

# Dronar: Obstacle Echolocation Using Drone Ego-Noise

Henrik Nilsson, Joakim Rydell, Anton Kullberg and Gustaf Hendeby

The self-archived postprint version of this conference paper is available at Linköping University Institutional Repository (DiVA):

<https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-206892>

N.B.: When citing this work, cite the original publication.

Nilsson, H., Rydell, J., Kullberg, A., Hendeby, G., (2024), Dronar: Obstacle Echolocation Using Drone Ego-Noise, *Proceedings of 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, 184-188.

<https://doi.org/10.1109/ICASSPW62465.2024.10627342>

Original publication available at:

<https://doi.org/10.1109/ICASSPW62465.2024.10627342>

Copyright:

<http://www.ieee.org/>

©2024 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

# DRONAR: OBSTACLE ECHOLOCATION USING DRONE EGO-NOISE

Henrik Nilsson<sup>1†</sup>, Joakim Rydell<sup>1</sup>, Anton Kullberg<sup>2</sup>, Gustaf Hendeby<sup>2</sup>

<sup>1</sup> Swedish Defence Research Agency (FOI), Linköping, Sweden

Email: {firstname.lastname}@foi.se

<sup>2</sup> Dept. Electrical Engineering, Linköping University, Sweden

Email: {firstname.lastname}@liu.se

## ABSTRACT

A method for obstacle detection using the sound that a drone naturally emits is proposed. The sound emitted from a vehicle, *ego-noise*, is often considered a complicating factor for mission fulfilment, without purpose. The idea in this paper is to utilise this ego-noise for obstacle detection, being the first to perform practical experiments of this. Adding a few microphones to the vehicle, the ego-noise is utilised as the sound source for echolocation. The method consists of auto-correlating the received signals to estimate echo delays, using the known array geometry and signal propagation speed to relate delays to distances, and then beamforming to position targets. A proof-of-concept has been constructed, and promising results are presented for experiments in a controlled environment.

**Index Terms**— Multi-rotor, Drone, Echolocation, Ego-noise, Obstacle

## 1. INTRODUCTION

Drones, and specifically multi-rotor aerial vehicles (MRVs), are ever more present in society. From autonomous delivery to inspection and more, MRVs see further use for many industrial and commercial applications. As use increases, further requirements are put on safety. A crucial part of MRV safety is obstacle avoidance, and many such methods exist. Obstacles range from small debris to brick walls, and to be able to avoid an obstacle, one must be able to locate it.

Most commonly, light detection and ranging (LIDAR) or traditional cameras are used for the task of obstacle localisation, but using multiple approaches is of interest when there are requirements for resiliency and redundancy. Audio-based solutions are of interest as they mitigate some of the shortcomings of light based solutions, such as the problem of occlusion and detecting transparent objects. However, the strong influence of ego-noise on measurements make MRVs a difficult platform for auditory methods and many existing methods have focused on characterising and limiting the effect of ego-noise [2–4].

Alternatively, the ego-noise can be used as a sound source for echolocation, something that has yet to be explored in practice. Beyond the merit of already being an intrinsic part of a MRV, ego-noise also circumvents the problem of strong ego-noise influence as this becomes an asset. An ego-noise-based approach can be advantageous when compared to other active approaches as the MRV does not emit any additional signature when compared to a MRVs normal operation.

The problem of locating an obstacle can, e.g., be solved by finding the position that maximises the likelihood of the obstacle being at that location. In this formulation, the localisation task boils down to an optimisation problem. In the context of beamforming, this is known as steered response-power (SRP). In the SRP beamforming framework, the localisation problem is usually solved by doing grid search over a set of positions.

We show that locating obstacles using ego-noise from a MRV is possible using an auto-correlation SRP beamforming approach. The main contribution of this paper is the proposed solution to the localisation problem, circumventing the problem of low signal-to-noise ratio (SNR) due to ego-noise, along with the first practical validation of ego-noise echolocation.

### 1.1. Related work

The ideas used in this work are at this point well established in the literature. In the field of radar, the presented setup mirrors that of multi-static radar [5]. Beamforming has been used as a term since the 1960s and many techniques are summarised in [6]. Relying on the auto-correlation of a signal to determine the echo distance is not a new idea, being part of patent [7] and used in many works. Generalised cross-correlation (GCC) was formalised in [8], describing previously used techniques in terms of a weighting function, such as the smoothed coherence transform (SCOT) and the phase transform (PHAT) introduced in [9]. DiBiase built upon these techniques in [10], combining SRP with PHAT (SRP-PHAT) and showing that it is a natural generalisation of GCC-PHAT.

Saqib and Jensen (along with others) have explored auditory range sound for mapping and localisation in a series of works [11–15]. [11] proposes an expectation maximisation (EM) method for time of arrival (TOA) and direction of arrival (DOA) estimation of acoustic reflectors using an active approach. [12] gives another active TOA estimation approach using a model of early reflections resulting in a statistically optimal nonlinear least squares (NLS) estimator. Further building upon this, they perform real-world tests of their solution on a robot platform (as opposed to their previously simulated result) [13]. With their latest work [15], they implement a complete simultaneous localisation and mapping (SLAM) solution for acoustic reflector localisation.

Inspiring this work, in [14], they present a method for using ego-noise echo to sense an acoustic reflector, which is the only previous work to propose using ego-noise for localisation known to Saqib et al. and the authors of this work. The proposed time difference of echo (TDOE) estimator achieves (in simulation) accurate distance estimates for distances up to around two meters. However, the authors assume the direct path component of the ego-noise to be known, an assumption that is non-trivial to realise in a real system.

\*This work has received funding from the European Union Horizon 2020 program (grant agreement no. 101021957 – NIGHTINGALE).

<sup>†</sup>This work is based on the author Nilsson’s master’s thesis [1].

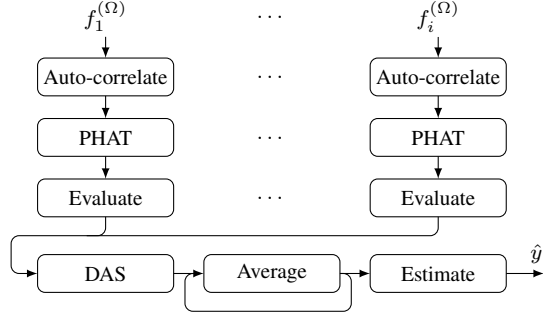


Fig. 1. Signal processing chain.

## 2. METHOD DESCRIPTION

Sound is recorded using an array of microphones and auto-correlation is applied. By evaluating the auto-correlations over a set of positions, i.e. using SRP beamforming, the location of obstacles can be extracted. Here, the assumptions are presented, and then the localisation solution in detail.

The signal processing chain is illustrated in Fig. 1, the nodes of which correspond to the *operations* described below. This can be viewed as an SRP beamformer with PHAT, which resembles SRP-PHAT proposed by DiBiase in [10], though the computational approach is different. Using the auto-correlation makes this technique a TOA instead of a time difference of arrival (TDOA) solution.

### 2.1. Signal Model and Assumptions

Throughout this paper, we assume a single echo-inducing obstacle. The scenario is depicted in Figure 2, where there may be multiple source signals  $s_j$ ,  $j = 1, \dots, N_s$ , and multiple microphones  $m_i$ ,  $i = 1, \dots, N_m$ . The signal for each microphone  $m_i(t)$  is given by

$$m_i(t) = \sum_j \frac{1}{d_{m_j,i}} s_j \left( t - \frac{d_{m_j,i}}{c} \right) + \frac{1}{d_{r_j} + d_{e_i}} s_j \left( t - \frac{d_{r_j} + d_{e_i}}{c} \right) + v. \quad (1)$$

Here,  $d_{m_j,i}$  is the distance between microphone  $i$  and source  $j$ ,  $d_{r_j}$  is the distance between the obstacle and source  $j$ ,  $d_{e_i}$  is the distance between the obstacle and microphone  $i$ . Further,  $c$  is the speed of sound in air ( $c \approx 343$  m/s) and the measurement noise  $v$  is distributed according to some prior  $v \sim p_v$ . The sound sources are approximated as point sources. The signal is thus modelled as the sum of a direct path and a time-shifted component of the source ego-noise signals  $s_j(t)$ .

### 2.2. Audio Processing

The audio is processed in *frames*, batches of length  $T$ . Each frame is a rectangular window of the continuous audio signal, processed individually and later combined. The frames are overlapping, denoted as a percentage of the overlap to total size ratio, e.g., 60% overlap. Frame  $\Omega$  is denoted as

$$f_i^{(\Omega)}(t) = m_i(o^{(\Omega)} + t), \quad t \in [0, T] \quad (2)$$

where  $o^{(\Omega)}$  is the start offset of frame  $\Omega$ .

The *auto-correlation* of each frame is given by

$$c_i^{(\Omega)}(t) = \int_{-\infty}^{\infty} f_i^{(\Omega)}(\tau) f_i^{(\Omega)}(\tau - t) d\tau. \quad (3)$$

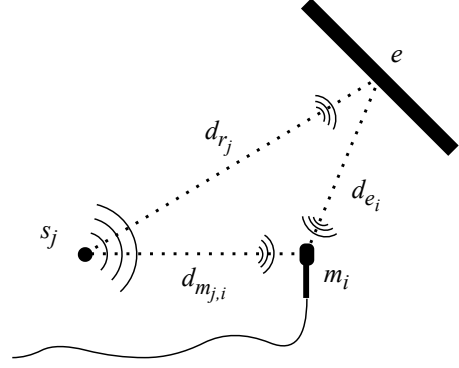


Fig. 2. Conceptual image of distances used in signal model.

The correlation is normalised such that the maximum is equal to 1, signifying perfect auto-correlation. The signal is also halved, we only look at  $c_i^{(\Omega)}(t)$ ,  $t \geq 0$ . Assuming no significant difference between the echo and direct path components (except the shift in time and signal attenuation), the auto-correlation of the signal according to (1) should show a spike at a time shift that corresponds to the travel time of the echo, along with the implicit spike at a time shift of zero. Any spikes that seem to originate from within the bounds of the array are of no interest, as with the implicit spike at time zero. For these reasons, this region close to the array is suppressed by zeroing the auto-correlation signals according to some threshold  $t_{th}$ .

$$c_i^{(\Omega)}(t) = 0, \quad \forall t < t_{th} \quad (4)$$

The auto-correlation  $c_i^{(\Omega)}(t)$  is whitened through the use of PHAT [9], i.e.,

$$\begin{aligned} C_i^{(\Omega)}(f) &= \mathcal{F}\{c_i^{(\Omega)}(t)\} \\ \psi_p(f) &= |C_i^{(\Omega)}(f)|^{-1} \\ w_i^{(\Omega)}(t) &= \mathcal{F}^{-1}\{\psi_p(f)C_i^{(\Omega)}(f)\}. \end{aligned} \quad (5)$$

### 2.3. Position Estimation

The whitened auto-correlation signal  $w_i^{(\Omega)}(t)$  is an indicator of how self-similar the audio signal is at time delay  $t$ . Assuming the speed of sound  $c$  to be known, the distance from the microphone to the cause of the echo can be estimated. We *evaluate* the auto-correlation signal for a set of time delays that correspond to a set of positions  $P = \{p = (x, y) : (x, y) \in \mathbb{R}^2\}$ . For this, we must calculate the distances  $d_{r_j}$ ,  $d_{e_i}$ ,  $d_{m_j,i}$  for the position. Each position, or time delay, can then be mapped to an echo intensity by

$$t_r = (d_{r_j} + d_{e_i} - d_{m_j,i})/c \quad (6)$$

$$O_i^{(\Omega)}(p) = w_i^{(\Omega)}(t_r). \quad (7)$$

Here,  $O_i^{(\Omega)} : \mathbb{R}^2 \mapsto \mathbb{R}$ , maps each position  $p$  to an echo intensity. Using the same coordinate system for all  $O_i^{(\Omega)}$ , the delay calculation of the canonical delay-and-sum (DAS) beamformer is incorporated in this step. Note that this technique generalises to three dimensions, but since the proof-of-concept microphones all lay in the same plane only two dimensions are considered here.

To produce the final position estimate, we first fuse the echo intensities from multiple microphones. DAS gives,

$$O_{DAS}^{(\Omega)}(p) = \sum_i O_i^{(\Omega)}(p). \quad (8)$$

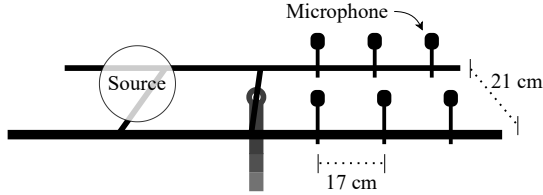


Fig. 3. Proof-of-concept array setup.

Note that other examples such as delay-multiply-and-sum (DMAS) exist, see e.g. [16]. For the sake of brevity, only DAS is presented here, but more alternatives are explored in [1]. To improve the SNR, multiple frames are *averaged*, i.e.,

$$O_{DAS}(p) = \frac{1}{N_{\Omega}} \sum_{\Omega} O_{DAS}^{(\Omega)}(p), \quad (9)$$

where  $N_{\Omega}$  are the number of frames. Finally, the maximum correlation is used as the position *estimate*, i.e.,

$$\hat{y} = \arg \max_{p \in P} O_{DAS}(p). \quad (10)$$

### 3. EXPERIMENTAL EVALUATION

The presented method is evaluated in an experiment campaign with scenarios of increasing realism. Firstly, a stand-alone microphone rig is used as a proof-of-concept, followed by experiments with a prototype system mounted on an MRV. Here, the most relevant results are provided along with a summarisation in Table 1. See [1] for more results.

#### 3.1. Experiment Setups

The proof-of-concept experiments were conducted using pro-grade recording equipment with microphones placed in a grid as shown in Fig. 3, where the placement of the sound source is also displayed. The microphone placement is assumed to be advantageous. Four sets of recordings were collected for offline processing, each consisting of six recordings of 60 s, each with a different obstacle location. For the proof-of-concept, the obstacle was a  $60 \times 60$  cm wooden board. The two first datasets were collected in an anechoic chamber with a loudspeaker playing a clean recording of a hovering MRV (Set 1) and an actual quadcopter with 6" rotors performing throttle modulations emulating flight as sound sources (Set 2). The procedure was then repeated in a conference room (Set 3 and 4, respectively), presumed to be a more challenging environment. Ground truth was provided by a Marvelmind Precise ( $\pm 2$  cm) indoor positioning system. The frame size was set to  $2^{13} = 8192$  samples (0.085 s) with 90% overlap between frames. The minimal distance was set to 1.2 m ( $t_{th} = \frac{1.2}{c}$  s).

For the prototype, a 16SoundsUSB kit [17] was used to record sound. The 16 microphones were mounted on two concentric circular carbon fibre frames (radius 0.7 m), offset in height (0.2 m) and in angle. The array was mounted to a custom quadcopter with 14" rotors, see Fig. 4. One dataset consisting of three recordings around 30 s long were collected for offline processing, during which the quadcopter hovered in front of a wall of corrugated metal. As in the proof-of-concept experiments, the frame size was set to  $2^{13} = 8192$  samples (0.085 s) with 90% overlap. However, the minimal distance was set to 2.4 m ( $t_{th} = \frac{2.4}{c}$  s).



Fig. 4. The prototype drone.

Table 1. Number of localisations with error  $< 0.6$  m and the respective average error.

Set	Succ. loc.	Avg. error (succ.)
Set 1	6/6	$\sim 0.087$ m
Set 2	4/6	$\sim 0.398$ m
Set 3	6/6	$\sim 0.129$ m
Set 4	3/6	$\sim 0.133$ m
Prototype	0/3	—

#### 3.2. Proof-of-Concept Results

The results from the proof-of-concept experiments are visualised in Figs. 5–7, where  $O_{DAS}(p)$  for a uniformly spaced grid is visualised. Note that the peaks in the auto-correlation of each respective microphone gives rise to the bright ellipses in the figures. For true echo peaks, the most likely obstacle position is the intersection of several of these ovals. In each figure, green pluses (+) denote microphones, red crosses (×) the sound sources, blue stars (★) the ground truth positions and purple circles (○) the estimates  $\hat{y}$ .

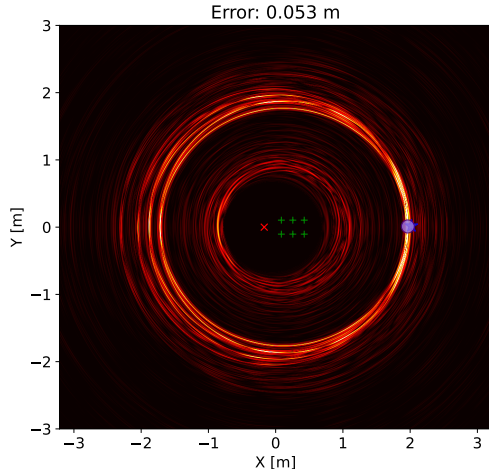
The anechoic chamber provided an ideal environment for this application, with minimal reverberations and ambient sound. Under these circumstances, and using the prerecorded hovering sound (Set 1), the introduced obstacle was easily identified in the resulting auto-correlation as illustrated in Fig. 5. Hence, under optimal circumstances, the results are very promising. Substituting the loudspeaker for a MRV (Set 2) yielded the results in Fig. 6. Even though a slightly worse SNR was observed, the localisation results are good, though slightly biased.

The environment in the conference room was more challenging, with plenty of surfaces that reflected the sound of the MRV. These less ideal circumstances had little impact on the ability to locate the obstacle using the prerecorded sound (Set 3) even though the SNR was worse. However, the results using the MRV as sound source (Set 4) was more impacted. See Fig. 7, where a spurious incorrect peak can be found in front of the array even though the true obstacle is located to the side. Nevertheless, these results were promising enough to motivate pursuing the prototype.

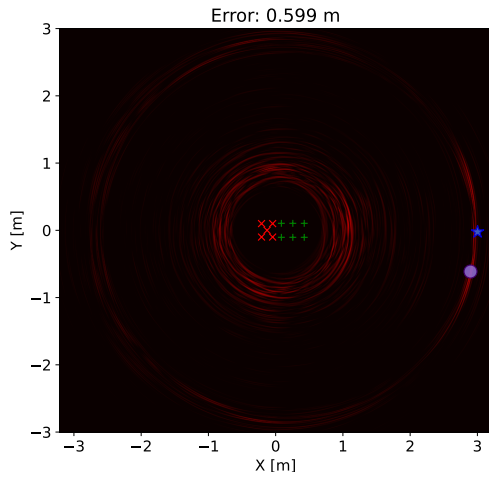
#### 3.3. Prototype Results

The results from the experiments with the prototype are exemplified in Fig. 8, and are annotated the same way as the proof-of-concept illustrations. The results this time are worse, and it is difficult to locate or even detect the obstacle.

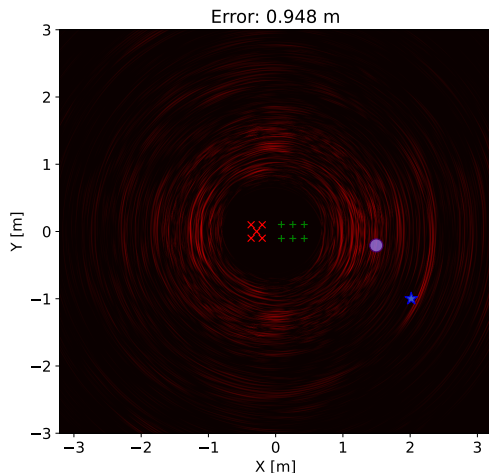
The results of the prototype experiment were less encouraging than the proof-of-concept results. There could be many reasons for



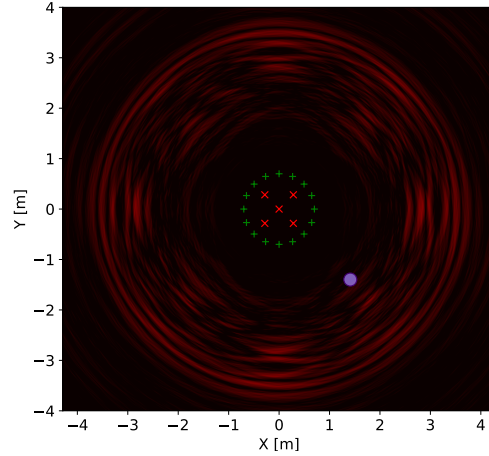
**Fig. 5.** Echo intensity plot from the proof-of-concept anechoic chamber, speaker source (Set 1) experiment with the obstacle at position (2,0).



**Fig. 6.** Echo intensity plot from the proof-of-concept anechoic chamber, drone source (Set 2) experiment with the obstacle at position (3,0).



**Fig. 7.** Echo intensity plot from the proof-of-concept conference room, drone source (Set 4) experiment with the obstacle at position (2,-1).



**Fig. 8.** Echo intensity plot from the prototype experiment with the obstacle approximately at position (-1,2.5).

this: e.g., the SNR is lower than in the other datasets, the quality of sound is slightly worse, corrugated metal wall provided a less optimal reflecting surface and the environment was less ideal. Still, some correlations can be seen, and an improved physical design in combination with more advanced signal processing could potentially improve the results substantially.

#### 4. CONCLUSION AND DISCUSSION

We show that localisation of obstacles using ego-noise is possible, for the first time in practice. The proposed method for obstacle echolocation using drone ego-noise was shown in proof-of-concept experiments to locate obstacles accurately enough to be useful for a MRV, especially in the more idealised cases (Sets 1–3). The more challenging Set 2 and 4 sometimes failed by indicating the obstacle on the opposite side of the array, at the boundary of the region close to the MRV in which detections were suppressed. This could be a bias resulting from the design of the proof-of-concept array. Increasing the dead zone might alleviate this issue, but at the price of not being able to locate obstacles close to the MRV.

In general, the sets that used a real MRV as a sound source (Set 2 and 4) were more challenging. Reasons could range from prop-wash to the MRV is not well approximated as a set of point sources, smudging the results, as well as poor measurement of microphone/rotor location. With the presented setup, extrapolating from the achieved results, the limit of useful range seems to be close to the maximum tested, say 3–5 meters.

There are multiple areas of future studies that were left as out of scope for this work. For example, firstly, improvements are to be expected with optimised array design. The approach could also be made much more computationally efficient by making a more intelligent choice of positions to evaluate, for instance combining the presented approach with a particle filter as with steered beamforming (SBF) [18]. Thirdly, a characterisation of what sound qualities constitutes good noise for these purposes would help with rotor selection and more.

Further work is needed to have a fully functional system as the prototype failed to locate the obstacle. Further analysis of the reasons for this is required, but it is probably a combination of the hardware used and a need for more advanced signal processing to suppress real-world phenomena such as background noise and prop-wash.

## References

- [1] Henrik Nilsson, “DRONAR: Obstacle echolocation using ego-noise,” M.S. thesis, Linköping University, 2023.
- [2] Lin Wang and Andrea Cavallaro, “Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles,” in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016, pp. 152–158.
- [3] Lin Wang and Andrea Cavallaro, “Deep-learning-assisted sound source localization from a flying drone,” *IEEE Sensors Journal*, vol. 22, no. 21, pp. 20828–20838, 2022.
- [4] Dmitrii Mukhutdinov, Ashish Alex, Andrea Cavallaro, and Lin Wang, “Deep learning models for single-channel speech enhancement on drones,” *IEEE Access*, vol. 11, pp. 22993–23007, 2023.
- [5] V. S. Chernyak, *Fundamentals of Multisite Radar Systems*, Routledge, London, 1st edition edition, 1998, Ebook published on 25 October 2017.
- [6] B.D. Van Veen and K.M. Buckley, “Beamforming: a versatile approach to spatial filtering,” *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [7] Nigel Preston, “Method for determining echo distance using autocorrelation in time of flight ranging systems,” US20040179428A1, 2003, Siemens AG.
- [8] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.
- [9] G. C. Carter, A. A. Nuttall, and P. G. Cable, “The Smoothed Coherence Transform (SCOT),” Tech. Rep. TC-159-72, Naval Underwater Systems, August 8 1972.
- [10] Joseph Hector DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*, Ph.D. thesis, Brown University, 2000.
- [11] Jesper Rindom Jensen, Usama Saqib, and Sharon Gannot, “An EM method for multichannel toa and doa estimation of acoustic echoes,” in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, pp. 120–124.
- [12] Usama Saqib and Jesper Rindom Jensen, “Sound-based distance estimation for indoor navigation in the presence of ego noise,” in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.
- [13] Usama Saqib and Jesper Rindom Jensen, “A model-based approach to acoustic reflector localization with a robotic platform,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 4499–4504.
- [14] Usama Saqib, Antoine Deleforge, and Jesper Rindom Jensen, “Detecting acoustic reflectors using a robot’s ego-noise,” in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 466–470.
- [15] Usama Saqib and Jesper Rindom Jensen, “A framework for spatial map generation using acoustic echoes for robotic platforms,” *Robotics and Autonomous Systems*, vol. 150, pp. 104009, 2022.
- [16] Giulia Matrone, Alessandro Stuart Savoia, Giosuè Caliano, and Giovanni Magenes, “The delay multiply and sum beamforming algorithm in ultrasound b-mode medical imaging,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 940–949, 2015.
- [17] Cédric Godin, Dominic Létourneau, Vincent-Philippe Rhéaume, and François Michaud, “16SoundsUSB,” <https://github.com/introlab/16SoundsUSB>, Visited: 2023-11-29.
- [18] Darren B. Ward and Robert C. Williamson, “Particle filter beamforming for acoustic source localization in a reverberant environment,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, vol. 2, pp. II–1777–II–1780.