Localization of RW-UAVs Using Particle Filtering Over Distributed Microphone Arrays

Jean-Samuel Lauzon, François Grondin, Dominic Létourneau, Alexis Lussier Desbiens, François Michaud

***Abstract*— Rotary-Wing Air Vehicles (RW-UAVs), also referred to as drones, have gained in popularity over the last few years. Intrusions over secured areas have become common and authorities are actively looking for solutions to detect and localize undesired drones. The sound generated by the propellers of the RW-UAVs is powerful enough to be perceived by a human observer nearby. In this paper, we examine the use of particle filtering to detect and localize in 3D the position of a RW-UAV based on sound source localization (SSL) over distributed microphone arrays (MAs). Results show that the proposed method is able to detect and track a drone with precision, as long as the noise emitted by the RW-UAVs dominates the background noise.**

I. INTRODUCTION

The field of civil drones, especially the Rotatory-Wing Air Vehicles (RW-UAVs), has expended rapidly lately, notably because they are easy to control and are low-cost. Unlike Fixed Wing Air Vehicles, RW-UAVs can perform stationary flight and precise maneuvers, which are useful in many applications such as videography and cartography [1], [2]. RW-UAVs can also be used to carry parcels at high speed, either with the control of a pilot, or autonomously following a trajectory using GPS and other sensors [3], [4]. However, they can also be used for privacy violation, spying, vandalism and especially for smuggling objects in jails. Authorities are thus looking for methods to detect and localize undesirable drones that fly over secured areas.

The sound produced by the propellers of RW-UAVs can be used to detect such drones. To our knowledge and probably because of the novelty of domestic RW-UAVs, very few research papers have yet addressed RW-UAV detection, with very limited results regarding performance and evaluation. A drone detection approach based on its acoustic signature has been proposed by Mezei et al. [5], with proof-of-concept results in laboratory conditions. DroneShield inc. and Drone-Detector inc. use a single-microphone to try to detect the acoustic signature of a drone's brushless motors and propellers. However, multiple microphones are required to determine the position of a drone. Acoustic cameras based on large microphone arrays (MAs) (over 100 microphones) are used to project sound power levels generated by RW-UAVs on a 2D image [6], [7]. Square Head/Norsonic inc. uses an acoustic camera made of 128 microphones or more, and Panasonic combines a 32-microphone array with a pan-tilt-camera. Such devices provide sound source localization (SSL) as the instantaneous direction of arrival (DOA) of sound, i.e., the 2D position (elevation, azimuth) of the sound sources. This is however insufficient to determine the 3D position of the sound source. In fact, at least two MAs spaced by a known distance are required to triangulate the 3D position of the sound source, as illustrated in Fig. 1. This paper demonstrates how multiple MAs can be used to derive the 3D position of a RW-UAV from the distributed MAs SSL data.

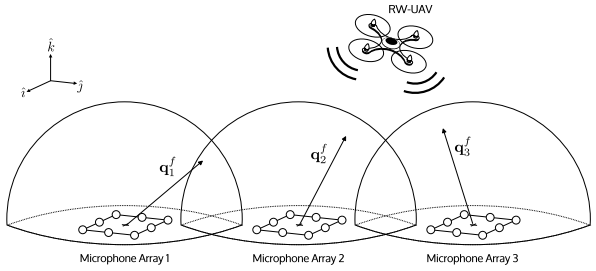

Fig. 1. SSL from distributed MAs

Drone detection field trials are challenging: they require large areas, good weather conditions, and a wide range of diverse noise conditions (wind, sound from the surrounding environments). This challenge involves conducting in-depth experiments with different SSL approaches (such as Generalized Cross-Correlation with Phase Transform method (GCC-PHAT) [8], [9], [10], [11], [12] or other variants [13], Multiple Signal Classification based on Standard Eigenvalue Decomposition (SEVD-MUSIC) [14], [15], [16], Multiple Signal Classification based on Generalized Eigenvalue Decomposition (GEVD-MUSIC) [17], Multiple Signal Classification based on Generalized Singular Value Decomposition (GSVD-MUSIC) [18]), MAs's positioning, data fusion approaches and other experimental condition. Before investigating all these conditions, in this paper we demonstrate the feasibility of doing 3D RW-UAV SSL using particle filtering to combine data from distributed MAs.

We chose to use the ManyEars framework [8] to implement SSL (elevation, azimuth) on distributed MAs, to which particle filtering (PF) is used to derive 3D SSL of a RW-UAV. ManyEars also uses particle filtering for sound source tracking (SST) of multiple sound sources using data from

This work was supported by the Fonds de recherche du Québec - Nature et technologies (FRQNT), and the Fondation de l'Université de Sherbrooke.

J.-S. Lauzon, F. Grondin, D. Létourneau, F. Michaud are with the Department of Electrical Engineering and Computer Engineering, and A. Lussier Desbiens is with the Department of Mechanical Engineering. This work is conducted at the Interdisciplinary Institute for Technological Innovation (3IT), 3000 boul. de l'Université, Québec (Canada) J1K 0A5, {Jean-Samuel.Lauzon,Francois.Grondin2, Dominic.Letourneau, Alexis.Lussier.Desbiens, Francois.Michaud}@USherbrooke.ca

a single MA, to estimate simultaneous the 2D positions (elevation, azimuth) of these sound sources. In this paper, PF is used to estimate the 3D position of one sound source, i.e., a RW-UAV, based on the DOAs provided by the MAs, distributed over an area at known positions. PF uses the redondancy in the information from the direction given by each MAs to narrow the distribution of particles around the true position of the drone.

This paper is organized as follows. Section II presents the overall system, along with how PF is used. Section III describes the experimental setup, followed by Section IV with the results obtained localizing a Parrot Bebop 2 drone in outdoor conditions.

II. Particle Filtering for 3D SSL Over Distributed MAs

Figure 2 illustrates the block diagram of our system using PF for 3D SSL using K MAs, each performing SSL to derive the DOA, represented as a unit vector q pointing in the direction of the loudest sound source and its energy e. As illustrated by Fig. 1, a virtual hemisphere around each MA is scanned and the sum of microphone pair cross-correlation is computed for each point on the surface. The point with the greatest magnitude corresponds to the DOA of sound, and the associated magnitude provides insights regarding the confidence in the beamformer output. All DOAs are transmitted to a central processing node, where time synchronization is performed to cope with transmission delay. Using the Network Time Protocol, synchronization can be achieved within a few milliseconds of accuracy [19]. For the current application, since the sound source maximum speed is in the order of a few tens of meters per second, this synchronization accuracy is sufficient. For each time frame f, the observation vectors $\mathbf{q}_k^f$ and its energy value e_k^f are sent to the PF module that returns the estimated 3D position $\hat{\mathbf{x}}^f$ of the sound source.

The main goal of PF for 3D SSL is to estimate the probability density function (PDF) of the position of the sound source using a finite set of particles. We chose to use PF rather than Kalman filtering because the states are modeled according to a non-gaussian PDF. Each particle has different parameters that are used to predict its new state. Each set of new observations allows them to be weighted according to how well they can represent the sound source. Following this logic, PF consists of the following elements: prediction model, instantaneous probability, observation assignation, particle filter instantiation, particle filter destruction, particle weight updates, and resampling.

A. Prediction Model

An excitation-damping model (similar to the one proposed in [9]) is used to predict the position of each particle h, as given by:

$$\dot{\mathbf{x}}_h^f = a_h^f \dot{\mathbf{x}}_h^{f-1} + b_h^f \mathbf{F}_x \tag{1}$$

$$\mathbf{x}_h^f = \mathbf{x}_h^{f-1} + \Delta T \dot{\mathbf{x}}_h^f \tag{2}$$

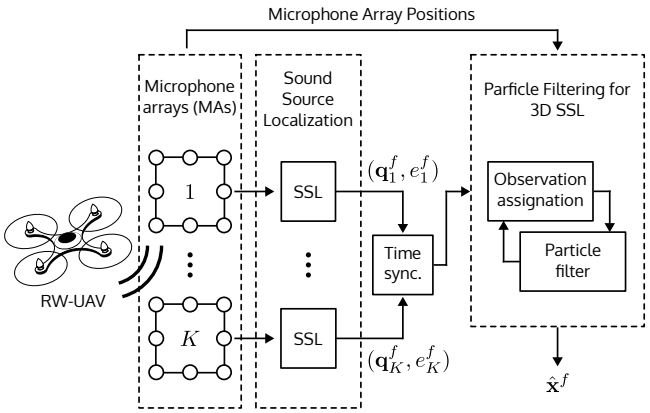

Fig. 2. Block diagram of PF for 3D SSL over distributed MAs

$$a_h^f = e^{-\alpha_h^f \Delta T} \tag{3}$$

$$b_h^f = \beta_h^f \sqrt{1 - (a_h^f)^2} \tag{4}$$

with $\mathbf{x}_h^f$ being the 3D position and $\dot{\mathbf{x}}_h^f$ the velocity of a particle, ΔT is a constant representing the time interval (in second) between two consecutive frames, $\mathbf{F}_x$ is a random variable generated using a multivariate standard normal distribution ($\mathbf{F}_x \sim \mathcal{N}_3(\mathbf{0}, \mathbf{I}_3)$), and α_h^f and β_h^f are parameters chosen to model the motion of a drone. The expression $\mathbf{F}_x$ represents the process noise in the speed of the source between each update. Variables a_h^f and β_h^f stand for the damping factor and the excitation factor, respectively. The particles can be in one of three motion states: Stationary, Constant Velocity, or Acceleration. When the filter is instantiated and when resampling is done, each particle is given a random motion state according to a given PDF.

B. Instantaneous Probability

The probability P_k^f that each observation $(\mathbf{q}_k^f, e_k^f)$ is generated by an active sound source, without being a false detection, is obtained from e_k^f and the threshold E_T:

$$P_k^f = \begin{cases} (e_k^f/E_T)^2/2, & e_k^f \leq E_T \\ 1 - (e_k^f/E_T)^{-2}/2, & e_k^f > E_T \end{cases} \tag{5}$$

C. Observation Assignation

Each observation $(\mathbf{q}_k^f, e_k^f)$ is assigned to a state: a false detection (hypothesis H_1), a new source (hypothesis H_2) or the currently tracked source (hypothesis H_3). The state hypothesis variable $\phi_k^f(c)$ and its possible values are listed in (6), with Φ_c^f representing the realization of the scenario c as given by (7). For instance, the realization of the scenario $\Phi_c^f = \{1, 3, 3\}$ indicates that the observation ($\mathbf{q}_1^f$,e_1^f) is a false detection, and the observations ($\mathbf{q}_2^f$,e_2^f) and ($\mathbf{q}_3^f$,e_3^f) correspond to the source being tracked. There are $C = (S+2)^K$ different possible assignation scenarios, where S

stands for the number of source detected ($S = 0$ or $S = 1$).

$$\phi_k^f(c) = \begin{cases} 1, & H_1 : \text{(False Detection)} \\ 2, & H_2 : \text{(New Source)} \\ 3, & H_3 : \text{(Existing Source)} \end{cases} \tag{6}$$

$$\Phi_c^f = \left\{\phi_1^f(c), \ldots, \phi_K^f(c)\right\} \tag{7}$$

The probability $P(\Phi_c^f|\mathbf{Q}^{1:f})$ of the occurrence of an assignation scenario given the observations is obtained using the Bayes rule:

$$P(\Phi_c^f|\mathbf{Q}^{1:f}) = \frac{P(\mathbf{Q}^{1:f}|\Phi_c^f)P(\Phi_c^f)}{\sum_{c=1}^{C} P(\mathbf{Q}^{1:f}|\Phi_c^f)P(\Phi_c^f)} \tag{8}$$

The expression $\mathbf{Q}^{1:f}$ stands for all the observations from the frame 1 to the current frame f. Observations are assumed to be conditionally independent and can therefore be represented as in (9). The expression $\mathbf{q}_k^{1:f}$ consists of all the observations of the vector $\mathbf{q}_k^f$ from frame 1 to the current frame f. The same hypothesis is assumed for the individual assignations, leading to (10).

$$P(\mathbf{Q}^{1:f}|\Phi_c^f) = \prod_{k=1}^{K} p(\mathbf{q}_k^{1:f}|\phi_k^f(c)) \tag{9}$$

$$P(\Phi_c^f) = \prod_{k=1}^{K} p(\phi_k^f(c)) \tag{10}$$

For the False Detection and New Source hypotheses, the conditional probability $p(\mathbf{q}_k^{1:f}|\phi_k^f)$ in (11) is uniform over the area of the virtual unit hemisphere. For the Existing Source hypothesis, the conditional probability depends on the previous weights of the particle filter ω_h^{f-1}, which corresponds to the probability the sound source is at the position of the particle given the observations, that is $p(\mathbf{x}_h^{f-1}|\mathbf{q}_k^{1:f-1})$. The probability that the observation occurs given the current position of the particles is expressed by $p(\mathbf{q}_k^f|\mathbf{x}_h^f)$:

$$p(\mathbf{q}_k^{1:f}|\phi_k^f(c)) = \begin{cases} 1/2\pi, & \phi_k^f(c) = 1, 2 \\ \sum_{h=1}^{H} \omega_h^{f-1} p(\mathbf{q}_k^f|\mathbf{x}_h^f), & \phi_k^f(c) = 3 \end{cases} \tag{11}$$

The prior probability $p(\phi_k^f(c))$ depends on the *a priori* probabilities that a new source appears (P_{new}) and that a false detection occurs (P_{false}), and on the probability P_k^f defined previously:

$$p(\phi_k^f(c)) = \begin{cases} (1 - P_k^f)P_{false}, & \phi_k^f(c) = 1 \\ P_k^f P_{new}, & \phi_k^f(c) = 2 \\ P_k^f, & \phi_k^f(c) = 3 \end{cases} \tag{12}$$

The probabilities that the observation $\mathbf{q}_k^f$ is associated to each state are given by (13). The expression $\delta_{x,y}$ refers to Kronecker delta. The expressions $P_{1|\mathbf{q}_k^f}$, $P_{2|\mathbf{q}_k^f}$ and $P_{3|\mathbf{q}_k^f}$ are normalized such that $P_{1|\mathbf{q}_k^f} + P_{2|\mathbf{q}_k^f} + P_{3|\mathbf{q}_k^f} = 1$.

$$P_{u|\mathbf{q}_k^f} = \sum_{c=1}^{C} \delta_{u,\phi_k^f(c)} P(\Phi_c^f|\mathbf{Q}^{1:f}) \quad 1 \leq u \leq 3 \tag{13}$$

D. Particle Filter Instantiation

PF is initialized when $P_{2|\mathbf{q}_k^f} > T_{new}$ for all MAs. To remove false detections caused by sporadic high energy noise, the previous condition needs to be met over F_{new} consecutive frames. When this happens, each particle has its position $\mathbf{x}_h^f$ and velocity $\dot{\mathbf{x}}_h^f$ drawn from multivariate Gaussian distributions, given by:

$$\mathbf{x}_h^f \sim \mathcal{N}_3(\boldsymbol{\mu}_{pos}, \boldsymbol{\Sigma}_{pos}) \tag{14}$$

$$\dot{\mathbf{x}}_h^f \sim \mathcal{N}_3(\boldsymbol{\mu}_{vel}, \boldsymbol{\Sigma}_{vel}) \tag{15}$$

The mean vector $\boldsymbol{\mu}_{vel}$ and the covariance matrices $\boldsymbol{\Sigma}_{pos}$ and $\boldsymbol{\Sigma}_{vel}$ are chosen to model the flying behavior of a drone. The covariance matrices $\boldsymbol{\Sigma}_{pos}$ and $\boldsymbol{\Sigma}_{vel}$ are diagonal and have variances of σ_{pos}^2 and σ_{vel}^2, respectively:

$$\boldsymbol{\Sigma}_{pos} = \sigma_{pos}^2 \mathbf{I}_3 \tag{16}$$

$$\boldsymbol{\Sigma}_{vel} = \sigma_{vel}^2 \mathbf{I}_3 \tag{17}$$

The parameter $\boldsymbol{\mu}_{pos}$ corresponds to the best estimation of the actual drone position. This position should lie at the intersection of the DOAs from all MAs. However, the DOAs obtained from the observation vectors $\mathbf{q}_k^f$ do not usually intersect perfectly, as there is always an error in the measured position and orientation of the MAs, and the estimated DOA direction. The Ray to Ray algorithm [20], as illustrated by Fig. 3, is used to determine the shortest distance between two skew lines. It is applied to find the closest point in 3D ($\mathbf{Z}_{ab}^f$) to the intersection of each pair of DOAs, $\mathbf{q}_a^f$ and $\mathbf{q}_b^f$, where $\mathbf{L}_a$ and $\mathbf{L}_b$ stand for the position of MAs $k = a$ and $k = b$, respectively. This solution is appealing by its simplicity since the PF can initialize its particles with a raw estimate of the source position. The optimal solution would be to minimize the angle between the observations and the estimated point. Although more precise, this method would be computationally costly and would make little difference in the overall tracking precision.

For K MAs, there are $K(K-1)/2$ pairs, and the estimated position vector $\boldsymbol{\mu}_{pos}$ is obtained using:

$$\boldsymbol{\mu}_{pos} = \frac{2}{K(K-1)} \sum_{a=1}^{K} \sum_{b=a+1}^{K} \mathbf{Z}_{ab}^f \tag{18}$$

The nearest point $\mathbf{Z}_{ab}^f$ is obtained by projection:

$$\mathbf{Z}_{ab}^f = \frac{\mathbf{L}_a + G_{ab,a}^f \mathbf{q}_a^f + \mathbf{L}_b + G_{ab,b}^f \mathbf{q}_b^f}{2} \tag{19}$$

The difference between the position of two different MAs a and b is given by (20) and the scalar values $G_{ab,a}^f$ and $G_{ab,b}^f$ are defined in (21) and (22).

$$\mathbf{d}_{ab} = \mathbf{L}_a - \mathbf{L}_b \tag{20}$$

$$G^f_{ab,a} = \frac{(\mathbf{q}^f_a \cdot \mathbf{q}^f_b)(\mathbf{q}^f_b \cdot \mathbf{d}_{ab}) - (\mathbf{q}^f_b \cdot \mathbf{q}^f_b)(\mathbf{q}^f_a \cdot \mathbf{d}_{ab})}{(\mathbf{q}^f_a \cdot \mathbf{q}^f_a)(\mathbf{q}^f_b \cdot \mathbf{q}^f_b) - (\mathbf{q}^f_a \cdot \mathbf{q}^f_b)^2} \quad (21)$$

$$G^f_{ab,b} = \frac{(\mathbf{q}^f_a \cdot \mathbf{q}^f_a)(\mathbf{q}^f_b \cdot \mathbf{d}_{ab}) - (\mathbf{q}^f_a \cdot \mathbf{q}^f_b)(\mathbf{q}^f_a \cdot \mathbf{d}_{ab})}{(\mathbf{q}^f_a \cdot \mathbf{q}^f_a)(\mathbf{q}^f_b \cdot \mathbf{q}^f_b) - (\mathbf{q}^f_a \cdot \mathbf{q}^f_b)^2} \quad (22)$$

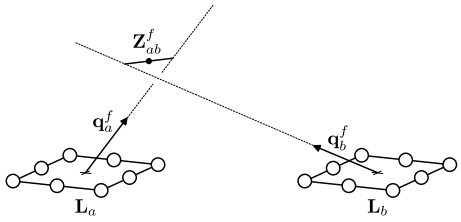

Fig. 3. Ray to Ray shortest distance algorithm

The motion states of the particles are chosen randomly, and weights ω^f_h are given a constant value of $1/H$.

E. Particle Filter Removal

To find out if the source is no longer active, a threshold T_{remove} is fixed. When the probability $P_{1|q^f_k}$ stays under this threshold for at least F_{remove} frames, PF is no longer updated and tracking stops ($S = 0$).

F. Particle Weights Update

At each new frame, the state of the particles are predicted and their weights are updated iteratively according to (23). Since each observation is independent, the probability $p(\mathbf{x}^f_h|\mathbf{Q}^f)$ of the particle $\mathbf{x}^f_h$ being the source of the observations $\mathbf{Q}^f$ is given by (24). The first part of the expression allows the particles to survive even if there is a false detection.

$$\omega^f_h = \frac{\omega^{f-1}_h p(\mathbf{x}^f_h|\mathbf{Q}^f)}{\sum_{h=1}^{H} \omega^{f-1}_h p(\mathbf{x}^f_h|\mathbf{Q}^f)} \quad (23)$$

$$p(\mathbf{x}^f_h|\mathbf{Q}^f) = \prod_{k=1}^{K} \left(\left(1 - P_{3|\mathbf{q}^f_k}\right) \frac{1}{H} + P_{3|\mathbf{q}^f_k} p(\mathbf{x}^f_h|\mathbf{q}^f_k) \right) \quad (24)$$

The probability $p(\mathbf{x}^f_h|\mathbf{q}^f_k)$ that a particle position fits the observations is obtained from the following normalization:

$$p(\mathbf{x}^f_h|\mathbf{q}^f_k) = \frac{p(\mathbf{q}^f_k|\mathbf{x}^f_h)}{\sum_{h=1}^{H} p(\mathbf{q}^f_k|\mathbf{x}^f_h)} \quad (25)$$

The expression $p(\mathbf{q}^f_k|\mathbf{x}^f_h)$ defined in (26) represents the probability that observation is realized given the position of the source in the particle h:

$$p(\mathbf{q}^f_k|\mathbf{x}^f_h) = \frac{1}{\sqrt{2\pi\sigma^2_\theta}} \exp\left(\frac{(\theta^f_{k,h} - \mu_\theta)^2}{2\sigma^2_\theta} \right) \quad (26)$$

The difference in angle between the particle $\mathbf{x}^f_h$ and the actual observation vector $\mathbf{q}^f_k$ is used to find a deviation angle $\theta^f_{k,h}$ in radians, obtained with the dot product projection:

$$\theta^f_{k,h} = \arccos \frac{\mathbf{q}^f_k \cdot \mathbf{x}^f_h}{|\mathbf{q}^f_k||\mathbf{x}^f_h|} \quad (27)$$

G. Resampling and Estimated Position

The estimated position $\hat{\mathbf{x}}^f$ is given by the sum of the product of the weights and position of particles, as expressed by (28).

$$\hat{\mathbf{x}}^f = \sum_{h=1}^{H} \mathbf{x}^f_h \omega^f_h \quad (28)$$

Resampling is required when the weight diversity goes below a threshold level proportional to the number of particles, as given by (29).

$$\left[\sum_{h=1}^{H} (\omega^f_h)^2 \right]^{-1} < \alpha H \quad (29)$$

III. EXPERIMENTAL SETUP AND METHODOLOGY

For the experiments, we used three 8-microphone MAs positioned on the ground. Figue 4 shows the microphone configuration for each MA. The audio card 8SoundsUSB [8], [21] is used to perform sound acquisition on each microphone at a sample rate of 48000 samples/sec. ManyEars, an open source framework for SSL, SST and sound source separation[1], is used to perform SSL on each MA. Raw data on each MA are recorded and processed offline to make analysis easier.

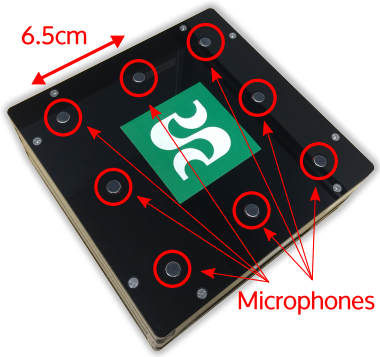

Fig. 4. Microphone array

Different trials were conducted using a DJI Phantom 2 drone and a Parrot Bebop 2 in controlled outside conditions near our research facility, to tune the parameters of the system. Table I presents the coefficients used for the excitation of particles. Table II summarizes the parameters used for PF. The energy threshold E_T, as well as P_{new}, P_{false} and α were set to the same values used in [8] as these parameters ensure robust performances with different 8-microphone array geometries. The number of frames F_{new} and F_{remove}, as well as the probability thresholds T_{new}

[1] https://sourceforge.net/projects/manyears/

and T_{remove}, were set empirically to filter spontaneous brief noise bursts and ultimately prevent false detections. The number of particles H was chosen as a compromise between calculation time and particle diversity. The parameters of the Gaussian distribution σ^2_{pos} and σ^2_{vel} were set to estimate the position and speed distribution of a drone after a detection. The value $\boldsymbol{\mu}_{vel}$ is a zero vector to provide a speed distribution in all possible directions. The parameters σ^2_θ and μ_θ were set empirically to best map the relationship between the observations and the expected results.

TABLE I
PARTICLE MOTION STATE PARAMETERS

Motion state	α^f_h	β^f_h	Probability
Stationary	2	0.05	10%
Constant velocity	0.5	3	40%
Acceleration	1.5	6	50%

TABLE II
SSL PARAMETERS

Parameter	Value	Parameter	Value
E_T	600	T_{new}	0.75
P_{new}	0.005	T_{remove}	0.3
P_{false}	0.05	F_{new}	10
H	500	F_{remove}	10
σ^2_{pos}	25	α	0.7
σ^2_{vel}	25	$\boldsymbol{\mu}_{vel}$	$\overrightarrow{0}$
σ^2_θ	0.0961	μ_θ	0

IV. RESULTS

To illustrate the feasibility of using PF to derive 3D SLL using distributed MAs in a realistic setup, we present data from a trial conducted outside in a meadow, next to a busy road. The MAs were placed in a triangle configuration on the ground, each separated by 10 m, as shown by Fig. 5. A Parrot Bebop 2 drone was used to fly above the three microphone arrays following various trajectories. A GPS onboard provides the baseline for the trajectory of the drone.

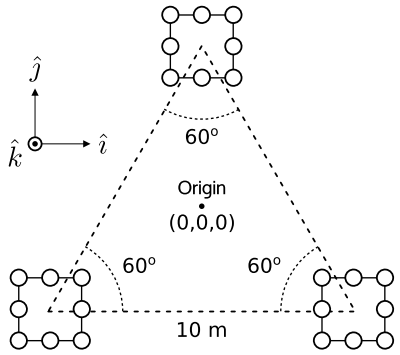

Fig. 5. Positions of the microphone arrays

Figure 6 presents the baseline trajectories in the x, y, and z directions (which are parallel to the $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ unit vectors). Results show that the system tracks accurately the drone in time segments A, C, E, G, I and K. Tracking is performed with precision in the X and Y directions (i.e., in the $\hat{\imath}$-$\hat{\jmath}$ plane). Precision decreases in the Z direction as elevation increases. This is explained by the far-field effect that is more prevalent in the Z direction with the current MAs disposition. For segments B, D and H, the sound of the drone became lower than the background noise, which explains why drone tracking stops. More specifically, in segment D, a noisy truck drove close to the field where the experiment was performed. In segment J, the operators, who stood next to the MAs, talked to each other, and speech interfered with the sound of the drone. In segment L, a motorized vehicle drove close to the MAs, and became the loudest sound source tracked.

V. CONCLUSION

This work demonstrates that RW-UAV 3D SSL is feasible using particle filtering from distributed MAs. To only evaluate the capability of combining SSL data from distributed MAs to derive RW-UAV 3D positions, it assumes that the loudest sound source is from a RW-UAV and that it dominates the background noise. This allows us to evaluate SSL 3D performance without using other means to filter sound sources. The next step is therefore to refine the approach by adding specific features to improve robustness to noise. For instance, various methods have been proposed to estimate the background noise [22], [23], [24] and generate time-frequency masks that make GCC-PHAT more robust to noise [25], [26], [27]. In future work, we plan to:

- Apply a time-frequency mask to reduce the contribution of interfering signals when SSL is performed;
- Perform SST locally on each MA such that both multiple RW-UAVs and interfering sound sources are tracked;
- Perform multi-source tracking to disregard the sound sources that are under a minimum elevation (i.e., are not flying in the sky), to make 3D SSL more robust to interfering sound sources from the ground;
- Build or use a large dataset of drones in flight to infer the motion state probabilities instead of empirically finding them.
- Assess performance in terms of robustness, precision and processing load.

REFERENCES

[1] D. Santano and H. Esmaeili, "Aerial videography in built heritage documentation: The case of post-independence architecture of malaysia," in *Proc. IEEE Int. Conf. Virtual Systems & Multimedia*, 2014, pp. 323–328.

[2] M. Sanfourche, B. Le Saux, A. Plyer, and G. Le Besnerais, "Cartographie et interprétation de l'environnement par drone," in *Colloque scientifique francophone Drones et moyens légers aéroportés d'observation*, 2014.

[3] P. Brisset, A. Drouin, M. Gorraz, P.-S. Huard, and J. Tyler, "The Paparazzi solution," in *2nd US-European Competition and Workshop on Micro Air Vehicles*, 2006.

[4] D. Floreano and R. J. Wood, "Science, technology and the future of small autonomous drones," *Nature*, vol. 521, no. 7553, pp. 460–466, 2015.

[5] J. Mezei, V. Fiaska, and A. Molnar, "Drone sound detection," in *Proc. IEEE Int. Symp. Computational Intelligence & Informatics*, 2015, pp. 333–338.

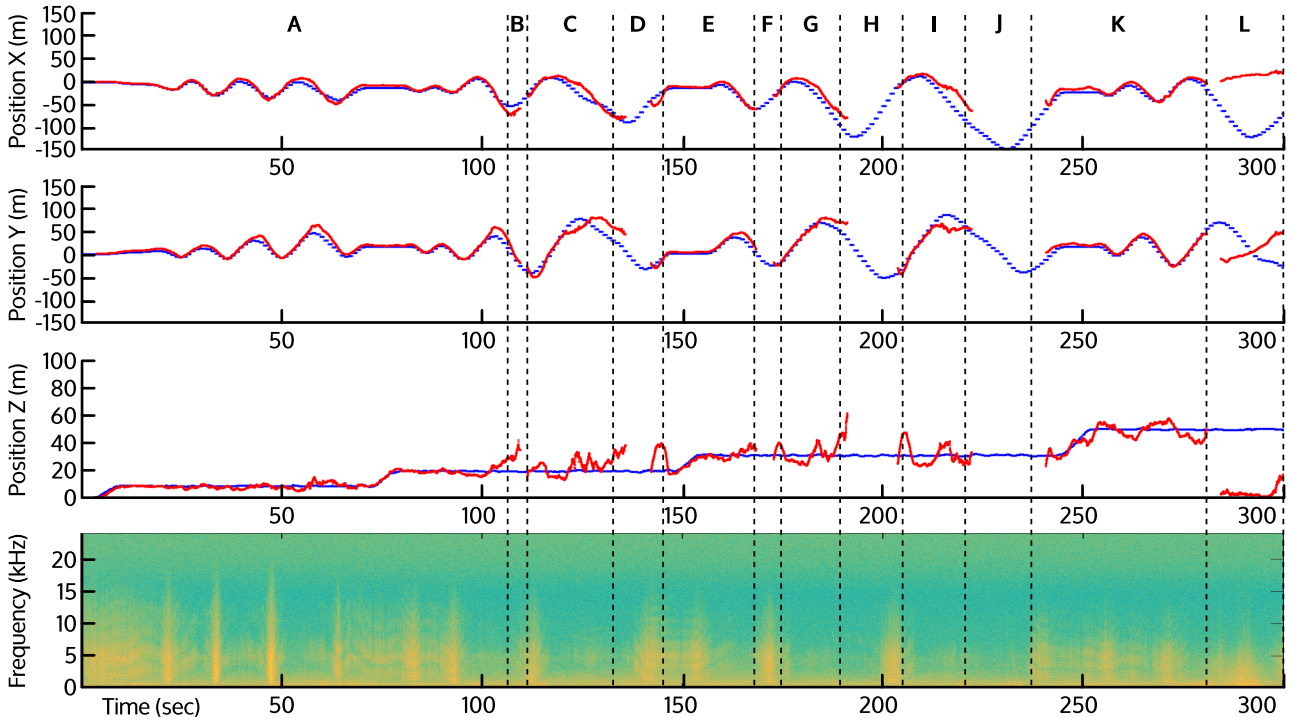

Fig. 6. GPS positions of the Parrot Bebop 2 drone (baseline, in blue) and tracked position (in red) over time. Sound interferences are observed in segments D (truck), J (speech) and L (motorized vehicle).

[6] J. Busset, F. Perrodin, P. Wellig, B. Ott, K. Heutschi, T. Rühl, and T. Nussbaumer, "Detection and tracking of drones using advanced acoustic cameras," in *Proc. SPIE*, vol. 9647, 2015.

[7] A. Zunino, M. Crocco, S. Martelli, A. Trucco, A. Del Bue, and V. Murino, "Seeing the sound: A new multimodal imaging device for computer vision," in *Proc. IEEE Int. Conf. Computer Vision Workshop*, 2015, pp. 693–701.

[8] F. Grondin, D. Létourneau, F. Ferland, V. Rousseau, and F. Michaud, "The ManyEars open framework," *Autonomous Robots*, vol. 34, no. 3, pp. 217–232, 2013.

[9] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems*, vol. 2, 2003, pp. 1228–1233.

[10] J.-M. Valin, F. Michaud, B. Hadjou, and J. Rouat, "Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach," in *Proc. IEEE Int. Conf. Robotics & Automation*, vol. 1, 2004, pp. 1033–1038.

[11] J.-M. Valin, F. Michaud, and J. Rouat, "Robust 3D localization and tracking of sound sources using beamforming and particle filtering," in *Proc. IEEE Int. Conf. Acoustics Speech & Signal Processing*, vol. 4, 2006, pp. 841–844.

[12] ——, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," *Robotics & Autonomous Systems*, vol. 55, no. 3, pp. 216–228, 2007.

[13] A. Badali, J.-M. Valin, F. Michaud, and P. Aarabi, "Evaluating real-time audio localization algorithms for artificial audition on mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems*, 2009.

[14] K. Nakadai, T. Takahashi, H. G. Okuno, H. Nakajima, Y. Hasegawa, and H. Tsujino, "Design and implementation of robot audition system 'HARK' – Open source software for listening to three simultaneous speakers," *Advanced Robotics*, vol. 24, no. 5-6, pp. 739–761, 2010.

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

[16] C. T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems*, 2009, pp. 2027–2032.

[17] K. Nakamura, K. Nakadai, F. Asano, and G. Ince, "Intelligent sound source localization and its application to multimodal human tracking," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems*, 2011, pp. 143–148.

[18] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution sound source localization for robots," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems*, 2012, pp. 694–699.

[19] D. L. Mills, "Internet time synchronization: The network time protocol," *IEEE Trans. Communications*, vol. 39, no. 10, pp. 1482–1493, 1991.

[20] P. Schneider and D. H. Eberly, *Geometric Tools for Computer Graphics*. Morgan Kaufmann, 2002.

[21] D. Abran-Côté, M. Bandou, A. Béland, G. Cayer, S. Choquette, F. Gosselin, F. Robitaille, D. T. Kizito, F. Grondin, and D. Létourneau, "USB synchronous multichannel audio acquisition system," Technical report, Dept. Elec. Eng. & Comp. Eng., Université de Sherbrooke, 2014.

[22] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403–2418, 2001.

[23] ——, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12–15, 2002.

[24] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.

[25] K. W. Wilson and T. Darrell, "Learning a precedence effect-like weighting function for the generalized cross-correlation framework," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2156–2164, 2006.

[26] F. Grondin and F. Michaud, "Time difference of arrival estimation based on binary frequency mask for sound source localization on mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robot & Systems*, 2015, pp. 6149–6154.

[27] ——, "Noise mask for TDOA sound source localization of speech on mobile robots in noisy environments," in *Proc. IEEE Int. Conf. Robotics & Automation*, 2016, pp. 6149–6154.