POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

RO-BAT: A bat-inspired approach on mobile robot navigation using Direction of Arrival estimation

Laurea Magistrale in Music and Acoustic Engineering - Ingegneria Acustica e Musicale

Author: Alberto Doimo

Advisor: Prof. Fabio Antonacci

Co-advisor: Dr. Thejasvi Beleyur, Prof. Dr.-Ing. Heiko Hamann, Dr. Andreagiovanni Reina

Academic year: 2023-2024

1. Introduction

This thesis represents the initial stage of the RO-BAT research project, hosted by the Centre for the Advanced Study of Collective Behaviour (CASCB) at the University of Konstanz. This project aims at building fully autonomous echolocating robots capable of sound-based navigation, inspired by the "Cocktail Party Problem" observed in biology. This refers to one of the fundamental challenges in sensory processing, both in nature and artificial systems. It describes a scenario in which individuals—whether animals or robots—must identify and interpret multiple acoustic signals in a noisy environment [1]. Bats, for instance, are able to process scenarios involving overlapping inputs and navigate effectively within a swarm using ultrasound pulses. This has inspired researchers to explore similar approaches in robotic navigation.

Swarm robotics, which involves groups of autonomous robots working cooperatively to accomplish shared tasks, could greatly benefit from echolocation-based sensing. In traditional robotic systems, visual or infrared (IR) sensors are commonly used; however, they often fail in challenging environments such as bright light, smoke, darkness, or fog and scale poorly in densely packed conditions. The use of echolocation presents a low-cost, low-energy solution suitable for lightweight, mobile robots, offering an effective way of enabling real-time sound source localisation on resource-constrained platforms, typically used in swarm robotics.

As presented in [4], many works focus on the sound source localisation problem in robotics, while a few present bio-inspired single echolocation solutions based on ultrasounds. However, none of them, to our knowledge, address the problem of overlapping signals in active locating agents. This thesis develops sound-based localisation in swarms, with a compact robot, which is the foundation for future advancements in active echolocation.

To establish baseline capabilities for active echolocation, three well-known Direction of Arrival (DOA) algorithms have been tested. As suggested in [4], the Generalised Cross Correlation with Phase Transform (GCC PHAT) [5] was chosen due to its simplicity and lightweight implementation, making it suitable for resource-limited robots. The Steered Response Power with Phase Transform (SRP PHAT) [3] was

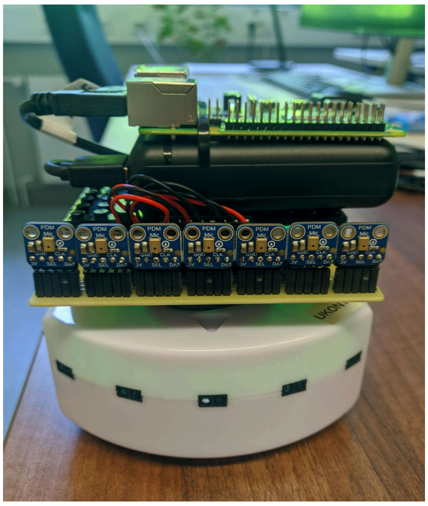

Figure 1: Photo of the ro-bat used in the tests with the array V1 of seven PDM microphones.

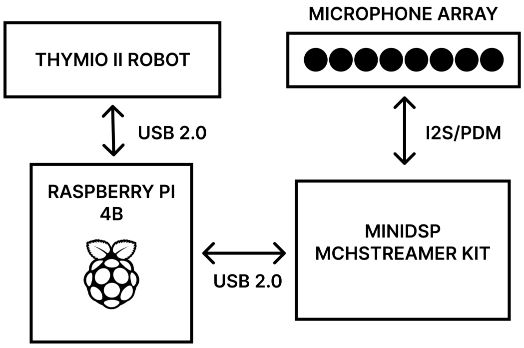

Figure 2: Overview of the main components mounted on the ro-bat.

selected for its capability to detect multiple sources while remaining computationally efficient. Finally, the Multiple Signal Classification (MUSIC) [2] algorithm was included for its well-known capability to perform multi-DOA detection, though it has higher computational requirements.
These algorithms are adapted to perform real-time processing of signals from Micro-Electro-Mechanical Systems (MEMS) digital microphones. They are connected using data formats, such as Pulse Density Modulation (PDM) and Inter-IC Sound (I^2S). MEMS microphones are used because they enable low-cost, low-power and compact designs, while still providing accurate sound field recordings.
In this thesis thesis the algorithms are tested and compared on real robots, on a challenging multiple sound-source environment.

2. Realisation

The robot, called ro-bat, was developed and realised by selection, design, prototyping and software development of all its parts. Hardware components were selected to ensure the system's compatibility with small, mobile robots, enabling real-time data processing with limited computational resources. The software was kept simple and accessible, to enable cross-platform compatibility. Here's an overview on the main hardware components:
The **Thymio II Robot** was chosen as the mobile platform for this project. It was selected because of its small size, cost and availability in the lab, enabling scalability, and suitability for swarm-based studies. This commercial solution allows the focus on the audio components without the need for custom robot development. Key features include infrared sensors for collision avoidance, ground detection, dual motors for ground movement, USB connectivity, capacitive buttons, a speaker and programmable RGB LEDs. Thymio II's advanced sensors and compatibility with the Python language, through the `thymiodirect` library, make it ideal for research applications.
Raspberry Pi single-board computer: The Thymio II platform is used in combination with a Raspberry Pi 4B, which significantly extends the ro-bat's processing power and enables wireless communication via Wi-Fi. The Thymio robot and the Raspberry Pi are connected via USB 2.0, with both devices powered by an external battery. The battery and Raspberry Pi are positioned on top of the robot (as in Figure 1), along with the sound card, which is also connected and powered via USB 2.0. This arrangement ensures the system remains compact and easy to set up.
The **MCHStreamer Sound Card** from MiniDSP serves as the core of audio signal processing. It reduces the load on the Raspberry Pi processor while extending its audio input and output capabilities. It supports various multichannel data formats (e.g., I^2S, PDM, TOSLINK, ADAT, S/PDIF (coax), TDM, DSD) and features an XMOS XCore 200 processor for data management via USB 2.0. The device complies with USB Audio Class 2.0 and works seamlessly across multiple platforms (PC, Mac, Linux, iOS, Android).

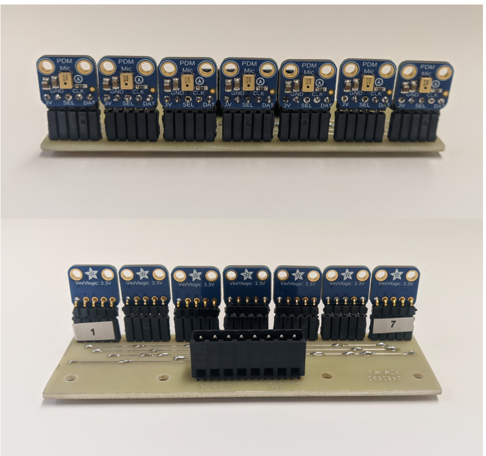

Figure 3: V1 PDM microphone array composed of seven MP34DT01-M MEMS breakout boards from Adafruit and custom connection board.

The sound card powers and synchronises MEMS microphones during data acquisition, with firmware options tailored for different formats: *AllRate* for I^2S and *PDM* for PDM. It supports sampling rates from 8 kHz to 384 kHz, making it suitable for ultrasonic applications. The MCH Streamer Kit, with its compact dimensions ($13 \times 40 \times 62$ mm) and cost-effective design, is ideal for swarm applications, delivering high performance at a competitive price.

MEMS Microphone Arrays: The microphone array geometry was developed and refined through successive prototypes. The linear shape for the array has been selected as the best compromise for this application, as it simplifies geometric calculations and keeps it compact when compared, for example, to circular arrays. Three linear array versions (V0,V1 and V2) have been developed, featuring different MEMS microphones and inter-element distances to achieve the optimal configuration. Since the selected microphones were already digital, no additional analog to digital converter was required between them and the sound card.

The **V0** uses eight Adafruit breakout boards to easily connect SPH0645LM4H-B I^2S bottom-port omnidirectional MEMS microphones to the sound card. The inter-distance is equal to 18 mm and the total length of the array of 126 mm.

The **V1** in Figure 1 and 3 is similar to the previous version, remaining a linear omnidirectional MEMS array but with key differences. The inter-distance is now reduced to 15 mm and the it uses only seven microphones, with a total length of 90 mm. It has been involved in the experimental testing of the ro-bat as described in Section 3 thanks to its form factor, which fits the Thymio II width and the MP34DT01-M PDM top-port MEMS microphones. These resulted to be more sensitive with respect to I^2S SPH0645LM4H-B MEMS and gave better results during testing.

The **V2**, features the same eight I^2S microphones of array V0 and it was created to test the ultrasound capabilities of the I^2S protocol. This small array features a denser positioning of microphones to better sample high frequency signals. The inter-distance is equal to 3 mm and the array measures 21 mm in total. It has only been used in pre-tests because of its noisy output at high frequency, but it represents the future direction towards the ultrasound echolocation.

The implementation of the Direction of Arrival (DOA) estimation algorithms is divided into three main steps:

Data Input Buffer: Incoming audio data is collected and stored in a buffer, that ensures correct real-time processing and accurate DOA estimation. The `sounddevice` library has been used for devices and buffers management.

DOA Computation: Buffered audio data is analysed and compared between the algorithms. In particular, the DOA is calculated in real time for SRP PHAT and MUSIC by adapting `pyroomacoustics` library.

Navigation and Avoidance: Based on the DOA results, navigation algorithms guide the ro-bat's movement, enabling real-time trajectory adjustments to avoid sound-emitting obstacles and perform collision avoidance.

3. Experimental testing

The experiments were conducted on real robots in the labs at the University of Konstanz. The setup included:

- **Arena:** A dark carpet with white tape markers for a 1.4×2.2 m area providing good grip for the robot wheels.
- One **Ro-bat** equipped with V1 microphones array, MCHStreamer sound card, and Raspberry Pi 4B as in Figure 1.
- **Sound Obstacles:** Five static Thymio robots emitting the same 30-minutes white noise audio from their loudspeaker.
- **Overhead Camera** for recording and

Run	Parameters	GCC-PHAT	SRP-PHAT	MUSIC
1	Time [min:sec]	8:27	8:16	8:33
	Sampling frequency [kHz]	16	16	8
2	Total time [min:sec]	8:33	8:19	7:45
	Sampling frequency [kHz]	32	32	8
3	Total time [min:sec]	8:18	8:14	7:12
	Sampling frequency [kHz]	32	32	8
4	Total time [min:sec]	10:26	15:29	4:48
	Sampling frequency [kHz]	32	32	8

Table 1: Duration and sampling frequencies used in the experimental tests.

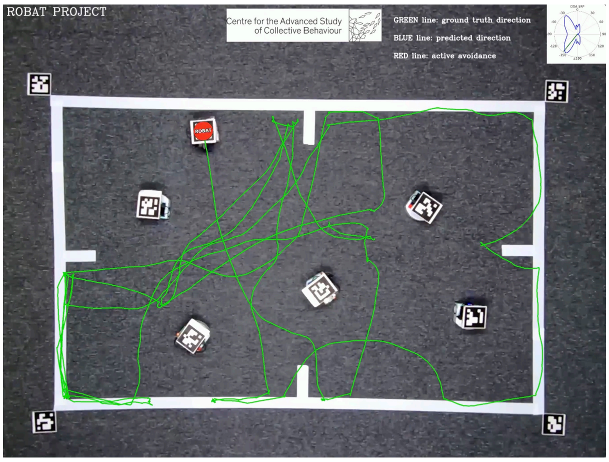

Figure 4: Example of one of the tracked experiments. The polar plot, visualising the DOA estimates versus the ground truth directions, is shown in the top right corner. The video containing all the experiments is available at URL: `https://youtu.be/FPZS9TciNro`.

tracking ro-bat movements with videos, synchronised with audio clips from the ro-bat.

- **ArUco Markers** used for object tracking and orientation in the arena.
- **External PC** used to remotely operate the Raspberry Pi via Wi-Fi.

In order to create an acoustic challenging environment for the ro-bat, sound sources were positioned randomly on the arena with a different configuration for each of the four testing runs. The test was approximately 25 minutes long for each run, conducted with the parameters detailed in Table 1. Video and audio data from the experiments were synchronised manually using claps. The footage was post-processed using OpenCV (Open Source Computer Vision Library) to detect ArUco markers (square pattern with black and white tiles) used to track positions and orientations. Figure 4 shows an example of the ro-bat being tracked while marking its trajectory. The Direction of Arrival (DOA) algorithms were recomputed using the audio recorded by the ro-bat, replicating the experimental conditions when estimating the directions of sound sources.

4. Results

The error between the estimated angle and the ground truth angle, spanning from 0 to 360 degrees, is shown in Figure 5. Data from the four experimental runs are combined into a dataset for each of the three algorithms tested, and the results are compared based on their mean values and standard deviations. The mean error is presented for all possible angles around the robot, starting from 0 (front), progressing counter-clockwise through 90 (left), 180 (back), and 270 degrees (right). The three algorithms perform similarly in terms of mean error, with the minimum values observed at 0 degrees and 180 degrees, corresponding to the front and back of the robot. All methods detect sound sources within the frontal 70-degree sector, achieving an average error below 30 degrees. Among the three algorithms, GCC-PHAT demonstrates the best performance, with errors on average 15 degrees lower than MUSIC and 5 to 10 degrees lower than SRP-PHAT.

Performance diminishes significantly for side an-

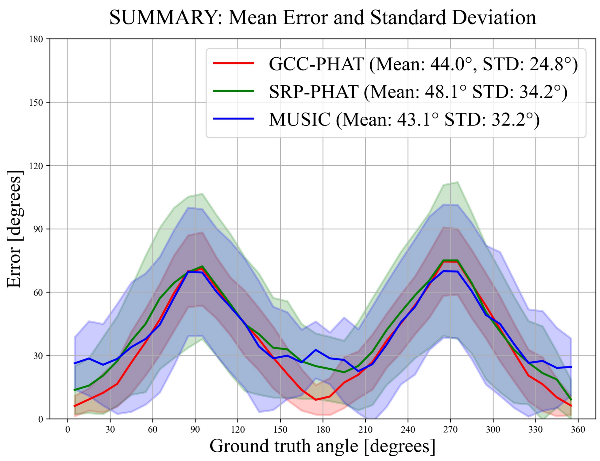

Figure 5: Comparison of the distribution of the mean error and standard deviation along the ground truth axis. All algorithms performed similarly on average, but with different spread of the data. GCC-PHAT performs better at 0 (front) and 180 (back) degrees and has a lower standard deviation with respect to SRP-PHAT and MUSIC.

gles (90 degrees and 270 degrees) across all algorithms. In particular, based on the scatter plots of the results, GCC-PHAT reaches errors of up to 120 degrees, while both SRP-PHAT and MUSIC reach a maximum error of 150 degrees, indicating a large variability in the estimations. This limitation arises from the linear microphone array's geometry and the tracking procedure, which occasionally biases the ground truth calculations. Specifically, the closest obstacle is used as the ground truth reference during video tracking, as single DOA estimation is based on the loudest signal. However, large estimated errors occur in some cases because the robot does not always consider the closest sound source as the loudest. Instead, it prioritizes other directions due to the obstacle configuration within the arena.

Additionally, the scatter plots shown that errors are positioned around 30 degrees apart at each ground truth angle for SRP-PHAT, meaning that the estimation is unable to resolve angles more finely. A similar pattern is shown in the error data for MUSIC, which splits into two main parts. This patterns of data suggest limited angular resolution for both these algorithms.

Figure 6 instead looks only at the error distribution across algorithms, without referring to the ground truth angles.

While the mean is similar, GCC-PHAT (left) has a more uniform error distribution from 0 to 90 degrees, with only few extreme values. In contrast, SRP-PHAT (centre) and MUSIC (right) show denser distributions below 30 degrees, but have larger tails above 90 degrees, which mean a grater variability in the estimated angles. From these distributions, some considerations about the spacial accuracy can be made. GCC-PHAT (left) has the largest concentration of errors around 45 degrees, with very few values above 90 degrees. On the contrary, SRP-PHAT and MUSIC have an higher concentration of values around 10 degrees, but the error extend above 90 up to 160 degrees. The shape varies gradually from left to right: MUSIC shows the slimmest distribution of the three from 60 to 90 degrees, while GCC-PHAT is instead the widest in this region. SRP-PHAT is positioned between the two. The shape of the error distribution, shows a gradual tendency of this algorithms to estimate the DOA more precisely, going from the fastest and easier to compute (GCC-PHAT) to the slowest and most complex (MUSIC), reflecting the expected speed-performance trade off.

The collision avoidance task was evaluated by looking at the total number of collisions between the ro-bat and the obstacles. With GCC-PHAT and SRP-PHAT, the ro-bat had no collisions during the experiments, while MUSIC caused bumping into obstacles in 15 out of 33 interactions, achieving a collision probability of 45.5%. These results highlight the challenges of MUSIC, which, despite its theoretical advantage in spatial estimation, suffers from high variability in error distribution and slower response times, making it unsuitable for this particular implementation.

Overall, GCC-PHAT offers faster and more reliable direction of arrival estimation due to its significant speed advantage—approximately 40 times faster than MUSIC and 16 times faster than SRP-PHAT—though its accuracy is limited. Conversely, SRP-PHAT and MUSIC demonstrate slightly better estimation accuracy but require further optimization for practical applications where computational speed and reliability are critical.

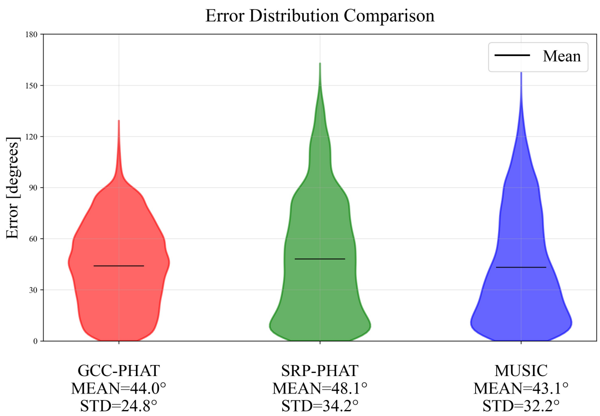

Figure 6: Distribution of the angular estimation error compared between different Direction of Arrival (DOA) algorithms. The mean error is almost the same for the three algorithms. The distribution is more uniform in GCC-PHAT (left) from 0 to 90 degrees, compared to SRP-PHAT (centre) and MUSIC (right), that show denser distributions below 30 degrees but larger tails.

5. Conclusions

In this thesis, the development of a biologically inspired navigation system for small mobile robots, using Direction of Arrival (DOA) estimation, is presented. The project involved the design and testing of various hardware components, including MEMS microphones and microphone arrays, to create a cost-effective and scalable solution. Three DOA estimation algorithms (GCC-PHAT, SRP-PHAT, and MUSIC) were implemented and tested for real-time functionality on small robotic platforms. A navigation algorithm was developed to assess the system's effectiveness in controlled experiments.

The algorithms displayed distinct strengths and limitations. GCC-PHAT excelled in speed, but with a large error at side angles, while SRP-PHAT and MUSIC demonstrated overall lower error distribution, but at the cost of higher variability and slower processing times. MUSIC's performance, in particular, was hindered by its computation time, making it unsuitable for real-time applications in this context. GCC-PHAT and SRP-PHAT successfully supported real-time obstacle detection and avoidance, although with trade-offs in speed and accuracy.

The thesis demonstrated the feasibility of sound-based navigation for small robots and established a foundation for further research. Future work could explore alternative array geometries, enhance DOA algorithms for efficiency and robustness, and enable the detection of multiple sound sources. This technology has promising applications in robotics, particularly for low-visibility environments, such as search-and-rescue missions, and in biological studies to simulate and understand animal behaviour.

References

[1] Thejasvi Beleyur and Holger R. Goerlitz. Modeling active sensing reveals echo detection even in large groups of bats. *Proceedings of the National Academy of Sciences*, 116(52):26662–26668, December 2019. Publisher: Proceedings of the National Academy of Sciences.

[2] Jacob Benesty, Jingdong Chen, and Yiteng Huang. *Microphone Array Signal Processing*. Springer Science & Business Media, March 2008.

[3] Joseph Hector Dibiase. *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*. PhD thesis, August 2000.

[4] Caleb Rascon and Ivan Meza. Localization of sound sources in robotics: A review. *Robotics and Autonomous Systems*, 96:184–210, October 2017.

[5] Ivan Jelev Tashev. *Sound capture and processing: practical approaches*. Wiley, Chichester, 2009.