Delft University of Technology

Lightweight Audio Source Localization for Swarm Robots

Majid, A.Y.; van der Horst, C.; van Rietbergen, T.L.; Zwart, D.J.; Venkatesha Prasad, R.R.

DOI
10.1109/CCNC49032.2021.9369572

Publication date
2021

Document Version
Accepted author manuscript

Published in
2021 IEEE 18th Annual Consumer Communications and Networking Conference, CCNC 2021

Citation (APA)
Majid, A. Y., van der Horst, C., van Rietbergen, T. L., Zwart, D. J., & Venkatesha Prasad, R. R. (2021). Lightweight Audio Source Localization for Swarm Robots. In *2021 IEEE 18th Annual Consumer Communications and Networking Conference, CCNC 2021* Article 9369572 (2021 IEEE 18th Annual Consumer Communications and Networking Conference, CCNC 2021). IEEE. https://doi.org/10.1109/CCNC49032.2021.9369572

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Lightweight Audio Source Localization for Swarm Robots

Amjad Yousef Majid
Delft University of Technology
Delft, The Netherlands
a.y.majid@tudelft.nl

Casper van der Horst
Delft University of Technology
Delft, The Netherlands
c.vanderhorst@student.tudelft.nl

Tomas van Rietbergen
Delft University of Technology
Delft, The Netherlands
t.l.vanRietbergen@student.tudelft.nl

David Johannes Zwart
Delft University of Technology
Delft, The Netherlands
d.j.zwart@student.tudelft.nl

R Venkatesha Prasad
Delft University of Technology
Delft, The Netherlands
r.r.venkateshaprasad@tudelft.nl

***Abstract*—Audio-based localization forms a potential solution for the coordination of resource-restricted (especially in the wireless spectrum) swarm robots while offering several interesting advantages. This paper proposes lightweight acoustic source localization methods aimed towards swarm robots with limited hardware capabilities. By leveraging recent advances in hardware technology the computational complexity of localizing a sound source can easily be handled. Using tiny microphones with a wake-on-sound feature the angle of arrival of a sound signal can be determined with minimal signal processing. The plane cutting algorithm is introduced which uses an audio-based localization approach built around the wake-on-sound feature and Valin's algorithm is adapted to exploit the wake-on-sound microphones. Hardware experiments were performed to determine the triggering accuracy of the microphones and the reliability of an energy-based distance estimator. Furthermore, the performance of the lightweight localization algorithm was investigated in Matlab and compared to the state of the art pure software-based approaches. This work will open up multitudes of innovation in the near future.**

I. Introduction

The value of swarm robotic systems lays in utilizing many simple and cheap robots to tackle complex tasks [1, 2, 3, 4]. The potential of these simple robots hinges on their coordination and collaboration capabilities [5]. For example, cooperative and collaborative swarms can construct 3D structures [6], transport relatively heavy objects [7] or optimize the search of an unknown terrain [5].

For these robots to operate *efficiently* they need to be able to locate their neighboring robots. Relative localization can be computationally expensive (e.g., vision-based localization [8] and cross-correlation audio-based localization methods [9]), which conflicts with the principle of simple robot design for scalable swarms.

To address the above important requirements of simplicity and effectiveness in building robot swarm applications, we propose an acoustic-based relative localization system with lightweight computational demand. Our system is capable of determining the direction of arrival (DOA) of an audio signal without the need for signal processing. Also, it takes advantage of the problem setup to determine the relative distances between the robots based on the Received Signal Strength (RSSI) of the audio signal.

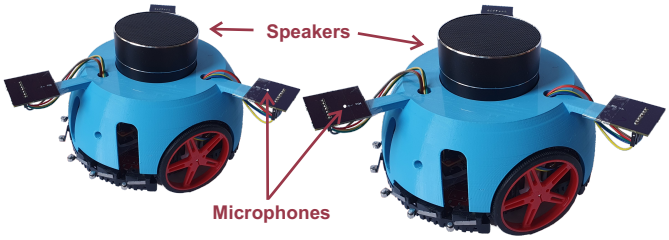


Fig. 1: Audio-based relative localization system mounted on mobile robots.

Although different signal types can be utilized for relative localization (e.g., optical signals [10] or RF transmissions [11]) audio signals have several compelling features: Audio signals propagate in darkness, humans can perceive audio making robots more socially acceptable, RF spectrum is heavily used in the ambiance and thus experiences a lot interference, and audio does not require line-of-sight or any infrastructure. Therefore, this work focuses on using acoustic signals for relative localization tasks.

Challenge. In this paper we deal with one important question: *What does it take to develop a lightweight audio-based relative positioning estimator for simple swarm robots?* The challenge here is to find an end-to-end solution for the problem at hand and also provide in-depth experimental results using real hardware (Figure 1).

Contributions. By taking advantage of the recent advancements in microphone hardware technology we design a computationally lightweight sound source localizer for swarm robots. This localizer can estimate the DOA of an audio signal without the need for signal processing. Also, it uses the RSSI values of the received signal for estimating the distances between the robots. Lastly, we show that wideband acoustic signals provide much more reliable distance estimations than narrowband signals.

II. Related Work

Acoustic source localization methods for robots can be divided into three main categories: temporal-, beamforming-, and machine learning-based approaches [12].

Having multiple microphones a receiver can estimate the azimuth (and elevation) angle of an arriving sound signal by correlating the received signals or exploiting their spatial and spectral characteristics [12]. For example, Valin et al. [13] mounted eight microphones on a mobile robot and used the Time Difference Of Arrival (TDOA) between the signals received by the microphones to estimate the direction of a single sound source in the azimuth plane.

Grondin and Michaud [14] presented a novel sound source localization and tracking method that requires 16 (or eight) microphones. The authors employed several techniques to reduce the computation load. For example, a coarse 3D audio scanning for a preliminary estimation of the source location which is then used by a higher resolution scanner to provide more accuracy, TDOA uncertainty modeling, and Kalman filtering.

In reverberant environments, beamforming-based localization methods reduce emitting sidelobes from sound sources resulting in directionality and reduced echoes [15, 16]. This provides a robust alternative to the TDOA estimation approach [17] by using microphone redundancy as the main leverage. Valin et al. [18] used beamforming coupled with particle filtering to enable robust 3D sound source localization and tracking, whilst still limiting the number of microphones used. Despite the improved accuracy of the algorithm for speech and noise bursts, it underperforms in detecting certain audio signals types. The beamforming approach also suffers from a high computational load. This is also the reason why the solution from Grondin and Michaud [14], as mentioned before, was still based on producing reliable TDOA with a light computational load.

Deep learning excels in the areas where the relationship between the observed signal and the underlying processes cannot be captured with simple models. Acoustic source localization is no exception [19, 20, 21, 22]. For example, Adavanne et al. [23] show how convolutional recurrent neural networks can aid in making the DOA detection robust against overlapping sound signals. As with beamforming, however, the computational load of trained neural networks for heterogeneous robots is a major issue as these robots cannot perform the required computation at an acceptable speed.

Our work aims to provide a computationally light algorithm for estimating a sound source. As such, we focus our attention on detecting a single acoustic source using the Time of Arrival (TOA) and RSSI as the metrics for locating the sound source.

III. System Overview and Modeling

A. System Overview

Our proposed localization system comprises an array of microphones with the Wake-on-Sound [24] feature (for 2D localization a minimum of three microphones is required); a speaker, to enable relative localization between the robots of a swarm; and a microcontroller (MCU) to run the localization algorithm. The microphones interrupt the MCU on an arriving sound signal. Thus, the MCU can directly record the time of arrival (TOA) of the signal using its internal clock. As the other microphones trigger, a set of TOAs is generated which is then used to estimate the Direction of Arrival (DOA) of the received signal. After notifications from the microphones, the MCU sets one of the microphones in recording mode to record the signal for energy determination. There are two major benefits to this approach: low cost and minimal signal processing. The microphone we used costs less than 2€. More importantly, this setup greatly reduces the computational power required by the MCU as it does not require any signal processing to determine the DOA of the sound signal (cf., CPU expensive cross-correlation in traditional systems) and only requires a single sound signal to be processed for distance estimation.

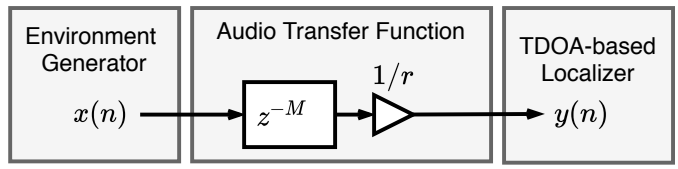


Fig. 2: The main components of an audio-based localization simulator using TDOA. Input audio signal $x(n)$ is delayed by z^{-M} representing M samples, attenuation is represented by $1/r$ (r being distance) and the result is a time-discrete signal $y(n)$.

B. Modeling

We have developed a modular MATLAB-based simulator to determine the accuracy and applicability of different audio-based localization algorithms for a swarm of mobile ground robots. Figure 2 illustrates the main components of this simulator.

1) Environment Generator: The first step of the model comprises defining the environment in which the simulation is done. Here the locations of the sound beacons, robots, and the microphones on the robots are set. The propagation medium represents air, and the user can feed the simulator with acoustic signals of arbitrary forms.

2) Audio Transfer: For each beacon-microphone pair, a transfer function is derived based on the Euclidean distance between them. The beacons are modeled as point sources. This results in an amplitude gain of the signal of $\frac{1}{r}$ (with r being the distance in meters) and a delay of $\frac{r*F_s}{c}$ samples (with F_s as the sampling frequency and c as the speed of sound). Consequently, the following transfer function [25],

$$H(z) = \frac{1}{r \cdot z^{\frac{r \cdot F_s}{c}}} \tag{1}$$

models the sound signal propagation at any given location. At the microphones, the superimposed sound signal is considered in the case of simultaneous transmissions, and the transfer of the recorded signal to the MCU is simulated by downsampling it. Additionally, the simulator features a noisy channel and

digital filters which allows us to observe the effects of different noise levels and filtering delays on determining the TDOA.

3) Determining the Time Difference of Arrival: We can distinguish two approaches for finding the TDOA of sound signals. First, a hardware-assisted approach that relies on the Wake-on-Sound feature of the PMM-3738-VM1010-R microphone [24] to determine the Time of Arrivals (TOAs) of the received signals. Then using the generated set of TOAs, the TDOAs can be found. The datasheet of the microphone [24] does not characterize the triggering delay of the microphone. This delay plays a crucial role in determining the validity of this approach. However, the datasheet states that an MCU can turn on the microphone within 200 μs after an alert from the microphone. Therefore, a uniformly distributed random delay in the range of $[50\ \mu s, 200\ \mu s]$ is added to the simulation. Second, a pure software approach where the highest peaks generated by cross-correlating the received signals determine the TDOAs.

4) Estimating Sound Source Direction: Here we introduce the Cutting The Plane (CTP) algorithm which is a hardware-assisted audio source localization algorithm. And, for comparison, we briefly describe a software-based algorithm is that is conceived by Valin et al. [13].

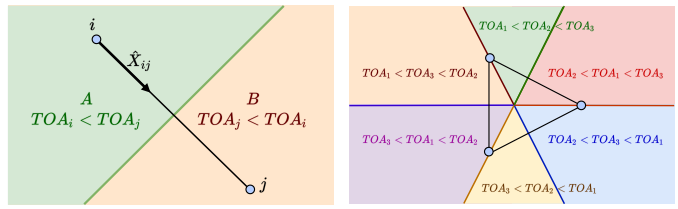
Cutting the plane algorithm. The CTP algorithm was designed as a low-computation algorithm to find a coarse directional estimation based on which microphone of a pair receives the signal first. This algorithm was designed with the Wake-on-Sound behavior of the microphones in mind.

For each pair of microphones i and j the perpendicular bisector divides the plane into two regions A and B (Figure 3a). All points in region A are closer to microphone i than to microphone j and vice versa. This means that a sound from a beacon placed in region A will reach microphone i before reaching microphone j. This means that comparing the times of arrival (TOAs) of the sound for microphones i and j (TOA_i and TOA_j respectively) will cut the plane of possible beacon locations in half. If we define the vector $\hat{X}_{ij}$ as the unit vector from i to j, then we can estimate the direction vector towards the beacon $\bar{u}$ as,

$$\bar{u} = \begin{cases} \hat{X}_{ij} & \text{if } TOA_j < TOA_i \\ -\hat{X}_{ij} & \text{otherwise.} \end{cases} \tag{2}$$

The full CTP algorithm is depicted in Algorithm 1. When N microphones are uniformly distributed around a circle and $N \geq 3$, the number of sections is equal to $2N$ and the accuracy of the estimated direction is $\pm 180°/N$. Although this algorithm might not provide high direction estimation accuracy due to hardware limitations, its strength lies in its simplicity. When combined with the wake-on-sound feature of the PMM-3738-VM1010-R microphones an MCU merely needs to compare the clock values of the arrivals of the triggers from the microphones and add or subtract predefined values.

Valin's algorithm. A more accurate but computationally heavy direction estimation method is proposed by [13]. The concept is made for audio-based localization with microphone arrays, which have small dimensions relative to the distance



(a) Perpendicular bisector cutting the plane of possible beacon locations in half based on TOA.

(b) Sections created by the CTP algorithm when three microphones are uniformly distributed around a circle.

Fig. 3: Cutting the plane between two microphones and the sections created with a three microphone setup.

Algorithm 1 Cutting the plane algorithm

Output: $\bar{u}$ (sound direction estimate)

```
ū ← 0̄
for i ← 1 to N_mics − 1 do
    for j ← i + 1 to N_mics do
        if TOA_j < TOA_i then
            ū ← ū + X̂_ij
        else
            ū ← ū − X̂_ij
        end if
    end for
end for
```

to the beacon. The estimation (for the two-dimensional case) is based on equation 3. Where x_i and y_i stand for the x and y coordinates of a microphone i in relation to the robot, u and v are the x and y components of the vector that points toward the beacon. Lastly, c being the speed of sound and ΔT_{ij} is the TDOA between microphones i and j which can be found as,

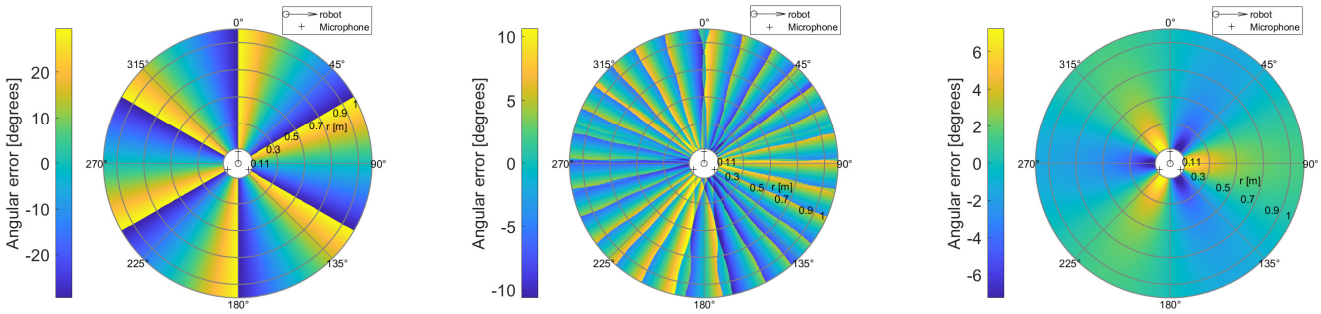
$$\Delta T_{ij} = TOA_i - TOA_j$$

$$\begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) \\ (x_3 - x_1) & (y_3 - y_1) \\ \vdots & \vdots \\ (x_N - x_1) & (y_N - y_1) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} c\Delta T_{12} \\ c\Delta T_{13} \\ \vdots \\ c\Delta T_{1N} \end{bmatrix} \tag{3}$$

or in vector notation:

$$\mathbf{X} \cdot \bar{u} = c\ \overline{\Delta T}. \tag{4}$$

To determine the direction vector $\bar{u}$, the pseudo inverse of $\mathbf{X}$ is required. Fortunately, since the locations of the microphones are known beforehand this inverse can be computed offline. This operator is used to calculate the output unit vector $\bar{u}$, the input data can be scaled – based on this unit normalization. Therefore, inputs can be in the form of clock ticks or μs without change. Furthermore, Valin's algorithm states that the estimation is based on the far-field assumption [13]. This assumption states that the distance to the beacon is much greater than the distance between the microphones. This assumption should be valid for swarm robots since the



(a) (b) (c)

Fig. 4: (b)Valin's with cross-correlation TDOA for direction determination; and (c) Valin's algorithm with Wake-on-Sound TDOA. For all setups three microphones were uniformly distributed around a circle with a radius of 10cm.

microphones are placed on robots that try to locate beacons outside of themselves.

5) Result: Different simulations were run to determine the characteristics of the two algorithms. Figure 4 shows the results from three experiments. These experiments show how the errors of the estimations are distributed for beacons on different locations around the microphones. For each simulation, the three microphones are uniformly placed around a circle with a radius of 10 cm. Primarily, Figure 4a shows the behavior of the CTP algorithm with three microphones. The borders between the different sections are very clear, as the error jumps from one extreme to another. As predicted the error is in the $\pm 30°$ range. Obviously, the estimation accuracy of this algorithm depends on the accuracy of the Wake-on-Sound feature of the microphone. However, for many swarm applications even coarse localization can be quite beneficial.

Figure 4b shows the performance of Valin's algorithm when the sound of the microphones is recorded at 10kHz, where the TDOAs are determined by cross-correlating the signals. It also shows that certain rays emerge on which the estimations are placed. These rays are not uniformly divided, and their number depends on the number of microphones, the sampling frequency of the recording, and the distances between the microphones. Therefore, more microphones, higher sampling frequency, and larger distances between microphones all increase the number of rays and thereby increase the accuracy. In this setup Valin's algorithm reaches an accuracy of $\pm 10°$ Lastly, the Figure 4c shows the results of how well Valin's algorithm would fare if the TDOA determination was handled via the Wake-on-Sound triggers. This would allow for a much higher resolution in the TDOA than would be possible with the cross-correlation method. Whilst also having a lot less computational overhead than having to record multiple microphones at once and then having to use cross-correlation over all those samples. This higher resolution of the TDOA also results in the disappearance of the rays which were visible in Figure 4b. Figure 4c also shows six spots near the center which have notably worse accuracies when compared to the rest of the figure. This is due to the far-field assumption not being valid this close to the microphones.

Valin's algorithm in conjunction with TDOAs determined by the Wake-on-Sound triggers looks very promising as a sound direction estimator for swarm robots since it results in good accuracy whilst not being computationally demanding. However, the usability of this algorithm depends on the characteristics of the hardware-generated TDOAs which will be tested next.

IV. Evaluation

In this section we present thorough experiments done on the real hardware.

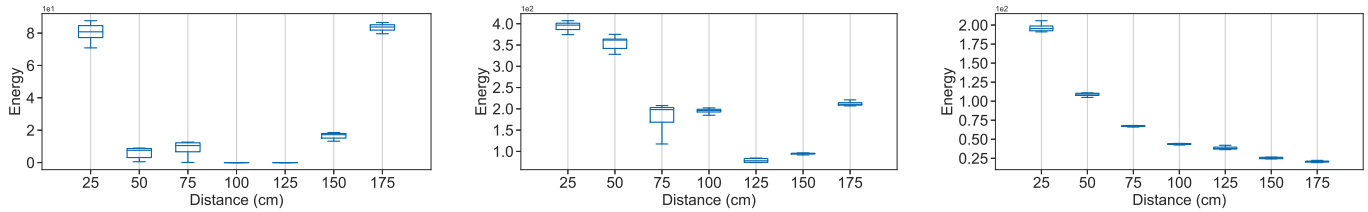
A. Direction of Arrival Detection Accuracy

To test the feasibility of using the Wake-on-Sound feature of the PMM-3738-VM1010-R microphone [24] for the TDOA and DOA estimation we set two microphones 30 cm apart. Then, we placed a sound source at a fixed distance from the centroid of the microphones. The angle of the line through the microphones relative to the line through the centroid and the sound source was varied from 0 to 180°in steps of 45°. At each angle, a sound signal was played 20 times. The time difference of arrival of the signals was captured using a MSP432 MCU [26] connected to the microphones triggering pins. Also, to control the mode of the microphones (i.e., zero-energy-listening or record) the MSP432 was used. When a sound is detected the microphones interrupt the MCU and enter the record mode. The MCU then resets the microphones to zero-energy-listening mode after a predefined amount of time, allowing them to trigger on the next sound signal.

The results show that the measured TDOAs can vary significantly which makes estimating accurate DOA challenging. However, the absence of negative values show that the microphone closer to the sound source constantly triggers first. Therefore, we can use CTP algorithm for reliable (but coarse) DOA estimation.

B. Distance Estimation

To study the effectiveness of using an energy-based acoustic distance estimator we placed a speaker and a microphone in room settings at distances ranging from 25-175 cm with 25 cm



(a) Pure acoustic signal of 440 Hz. (b) Pure acoustic signal of 650 Hz (c) White noise acoustic signal

Fig. 5: Acoustic energy received by a microphone placed at different distances from the source. Figures 5a and 5b snapshot the system behavior when narrowband signals are emitted. Figure 5c shows this relationship when wideband signal is used.

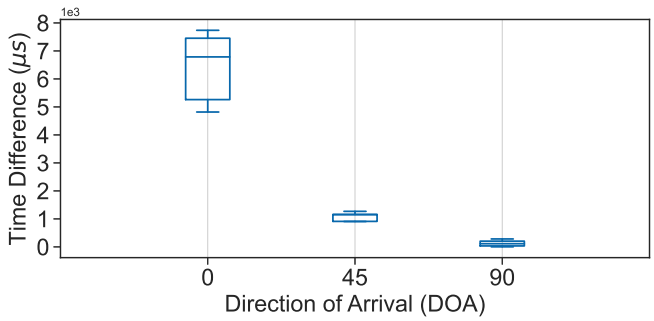


Fig. 6: The time difference of arrival of audio signals at two microphones that are 10 cm apart.

increments. For each trial, the source emitted an acoustic signal with particular characteristics (Figure 5). The receiver recorded the detected signal for three seconds. The energy level of this recording was then determined. This process was repeated 20 times. Inspecting Figure 5 shows that narrowband signals do not provide reliable or consistent results, for example, the received energies at 25 cm and 175 cm for 440 Hz tone are indistinguishable (Figure 5a). However, wideband signals provide robust energy-based distance estimation (Figure 5c).

V. Discussion

a) Triggering threshold: One limitation of the current hardware setup is the fixed threshold of the wake-on-sound feature. However, This limitation can be mitigated by adding a simple circuit to enable digital control of the wake-on-sound threshold.

b) Localization signals or noise?: In order to differentiate between the localization signals and irrelevant ambient noise, a form of signal identification is necessary. One way of achieving this is to transmit a train of pulses with known time interval between the pulses. An obvious downside of this approach is the additional time needed for signal identification. This delay, however, can be tolerable in swarm robotic applications as the speed of the robot is limited.

c) Simultaneous transmissions: The distributed fashion in which swarm robots operate implies that multiple robots might send out sound pulses simultaneously causing a collision. As the current localization method is unable to distinguish multiple signal sources, transmission overlaps pose a problem analogous to the medium access control (MAC) problem seen in wireless networks. Hence, MAC protocols (e.g., CSMA) can offer a solution (e.g., robots must listen before talk).

VI. Conclusion

In this paper we investigated the potential use of small microphones to equip simple swarm robots with audio-based localization capabilities. The wake-on-sound feature makes it possible to determine the angle of arrival of a sound signal without the need for signal processing. The plane cutting algorithm was introduced for audio localization and Valin's algorithm was adapted to work with the new hardware capabilities for comparison. Furthermore, experiments on the triggering accuracy of the microphones showed that they provide acceptable direction estimation. Lastly, it was shown that for energy-based distance estimation wideband signals surpass narrowband signals and provides reliable estimates.

Acknowledgment

This work has been undertaken in the Internet of Swarms project sponsored by Cognizant Technology Solutions and Rijksdienst voor Ondernemend Nederland under PPS O&I.

References

[1] G. De Croon, K. De Clercq, R. Ruijsink, B. Remes, and C. De Wagter, "Design, aerodynamics, and vision-based control of the delfly," *International Journal of Micro Air Vehicles*, vol. 1, no. 2, pp. 71–97, 2009.

[2] J. Y. Kim, T. Colaco, Z. Kashino, G. Nejat, and B. Benhabib, "mroberto: A modular millirobot for swarm-behavior studies," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 2109–2114.

[3] M. Rubenstein, A. Cornejo, and R. Nagpal, "Programmable self-assembly in a thousand-robot swarm," *Science*, vol. 345, no. 6198, pp. 795–799, 2014.

[4] D. Pickem, M. Lee, and M. Egerstedt, "The gritsbot in its natural habitat-a multi-robot testbed," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4062–4067.

[5] H. Hamann, *Swarm robotics: A formal approach*. Springer, 2018.

[6] K. H. Petersen, R. Nagpal, and J. K. Werfel, "Termes: An autonomous robotic system for three-dimensional collective construction," *Robotics: science and systems VII*, 2011.

[7] A. Martinoli and F. Mondada, "Collective and cooperative group behaviours: Biologically inspired experiments in robotics," in *Experimental Robotics IV*. Springer, 1997, pp. 1–10.

[8] J. Faigl, T. Krajník, J. Chudoba, L. Přeučil, and M. Saska, "Low-cost embedded system for relative localization in robotic swarms," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 993–998.

[9] M. Basiri, F. Schill, D. Floreano, and P. U. Lima, "Audio-based localization for swarms of micro air vehicles," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 4729–4734.

[10] M. Saska, T. Baca, J. Thomas, J. Chudoba, L. Preucil, T. Krajnik, J. Faigl, G. Loianno, and V. Kumar, "System for deployment of groups of unmanned micro aerial vehicles in gps-denied environments using onboard visual relative localization," *Autonomous Robots*, vol. 41, no. 4, pp. 919–944, 2017.

[11] A. Kohlbacher, J. Eliasson, K. Acres, H. Chung, and J. C. Barca, "A low cost omnidirectional relative localization sensor for swarm applications," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE, 2018, pp. 694–699.

[12] J. M. Vera-Diaz, D. Pizarro, and J. Macias-Guarasa, "Towards end-to-end acoustic localization using deep learning: From audio signals to source position coordinates," *Sensors*, vol. 18, no. 10, p. 3418, 2018.

[13] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, vol. 2. IEEE, 2003, pp. 1228–1233.

[14] F. Grondin and F. Michaud, "Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations," *Robotics and Autonomous Systems*, vol. 113, pp. 63–80, 2019.

[15] M. Cobos, M. García-Pineda, and M. Arevalillo-Herráez, "Steered response power localization of acoustic passband signals," *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 717–721, 2017.

[16] W. Ma, H. Bao, C. Zhang, and X. Liu, "Beamforming of phased microphone array for rotating sound source localization," *Journal of Sound and Vibration*, vol. 467, p. 115064, 2020.

[17] J. H. DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*. Brown University Providence, RI, 2000.

[18] J.-M. Valin, F. Michaud, and J. Rouat, "Robust 3d localization and tracking of sound sources using beamforming and particle filtering," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 4. IEEE, 2006, pp. IV–IV.

[19] S. Chakrabarty and E. A. Habets, "Multi-speaker localization using convolutional neural network trained with noise," *arXiv preprint arXiv:1712.04276*, 2017.

[20] R. Takeda and K. Komatani, "Unsupervised adaptation of deep neural networks for sound source localization using entropy minimization," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2217–2221.

[21] E. L. Ferguson, S. B. Williams, and C. T. Jin, "Sound source localization in a multipath environment using convolutional neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2386–2390.

[22] L. Comanducci, F. Borra, P. Bestagini, F. Antonacci, S. Tubaro, and A. Sarti, "Source localization using distributed microphones in reverberant environments based on deep learning and ray space transform," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2238–2251, 2020.

[23] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 34–48, 2019.

[24] pui audio, vesper. (2020) PMM-3738-VM1010-R: A zeropower listening™ piezoelectric MEMS microphone. [Online]. Available: shorturl.at/pACDO

[25] J. O. Smith, *Physical audio signal processing: For virtual musical instruments and audio effects*. W3K publishing, 2010.

[26] Texas Instruments. (2020) MSP-EXP432P401R: Microcontroller launchpad. [Online]. Available: shorturl.at/dvOSY