On-Board Relative Bearing Estimation for Teams of Drones Using Sound

Meysam Basiri, Felix Schill, Pedro Lima, and Dario Floreano

Abstract—In a team of autonomous drones, individual knowledge about the relative location of teammates is essential. Existing relative positioning solutions for teams of small drones mostly rely on external systems such as motion tracking cameras or GPS satellites that might not always be accessible. In this letter, we describe an onboard solution to measure the 3-D relative direction between drones using sound as the main source of information. First, we describe a method to measure the directions of other robots from perceiving their engine sounds in the absence of self-engine noise. We then extend the method to use active acoustic signaling to obtain the relative directions in the presence of self-engine noise, to increase the detection range, and to discriminate the identity of robots. Methods are evaluated in real world experiments and a fully autonomous leader-following behavior is illustrated with two drones using the proposed system.

Index Terms—Aerial Robotics, Swarms, Localization.

I. Introduction

TEAMS of drones have many potential applications, such as, a rapidly deployable communication network for disaster areas [1], sensing and mapping of chemical clouds [2], searching for forest fires [3], aerial surveillance [4], aerial transportation [5], building constructions [6] and detecting targets of interest [7]. Rapid progress in the design and control of multi-drone systems has been observed over the past years [6], [8].

Relative positioning is the problem of gaining information about the position of other robots by individuals in a robotic group. This information is necessary for many different tasks such as inter-robot collision avoidance [9] and formation control [10]. Due to the strict payload constraints of small drones, there is a lack of technological possibilities that could provide relative positioning without relying on any external systems. Hence most solutions depend on an external positioning system, such as motion tracking cameras or GPS satellites, and a communication network to share the position of robots with each other. Despite the good accuracy of such solutions they are highly dependent on the availability of the external system. Motion tracking cameras are impractical for use in unprepared

M. Basiri, F. Schill, and D. Floreano are with the Laboratory of Intelligent Systems, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: meysam.basiri@epfl.ch; felix.schill@epfl.ch; dario.floreano@epfl.ch).

P. Lima is with the Institute for Systems and Robotics, Técnico Lisboa, 1049-001 Lisbon, Portugal (e-mail: pal@isr.tecnico.ulisboa.pt).

or large environments and GPS signals might be unavailable, such as in indoor or cluttered areas, or unreliable due to jamming or spoofing.

An alternative approach to relative positioning is to directly measure this information using on-board sensors. Onboard solutions are mostly developed for ground robots in 2D using sensors such as laser range scanners, infrared sensors and cameras [11], [12]. 3D laser range scanners are expensive and bulky instruments that are not suitable for small drones. A 3D laser scanner with a limited field of view of 30x40 degrees has been used in [13] on an outdoor helicopter with a payload of 29kg. Mini laser range finders have been used by [14], [15], for detection of large obstacles located directly in front of a drone. However, due to their single point/planar detection ability they are not suitable for measuring position of other robots in 3D. Radar-based systems, such as the Traffic Collision Avoidance system (TCAS), are used for relative positioning between commercial aircrafts. TCAS systems are not suitable for drones as they are bulky and expensive. More recently, advancements in radio communication modules with highly accurate, synchronized clocks are used for the time-of-flight range-only measurement in cooperative wireless sensor nodes [16]. But so far, these modules have not been tested on moving drones. Doppler radar transducers with a total weight of 300 grams have been used in [17] to detect the relative bearing of colliding obstacles. However, the small field of view (30°), low angular resolution (15°) and small operating range (10 meters) are some of the major drawbacks of this system. Optical sensors have been shown effective for obstacle avoidance and altitude control in drones [18]. Few works investigate use of cameras on UAVs for measuring the relative direction of large moving aircrafts relative to the background scene [19], [20]. Limited field of view, high dependency on visual contrast and light conditions, failure when the target is stationary or located on non-uniform or cluttered backgrounds, high computational demand for processing high resolution and high frame rate images [20] are some of the drawbacks of this approach. To the best of our knowledge, the only onboard relative positioning systems demonstrated on real drones are based on infra-red (IR) sensors [21], [22] for indoor flying robots. The sensor suite proposed by [22] has been shown to provide accurate range and bearing up to 10 meters of range. It consist of a total of 208 IR sensors and it has a weight of ≈ 400 g, a size of 50×50 cm and a power consumption of 10 watts, and thus not suitable for small drones.

Sound-based relative positioning have mostly been used in underwater robotic teams [23]. Sound is also used by many birds and insects for relative positioning, such as flight calls

of nocturnal migratory birds [24] and phonotaxis behaviour in insects [25]. Sound-based relative positioning for teams of drones could have several potential advantages. It would be independent of illumination and visibility conditions and it could operate at night time. It would rely on microphone sensors, which are low cost, lightweight, small and passive sensors, that are very suitable for small drones. Omnidirectional microphones allow full 3D coverage needed for aerial robots. In addition, since sound waves are capable of overcoming obstacles it could provide information through foliage and overcome occlusions by other robots. Furthermore, since the engine of most flying platforms already produce sound in flight, this sound could be exploited for passively obtaining the relative position of team-mates and of other non-cooperative flying platforms. This could be further exploited to detect and locate many other interesting acoustic targets, such as the sound of a safety whistle in a search and rescue mission [26].

This letter describes an onboard audio-based system that allows individual drones in a team to measure the relative bearing of other drones. This bearing information could be used to achieve tasks such as collision avoidance [27], formation control [28] and cooperative positioning [29]–[31]. Two solutions are presented and separately validated. The first solution (Section II) consists of passively monitoring the sound of the engines of other drones. The second solution (Section III) introduces active acoustic signaling (chirping) for improving the performance. The two methods are validated with real robots (Section IV) and an audio-based leader-following behavior is demonstrated with two drones. Conclusions and future works are discussed in Section V.

II. PASSIVE AUDIO-BASED SYSTEM

This section explains a passive onboard system to measure the relative direction of other robots based entirely on the sound of their engines. Exploiting the already available engine sounds results in a highly energy-efficient solution. This method was found to be suitable when there is either no self-engine noise present, or when the self-engine noise is different, and it can be separated from the target engine sound. This method could potentially be employed in groups of fixed-wing drones, as they can glide with their engines reduced or turned off, or in groups of rotor-craft drones that can rest on ground or attach to ceilings or walls, in order to listen to sound of their flying teammates. Localizing the engine sound of distant robots in presence of an engine with the same sound characteristics that is only few centimeters away from the microphones, hence resulting in an extremely low signal to noise ratio (SNR), is a challenging problem that is beyond the scope of this work. Furthermore, since the detection range for this system depends on the signal to noise ratio, the suitability of this approach depends on the target sound levels and noise level of the environment. Note that the active acoustic signaling method described in Section III can obtain the relative directions in presence of self-engine noise and is more robust to environment noises.

Figure 1 shows a block diagram with the main units of the passive bearing measurement system. An explanation of each
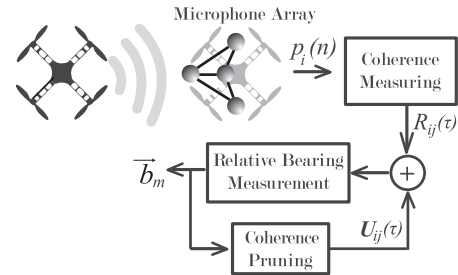

Fig. 1. Schematic diagram of the proposed passive audio-based relative bearing measurement system illustrating the main parts of the system.
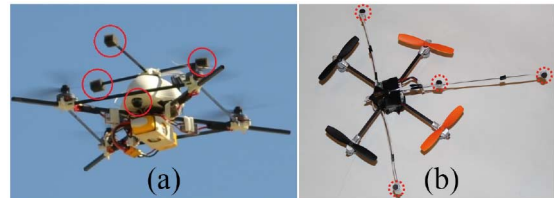

Fig. 2. Pictures of the two drones used in this work, with the microphones indicated by red circles. (a) small drone (440 grams) with a microphone array (21 grams) used for outdoor experiments (b) pocketsize drone (31.4 grams) with a microphone array (3.5 grams) used for indoor experiments.

unit is provided in this section. Some of the units are also used in the active method.

A. Microphone Array

An onboard microphone array is used to simultaneously measure sound waves at different points in space. A minimum number of four microphones, if not all placed on the same plane, is required to compute the direction of sound sources in 3D without ambiguity. This number is used here to minimize the hardware and computational complexity. However, due to the small size of these sensors, more sensors could be added to increase the robustness. There are no strict constraints on the sensor position or geometry of the array, making it suitable for any shape of drone, and only the microphone positions relative to each other must be known with good accuracy. In this work, we use two different types of array geometries, as shown in Figure 2, where a flat triangular array with a total weight of 3.5 grams is used on a pocket-size drone for testing the passive method and a tetrahedral microphone array is used later for testing the active method, providing approximately equal performance in all directions [32]. The flat microphone array results in an ambiguity whether the measured 3D-direction is located above or below the array that is resolved by the drone since it is measuring the directions while resting on a surface.

Sound waves are picked up by the microphones and are converted into electric signals. These signals are then amplified and filtered using basic analogue filters to remove unwanted noise. Signals from all microphones are simultaneously sampled and converted into digitized form with a sampling frequency F_s. After a discrete sequence of N samples is obtained from every microphone they are passed to the coherence measuring unit. The length N of the sequence should be chosen as a trade-off between the stability, obtained with large values of N, and

the tracking speed. In this work, we use $N = 1024$ samples to meet the low memory and computational power of common micro-controllers.

B. Coherence Measuring Unit

The coherence measuring unit is based on the time difference of arrival methods (TDOA) used in sound localization literature [33]. This unit starts by measuring the similarity between every pair of microphone signals, as a function of time delay τ applied to one of them, using:

$$R_{ij}(\tau) = FFT^{-1}\left(\frac{FFT(p_i(n)).FFT^*(p_j(n))}{W}\right) \quad (1)$$

where FFT is the N-point Fast Fourier Transform, FFT^{-1} is the inverse FFT, FFT^* denotes the complex conjugate of the FFT results, $p_i(n)$ is the digitized sequence of microphone i with $n = 1, \ldots, N$ and τ is the correlation lag in samples. W is the spectral weighting function for improving the similarity analysis that is described later. Note that for $W = 1$, function (1) is a general cross correlation that is computed in frequency domain. Only the value of $R_{ij}(\tau)$ for all the probable discrete time delays in the range $\pm\tau_{\max}$ is of interest and is stored for further processing. The maximum time delay $\tau_{\max}$ is limited to the sampling frequency F_s and the distance d_m between the two microphones:

$$\tau_{\max} = \frac{d_m}{c}F_s \quad (2)$$

where c is the speed of sound. Investigating the similarity degree only for a set of integer delays limits the resolution of the coherence measurements, and hence the bearing measurements, to the sampling frequency and the microphone baseline distance. In this work a sampling frequency of 40 kHz was used, that allowed the digitization to be performed by the same micro-controller performing the computations and avoid the need for any additional hardware. A large microphone baseline of 18 cm suitable for the size of our robots were also used. However, to provide more resolution to the similarity measurements, a cubic spline interpolation with a factor of 10 was performed on the values of $R_{ij}(\tau)$.

One limitation of using the general cross correlation method, i.e. $W = 1$, for measuring the similarity is that results are strongly dependent on the statistical properties of the sound signal and it usually leads to wide cross correlation peaks. A weighting function W can improve the robustness by performing frequency bin weighting. Many different weighting functions are used in the literature [34]. The PHAT weighting [35] is one of the most popular weighting functions showing robust localization performance, particularly against reverberations:

$$W_{PHAT} = |\text{FFT}(p_i(n))|\,|\text{FFT}(p_j(n))| \quad (3)$$

PHAT whitens the cross-spectrum, giving equal contribution to all frequencies, to rely only on phase information of the signal leading to much sharper correlation peaks. One drawback of this method is that frequencies dominated by the noise are also considered equally. A modified version of the PHAT
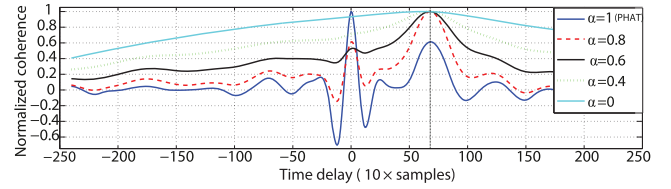

Fig. 3. Coherence measure $R_{ij}(\tau)$ between signals of a microphone pair, experiencing the sound of a drone in a room, computed using weighting function (4) and for different values of α. The solid vertical line indicates the true time delay.

weighting was used here instead to only take into account the frequency bandwidth where the sound is mostly present while de-emphasizing the dominant frequencies.

$$W = \chi\left(|\text{FFT}(p_i(n))|\,|\text{FFT}(p_j(n))|\right)^\alpha \quad (4)$$

where

$$\chi = \begin{cases} 1 & f_{\min} < f < f_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$f_{\min}$ and $f_{\max}$ are the minimum and maximum frequencies of the bandwidth. α is chosen between [0,1] and is used to control the trade-off between full whitening and no whitening of frequencies inside the defined range. The larger the values of α, the sharper are the peaks in the resulting coherence measurements. However, a large α could lead to less robustness in case of low signal to noise ratios. A larger frequency range would also result in sharper correlation peaks.

Figure 3 shows a comparison between PHAT weighting ($\alpha = 1$), $\alpha = 0.8$, $\alpha = 0.6$, $\alpha = 0.4$ and general cross correlation ($\alpha = 0$), using the entire frequency range, for a pair of microphones experiencing sound of a pocket-size drone flying in a room. The PHAT weighting results in an erroneous global peak that does not correspond to the true bearing. This is because the engine sound of drone is absent in a wide range of frequencies and the noise in these frequencies are considered equally in computing the coherence. This problem can be resolved by reducing the value of α to put more emphasis on the dominant frequency bins, with the expense of obtaining wider peaks.

For best performance, the values of α, $f_{\min}$ and $f_{\max}$ can be selected based on the sound specifications. The instantaneous frequency spectrum of the measured sound sequence could be used to compute these values prior to measuring the coherence using thresholding. The sound spectrum for when no robots are present could be used as a good reference for obtaining the values. Figure 4 shows the frequency spectrums of two sound sequences, when "no drone" and when "a drone" was present in a room. A suitable frequency range can be obtained from comparing the two spectrums to find the range where the robot sound is mostly present. Also since in the defined range there are very few frequency bins that do not contain the sound of the robot, $\alpha = 1$ would be suitable to consider equally all of the frequencies of this range.

The instantaneous frequency spectrum of the measured sequences is also employed to detect the presence of robots in the vicinity and to distinguish the sound of a robot from other sound sources that might be present. For this, a template
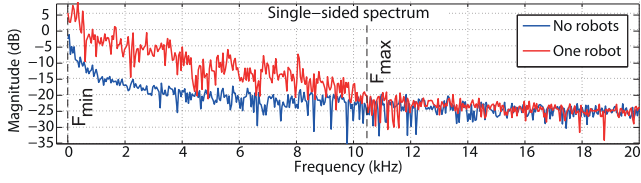
Fig. 4. Frequency spectrums of sound sequence in two cases of "no drone" and "one drone" recorded inside a room'. This information can be used to find the parameters of the proposed weighting function

Fig. 5. Coherence measurement plots in an experiment with one perceiving drone and two target rotorcraft drones, illustrating the pruning step

of the robot's sound frequency spectrum is stored in the memory and is continuously correlated with one of the microphone signals, prior to the coherence measurements. A good correlation, defined by a threshold found empirically, indicates that the sound of the desired robot is present.

C. Relative Bearing Measurement Unit

Upon finding the similarity measures R_{ij} from (1) for all microphone pairs ij, a search for the most likely source direction $\overrightarrow{b}_m$ is performed. This is the direction that maximizes the sum of the coherence measurements from all pairs.

$$\overrightarrow{b}_m = \arg\max_{\overrightarrow{b}} \sum_{i,j} R_{ij}(\tau_{\overrightarrow{b}ij}) \qquad (6)$$

where $\tau_{\overrightarrow{b}ij}$ is the expected time delay if the source was in the direction $\overrightarrow{b}$, and is computed from the coordinates of microphones i and j in the body reference frame.

$$\tau_{\overrightarrow{b}ij} = \frac{\overrightarrow{x}_{ij} \cdot \overrightarrow{b}}{c} \qquad (7)$$

where $\overrightarrow{x}_{ij}$ is a vector connecting microphone i to j and c is the speed of sound. A spherical geodesic grid, defined on a unit sphere, is used to search for the most likely source direction among the set of all potential directions. Each grid point represents a direction vector $\overrightarrow{b}$ that starts at the origin and ends at that grid point. For better resolution a weighted averaging between the grid point with the highest coherence value and its six adjacent points, based on their coherence values, is used as the final estimated target direction.

D. Coherence Pruning Unit

The previous units described a method for locating the direction of a single neighboring robot by finding the direction with the maximum coherence among all pairs. In the case of multiple neighboring robots, this method will provide the direction of the dominant sound source that is exhibiting the highest coherence in the similarity measurements. For a homogeneous group of drones with the same sound characteristics, the dominant sound source will correspond to the nearest neighbor. However, in practice, this might not always be the case as the engine sounds are varying with throttle power. The Coherence pruning unit is used to also obtain the bearing information of other existing robots that are potentially masked behind the sound of the dominant robot. Inspired by the work of [36], the idea is to
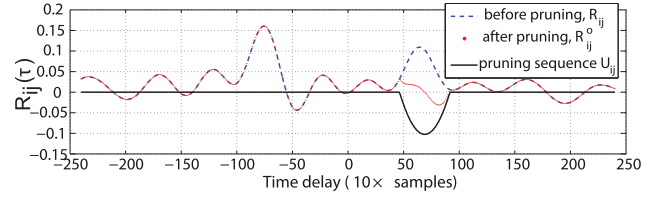
de-emphasize the effect of the dominant robot in the similarity measurements to locate other robots with weaker coherences.

This unit initially uses Equation (7) to compute the time delay $\tau_{\overrightarrow{b}_m ij}$, for every microphone pair, corresponding to the obtained bearing measurement $\overrightarrow{b}_m$. This time delay is then used to generate a pruning sequence $U_{ij}(\tau)$ that is added to the similarity measurements $R_{ij}(\tau)$ to de-emphasize the existing peak of the dominant source at time lag $\tau_{\overrightarrow{b}_m ij}$.

$$R_{ij}^o(\tau) = R_{ij}(\tau) + U_{ij}(\tau) \qquad (8)$$

$$U_{ij}(\tau) = \begin{cases} B(\tau) & B(\tau) > 0 \\ 0 & B(\tau) \leq 0 \end{cases}$$

$$B(\tau) = \frac{1}{L}\left(\tau - \tau_{\overrightarrow{b}_m ij}\right)^2 - R_{ij}(\tau_{\overrightarrow{b}_m ij}) \qquad \tau \in [-\tau_{\max}, \tau_{\max}]$$

where $B(\tau)$ is a second order polynomial sequence and L is a constant that defines the sharpness of this polynomial. The value of L is chosen to produce a polynomial with approximately equal sharpness as the coherence peaks. Figure 5 shows plots of a microphone pair's coherence measurements in an experiment with one perceiving drone and two target drones, before and after the coherence pruning was preformed. Note that the localized dominant sound source does not necessarily correspond to the dominant peaks in the coherence measurements, as shown in this Figure. The pruned coherence measurements R_{ij}^o are then passed into the bearing measurement unit to search again for the most probable sound direction. This procedure is repeated similarly for locating other existing robots that might also be present. Experiments discussed in Section IV-A show that up to two pruning iterations can be performed to acquire an accurate direction for the three most dominant target robots. A large drop in the precision is observed after the third pruning iteration,i.e. for obtaining the direction of the fourth dominant robot. However, in reality, since the dominance of robots alter in time due to the change in the engine sound and the movements of robots it is potentially possible to track the direction of more robots throughout time with only two pruning iterations and a memory based algorithm.

III. ACTIVE AUDIO-BASED SYSTEM

A passive audio-based relative bearing measurement system, based on perceiving the engine-sound of robots, was described in the previous section. Although this method has many advantages, some drawbacks such as the no self-engine noise constraint, the detection range depending on the target's sound level, instantaneous detection of up to three robots, and
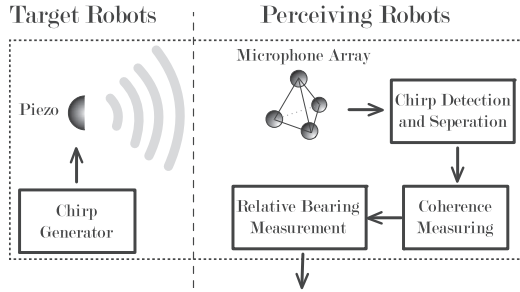
Fig. 6. Schematic diagram of the proposed relative positioning system illustrating main parts of the system.
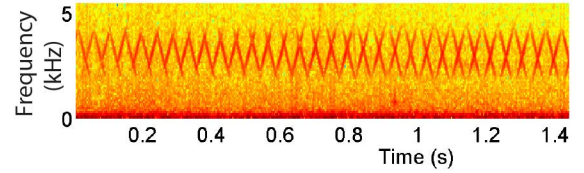

Fig. 7. Spectrogram of an in-flight sound recording with one perceiving drone and two chirping drones (One produced up-sweep chirps, 1.7 kHz to 4.7 kHz, and the other produced down-sweep chirps, 4.7 kHz to 1.7 kHz).

not being able to identify robot identities motivated research for a better solution. This section explains a method based on active sound signaling where robots generate unique sounds to assist each other in finding the relative bearing information despite the loud self-engine noise.

Figure 6 shows the block diagram of the active audio-based relative bearing measurement system. The system is divided into two main parts of "Target robot" and "Perceiving robot" states to illustrate the main units of the system at each state. In the target robot state, the "Chirp Generator" of a robot generates unique chirp sound of predefined rate and frequency. In the perceiving state, sound waves are picked up by an onboard microphone array and are continuously checked by the "Chirp Detection and Separation" unit for existence of chirps in the sound mixture. When a full chirp is detected, it is filtered out from the sound mixture and is passed to the "Coherence Measuring" unit. The coherence measuring unit, previously described in Section II-B, obtains a measure of similarity between the chirps among all of the microphone pairs. The frequency range used in the coherence measuring, defined by f_{min} and f_{max} in Equation (5), is chosen as the minimum and maximum frequencies of the perceived chirp. The value of $\alpha = 1$ is used to consider equally all the frequencies within the chirp's range of frequencies. The resulting coherence measurements and the knowledge of the microphone array's geometry is then used by the Relative Bearing Measurement unit, explained in Section II-C, to estimate a measure of the target's direction. A more detailed explanation of the units that were not explained previously is presented here.

A. Chirp Generator

Piezoelectric transducers are simple, inexpensive and lightweight devices, suitable to be used on drones, that generate sound by converting electrical signals into mechanical vibrations. A loud sound wave can be produced if the input signal frequencies are close to the resonance frequency of the piezo element. To generate a loud sound that is needed for achieving a long detection range, particularly suitable for outdoor operations, and to avoid the problem of ambiguous bearing measurements related with localization of narrowband sounds [26], a piezo transducer is used on the robots to generate band limited linear chirp signals.

The chirp generating unit of every target robot generates linear chirps with a predefined and unique chirp rate. Since an entire chirp is used by the perceiving robots for computing a single bearing measurement, the time interval between the chirps can be chosen in accordance with the required measurement rates. To generate continuous chirps, the frequency of the input sine-wave signal to the piezoelectric element is varied in time:

$$f(t) = F_{str} + \frac{F_{end} - F_{str}}{\Delta T} \bmod \left(\frac{t}{\Delta T} \right) \qquad (9)$$

where F_{str} and F_{end} are the starting and ending frequencies of the chirp that is chosen differently for every robot, and ΔT is the chirp duration that is equal to the sampled sequence, i.e. $\Delta T = N/F_s$. For better detection against the self-engine noise the chirp's frequency range is chosen in the range where the engine sound is less dominant, particularly in multi-kilohertz range. Figure 7 shows the spectrogram of a sound recording, performed with three drones, i.e. one perceiving drone and two target drones chirping continuously.

B. Chirp Detection and Extraction

This unit is responsible for the detection and extraction of a chirp in the perceived sound mixture. For this purpose, one of the microphone signals are continuously checked to find the existence of chirps in the sound mixture. The presence of a chirp is detected by template matching, where a continuous cross correlation of the sound mixture with the template of the desired chirp determines the existence and the time segment that contains the chirp.

After a chirp is detected, it is filtered out from other sounds and overlapping chirps that might also be present in the selected time segment. For this purpose, initially the time segment from all of the microphone signals are passed through a band-pass filter to remove the unwanted low and high frequency noises that are outside the chirp's frequency range (see Figure 8.b). Furthermore, Fractional Fourier transform (FRFT) [37] is used to remove noise within the frequency range of the chirp. Unlike Fast Fourier transform (FFT), the FRFT provides a compact representation of chirp signals, which makes it possible to remove the noise inside the same frequency region as the chirp, that cannot be removed with traditional frequency domain filters.

To represent the detected chirp in its most compact form, the Fractional Fourier transform (FRFT) of the time window containing the entire chirp is computed with an FRFT order of α obtained by the following equation.

$$\alpha = \frac{2}{\pi} \varphi = \frac{2}{\pi} \tan^{-1} \left(\frac{f_s}{F_{end} - F_{str}} \right) \qquad (10)$$
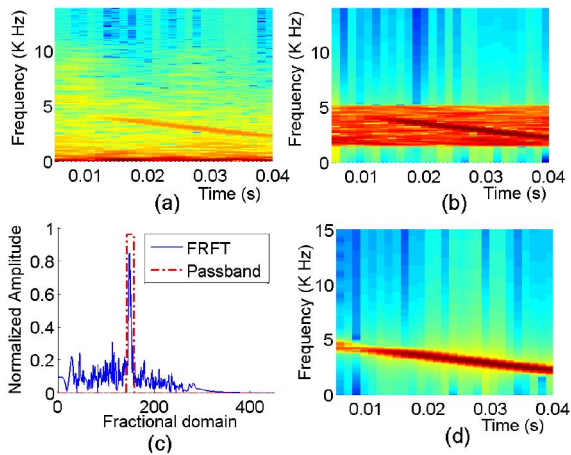
Fig. 8. In-flight sound of a chirping drone recorded by a perceiving drone in different steps of the chirp extraction procedure (a) Spectrogram of a detected chirp (b)Spectrogram of the signal after band-pass filtering (c) FRFT transform of the band-passed chirp and the corresponding passband region (d) Spectrogram of the final filtered chirp.
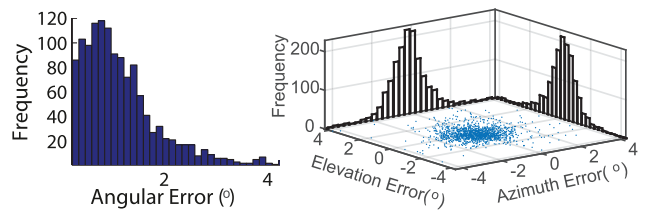

Fig. 9. Histogram of the angular error between the true and the measured 3D-direction vectors along with the 2D distribution and histograms for the errors in azimuth and elevation.
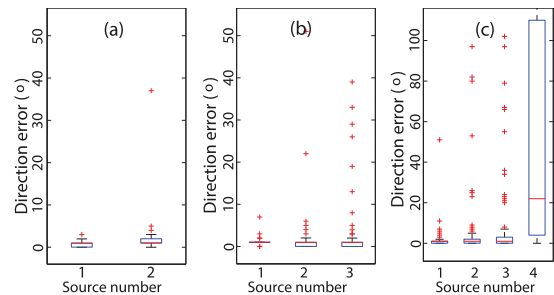

Fig. 10. Relative direction measurement error in experiments with multiple targets. a) Two targets b) Three targets c) Four targets

where F_{str} and F_{end} are the start and end frequencies of the perceived chirp in Hertz and f_s is the sampling frequency. Equation (10) was derived from the geometrical relationship between the chirp rate and FRFT order for time-frequency discretized chirp signals provided by [38].

Upon computing the FRFT, the result contains an impulse-shaped peak that corresponds to the desired chirp. This chirp is then filtered out from other sounds, by only retaining the bin with the highest peak along with its few nearby bins and setting all other bins to zero (illustrated in Figure 8.c). The ratio of the peak value to the mean value of all zeroed bins prior to zeroing provides a good measure for the quality of the perceived chirp and the reliability of the obtained bearing measurement. Finally, the filtered chirp in the FRFT domain is transformed back to the time domain by computing the inverse FRFT. Steps and result of the chirp extraction procedure for a detected chirp in a real world experiment with two flying drones, one chirping drone and one observing drone, is illustrated in Figure 8.

IV. EXPERIMENTS AND RESULTS

Experiments were performed to test the proposed passive and active systems on real drones. Two different types of drones (Figure 2) were used in these experiments. A small embedded circuit board, equipped with an Atmel AVR32 microcontroller, was developed and used on the drones for sampling the microphone signals and for the computations.

A. Passive System

The pocket size drones (Figure 2b) were used indoors to test the passive method. In first experiments, a target drone was flown manually inside an empty room, with dimensions $(6 \times 3.5 \times 3)$ meters, where a perceiving drone was resting on the ground in the center of the room with engines off. A motion tracking system was used for measuring the true robot positions and to compute the actual 3D relative bearings between

robots. No other major sources of acoustic noise was present inside the room, however, the sound of the cooling fans belonging to the 8 tracking cameras and two computers were heard clearly. Accurate bearing measurements were obtained despite the relatively quiet engine of these robots. Figure 9 shows the angular error between the measured and the true 3D-direction vectors, computed with vector dot product, and a 2D representation of the error distribution using the azimuth and elevation angles between the vectors. A root mean square error of 1.39 degrees was computed for the angular error.

Experiments were also performed to test the performance of the system and the coherence pruning when experiencing sound from multiple targets. Figure 10 shows box plots of the angular error in the cases of two, three and four targets. In these plots, the target number n is the target that is localized after $n-1$ pruning iterations. In all experiments up to three targets were localized with a good accuracy. However, the precision dropped significantly after the third pruning step resulting in poor measurements for the direction of the fourth target. The number of existing targets did not have a major effect on the localization for the first three dominant targets, with only a minor increase in the number of outliers.

B. Active System

Outdoor experiments were performed with the drones shown in (Figure 2a), to test the active system in presence of self engine noise. An array of 4 microphones was used on a perceiving robot and the target robots were equipped with a piezo transducer. The array consisted of three microphones forming a triangle of edge length 18cm, each placed between the propellers, and the fourth microphone was placed approximately 5.2cm under the center of the triangle. Although wind noise is mostly present in the low frequencies (< 200 Hz) and is
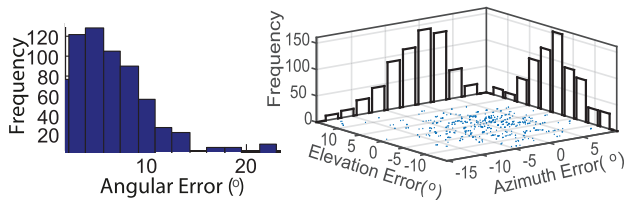
Fig. 11. Histograms for the angular, azimuth and elevation errors between the GPS-based and the audio-based direction vectors, for two flying drones.

Fig. 12. Number of incorrectly localized targets for the cases of 6,8,10, 12 and 14 targets with overlapping chirps

removed by the bandpass filter, but microphones were covered with a small piece of foam to prevent potential signal clipping due to strong winds.

Figure 11 illustrate the histogram of the angular difference between the GPS based and audio-based relative bearing measurements for an experiment where a perceiving drone was flown in the vicinity of a continuously chirping and hovering drone. Experiments showed the success of detection, filtering and localizing the chirp sound of target robots in presence of the self-engine noise. A detection range of ≈ 60 meters was achieved in flight, however this range increased to about ≈ 150 meters if the perceiving robot rested on the ground with its engines off.

Outdoor experiments for up to two target drones chirping continuously with overlapping chirps (Figure 7) showed similar performances. To test the performance in the case of more than two overlapping chirps, experiments were conducted in simulation. In these experiments a free-field sound-wave propagation model was used to simulate the sound at the point of a microphone from randomly positioned sources. The sound received by a microphone was simulated by superimposing the waves from all sources, based on the principle of superposition of sound waves. Every source generated a continuous chirp sequence, with unique chirp rates but with equal chirp duration, to ensure overlapping of all chirps. The chirp sequence for every source was randomly started in every experiment to simulate the fact that the chirp timing between robots are not synchronous and that chirps from different robots could overlap at different times. A microphone array with the same dimension as the one on the real drone was used for the simulations. Experiments showed that for large number of overlapping chirps some chirps might not be detected or correctly localized. Figure 12 shows the number of sources that were incorrectly localized for the cases of 6, 8, 10, 12 and 14 overlapping chirps, each computed from 100 experiment runs with random overlapping configurations. Note that in practice it is possible to reduce the number of overlapping chirps, by reducing the rate of chirping of robots, or by using different piezos and spreading the chirps in different frequency ranges.

In addition, a leader-follower motion coordination between two drones was implemented to further test the onboard measurement system. For this, a reactive proportional controller was implemented on a drone to autonomously follow a chirping drone based entirely on the instantaneous audio-based 3D bearing measurements. The heading (i.e. yaw) of the follower was directly controlled by the relative azimuth measurements to try orient it towards the leader. While a pressure sensor was used to
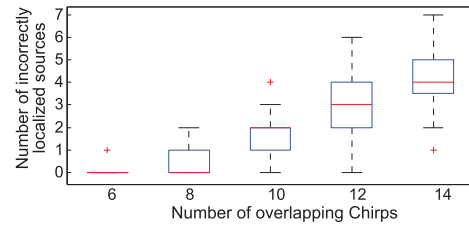

Fig. 13. Photo showing an instance of the audio-based leader-follower experiments. Videos at http://lis.epfl.ch/ABSMAV
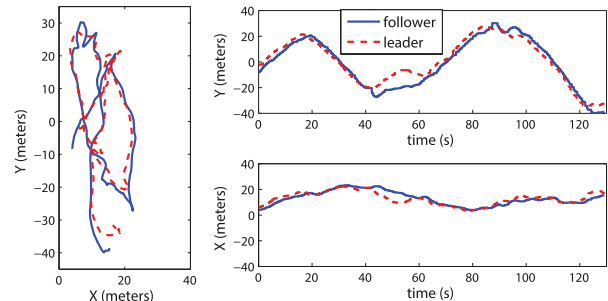

Fig. 14. GPS Path of drones for the leader-follower experiment.

maintain a fixed altitude, the speed of the follower was directly controlled by the relative elevation measurements to try getting closer to the target, i.e. $Speed \propto (90^o - |elevation^o|)$. Note that this method requires the leader and follower to be on different altitudes, otherwise the follower can reach and collide with the leader. Figure 13 shows an instance of the leader-following experiment. Figure 14 shows the GPS path of the two drones, illustrating the success of the fully autonomous follower to follow the leader while using the 3D sound-based bearing information.

V. CONCLUSIONS

This letter described an onboard system for measuring the relative bearing among robots in a team of drones, where a passive and an active method was presented. Both methods used a lightweight array of microphones to measure the direction of sound. The passive method relied on the available engine sound of robots and it showed a good accuracy in measuring the direction of up to three drones in the absence of self engine noises. The active method relied on perceiving acoustic chirps emitted by robots that allowed obtaining the relative bearings in the presence of self-engine noise, to increase the detection range and to discriminate the identity of robots. A fully autonomous leader following behavior was demonstrated

on real drones using the on-board audio-based system. Design of new acoustic sensors, wind protectors and anti vibration mounts more suitable for drones, research on self-engine noise cancellation both at the mechanical and computational levels, exploiting other acoustic cues such as intensity or Doppler shift of frequency, integration of the audio based system with other technologies such as vision or radio waves, and combining the bearing measurements with predictions estimated from the motion model of wirelessly-connected drones using a nonlinear filtering framework, are some future work aiming at improving relative positioning.

REFERENCES

[1] S. Hauert, S. Leven, J. Zufferey, and D. Floreano, "Communication-based swarming for flying robots," in *Proc. Int. Conf. Rob. Autom. Workshop Netw. Sci. Syst.*, Anchorage, AL, USA, 2010, pp. 15–20.

[2] J. Oyekan and H. Huosheng, "Toward bacterial swarm for environmental monitoring," in *Proc. IEEE Int. Conf. Autom. Logist. (ICAL'09)*, 2009, pp. 399–404.

[3] L. Merino, F. Caballero, J. R. Martínez-de Dios, J. Ferruz, and A. Ollero, "A cooperative perception system for multiple UAVS: Application to automatic detection of forest fires," *J. Field Rob.*, vol. 23, nos. 3–4, pp. 165–184, 2006.

[4] R. W. Beard, T. W. McLain, D. B. Nelson, D. Kingston, and D. Johanson, "Decentralized cooperative aerial surveillance using fixed-wing miniature UAVS," *Proc. IEEE*, vol. 94, no. 7, pp. 1306–1324, Jul. 2006.

[5] N. Michael, J. Fink, and V. Kumar, "Cooperative manipulation and transportation with aerial robots," *Auton. Robots*, vol. 30, no. 1, pp. 73–86, 2011.

[6] Q. Lindsey, D. Mellinger, and V. Kumar, "Construction with quadrotor teams," *Auton. Robots*, vol. 33, no. 3, pp. 323–336, 2012.

[7] F. Ruini and A. Cangelosi, "Extending the evolutionary robotics approach to flying machines: An application to MAV teams," *Neural Netw.*, vol. 22, no. 5, pp. 812–821, 2009.

[8] A. Kushleyev, D. Mellinger, C. Powers, and V. Kumar, "Towards a swarm of agile micro quadrotors," *Auton. Robots*, vol. 35, no. 4, pp. 287–300, 2013.

[9] R. Carnie, R. Walker, and P. Corke, "Image processing algorithms for UAV "sense and avoid"," in *Proc. IEEE Int. Conf. Rob. Autom. (ICRA)*, May 2006, pp. 2848–2853.

[10] M. Basiri, A. Bishop, and P. Jensfelt, "Distributed control of triangular formations with angle-only constraints," *Syst. Control Lett.*, vol. 59, no. 2, pp. 147–154, 2010.

[11] J. Pugh and A. Martinoli, "Relative localization and communication module for small-scale multi-robot systems," in *Proc. IEEE Int. Conf. Rob. Autom. (ICRA)*, May 2006, pp. 188–193.

[12] A. Howard, M. J. Matarić, and G. S. Sukhatme, "Localization for mobile robot teams: A distributed MLE approach," in *Experimental Robotics VIII*. New York, NY, USA: Springer, 2003, pp. 146–155.

[13] S. Scherer, S. Singh, L. Chamberlain, and S. Saripalli, "Flying fast and low among obstacles," in *Proc. IEEE Int. Conf. Rob. Autom. (ICRA)*, 2007, pp. 2023–2029.

[14] J. B. Saunders, B. Call, A. Curtis, R. W. Beard, and T. W. McLain, "Static and dynamic obstacle avoidance in miniature air vehicles," in *Proc. AIAA Infotech Aerosp.*, 2005.

[15] C. Kownacki, "Obstacle avoidance strategy for micro aerial vehicle," in *Advances in Aerospace Guidance, Navigation and Control*. New York, NY, USA: Springer, 2011, pp. 117–135.

[16] S. Lanzisera, D. T. Lin, and K. S. Pister, "RF time of flight ranging for wireless sensor network localization," in *Proc. IEEE Int. Workshop Intell. Solutions Embedded Syst.*, 2006, pp. 1–12.

[17] A. Viquerat, L. Blackhall, A. Reid, S. Sukkarieh, and G. Brooker, "Reactive collision avoidance for unmanned aerial vehicles using doppler radar," in *Field and Service Robotics*. New York, NY, USA: Springer, 2008, pp. 245–254.

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

[19] J. Utt, J. McCalmont, and M. Deschenes, "Development of a sense and avoid system," in *Proc. AIAA Infotech Aerosp.*, 2005.

[20] L. Mejias, S. McNamara, J. Lai, and J. Ford, "Vision-based detection and tracking of aerial targets for UAV collision avoidance," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2010, pp. 87–92.

[21] C. Melhuish, J. Welsby, and P. Greenway, "Gradient ascent with a group of minimalist real robots: Implementing secondary swarming," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2002, vol. 2, pp. 509–514.

[22] J. Roberts, T. Stirling, J.-C. Zufferey, and D. Floreano, "2.5d infrared range and bearing system for collective robotics," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2009, pp. 3659–3664.

[23] N. Kottege and U. Zimmer, "Relative localisation for AUV swarms," in *Proc. Symp. Underwater Technol. Workshop Sci. Use Submarine Cables Relat. Technol.*, 2007, pp. 588–593.

[24] A. Farnsworth, "Flight calls and their value for future ornithological studies and conservation research," *Auk*, vol. 122, no. 3, pp. 733–746, 2005.

[25] P. Muller and D. Robert, "A shot in the dark: The silent quest of a free-flying phonotactic fly," *J. Exp. Biol.*, vol. 204, no. 6, pp. 1039–1052, 2001.

[26] M. Basiri, F. Schill, P. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2012, pp. 4737–4742.

[27] P. Angelov, C. D. Bocaniala, C. Xideas, C. Patchett, D. Ansell, M. Everett, and G. Leng, "A passive approach to autonomous collision detection and avoidance in uninhabited aerial systems," in *Proc. 10th Int. Conf. Comput. Model. Sim*, 2008, pp. 64–69.

[28] A. Franchi, C. Masone, H. H. Bülthoff, and P. R. Giordano, "Bilateral teleoperation of multiple UAVS with decentralized bearing-only formation control," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2011, pp. 2215–2222.

[29] X. S. Zhou *et al.*, "Determining 3-D relative transformations for any combination of range and bearing measurements," *IEEE Trans. Rob.*, vol. 29, no. 2, pp. 458–474, Apr. 2013.

[30] M. Basiri, F. Schill, D. Floreano, and P. U. Lima, "Audio-based localization for swarms of micro air vehicles," in *Proc. IEEE Int. Conf. Rob. Autom. (ICRA)*, 2014, pp. 4729–4734.

[31] M. Basiri, F. Schill, D. Floreano, and P. U. Lima, "Audio-based relative positioning system for multiple micro air vehicle systems." in *Proc. Rob.: Sci. Syst.*, Berlin, Germany, 2013.

[32] J.-S. Hu, C.-M. Tsai, C.-Y. Chan, and Y.-J. Chang, "Geometrical arrangement of microphone array for accuracy enhancement in sound source localization," in *Proc. IEEE 8th Asian Control Conf. (ASCC)*, Kaohsiung, Taiwan, 2011, pp. 299–304.

[33] J. C. Murray, H. Erwin, and S. Wermter, "Robotics sound-source localization and tracking using interaural time difference and cross-correlation," in *Proc. NeuroBotics Workshop*, 2004, pp. 89–97.

[34] J. Perez-Lorenzo, R. Viciana-Abad, P. Reche-Lopez, F. Rivas, J. Escolano, "Evaluation of generalized cross-correlation methods for direction of arrival estimation using two microphones in real environments," *Appl. Acoust.*, vol. 73, no. 8, pp. 698–712, 2012.

[35] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[36] A. Brutti, M. Omologo, and P. Svaizer, "Localization of multiple speakers based on a two step acoustic map analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2008, pp. 4349–4352.

[37] V. Namias, "The fractional order fourier transform and its application to quantum mechanics," *IMA J. Appl. Math.*, vol. 25, no. 3, pp. 241–265, 1980.

[38] C. Capus and K. Brown, "Short-time fractional fourier methods for the time-frequency representation of chirp signals," *J. Acoust. Soc. Amer.*, vol. 113, no. 6, pp. 3253–3263, 2003.