate several phenomena: drones starting or stopping to fly, drones leaving the monitored area, multiple sound sources on the drones that are sufficiently afar from each other to be detected as individual maxima, or changes in the flying configuration such that in the observed frequency band not enough energy is present to be considered as source by the algorithm. Furthermore, a sudden change of the velocity may lead to the predicted position to be too far from the actual one, so that the flight path of one drone may be cut into several trajectory pieces.

Several views of the trajectory of Scenario 1 are shown in Fig. 7. The recording started shortly after take-off, when the drone is hovering for a couple of seconds before it climbs another 50 cm and commences its path from FATO site A to C, where it lands. As can be seen in the side view from B towards D, the horizontal part of the path is not completely straight. This may either be caused by the drone adjusting its path when its sensors encounter the array or a slight positional error due to an incorrect sound propagation model – e.g. the flow caused by the drone itself is not accounted for. This particular characteristic is not observed in the other scenarios.

Additionally, the climb and landing paths appear to be somewhat slanted away from the array. This is not plausible and was also not physically observed. It can therefore be concluded that this is due to a mapping error by the algorithm, which may be caused by the elongated point spread function lateral sources exhibit at large angular deviations from the array axis, or by different visibility of noise-emitting parts of the drones from the microphones. In the other

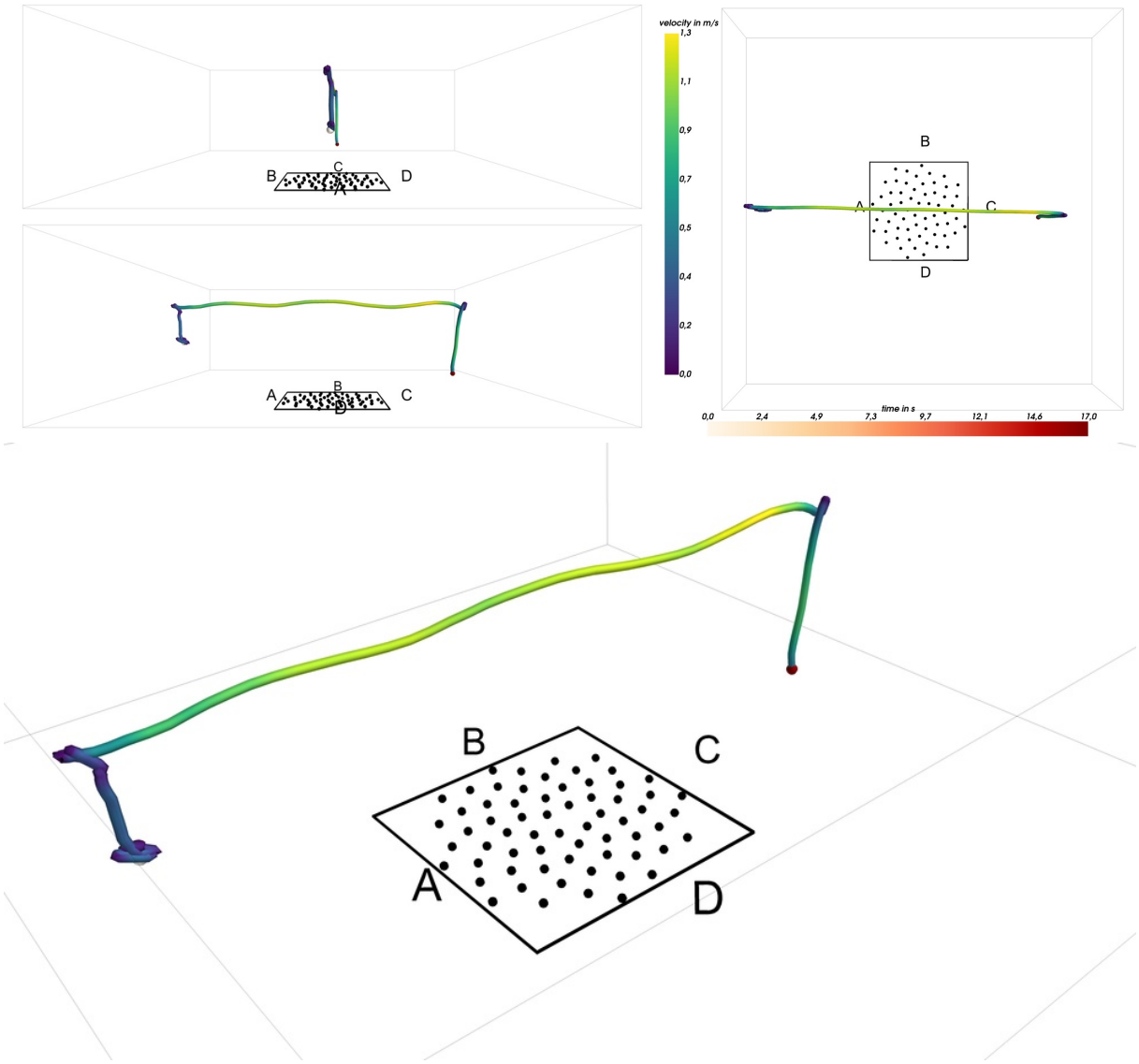

Figure 7: Scenario 1: Single drone flying from FATO A to FATO C.

scenarios, the apparent slanting of vertical climb or descent trajectories at lateral positions can be seen as well.

Scenario 2 is depicted in Fig. 8. The acoustic tracking starts with the drones being already in-flight to the opposite sites. The calculated trajectories are more straight than in the previous case, which may be influenced by the flight height of the two drones being (inadvertently) lower than that of the single drone.

Scenario 3 featured two opposite drones flying towards the center an performing a 90° left turn in front of each other (Fig. 9). Apart from different velocity profiles the drones' trajectories are very similar. It is visible that the drone flying from FATO A to B exhibited a more unsteady hovering behavior while performing the turn.

In general, all pre-programmed flight maneuvers performed as intended during the prelimi-

nary tests. However, the poor lighting conditions and the uneven floor in the anechoic chamber led to the drones repeatedly deviating from their projected behavior.

This becomes most apparent for Scenario 4, where one of the drones did not take off. As shown in Fig. 10, the remaining three drones performed as expected. However, the reconstructed trajectory of the drone flying from A to C, with its path being the most elevated of the three, shows several peculiarities. Firstly, the trajectory during hovering was split into three parts. Secondly, the path over the array seems to considerably vary in height, which was not visually observed. Aside from the flight height, this drone intersecting with the drone flying from D to B (in (x,y) positions) at almost the same time might impact the position detection as well.

Finally, two segments of the recording of Scenario 5 are presented in Figure 11. Each of the four drones was individually remote-controlled by a person, and flying was not constrained to the monitored area, which was however extended by 1 m in z direction. As is visible, the reconstructed path is interrupted several times[1]. Apparently, the filter configuration with fixed coefficients α and β is not sufficient for obtaining a continuous trajectory in this case. However, in general, the found trajectories are plausible and match the visual observations.

3.3 Computational cost

All calculations were done on a DELL XPS 13 9360 Laptop with 8 GB RAM and an Intel i5-8250U 4×1.60 GHz CPU. The detection of the source candidates with beamforming on the extensive grid is by far the computationally most costly part of the method, with a runtime of ≈ 8 s per time step.

Acceleration, e.g. for real time application, could be achieved by using an adaptive focus grid definition, less sensors, and by calculating the functional beamforming result with only one discrete frequency or for less time steps.

4 Conclusion

A method for positional tracking of UAS using microphone array measurements was presented. Although the resolution capability of a planar array is limited, it has been shown to be sufficient to track multiple drones flying in close vicinity to each other and even crossing paths. The used filter setup allowed a fast calculation, but led to short-time interruptions of some trajectories.

A comparison with position detection from video recordings has shown that acoustic positioning can perform similarly. Several flight scenarios were tested, ensuring the robustness of the method. Therefore, this technique can be applied for cases where the trajectory of moving sound-emitting objects is of interest, be it as redundant unit to video monitoring or stand-alone system.

Furthermore, the reconstructed trajectories could serve as basis for time domain beamforming methods to isolate the signal of individual drones, which could be realized with the same array.

[1] It should be noted that the two most-closely spaced dark-red points between the letters C and D in the upper plot do *not* mark a falsely detected trajectory interruption but an actual collision of two drones in mid-air.

References

[1] Y. I. Abdel-Aziz, H. M. Karara, and M. Hauck. "Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry." *Photogrammetric Engineering & Remote Sensing*, 81(2), 103–107, 2015. ISSN 0099-1112. doi:10.14358/PERS.81.2.103.

[2] B. D. O. Anderson and J. B. Moore. *Optimal filtering*. Prentice-Hall, Englewood Cliffs, N.J., 1979. ISBN 0136381227.

[3] G. Battista, P. Chiariotti, G. Herold, E. Sarradj, and P. Castellini. "Inverse methods for three-dimensional acoustic mapping with a single planar array." In *Proceedings of the 7th Berlin Beamforming Conference*, pages 1–23. 2018.

[4] A. Christian and R. Cabell. "Initial investigation into the psychoacoustic properties of small unmanned aerial system noise." *23rd AIAA/CEAS Aeroacoustics Conference, 2017*, (June), 2017. doi:10.2514/6.2017-4051.

[5] R. P. Dougherty. "Functional beamforming." In *Proceedings of the 5th Berlin Beamforming Conference*, pages 1–25. Berlin, 2014.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, USA, 2 edition, 2003. ISBN 0521540518.

[7] ICAO. "ANNEX 16 to the Convention on International Civil Aviation, Environmental Protection. Volume I - Aircraft Noise. 5th edition, incorporating Amendments 1-9." 2008.

[8] ISO. "ISO 3744:2010 – Acoustics – Determination of sound power levels and sound energy levels of noise sources using sound pressure – Engineering methods for an essentially free field over a reflecting plane.", 2010.

[9] B. E. Jackson, D. J. Evangelista, D. D. Ray, and T. L. Hedrick. "3D for the people: multi-camera motion capture in the field with consumer-grade cameras and open source software." *Biology Open*, 5(9), 1334–1342, 2016. doi:10.1242/bio.018713.

[10] M. I. A. Lourakis and A. A. Argyros. "SBA: A Software Package for Generic Sparse Bundle Adjustment." *ACM Trans. Math. Softw.*, 36(1), 2009. ISSN 0098-3500. doi:10.1145/1486525.1486527.

[11] E. Sarradj. "Three-Dimensional Acoustic Source Mapping with Different Beamforming Steering Vector Formulations." *Advances in Acoustics and Vibration*, 2012, 1–12, 2012. ISSN 1687-6261. doi:10.1155/2012/292695.

[12] E. Sarradj and G. Herold. "A Python framework for microphone array data processing." *Applied Acoustics*, 116, 50–58, 2017. ISSN 0003682X. doi:10.1016/j.apacoust.2016.09.015.

[13] D. H. Theriault, N. W. Fuller, B. E. Jackson, E. Bluhm, D. Evangelista, Z. Wu, M. Betke, and T. L. Hedrick. "A protocol and calibration method for accurate multi-camera field videography." *Journal of Experimental Biology*, 217(11), 1843–1848, 2014. ISSN 0022-0949. doi:10.1242/jeb.100529.

[14] J. Treichel and S. Körper. "Untersuchung der Geräuschemission von Drohnen." *Lärmbekämpfung*, 14(4), 108–114, 2019.

[15] P. Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature methods*, 2020. ISSN 1548-7105. doi:10.1038/s41592-019-0686-2.

[16] P. D. Welch. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms." *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73, 1967. ISSN 00189278. doi:10.1109/TAU.1967.1161901.

[17] H. S. Wu, Q. Zhao, D. Zou, and Y. Q. Chen. "Automated 3D trajectory measuring of large numbers of moving particles." *Optics Express*, 19(8), 7646, 2011. ISSN 1094-4087. doi:10.1364/OE.19.007646.

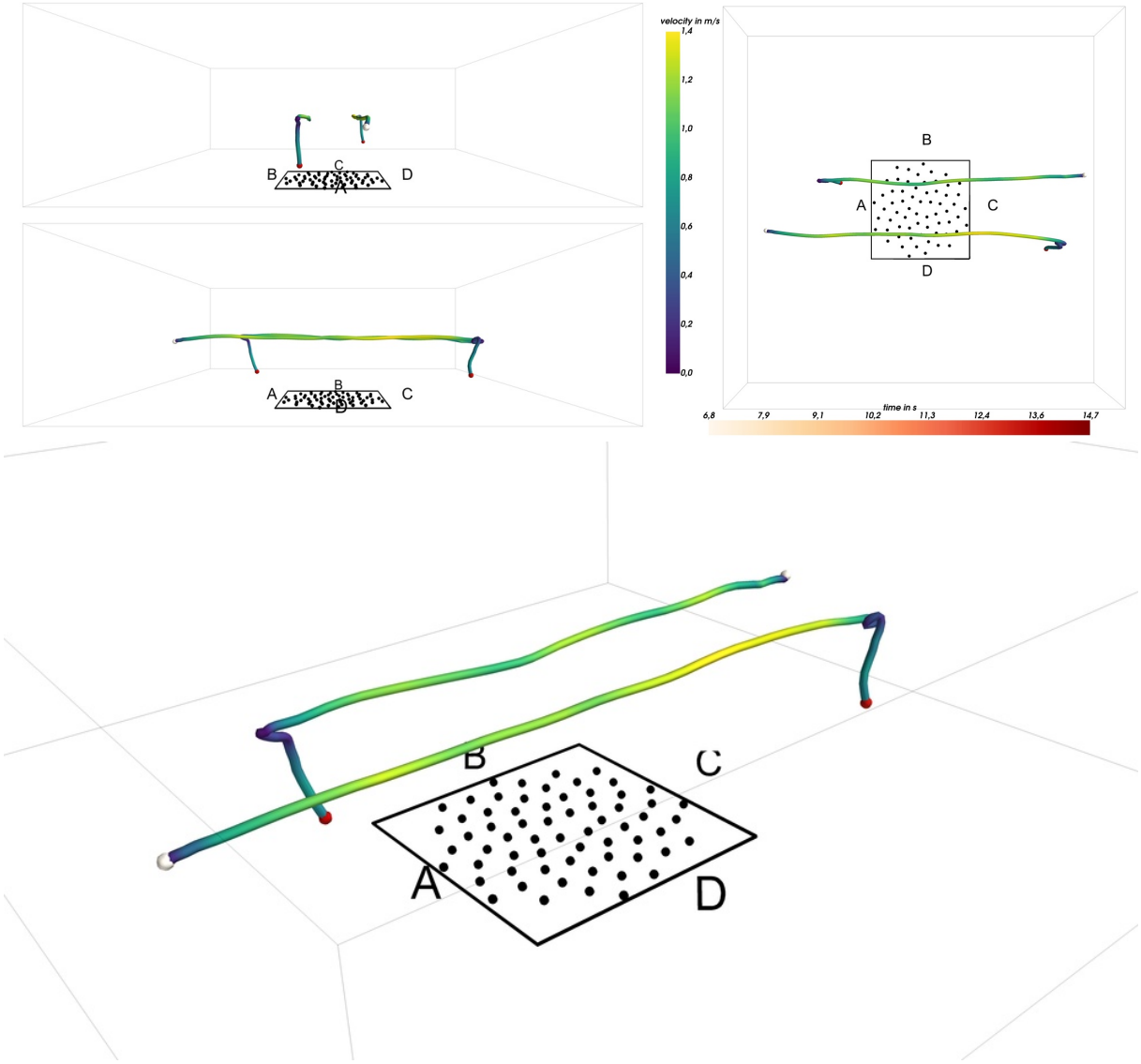

Figure 8: Reconstructed trajectories of Scenario 2: Two drones at FATO A/C flying in opposite direction towards FATO C/A.

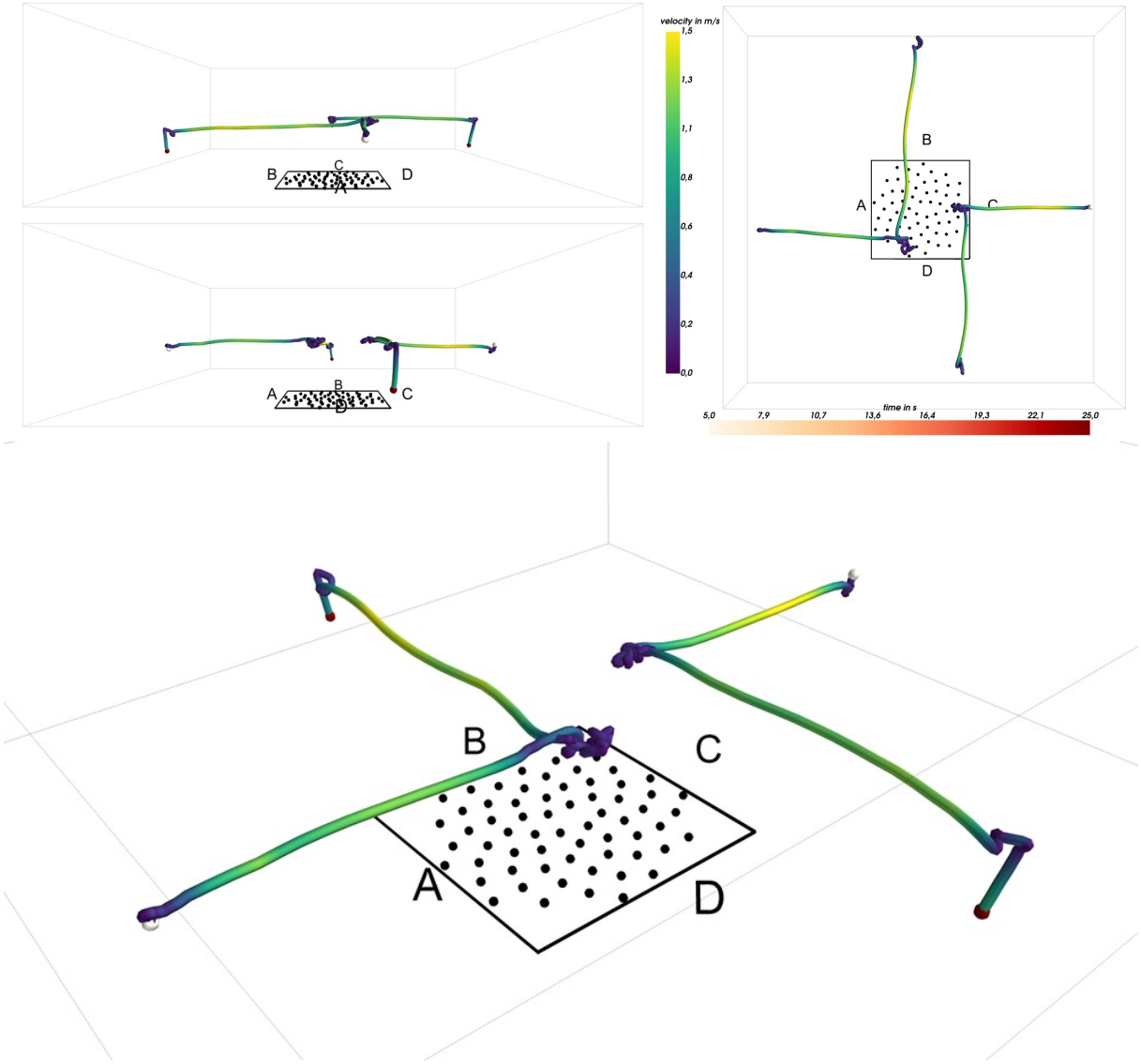

Figure 9: Reconstructed trajectories of Scenario 3: Two drones starting from opposite FATOs A/C, flying towards array, where they perform a left turn and fly towards FATOs B/D.

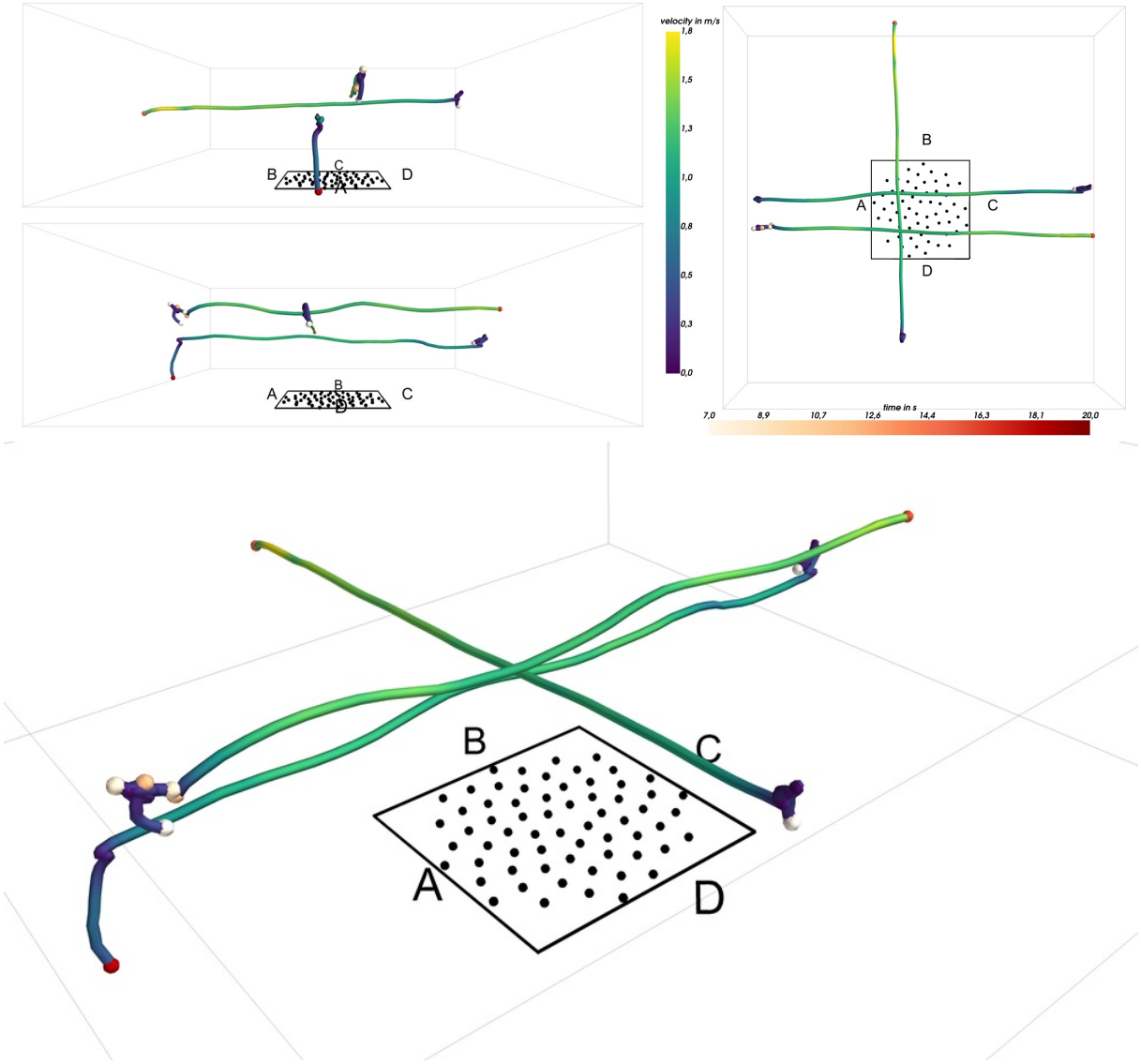

Figure 10: Reconstructed trajectories of Scenario 4: Three drones from FATOs A/C/D flying to the opposite FATO.

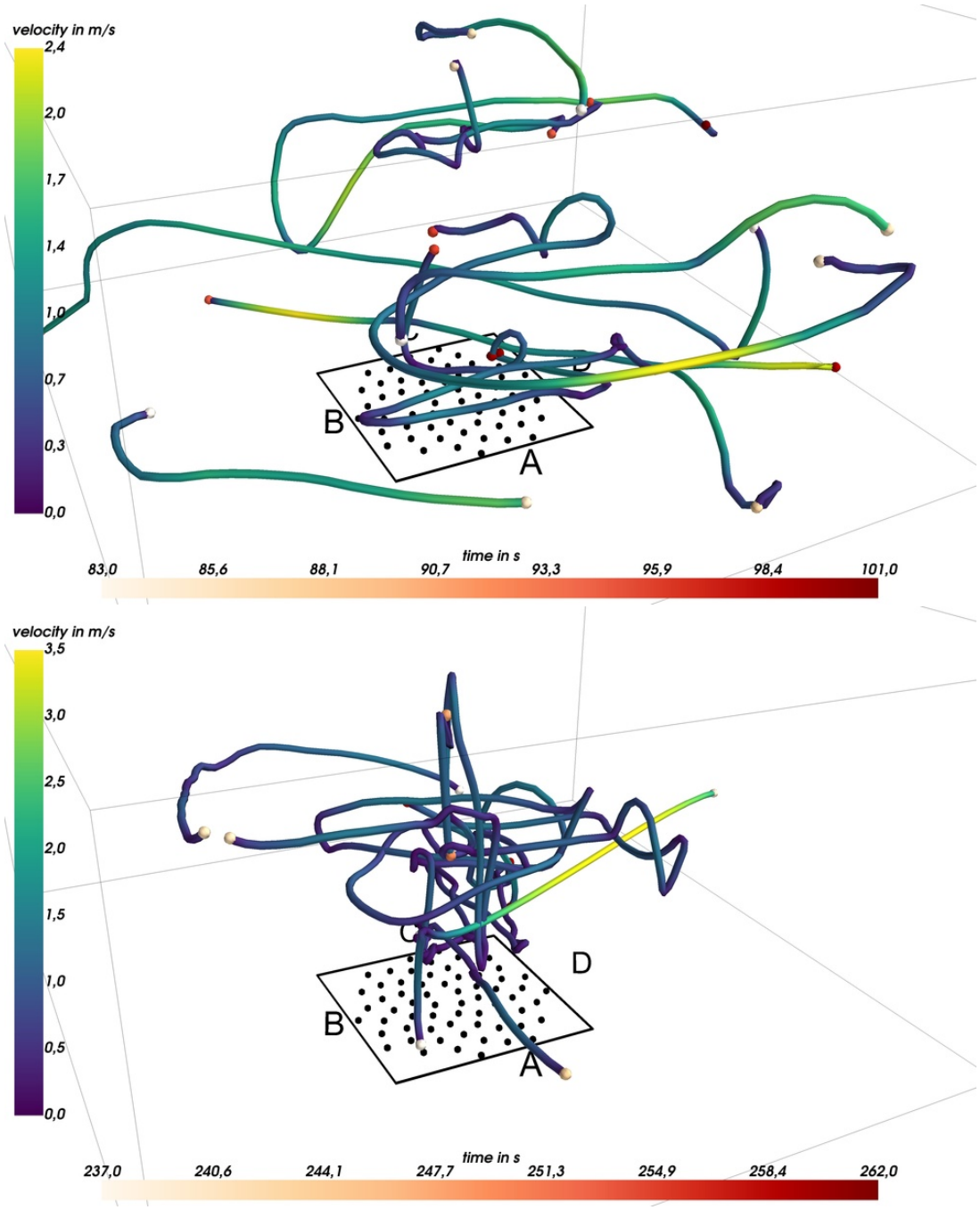

Figure 11: Reconstructed trajectories of Scenario 5, two different time segments: Four drones flying remote-controlled and arbitrarily.