

# Robust Acoustic Source Localization of Emergency Signals from Micro Air Vehicles

Meysam Basiri<sup>1,2</sup>, Felix Schill<sup>1</sup>, Pedro U.Lima<sup>2</sup> and Dario Floreano<sup>1</sup>

**Abstract**—In search and rescue missions, Micro Air Vehicles (MAV's) can assist rescuers to faster locate victims inside a large search area and to coordinate their efforts. Acoustic signals play an important role in outdoor rescue operations. Emergency whistles, as found on most aircraft life vests, are commonly carried by people engaging in outdoor activities, and are also used by rescue teams, as they allow to signal reliably over long distances and far beyond visibility. For a MAV involved in such missions, the ability to locate the source of a distress sound signal, such as an emergency whistle blown by a person in need of help, is therefore significantly important and would allow the localization of victims and rescuers during night time, through foliage and in adverse conditions such as dust, fog and smoke. In this paper we present a sound source localization system for a MAV to locate narrowband sound sources on the ground, such as the sound of a whistle or personal alarm siren. We propose a method based on a particle filter to combine information from the cross correlation between signals of four spatially separated microphones mounted on the MAV, the dynamics of the aerial platform, and the doppler shift in frequency of the sound due to the motion of the MAV. Furthermore, we evaluate our proposed method in a real world experiment where a flying micro air vehicle is used to locate and track the position of a narrowband sound source on the ground.

## I. INTRODUCTION

The main objective of a search and rescue mission is to quickly locate and extract victims from the disaster situation. A search effort in rough outdoor terrain can be very time consuming and physically challenging, and keeping track of the positions of multiple rescue teams in a large area without communication infrastructure can be an additional problem. Autonomous Micro Air Vehicles (MAVs) can assist rescuers to faster locate victims in a large search area, and help coordinate rescue efforts by reporting the location of rescue teams to the mission coordinator [1]. They can directly reach potential target areas by flying over obstacles and inaccessible terrains, and hence achieve a faster area coverage than ground units.

Sound is one of the most important cues for locating people in a disaster situation. Sound waves travel in all directions and can be detected at long distances from the sound source, and beyond line of sight. During night time or in fog, dust, smoke, dense forests or otherwise cluttered environments, acoustic signals are far more reliable than

visual cues. This is the main motivation behind the use of safety whistles in most survival kits and disaster preparedness supplies offered today. Safety whistles are an inexpensive and effective method for emitting a distress signal in emergency situations, and are also commonly used for basic signalling when noise or distance makes voice communication difficult. They are often used by people engaging in outdoor activities, such as hikers, mountaineers, skiers, boaters and scuba divers, and are commonly provided with airplane life vests. These whistles enable the user to generate a very loud and clear narrow-band sound (usually a single frequency between 2-5 KHz), which can be perceived distinctly from long ranges and in noisy environments, without making the signaller hoarse and exhausted. There is usually little environmental noise in the multi-kilohertz range, which enhances detectability. However, a disadvantage of high-frequency, narrow-band sounds is the difficulty for human listeners to correctly locate the direction of the source [2].

The aim of this work is to develop an audio based localization system for MAVs to locate and track narrowband sources on the ground. Such a system can also provide a simple way of interaction between human operators and MAVs. For example, a rescue team could easily signal its position to the mission coordinator by using specific whistle signals, or could command robots to land in a desired spot.

Hearing has always been one of the key senses among humans and animals. Despite this, audition in robotics has not received great attention compared to vision, and most studies on this focus on speech recognition and localization of talkers for home, office, and humanoid robots [3] [4]. In most works, a technique inspired by animal hearing called Inter-aural Time Difference (ITD) (also known as Time Difference of Arrival TDOA) is used for localizing sound sources. This method measures the time delay caused by the finite speed of sound between the signals received by two microphones. While the complex hearing capabilities of animals achieve good performance with only one pair of acoustic sensors, technical systems often use arrays of microphones for assisting robots in locating broadband sound sources in the environment [5]. Design of new acoustic sensors that could possibly be used on MAVs have been investigated in some recent works [6] [7].

The TDOA is generally estimated by using cross-correlation of the signals to find the best match. A problem faced with this approach is that it requires the sound source to be a broadband source. This is because of the repetitive nature of narrowband sounds, which introduces an ambiguity in the time delay estimation, i.e. coherence testing among

\*This work was supported by a doctoral grant from FCT (SFRH/BD/51070/2010), EC FP-7 research funding mechanism under grant agreement no. 266470 and FCT project [PEst-OE/EEI/LA0009/2011].

<sup>1</sup> Laboratory of Intelligent Systems, Ecole Polytechnique Federale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: meysam.basiri, felix.schill, dario.floreano@epfl.ch)

<sup>2</sup> Institute for Systems and Robotics, Instituto Superior Tecnico, Lisboa, Portugal (e-mail: pal@isr.ist.utl.pt)

the signals from different microphones no longer provides a unique time delay for every microphone-pair. This problem is particularly pronounced for high frequencies. Reducing the distance between the microphones reduces ambiguities, but also reduces resolution. Our strategy to overcome this problem is to use this ambiguous information along with other sources of information, such as the dynamic behaviour of the robot and, if the source frequency is known, the Doppler shift in the sound frequency caused by the MAV's motion, in order to obtain a more reliable estimate. A particle filtering technique is employed, also known as sequential Monte Carlo method [8]. Particle filtering is considered a powerful tool for handling localization, navigation and tracking problems [9]. Few works show the effectiveness of particle filtering for tracking wide-band sound sources inside reverberant environments [10],[11]. This paper focuses on the tracking of narrowband and single-frequency sound sources, proposes a novel method to incorporate two types of acoustic information (TDOA and Doppler shift) and vehicle dynamics in a particle filter framework, and presents experimental results.

This paper is organized as follows: Section II describes the proposed method for localizing narrowband sound sources and explains the different parts of the localization system in detail. In Section III results of real experiments with the proposed method is provided, where a flying MAV is used to locate and track the location of a sound source.

## II. PROPOSED METHOD

This section explains our method for allowing a MAV to localize and track a target on the ground that emits a narrowband sound. Figure 1 presents the schematic diagram of the localization system consisting of three main units. The first unit is the "TDOA Localization" unit which is responsible for measuring probable directions of the target using the estimated time difference among the signals of an on-board microphone array. The second unit is the "Doppler Speed Estimation" unit which uses the measured sound frequency and the Doppler shift to obtain a measure of the relative speed between the target and the robot itself. The third unit is the particle filter tracker which uses the information obtained by the previous units along with the information about the position and motion of the robot itself, throughout time, to estimate and track the position of the target while eliminating false estimates. A more detailed explanation of each unit is presented in the following sections.

### A. TDOA Localization

Our approach for locating the direction of a target is based on the Time Delay of Arrival (TDOA) estimation technique. In this method the time delays between signals of spatially separated microphones are found using coherence measures (This is described later in section II-A.1). These time delays are then used along with knowledge of the microphone array's geometry to calculate the direction to a sound source. Figure 2 illustrates a simple sound field consisting of two microphones  $((i),(j))$  and a sound source  $S$ . Here, two

microphones that are  $d_m$  meters apart experience sound waves emitted from a sound source  $S$ . Considering the far field assumption ( $d_s \gg d_m$ ), it is possible to state that:

$$\cos \theta = \frac{c\tau_{ij}}{d_m} \quad (1)$$

where  $\tau_{ij}$  is the time difference of arrival between the two microphones,  $c$  is the speed of sound and  $\theta$  is the angle of incidence between the sound wave-front and the microphone pair's axis. Also, from the cosine law it is possible to derive:

$$\cos \theta = \frac{\vec{x}_{ij} \cdot \vec{u}}{d_m} \quad (2)$$

where  $\vec{u}$  is a unit vector pointing towards the sound source. By combining equations (1) and (2) we get

$$\vec{x}_{ij} \cdot \vec{u} = c\tau_{ij} \quad (3)$$

In three dimensional space and by considering a body-fixed coordinate system that is attached to the MAV's frame, the vectors  $\vec{u}$  and  $\vec{x}_{ij}$  are defined as

$$\begin{aligned} \vec{u} &= (u_x, u_y, u_z) \\ \vec{x}_{ij} &= (m_{ix} - m_{jx}, m_{iy} - m_{jy}, m_{iz} - m_{jz}) \\ &= (m_{ijx}, m_{ijy}, m_{ijz}) \end{aligned} \quad (4)$$

where  $(m_{ix}, m_{iy}, m_{iz})$  are the coordinates of microphone  $i$ . From (3) and (4) a linear system of  $N$  equations can be obtained when  $N$  different microphone-pairs are used:

$$A \times u = b \quad (5)$$

$$A = \begin{bmatrix} m_{12x} & m_{12y} & m_{12z} \\ m_{13x} & m_{13y} & m_{13z} \\ \vdots & \vdots & \vdots \\ m_{ijx} & m_{ijy} & m_{ijz} \end{bmatrix} \begin{matrix} 1 \\ 2 \\ \vdots \\ N \end{matrix} \quad u = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \quad b = \begin{bmatrix} c\tau_{12} \\ c\tau_{13} \\ \vdots \\ c\tau_{ij} \end{bmatrix}$$

The system of equations (5) can be used to compute a unique solution  $u$  if more than three microphone-pairs exist which are not all in the same plane. A value for  $u$  that best satisfies the system of equations (5) can be estimated using the linear least square method [12]

$$u_{ls} = (A^T A)^{-1} A^T b \quad (6)$$

to minimize the sum of the squared errors

$$\varepsilon_r = \sum (A u_{ls} - b)^2 \quad (7)$$

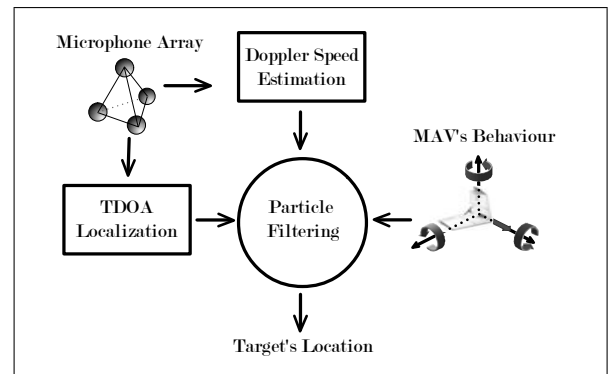


Fig. 1. Schematic diagram of the proposed localization system illustrating main parts of the system

1) *Estimation of TDOA*: As explained previously, for a pair of microphones inside a sound field, a time delay  $\tau_d$  exists between their signals. This time delay is dependent on the angle  $\theta$  between the sound wave-front and the microphone pair's axis and lies within the range

$$-\frac{d_m}{c} < \tau_d < \frac{d_m}{c} \quad (8)$$

where  $c$  is the speed of sound. Cross correlation is a commonly used technique for measuring the coherence and obtaining the time delay  $\tau_d$  between two signals. Cross correlation of two microphone signals each having a length of  $N$  samples can be computed by

$$R_{ij}(\tau) = \sum_{k=0}^{N-1} p_i[n] p_j[n - \tau]$$

where  $p_i[n]$  is the signal perceived by microphone  $i$  and  $\tau$  is the correlation lag in samples in the range expressed by equation (8). The correlation value  $R_{ij}(\tau)$  reaches a maximum value at  $\tau = \tau_d$ .

The shape of the cross correlation  $R_{ij}$  is dependent on the statistical properties of the sound source itself. If the sound source is an ideal white noise then  $R_{ij}$  is equal to an impulse function transposed by  $\tau_d$ . However, if the sound source is a band limited white noise,  $R_{ij}$  no longer has an impulse shaped peak but instead has a more broadened peak. The actual shape of the correlation  $R_{ij}$  can be described mathematically by [13]

$$R_{ij}(\tau) = P_a \frac{\sin(\pi B(\tau - \tau_d))}{\pi B(\tau - \tau_d)} \cos(2\pi f_0(\tau - \tau_d)) \quad (9)$$

where  $B$  is the bandwidth,  $f_0$  is the centre frequency, and  $P_a$  is the signal power. As the bandwidth approaches zero, the correlation  $R_{ij}$  becomes a periodic sine wave with peaks of the same amplitude and period of  $1/f_0$  and hence makes the identification of the peak corresponding to the correct time delay ambiguous. Figure 3 shows cross-correlations in a simple sound field case explained in figure 2, for three different sound sources. It can be seen that the narrowband sounds from a whistle and a piezo alarm show multiple peaks, introducing ambiguity in the direction estimation.

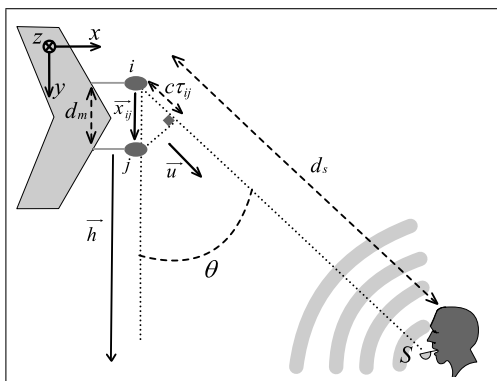


Fig. 2. Illustration of a sound field consisting of two microphones ( $i, j$ ) and a sound source  $S$ . The microphones are placed on a MAV and  $d_m$  meters apart from one another. Vector  $u$  is a unit vector in the body-fixed coordinate system  $(x, y, z)$  pointing towards the sound source  $S$ .

One method to tackle the time delay ambiguity problem for narrow-band sound sources is to decrease the distance between the microphones in order to avoid multiple cross correlation peaks. To only have a single cross-correlation peak, the microphone pair's inter-distance must satisfy

$$d_m < \frac{c}{2f_0} \quad (10)$$

However, the angular resolution of a microphone array for a given sampling frequency decreases for smaller distances between the microphones. We therefore propose a method which considers all the peaks above a certain threshold in the cross correlation of every microphone pair, and hence obtains a set of possible time delays for every pair. Corresponding time delays among different microphone pairs are then identified, from which probable directions of the sound source is obtained. This information is later fused with the Doppler speed estimate and the robot's behaviour in order to remove ambiguities and locate the sound source.

To identify the corresponding time delays among different microphone pairs we use equations (6) and (7) to compute the source direction  $u_{l,s}$  and error  $\varepsilon_r$  for all possible time delay combinations. Those values of  $u_{l,s}$  that have the lowest errors can be considered as a potential direction to the sound source. Hence a set of possible source directions are identified:

$$\{u_k, k = 1 : M\} \quad (11)$$

where  $M$  is equal to the number of corresponding peaks among microphone pairs.

### B. Relative Speed Estimation

The acoustic Doppler shift in frequency is defined as the change in frequency due to the relative motion between a sound source and an observer. The relationship between the frequency  $f_0$  of the sound source and the observed frequency  $f$  is defined by the equation

$$f = f_0 \frac{c \pm v_o}{c \mp v_s} \quad (12)$$

where  $v_o$  and  $v_s$  are the speeds of the observer and the sound source respectively, assuming that they are either directly

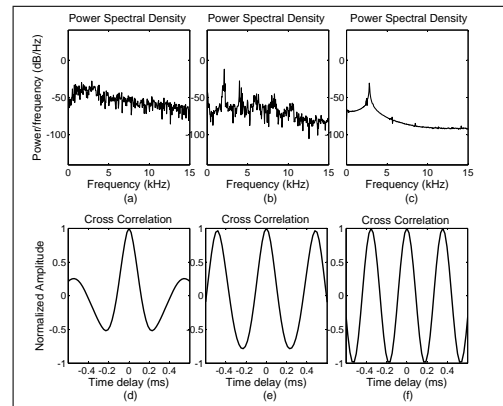


Fig. 3. Cross-correlations for three different cases of sound source. The power spectral density of every sound source is shown in graphs(a-c) and their corresponding cross correlation is shown below each graph (d-f). Three sound sources are: (a) air blower (b) emergency whistle (c) piezo alarm

approaching or receding from each other. In the problem proposed in this paper, the aim is to locate a stationary narrowband sound source of known frequency  $f_s$  that is located on the ground from a moving MAV. Equation (12) can be used to obtain an estimate of the relative speed between the MAV and the stationary sound source:

$$v_r = c\left(\frac{f}{f_0} - 1\right) \quad (13)$$

$f$  is measured by searching for the maximum peak in the power spectral density of the sound measurements inside the range  $f_0 - f_m < f < f_0 + f_m$ , where  $f_m$  is the maximum possible shift in frequency.  $f_m$  can be obtained from the absolute speed  $v_R$  of the MAV

$$f_m = \frac{v_R}{c} f_0 \quad (14)$$

### C. Particle-Filtering Localization and Tracking

The previous two sections described methods of providing an instantaneous, noisy and ambiguous information about a possible narrowband sound source in the environment. It is now required to estimate more reliably the position of this sound source by fusing all the available information together. Also information about the behaviour of the MAV in time, along with all measurements available up to the current time, could be employed to obtain a more accurate estimate. For this reason, we will use the particle filtering technique to recursively estimate the probability density of the target location. Using this method, all of the hypotheses about the target's position are represented as a set of particles with individual weights.

At time instant  $t$ , the direction to the sound source is modelled using a set of  $N$  particles of direction vectors  $u_i$  and weight  $w_i$ , where  $u_i = (u_{xi}, u_{yi}, u_{zi})$  is a unit vector in the body-fixed coordinate system that starts at the origin and points towards a direction.  $u_i$  can also be described in body-fixed spherical coordinate system  $(r, \angle\phi, \angle\theta)$  by:

$$u_i = (1, \phi_i, \theta_i) \quad i = 1, 2, \dots, N \quad (15)$$

where  $\phi_i$  is the azimuth defined in the range  $[-\pi, \pi]$  and  $\theta_i$  is the elevation defined in the range  $[-\pi/2, \pi/2]$ . A two dimensional state vector is specified for every particle:

$$S_i(t) = [\phi_i(t) \ \theta_i(t)] \quad (16)$$

The algorithm starts by forming an initial set of particles  $\{S_i(0), i = 1 : N\}$ . Particles either could be generated uniformly over the entire state space, or only over a desired part of the state space if some prior knowledge about the possible location of the target is available. In the proposed problem the initial state space is reduced to all vectors pointing towards the ground.

1) *Prediction*: In the prediction step, a set of new particles  $\tilde{S}_i(t)$  is predicted by propagating  $S_i(t-1)$  according to a probabilistic motion model. This is achieved by transforming the vectors  $u_i(t-1)$  to  $\tilde{u}_i(t)$  using the information gathered from MAV's on-board sensors indicating the change in the state of the MAV. In this work, we measure the change in

roll and pitch of the MAV using on-board gyroscopes and the airspeed and altitude using an absolute and a differential pressure sensor. As no compass is present on the MAV, the heading information is obtained from an on-board GPS sensor. We use a simple model for the prediction step which assumes that the robot has only forward motion (i.e. along the  $x$  axis on the body-fixed coordinate system), and the target is located on a flat ground (i.e. at zero height).

The prediction procedure starts by estimating a target distance  $d_i$  for every particle based on the MAV's current height and orientation:

$$d_i = \frac{h}{\cos \angle(\vec{u}_i(t-1), \vec{h})} \quad (17)$$

where  $\angle(\vec{u}_i(t-1), \vec{h})$  is the angle between vectors  $\vec{u}_i$  and the height vector  $\vec{h}$ . From (15) and (17), it is possible to construct a vector  $T_i$  that extends the unit vector  $\vec{u}_i$  until it reaches the ground.

$$\vec{T}_i(t-1) = (d_i, \phi_i, \theta_i) \quad (18)$$

Furthermore, this vector is updated by taking into account the translation and rotations of the body-fixed coordinate system due to the speed  $\vec{v}$  and the change in the yaw ( $\Delta\lambda$ ), pitch ( $\Delta\beta$ ) and roll ( $\Delta\alpha$ ) angles of the MAV.

$$\vec{T}_i(t) = R(\Delta\lambda, \Delta\beta, \Delta\alpha) \times (\vec{T}_i(t-1) - \vec{v}\Delta t) \quad (19)$$

where  $\Delta t$  is the time interval between the updates and  $R$  is a rotation matrix representing the rotations of the MAV. Finally a vector  $\tilde{u}_i$  is randomly selected from all possible direction vectors  $\vec{p}$  that satisfy the equation

$$\angle(\vec{T}_i, \vec{p}) = N(0, \sigma_A) \quad (20)$$

where  $N(0, \sigma_A)$  is a random number generated with a normal distribution of mean zero and standard deviation  $\sigma_A$ . The value of  $\sigma_A$  is chosen in relation with the accuracy of the model that is used.

2) *Update*: As previously explained, a set of  $M$  direction measurements from the TDOA localization unit (11) and one relative speed measurement  $v_r$  from the Doppler Estimation unit (13) is obtained at every time-step. In the update step, the likelihood of obtaining these measurements is investigated for every particle and particles are weighted based on this measure. For this investigation, we propose the likelihood function:

$$w_i = L_{Di} \times \sum_{k=1}^M L_{Tik} \quad (21)$$

where

$$L_{Di} = N(\varepsilon_{Di} : 0, \sigma_D) \quad L_{Tik} = N(\varepsilon_{Tik} : 0, \sigma_T)$$

$N(x : m, \sigma)$  stands for a normal distribution with mean  $m$  and standard deviation  $\sigma$  evaluated at  $x$ .  $\varepsilon_{Tik}$  is the angle between  $u_i$  and the  $k$ th TDOA measurement and  $\varepsilon_{Di}$  is the error between the computed relative speed  $\tilde{v}_{ri}$  and the measured relative speed value  $v_r$ .

$$\varepsilon_{Tik} = \angle(\vec{u}_k, \vec{u}_i) \quad \varepsilon_{Di} = |v_r - \tilde{v}_{ri}| \quad (22)$$



Fig. 4. Picture of the MAV platform [14] used for experimenting the proposed algorithm. Four microphones and an on-board digital sound recorder is used for recording sounds during flight.

where the computed relative speed  $\tilde{v}_{ri}$  is simply calculated using the MAV's velocity  $\vec{V}_R$  and the vector  $\vec{u}_i$  from

$$\tilde{v}_{ri} = \left\| \vec{V}_R \right\| \cos \angle(\vec{V}_R, \vec{u}_i) \quad (23)$$

The values of  $\sigma_T$  and  $\sigma_D$  in (21) reflect the confidence of Doppler-speed and TDOA measurements respectively and can be found empirically.

3) *Target Direction Estimation*: The direction to the target can be estimated at each time step from the probability density function represented by the particle set. For this a weighted mean of all particles' positions could be used. However, to avoid inaccurate estimations for situations with multi-modal distributions, a weighted mean of particles located in a local neighbourhood of the particle with the highest weight is used instead:

$$\bar{S}_T = \sum_{i=1}^K w_i S_i : \forall |S_i - S_{max}| < \xi \quad (24)$$

### III. EXPERIMENTS AND RESULTS

To test and verify the proposed algorithm, multiple real experiments were performed with the MAV platform shown in figure 4 for locating a stationary target on the ground that emits a narrowband sound. A microphone array consisting of four microphones is mounted on the MAV along with a digital sound recorder for recording the microphone signals. The microphones are positioned in a way to form a regular tetrahedron of edge length 10 cm. The MAV is equipped with an autopilot that allows it to fly fully autonomously to predefined waypoints. The orientation, altitude, air-speed and global positioning information of the MAV are measured using on-board sensors and are transmitted to a ground station. The MAV was controlled to fly within the visual range of a safety pilot while occasionally reducing or even turning off its engine to increase the detection range by



Fig. 5. (a) The safety whistle used in the experiment (commonly available in outdoor shops). It features two chambers, and emits two closely spaced, superimposed frequencies between 2kHz and 2.1 kHz. (b) A hand-held piezo alarm used in this work that emits a single frequency of 3.8 kHz.

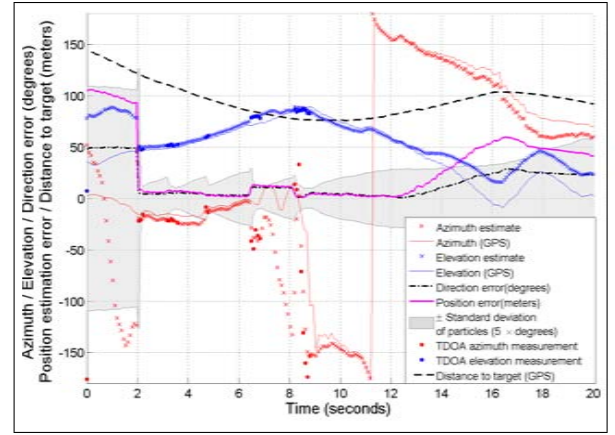


Fig. 6. Result of an experiment where a MAV is used to locate a human target on the ground who is occasionally blowing into a safety whistle.

increasing the signal to noise ratio. This reduction in the engine power is achieved automatically whenever the MAV is descending.

Two different sound sources were used for the experiments: a commonly available duo-tone safety whistle as shown in figure 5(a), and a 3.8kHz hand-held piezo alarm shown in figure 5(b). The sounds were triggered by a human experimenter in a known location. The whistle was blown in intervals of approximately 1 to 2 seconds.

Figure 6 shows the result of an experiment where the safety whistle is used as the target sound source. This figure illustrates the relative direction estimates of the flying MAV and the target position estimation error in meters during a time interval of the flight. These estimates are also compared against the measurements computed from absolute GPS positions and MAV orientation measurements. Here, the particle filtering algorithm is initialized after the MAV's motor input drops below a predefined threshold ( $N=100$ ). As soon as the first whistle is observed, particles converge toward the direction of the target. Furthermore, when there are no observations available or the MAV's motor input is above the predefined threshold, the particle filtering update step is no longer performed and hence only the probabilistic motion model of the MAV is responsible for the tracking. This results in the gradual increase in the spreading of the particles until the next observation is available. It can be seen that in this experiment, and after the last set of whistle sound observations, the target position is tracked correctly for several seconds but then suddenly the error starts to increase. This is because at this point the MAV start to perform a sharp 180 degrees turn to return to the home way-point. Note that although the microphone pair's inter distance and the frequency of the target sound does not satisfy equation (10), and hence 2 peaks could appear in the cross correlations, there are no ambiguities in the final TDOA observations. This is because the current geometry of the microphone array ensures that for this target frequency not all pairs experience this ambiguity simultaneously. Therefore in the proposed method, in the step were corresponding time delays among different pairs are identified, incorrect peaks are automatically eliminated and a single direction

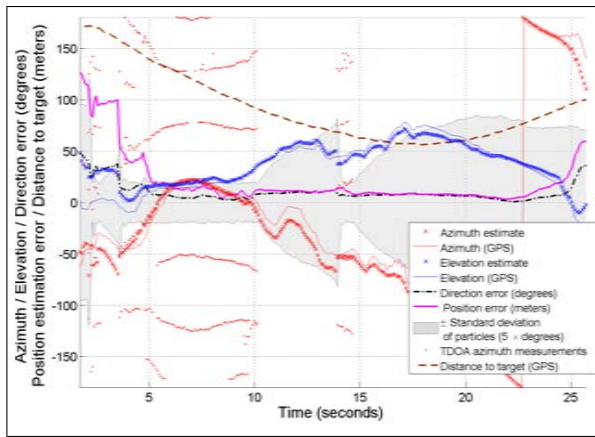


Fig. 7. Result of an experiment where a MAV is used to locate the position of a human target on the ground that is holding a piezo alarm (N=100).

measurement is obtained. Hence, in this experiment, as there are no ambiguities present in the TDOA measurement, the doppler based relative speed estimation step and knowledge of the target frequency is no longer necessary. The left plot in figure 8 illustrates this by providing a comparison between using and not using the doppler speed estimation in the filtering process (first and second error bar from the left).

To further test the performance of the system in situations where ambiguities are still present in the final TDOA observations, a set of experiments was performed in which a piezo alarm of frequency 3.8 kHz was used as the target source. Figure 7 shows the result of an experiment that illustrates the correct direction estimation despite the high ambiguity existing in the TDOA measurements. The raw azimuth estimates from TDOA alone are shown as small red dots; ambiguities in the form of 5-6 possible angles are clearly visible. The particle-filtered estimate quickly converges to the correct estimate. Furthermore, the target position is shown to be tracked correctly long after observations are no longer available. However, figure 8 shows that the knowledge of the target frequency and the doppler speed estimation is necessary to obtain a correct estimate when using the piezo alarm. A detection range beyond 150 meters was achieved with both the whistle and the alarm.

To move towards an embedded real time implementation of the proposed solution, and since the complexity of this method is highly dependent on the number of particles, further evaluations were performed to evaluate the effect of this parameter on the localization performance. Figure 8 shows a relatively small trade-off between the computational complexity and accuracy of the proposed algorithm. It can be observed that it is feasible to use as little as 20 particles with acceptable results, and therefore to minimize greatly the computational complexity of the proposed solution.

#### IV. CONCLUSION

This paper presents a solution to the problem of localizing narrowband sound sources on the ground from a micro air vehicle. The solution provided in this paper requires an on-board microphone array to measure probable directions of the sound source based on TDOA measurements, and on-board

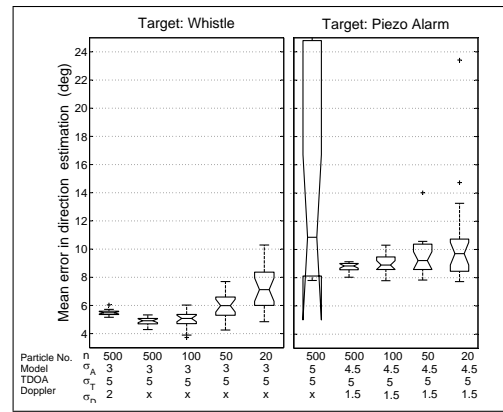


Fig. 8. Box plots of average error in direction estimation for 20 runs and on 15 seconds time interval during flight time, showing the effect of particle numbers and Doppler-speed estimations on the mean error.  $\sigma = x$  indicates that doppler speed estimations are not used in the update step.

sensors to obtain information about the state of the MAV. In case of ambiguities in the TDOA measurements due to the repetitive nature of narrow band sounds, knowledge of the vehicle dynamics, and furthermore the sound frequency and the Doppler shift of this frequency is used for resolving these ambiguities. The particle filtering technique used in this paper was shown to be well-suited for fusing these multiple sensors in order to achieve robust estimation of the position of the sound source.

#### REFERENCES

- [1] S. Hauert, S. Leven, J. Zufferey, and D. Floreano, "Communication-based swarming for flying robots," in *Proc. Intl. Conf. Robotics and Automation Workshop on Network Science and Systems*, 2010.
- [2] R. Stern, A. Zeiberg, and C. Trahiotis, "Lateralization of complex binaural stimuli: A weighted-image model," *The Journal of the Acoustical Society of America*, vol. 84, p. 156, 1988.
- [3] Y. Matsusaka, T. Tojo, S. Kubota, K. Furukawa, D. Tamiya, K. Hayata, Y. Nakano, and T. Kobayashi, "Multi-person conversation via multi-modal interface—a robot who communicate with multi-user," in *Sixth Eu Conference on Speech Communication and Technology*, 1999.
- [4] K. Nakadai, T. Lourens, H. Okuno, and H. Kitano, "Active audition for humanoid," in *AAAI-2000*. MIT Press, 2000, pp. 832–839.
- [5] J. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *IROS-2003*, vol. 2. IEEE, 2003, pp. 1228–1233.
- [6] F. Ruffier, S. Benacchio, F. Expert, and E. Ogam, "A tiny directional sound sensor inspired by crickets designed for micro-air vehicles," in *Sensors, 2011 IEEE*. IEEE, 2011, pp. 970–973.
- [7] I. de Bree, I. Wind, and I. Druyvesteyn, "Multi purpose acoustic vector sensors for battlefield acoustics."
- [8] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, 1996.
- [9] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001.
- [10] D. Ward, E. Lehmann, and R. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.
- [11] H. Asoh, F. Asano, T. Yoshimura, K. Yamamoto, Y. Motomura, N. Ichimura, I. Hara, and J. Ogata, "An application of a particle filter to bayesian multiple sound source tracking with audio and video information fusion," in *Proc. Fusion*. Citeseer, 2004, pp. 805–812.
- [12] C. Lawson and R. Hanson, *Solving least squares problems*. Society for Industrial Mathematics, 1995, vol. 15.
- [13] B. Ferguson, "Time-delay estimation techniques applied to the acoustic detection of jet aircraft transits," *The Journal of the Acoustical Society of America*, vol. 106, p. 255, 1999.
- [14] S. Leven, J. Zufferey, and D. Floreano, "A simple and robust fixed-wing platform for outdoor flying robot experiments," in *International symposium on flying insects and robots*, 2007, pp. 69–70.