POLITECNICO
MILANO 1863

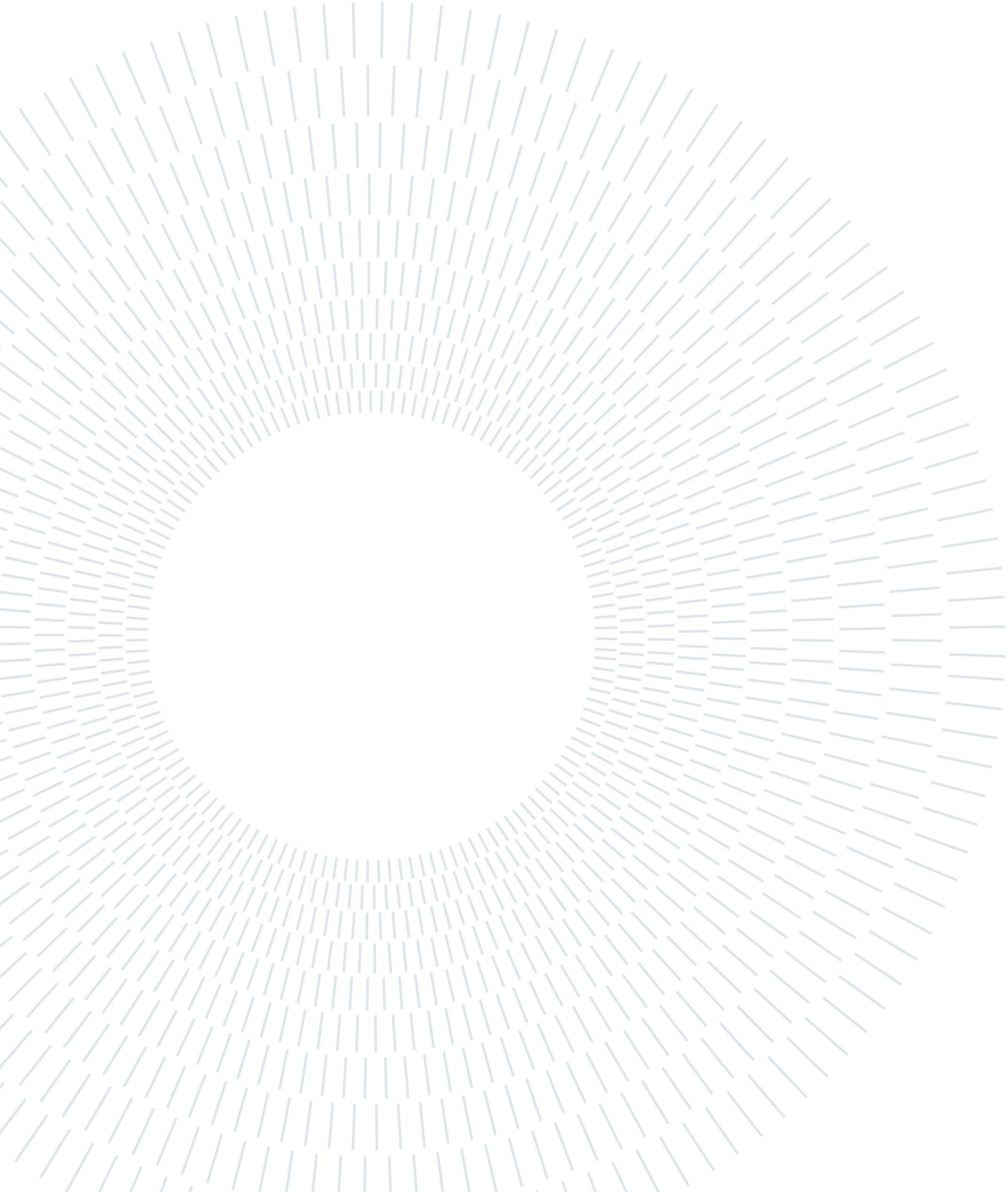
SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Echolocation for the RO-BAT Platform: Biomimetic Autonomous Mobile Robot Navigation Exploiting Ultrasonic Echoes

Tesi di Laurea Magistrale in
Music and Acoustic Engineering
Ingegneria Acustica e Musicale

Author: **Gabriele Baroli**

Student ID: 221120
Advisor: Prof. Fabio Antonacci
Co-advisors: Dr. Thejasvi Beleyur, Prof. Dr.-Ing. Heiko Hamann,
Dr. Andreagiovanni Reina
Academic Year: 2024-25



Abstract

Large groups of bats can achieve coordinated collective motion relying only on echolocation while experiencing the so-called Cocktail Party Problem: in order to echolocate they need to hear the echoes of their own call, but in groups the echoes overlap with conspecifics' emissions and the echoes thereof, resulting in masking and jamming effects. It is still unclear how they manage to maneuver in such conditions. To shed light on coordination mechanisms in crowded environments, biologists and engineers teamed up to implement the Ro-BAT platform as a robotic model of swarms of active sensing agents. Previous work realised a first prototype of a Ro-BAT capable of passive sound source localisation.

In this thesis I propose a second version of the Ro-BAT capable of active echolocation, equipped with an electrostatic ultrasonic transducer for emission, and an array of Micro-Electro-Mechanical System (MEMS) microphones specifically designed for ultrasound recording. I also implemented a signal processing pipeline for handling the incoming echoes of the emitted call to establish the relative position of nearby obstacles in terms of range and angle. For the angle, I tested three different Direction of Arrival (DOA) estimation algorithms, being them Delay and Sum (DAS), the Capon method, and Multiple Signal Classification (MUSIC), in static conditions. I then chose the most efficient one, i.e. DAS, to be included in the final pipeline for obstacle avoidance. The robot was then tested in lab experiments regarding its ability to autonomously navigate an arena with reflective obstacles only through echolocation.

The tests' results show that my bioinspired robot, mainly assembled with consumer-grade hardware, is capable of locating obstacles with a good level of accuracy thanks to the echoes of the robot's emission and steer its path accordingly to avoid collisions. The proposed solution can therefore be considered a starting point to be scaled to a full swarm of active sensing agents. This robotic model could then be used to perform control studies to deepen the understanding of collective behaviour mediated by acoustics in swarms of biological and robotic agents.

Keywords: ultrasound, echolocation, Direction of Arrival estimation, swarm robotics,

collective behaviour.

Abstract in lingua italiana

Gruppi numerosi di pipistrelli sono in grado di muoversi in modo coordinato sfruttando solamente l'ecolocalizzazione nonostante siano sottoposti al Problema del Cocktail Party: per ecolocalizzare devono necessariamente percepire gli echi del proprio richiamo, ma in gruppo questi sono spesso sovrapposti a richiami di altri pipistrelli, e agli echi derivanti. Questi segnali possono mascherare o disturbare la ricezione dei propri echi. Non è ancora stato spiegato del tutto come i pipistrelli riescano a destreggiarsi in condizioni simili. Per approfondire la ricerca su meccanismi di coordinazione in ambienti affollati, biologi ed esperti di robotica riuniti in un team hanno progettato la piattaforma Ro-BAT come un modello robotico di uno sciame di agenti che percepiscono in maniera attiva. Lavori precedenti si sono concentrati sulla realizzazione di un primo prototipo di Ro-BAT in grado di localizzare passivamente sorgenti sonore.

In questa tesi io presento una seconda versione del Ro-BAT in grado di ecolocalizzare, equipaggiato con un trasduttore elettrostatico per l'emissione di ultrasuoni e con una schiera di microfoni MEMS specificamente ideati per la registrazione di ultrasuoni. Ho anche implementato una catena di processamento dei segnali che permette al robot di stabilire la posizione relativa di ostacoli nei dintorni, in termini di distanza e angolo, tramite gli echi percepiti. Ho testato tre diversi algoritmi per la stima della Direzione di Arrivo in condizioni statiche, i quali sono Delay and Sum (DAS), il metodo di Capon, e Multiple Signal Classification (MUSIC). Il più efficiente dei tre, DAS, è stato integrato nel comportamento di evitamento degli ostacoli impiegato. Il robot è stato poi messo alla prova in esperimenti di laboratorio, testando la sua capacità di navigare autonomamente un'arena con ostacoli riflettenti solamente sfruttando l'ecolocalizzazione.

I risultati ottenuti nei test indicano che il mio robot bioispirato, assemblato in gran parte con materiali di qualità per il consumatore, è capace di localizzare ostacoli con un buon grado di accuratezza tramite gli echi della sua emissione. La soluzione proposta può quindi essere considerata come un punto di partenza per scalare verso un intero sciame di agenti che percepiscono in maniera attiva. Il risultante modello robotico può quindi essere usato per studi di controllo per approfondire la conoscenza del comportamento collettivo mediato tramite l'acustica in sciami di agenti biologici e robotici.

Parole chiave: ultrasuoni, ecolocalizzazione, stima della Direzione d'Arrivo, robotica degli sciami, comportamento collettivo.

Contents

1| Introduction

A fascinating and still disputed topic in biology is how large groups of echolocating bats can coordinate their collective movement while emerging from their roost location for foraging, sometimes, in groups comprising up to a few millions individuals [31]. Bats are active sensing agents, meaning that in order to extract information on their surroundings they probe the environment with short and powerful ultrasonic emissions and listen back to the echoes generated by reflectors around them. Therefore, they are expected to be very susceptible to signals which may mask or jam the echoes they use for echolocation.

This condition is a widely studied phenomenon in acoustics, and it is labelled as "the cocktail party problem" [4]. The intensity at which bats are expected to experience this disruptive phenomenon on their echolocation process has led researchers in the field to elevate this condition to the "cocktail party nightmare" [38]. Despite this, they still proceed in coordinated group flights on a daily basis, with few to no collisions reported.

Theoretical modelling has individuated bottlenecks in the sensory abilities of bats and other echolocating agents [5]: their ability to locate other individuals decreases for increasing group sizes. As a matter of fact, a robotic model of a group of active sensing agents can be a helpful tool to validate the theoretical studies, while also giving insights into real world mechanisms of which the full comprehension is still lacking. The Ro-BAT platform has been ideated with this twofold purpose in mind. On the one hand, modelling active sensing agents with robots lets researchers perform control studies on numerous parameters which could influence the efficacy of active sensing in physical swarms, as opposed to simulations. In simulated experiments, simplifying hypothesis are often introduced, with the risk of not fully encapsulating the complexities of the real world in their full extent. The data extracted from such work could then be used to explain biological mechanisms happening in swarms of which we do not have a full understanding yet. On the other hand, synthesising bioinspired behaviours for robots could lead to the development of new technologies, such as a cost-effective on-board localisation system for a coordinated collective motion mediated by acoustics.

1.1. Aim

After the successful implementation of a small robot capable of passive sound source localisation [9], the following step was represented by extending the capabilities of the very same robotic platform by equipping it with the necessary tools to perform active echolocation. Therefore the aim of this thesis is the development of ultrasound emission and reception devoted hardware, as well as the implementation of a signal processing pipeline which emulates the echolocation process performed by bats. The developed technology should then be deployed on the compact and flexible robotic platform already tested previously.

1.2. Hosting Institution

During my thesis work, I was hosted at the Centre for the Advanced Study of Collective Behaviour (CASCB), an interdisciplinary excellence research cluster based at the University of Konstanz that includes biologists, psychologists and roboticists with the common goal of pushing forward the study of the behaviour of individuals reunited in collectives of various nature, being them animal groups, human crowds, or decentralised robot swarms. I was co-supervised by Dr. Andreagiovanni Reina, leader of the Group Intelligence and Organisation research group, Dr. Thejasvi Beleyur, Principal Investigator of the newly founded Active Sensing Collectives research group and ideator of the Ro-BAT Project, and Prof. Dr.-Ing. Heiko Hamann, leader of the Cyber-Physical Systems research group. Dr. Beleyur provided the necessary competences in bioacoustics and animal behaviour, while both Dr. Reina and Prof. Hamann contributed to the thesis work with their robotics insights.

A significant contribution to the thesis work was also added by the Electronics Service Centre of the University Workshop: thanks to them it was possible to design and fabricate the needed electronics components in a relatively short span of time. Their support was invaluable for the good results of the work presented in these pages.

1.3. Original Contributions

The content of this thesis collocates itself in the greater landscape of the Ro-BAT Project at an initial step, in which the robotic model of active sensing agents is designed and prototyped. Pushing forward what was established in the previous work [9], I designed a microphone array based on Micro-Electro-Mechanical Systems technology, specifically intended for ultrasound recording. I also selected and tested the necessary hardware components to connect the electrostatic transducer to the audio interface already in use.

On the processing side, I developed the capabilities of the robot in a modular fashion. I coded the ranging mechanism, then the angle finding mechanism, and ultimately I combined the two together in a cohesive script for object localisation based on the echoes of the emitted call. For the angle, I selected and adapted to the system three well established methods for Direction of Arrival (DOA) estimation: Delay and Sum (DAS), Capon, and Multiple Signal Classification (MUSIC). I tested their performance in static conditions, and chose the most efficient one to be included in the final processing pipeline emulating echolocation.

Finally, I coded an obstacle avoidance behaviour for the robot that I tested in lab experiments. I organised an assessment of the performance of the robot with an autonomous navigation of an arena with reflecting obstacles, and extracted and analysed the data gathered in such experiments.

The developed robotic platform was displayed as a representative example of the bio-inspired robotics research carried out at the CASCB in numerous occasions to visiting researchers and institutions. It was also showcased with live demos at a PhD gathering at the University of Konstanz, and also at the annual spring retreat organised by the CASCB, with a general good reception from experts in the field of robotics and behavioural biology.

1.4. Thesis Outline

The rest of thesis is structured in the following chapters.

Chapter 2 introduces the theoretical background needed for the comprehension of the presented work, both in terms of problem statement and necessary tools and technologies which were employed in the development of the robot. A state of the art on the matter of bioinspired sonar sensors for robotics and autonomous echolocating robots is also presented.

Chapter 3 deals with the signal processing pipeline that was ultimately developed for accomplishing the task given at the start of the thesis. The methods employed are described in details, and an evaluation of the algorithms developed is given to the end of justifying the choices taken at the moment of the effective deployment on the robot for autonomous navigation.

Chapter 4 describes how the robot was implemented, giving a detailed description of the hardware pieces used and the final configuration reached. The Chapter contains also the characterisation procedures and results that were carried out on the final robot, in order to assess the performance of the various components that constitute the prototype built.

Chapter 5 presents the testing modalities and the obtained results after the tests conducted on the robot regarding its capability of navigating autonomously a controlled environment while avoiding obstacles encountered on its route.

Chapter 6, finally, reports the conclusions that can be drawn from this work, and the future developments of the project this thesis is a part of.

2 | Background and State of the Art

This Chapter is structured in two main sections. Section 2.1 will introduce the fundamental concepts which are necessary for the comprehension of the rest of the thesis. Section 2.2 will give instead an overview of works in the field of bioinspired acoustic sensors and robotic models of echolocating bats.

2.1. Background

The implementation of a robotic model of an echolocating bat implies the handling of two distinct phases of the process: the emission of a loud, wide-band ultrasonic call, and the recording and processing of the echoes generated by the presence of objects around the agent. For the emission stage, advancements in transducer technology for robotics has lead to the introduction of audio sources which are able to handle these strict operating conditions. For the reception stage, instead, two different strategies have been explored in literature: microphone arrays, and binaural bat-inspired microphones. A comprehensive analysis of the literature about both systems will be given in section 2.2. Microphone arrays were chosen for the development of this thesis, following the route traced in [9].

2.1.1. Ultrasound Properties and Echolocation

Ultrasound is defined as a pressure wave with frequencies which exceed the upper human hearing limit. Despite this being different from individual to individual, a commonly accepted definition states the beginning of the ultrasonic spectrum at around 20 kHz [26], which corresponds to a maximum wavelength of about 17 mm. It is known that planar pressure wavefronts undergo specular reflection when their wavelength is much smaller than the effective dimensions of the objects they hit, and so in most practical cases ultrasonic propagation is treated with a geometrical approach [26].

Echolocation in its simplest form is then suited for objects which once hit by an ultrasonic wavefront cause the immediate return of an echo back to the position of the emitter: this includes flat and smooth surfaces perpendicular to the direction of propagation of the wavefront, objects made out of convex surfaces, and objects which cause the scattering of the incident wave. To this merit, it has been proven that bats have a hard time around smooth and flat artificial surfaces, both horizontal and vertical. In [15] for example is shown that bats often try to drink from sheets of metal on the ground because they mistake them for water bodies. In [16] instead is studied how flat vertical surfaces act as a sensory trap for echolocating bats: such surfaces reflect the emitted echolocation calls away from the listener, causing the illusion of open flyways.

In order to develop a solid groundwork which might be expanded in the future, the robot I built deals with the simple case of echolocating objects which cause an immediate return of the echo to the listener, without considering more complicated scenarios in which the emitted call bounces more than once on nearby objects before getting back to the listener.

2.1.2. Microphone Arrays

Multiple microphones distributed in space which record the same acoustic field constitute an array of sensors. The advantage of having multiple microphones recording the same scene is that spatial information can be extracted from the array recordings, such as the direction from which a source arrives to the microphones. Conversely, it is also possible to focus the array directivity pattern in a certain direction to enhance the reception of the signals coming from that angle, while at the same time attenuating the unwanted interferers from other directions. The first operation is called Direction of Arrival (DOA) Estimation, while the second takes the name Beamforming.

Microphone arrays are usually classified based on their geometry. The sensors can be arranged in a line, which would constitute a linear array, or in more complex shapes which can also be two- or three-dimensional. In the remainder of the Chapter a linear geometry is considered, and in particular a uniform distribution of sensors in a line, which is commonly referred to as a Uniform Linear Array (ULA). Fig. 2.1 reports the situation in which a source located in the far field emits a signal which impinges on a ULA composed of M sensors. The planar wavefronts are perpendicular to the direction of propagation, and hit the microphones in the array each with a specific delay with respect to a microphone taken as a reference. The reference microphone is usually the first one on the left.

Figure 2.1: Source impinging from the direction θ on a Uniform Linear Array. The angle θ is measured from the axis of the array, and the microphone taken as a reference is the first one on the left. If the source is in the far-field, the wavefronts are planar, and the time delays between the arrival of the signal to each sensor in the array depend only on the inter-element distance and on the DOA.

2.1.3. Signal Model

Considering an array of M sensors and a single source in the far-field impinging on the array from a certain direction θ, the signal recorded by the k-th microphone in the array can be modelled as:

$$y_k(n) = h_k(n) * s(n - \tau_k) + e_k(n) \tag{2.1}$$

where $s(n)$ is the source signal measured at the source, $h_k(n)$ is the impulse response of the microphone, τ_k is the propagation delay from the position of the source to the sensor, and $e_k(n)$ is the additive noise present at the sensor. For a Uniform Linear Array, the relation between the propagation delay at the k-th channel and the direction of arrival of the source is the following:

$$\tau_k = (k-1)\frac{d\sin(\theta)}{c} \qquad \text{for} \quad \theta \in [-90^\circ; 90^\circ] \qquad \text{and } k = 1; \ldots; M \tag{2.2}$$

where d is the inter-element distance, θ is the source DOA, and c is the sound speed in air, considered equal to $343\,\mathrm{m\,s^{-1}}$. Now, the equation that models the array signal can be greatly simplified by assuming a narrow-band source, which is not always the case, especially for ultrasonic sources. Nonetheless, a wide-band source can be filtered with a filter bank to be separated in multiple narrow-band components, which can be processed individually. This justifies the use of a narrow-band model also for signals which are wide-band in nature.

A narrow-band signal with known carrier frequency ω_c takes the form:

$$s(n) = \alpha(n)\cos[\omega_c n + \phi(n)] \tag{2.3}$$

where $\alpha(n)$ and $\phi(n)$ are respectively the time-varying instantaneous amplitude and phase. The signal is analysed in the time-frequency domain by means of a Short Time Fourier Transform (STFT) employing a rectangular window (e.g. no windowing on the extracted frames) and a unitary hop-size, a method also known as sliding Discrete Time Fourier Transform (DTFT). The extracted frames are indicated as $s_t(n)$, and their DTFT is indicated as $S_t(\omega)$. If the attention is focused on the carrier frequency ω_c, the effect that the STFT has on a narrow-band signal is equivalent to a demodulation of a bandpass signal followed by a low-pass filtering, if $\alpha(n)$ and $\phi(n)$ are slowly-varying in the extracted frame. The result is a baseband signal of the form:

$$S_t(\omega_c) = \alpha(t)e^{j\phi(t)} \tag{2.4}$$

which is only function of time. The notation can be simplified by considering $s(t) \triangleq S_t(\omega_c)$, eliminating the frequency dependency.

Working in the STFT domain, the complete array model for a single source can be derived starting from:

$$y_k(t) = H_k(\omega_c)s(t)e^{-j\omega_c\tau_k} + e_k(t) \tag{2.5}$$

where $H_k(\omega_c)$ is the frequency response of the k-th sensor evaluated for $\omega = \omega_c$. This equation is the STFT of Eq. 2.1. Using a vectorial representation, the array signal for a single source can be written as:

$$\mathbf{y}(t) = \mathbf{a}(\theta)s(t) + \mathbf{e}(t) \tag{2.6}$$

where $\mathbf{y}(t) = [y_1(t), \ldots, y_M(t)]^T$, $\mathbf{a}(\theta) = [H_1(\omega_c)e^{-j\omega_c\tau_k}, \ldots, H_M(\omega_c)e^{-j\omega_c\tau_k}]^T$ and $\mathbf{e}(t) = [e_1(t), \ldots, e_M(t)]^T$.

Assuming ideal linear acoustic propagation, Eq. 2.6 can be extended to the case N of sources impinging on the array with N different DOAs as:

$$y(t) = As(t) + e(t) \tag{2.7}$$

where $A = [a(\theta_1); \dots; a(\theta_N)]$ is the $M \times N$ propagation matrix, and $s(t) = [s_1(t); \dots; s_N(t)]^T$ is the source vector.

The propagation vectors can be simplified by introducing the notion of spatial frequency: this quantity, defined as

$$\omega_s \triangleq \omega_c \frac{d\sin(\theta)}{c} \tag{2.8}$$

is directly related to the DOA of the source, and if substituted into the propagation vector formulation leads to

$$a(\theta) = \left[1; e^{-j\omega_s}; \dots; e^{-j(M-1)\omega_s}\right]^T \tag{2.9}$$

where the sensors' frequency responses were also assumed to be $H_1(\omega_c) = H_2(\omega_c) = \dots = H_M(\omega_c) = 1$. From the propagation vector expression in 2.9, it is clear how the array samples a complex sinusoid of frequency ω_s, with a unitary sampling frequency.

2.1.4. Spatial Aliasing and Effective Array Aperture

Since the array is effectively sampling a sinusoid with a given frequency, the Nyquist-Shannon theorem must hold also in this case to avoid the aliasing phenomenon, which introduces ambiguity in the representation of a certain DOA due to the appearance of grating lobes in the array directivity pattern.

The condition to be satisfied is:

$$|\omega_s| \leq \pi \tag{2.10}$$

Substituting to the spatial frequency its definition, and introducing the wavelength $\lambda = \frac{2\pi c}{\omega_c}$ the condition becomes:

$$\left|2\pi \frac{d\sin(\theta)}{\lambda}\right| \leq \pi \tag{2.11}$$

The anti-aliasing condition can be conveniently expressed as a function of the inter-element distance d:

$$d \leq \frac{\lambda}{2} \tag{2.12}$$

This condition establishes the minimum wavelength that won't be aliased by an array with a sensor distance d. The maximum frequency which won't cause spatial aliasing can therefore be written as:

$$f_{max} = \frac{c}{2d} \tag{2.13}$$

For larger frequencies it is then necessary to use densely packed microphone arrays, to make sure the spatial aliasing phenomenon is avoided.

Microphone arrays exhibit also an empirical lower frequency bound, given by the relationship between the wavelengths of the recorded signal and their aperture. Their functioning is based on capturing a phase difference between signals impinging on each sensor in the array. If the wavelength of an incoming signal is comparable to the array aperture, given by:

$$L = M \cdot (d - 1) \tag{2.14}$$

the phase difference between each channel becomes negligible since the wavefront vary too slowly. A good rule of thumb for an effective microphone array aperture states that the aperture should be at least twice the maximum wavelength recorded. This can be translated to a condition on the frequency of the recorded sounds as:

$$f_{min} = \frac{c}{2L} \tag{2.15}$$

The array designed for the active sensing Ro-BAT is composed of eight microphones with an inter-element distance of $2.7\,\mathrm{mm}$. The two frequency bounds are then:

$$\begin{cases} f_{min} = 9.1\,\mathrm{kHz} \\ f_{max} = 63.5\,\mathrm{kHz} \end{cases}$$

2.1.5. Direction of Arrival Estimation Problem Formulation

As stated previously, the spatial frequency of a signal is directly related to its Direction of Arrival (DOA). The problem of DOA estimation can then be formulated as a spatial spectral estimation of the impinging signals on the microphone array. Two popular methods for accomplishing this task are spatial filtering, and subspace based methods.

Spatial Filtering

One possible approach to DOA estimation is constituted by spatial filtering. This strategy consists in designing a linear filter which will combine the array signals in such a way that constructive interference will enhance signals arriving from directions of interest, and destructive interference will attenuate signals coming from unwanted directions. Delay and Sum [12] and the Capon method [7] are classical examples of such filters. In general, for an array of M elements the filter will be composed by M complex weights $h = [h_1, \ldots, h_M]^H$, where the superscript ^{H} indicates the hermitian transpose. The filtered signal takes the form:

$$y_F(t) = \sum_{k=1}^{M} h_k \; y_k(t) = h^H y(t) \tag{2.16}$$

The power of the filtered signal is then:

$$E \; |y_F(t)|^2 \; = h^H R h \tag{2.17}$$

where R is the covariance matrix of the array data, defined as:

$$R = E \; y(t) y^H(t) \tag{2.18}$$

The operator E indicates the statistical expectation. Since the array signal is finite in time, in most practical cases the covariance matrix is swapped for its sample estimate, defined as:

$$\hat{R} = \frac{1}{K} \sum_{t=1}^{K} y(t) y^H(t) \tag{2.19}$$

The power of the filtered signal constitutes a good indication of the energy coming from a certain direction , so by comparing the energy coming from all the directions it is possible to compute a spatial spectrum, whose peaks correspond to directions from which signals impinge on the array.

Subspace Based Methods

An alternative to spatial filtering for DOA estimation is represented by methods which separate the covariance matrix of the array data into a signal subspace, which is related to the signals impinging on the array, and in a noise subspace, related to the noise introduced by the sensors. These methods are based on a series of assumptions, which are:

- Exactly N sources impinge on the array from N distinct directions, thus leading to N distinct spatial frequencies. N must be also lower than M, which is the number of sensors in the array.
- The covariance matrix of the source signals $R_s = E \; s(t) s^H(t)$ is full-rank.
- The sensor noise is spatially white, with independent and identically distributed components having identical variance. The covariance matrix of the error signals is then $E \; e(t) e^H(t) \; = \; {}^2 I_M$. The noise is also uncorrelated to the source signals.

Under these assumptions, the covariance matrix of the array data can be written as the sum of two components, one related to the source signals and the other to the noise, as:

$$R = AR_sA^H + \sigma^2 I_M \tag{2.20}$$

By performing an eigenvalue decomposition and sorting the eigenvalues in ascending order, it is possible to express the covariance matrix of the array data as:

$$R = [U;V]\,\Lambda_y\,\left[U^H;V^H\right]^T \tag{2.21}$$

where the eigenvectors in U are associated to the eigenvalues $\{\lambda_1;\ldots;\lambda_N\}$ and the eigenvectors in V to the eigenvalues $\{\lambda_{N+1};\ldots;\lambda_M\}$. The diagonal matrix Λ_y contains the sorted eigenvalues of R. It is possible to prove that the eigenvectors in U are associated to the source signals, therefore spanning the signal subspace, and the eigenvectors in V are associated to the sensors' noise, giving origin to the noise subspace of the covariance matrix. Each of the two subspaces found possesses properties which can be leveraged to estimate the DOAs of the sources looked after. MUSIC [43] exploits the orthogonality of the noise subspace to the propagation vectors corresponding to DOAs of actual sources, while an example of a method that exploits properties of the signal subspace is constituted by ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [41].

The matrices U and V are obtainable if and only if the conditions reported above apply, and the sources signal power is higher than the noise power.

2.2. State of the Art

In recent years the rapid development of mobile robotics has lead to a vast exploration of sensing strategies with which a robot can acquire information on its surroundings to then plan its movement strategy. Among these solutions a vast majority is focused on cameras or light-based sensors, but also sound and in particular ultrasound emission and echoes detection is currently being investigated. In this latter case, though, almost all of the research made is dedicated to advanced sonar sensors for robotics and algorithms for Simultaneous Localisation and Mapping (SLAM) exploiting acoustic signals inspired by the echolocation capabilities of bats to various degrees. These works are mainly focused on strategies that individual agents can exploit, and it is not clear if the proposed solutions can scale to groups of agents. This is why a cheap and scalable platform such as the Ro-BAT was designed, with the goal of tackling the cocktail party problem in swarms of active sensing agents.

Section 2.2.1 will give an overview of bioinspired sensors for localisation and navigation based on acoustics. Most of the works presented in this Section come from the Cosys-Lab at the University of Antwerp [39], and constitutes a great bulk of work on bat-inspired sensory solutions for robotics.

In Section 2.2.2, instead, other works which present a full-fledged autonomous echolocating robot are reviewed, presenting the novelty and advancements that they introduced, and also underlying their strengths and weaknesses with respect to scalability.

2.2.1. Bio-Inspired Acoustic Sensors

Sonar sensing is surely not a novel area of research in engineering, but recent advancements in behavioural studies on bats and a refinement of the sonar technology involved have sparked new interest for what a bioinspired sensor could bring to the table. In mobile robotics, on the one side, new sensing strategies are explored for autonomous navigation in complicated environments. On the other side, robotic models could help in improving the knowledge on biological mechanisms which are not fully understood yet.

In [47] two different types of audio receivers for a bioinspired sonar sensor are evaluated: binaural microphones with 3D-printed replicas of bats' external ears are compared against microphone arrays in combination with digital beamforming. The comparison was based on an objective information theoretic performance measure that quantified their localisation capabilities in the presence of realistic noise. The conclusions drawn in the paper are that the two systems are more or less equivalent on the localisation performance side, but a possible advantage in the array solution, despite the increased complexity in managing more than two microphones, consists in the possibility to digitally steer the directivity pattern of the sensor in desired directions without resorting to mechanical modifications of the sensors, which would be needed in the case of two microphones encased in pinnae replicas.

Both binaural microphones and microphone arrays were used to extract information then fed to a SLAM model based on the mammalian hippocampus in subsequent works. In [45], a biomimetic sonar sensor composed of an electrostatic ultrasound transducer and two omnidirectional microphones encased in plastic replicas of a bat's pinnae is used as an exteroceptive sensor, while interoceptive sensory input is given by the robot's odometry retrieved by the motor commands. The echoes, which are filtered by an Echolocation Related Transfer Function (ERTF) given by the superimposed effect of the emitter's directivity and the spatial filtering introduced by the artificial outer ears, are then processed to extract a "cochleogram". This time-frequency representation of the recorded echoes is obtained by modelling the functioning of a bat's cochlea with a bank of gammatone filters. The obtained cochleograms are used as local view templates of the environment, which combined with the robot's odometry constitute the input to the SLAM algorithm. The result is a topological map of the environment. The paper's authors drove a robot equipped with such sensor in an unmodified office environment, and a geometrically consistent topological map was the result of the SLAM algorithm. The authors conclude that cochleograms are in general sufficiently descriptive of a particular local view to be used as a base for localisation and mapping of a large environment, despite still giving rise to ambiguity in some cases. Such a system would work in conditions of scarce visibility which are problematic for cameras or LIDAR sensors.

The other approach to SLAM [46] employs an array-based biomimetic sensor [48] consisting of the same emitter as in the previous work, but on the receiving end a planar random array of 32 condenser microphones is used in conjunction with a Delay and Sum beamformer. The output of this sensor is a so-called "energy scape" representation of the environment, where each point identified by three coordinates (range, azimuth and elevation) represents the energy reflected back to the sensor from that particular location. This sensor was exploited for obtaining both information on the nearby environment, and also odometry data based on the acoustic flow of consecutive snapshots acquired by the device. The sensor was mounted on an electric wheelchair and driven in a typical office environment. A geometrically consistent topological representation of the environment was obtained also for this case, by feeding to the SLAM algorithm the 3D energy scapes as local view templates, and the odometry data extracted from the acoustic flow model as the interoceptive sensory input.

Both sensors architectures were refined in subsequent works: in [27] a biomimetic sonar sensor based on the common big-eared bat (*Micronycteris microtis*) outer ears replicas is used to guide the autonomous navigation of a two-wheeled robot, in conjunction with a subsumption architecture whose layers are:

1. Collision avoidance: the robot rotates on itself until its pathway is free of reflectors. This behaviour is triggered by the presence of obstacles under a certain range threshold.
2. Obstacle avoidance: the robot proceeds on a curved trajectory distancing itself from nearby obstacles, moving in the opposite direction with respect to the highest amount of energy received between the two ears.
3. Straight drive: the robot proceeds forward at constant velocity along a straight line when its pathway is free of obstacles.

The input to the control system is constituted exclusively by data extracted from the echoes received by the two microphones inside the bat ears replicas. The robot was successful in performing an autonomous navigation of a cluttered environment. An interesting modification introduced to the sensor architecture was a baffle around the ultrasonic transducer, which enlarged its radiation pattern. The effect of baffles on ultrasonic transducers was studied in [20], concluding that good-quality 3D-printed baffles can enlarge the narrow directivity pattern of electrostatic transducers to near-omnidirectionality without significant loss in the output power.

In [22] instead, the array-based sensor architecture was refined by swapping the condenser microphones for digital Micro-Electro-Mechanical Systems (MEMS) microphones, reducing costs and need for extra Analog-to-Digital conversion (ADC) equipment, and the hardware implementation is completely embedded, constituting a plug-and-play sensor for robotic applications. As a beamforming technique, the Delay and Sum method was still used. The resulting sensor was then deployed on a mobile robot which was driven in a cluttered corridor to realise a 3D map of the traversed environment. The data obtained from the sensor was confronted with LIDAR measurements for validation. A reduced form factor version of the sensor explained above was realised in [53]. Contrary to the previous sensor model, this device works with narrow-band emissions and the Multiple Signal Classification (MUSIC) algorithm for retrieving the Direction of Arrival (DOA) of the echoes. This device architecture reduces production costs with respect to the previous sensor, but it suffers from the multipath propagation of the echoes, generating "ghost" reflectors. A more in-depth simulation study on such sensor is found in [54], where different DOA estimation algorithms were employed with the cited sensor architecture varying a number of parameters such as the Signal-to-Noise Ratio (SNR) of the received echoes, the number of snapshots taken for the same scenario, the spacing of two closely placed objects and the amount of objects which could be resolved by the sensor at the same time. The authors conclude that the most limiting aspect of sonar sensors is the need for a high dynamic range, since objects placed at sharper angles suffer from harsh attenuations due to both the directivity of the transducer employed, and the reduced aperture of the array. The SNR and the number of snapshots acquired do not present any significant effect on the localisation process, which is desirable since mobile robots need fast updating sensor readings to operate, often in noisy conditions.

One last frontier of research in the sensor architecture field is represented by [28], in which the aperture of an array based sensor is reduced by means of 3D-printed waveguides which effectively realise an array geometry that could otherwise be impossible to reach due to the physical dimensions of the MEMS microphones employed. This approach extends the frequency range at which ultrasonic microphone arrays can operate before incurring in the spatial aliasing phenomenon, but introduces a significant amount of attenuation in the signals recorded by the microphones, and requires a calibration step before operation for taking into account the phase shifts introduced by the baffles. A notable result of the integration of such baffles to microphone arrays is a directivity pattern resemblant of a bat Head-Related Transfer Function (HRTF), which is an interesting aspect for biological modelling researches.

2.2.2. Autonomous Echolocating Robots

In addition to bioinspired sensors, another branch of research is focused on the realisation of full-fledged robots which which can navigate autonomously relying on the information extracted from the environment by probing it with short ultrasonic emissions.

A first example of an autonomous robot whose navigation strategy is based on the echolocation capabilities of bats is represented by [57]. In this work, an autonomous robot was equipped with a minimal biomimetic sonar sensor, consisting of an ultrasonic transducer and two ultrasound sensors. The narrow-band signal emitted in combination with the reduced dimensions of the transducer led to a radiation pattern comparable to the one of a real bat, having a full width of 80-100 at 6 dB. Objects could be located in range by the time of arrival of the echoes, and in direction by the Time Difference of Arrival (TDOA) at the two receivers. The sensor unit was mounted on a servomotor which could steer the unit in the direction of the echoes received, to keep the localised obstacle in the sensor's field of view (FOV). This strategy was borrowed by bats, which were shown to turn their heads to focus specific reflectors previously identified. An Inter-Pulse Interval (IPI) control mechanism based on the nearest obstacle distance was also implemented. When navigating autonomously a rectangular arena with cylindrical obstacles, the head steering mechanism was proven to perform better than a fixed sensor unit aligned with the direction of motion, but the localisation of only the nearest obstacle did not perform satisfyingly, leading to numerous collisions. It was for this reason that the authors also implemented a Multi-Object-Detecting Navigation System (Multi-ODNS). This obstacle localisation method took into account multiple echoes, corresponding to more than one obstacle, and after locating each one a repulsion force vector was computed for each obstacle as well. The direction in which the robot steered was then a consequence of choosing to avoid the obstacle with the highest repulsion force vector. This new approach to obstacle avoidance demonstrated itself to be successful where the Mono-ODNS previously failed. This robot configuration presents a bio-inspired obstacle avoidance behaviour introducing also the possibility of mechanically steering the sonar sensor, which is a natural behaviour for bats. The emitter circuit is too simple, though, being only capable of emitting 40 kHz short tones. The flexibility to model species of Frequency Modulating (FM) bats is therefore excluded.

Another example of autonomous bio-inspired robot is constituted by [10]. The output of this work is a fully autonomous terrestrial robot which can map the 2D structure of a novel outdoor environment, label the obstacles as plants or non-plants, and then plan a path accordingly to navigate while avoiding obstacles. This all-terrain, rugged robot is equipped with an ultrasonic electrodynamic speaker, and two condenser microphones specialised for ultrasound recording. The sensor head is mounted on a gimbal which can rotate while maintaining stability. The A/D and D/A conversions are handled by a high-end Data Acquisition (DAQ) board. The robot moves forward in steps of 0:5 m, then stops for 30 s to acquire echoes from the environment by emitting three 10 ms sweeps from 100 kHz to 20 kHz in the directions 60, 0 and 60 with respect to the driving orientation to emulate the directivity of a bat's emission. The resulting echoes are processed to individuate the distance and the angle from nearby obstacles, and the extension of their contour. Audio features are also extracted from the received echoes to classify the reflectors encountered as plants or non-plants thanks to an artificial neural network. The robot was capable of mapping the position of obstacles in a novel environment and their contour with a satisfying degree of accuracy and planning its route accordingly without any recorded collision. The classification accuracy was also above the chance level of 50% of right decisions. This work implements an acoustic based autonomous navigation method which is biologically plausible, even though with evident limitations related to the movement speed of the robot and its emission strategy. Furthermore, the audio feature based classification of the type of obstacles encountered is an interesting introduction, which could be further exploited in other settings, also in a swarm context. The dimensions, slow reactiveness and the realisation cost of a single agent are not suited for swarm robotics, though.

In [30], instead, a robotic model for the avoidance of non-localisable obstacles by bats is presented. This work aimed at validating a minimal set of acoustic cues for reacting to a series of non-localisable echoes, meaning a group of highly overlapping echoes caused by complex reflectors such as the vegetation among which bats fly in nature. The cues taken into consideration are the Interaural Intensity Difference (IID), and the distance from the closest obstacle. Moreover, also two different navigation modalities are investigated together with the primary question, being those the Fixed Head strategy, and gaze steering according to a Delayed Linear Adaptive Law (DLAL). The robotic platform used consists of a differential drive robot mounting a narrow-band ultrasonic transducer and two microphones encased in stylised plastic replicas of a bat's pinnae. The sensor unit was mounted on a pan-tilt system which could emulate head movements observed in bats. The obstacle avoidance behaviour was tested both in a rectangular arena with complex reflectors which produced highly overlapping echoes, and in a cluttered corridor whose walls lined with egg cartons also reflected the robot's calls in a non localisable fashion. By comparing the two navigation strategies in conjunction with the set of cues chosen to extract spatial information from the environment, the authors conclude that the IID and the distance from the nearest obstacle is enough for achieving autonomous navigation, despite collisions were reported, even though only in the 10% of obstacle interactions. The robot was also able to complete the navigation of the corridor successfully. Contrary to [57], they found that acoustic gaze scanning leads to a worse performance than locking the sensing direction to the motion direction. This is due to the amount of angular information that can be extracted from overlapping echoes: in the case of highly cluttered environments steering the sensor unit does not improve the efficiency of the motion. The results reported in this work offer a solid strategy for a simple obstacle avoidance behaviour in heavily cluttered environments which relies on a minimal but robust set of cues extracted from the echoes received. This strategy is mainly applicable to robots and sensors whose receiving unit is constituted by a left and a right "ear". An application to microphone arrays would require further elaboration.

2.2.3. Summary and Conclusions

In this Section a series of bioinspired sensor architectures and robotic platforms based on acoustics were analysed and reviewed. The solutions presented are enormous sources of inspiration for biomimetic acoustic sensor architectures, but are not thought for swarms of echolocating agents: I instead argue that the presence of similar audio sources to the ones employed would jam the sensors to various degrees, rendering them ineffective in a collective of active sensing robots. They can however constitute a first approximation of the echolocation process that most bats exploit continuously during flight.

Another point of crucial importance is the construction complexity of such solutions: they all require the development of highly specialised hardware pieces and are generally expensive and cumbersome if compared to small robotic platforms which are often employed for swarm robotics, making their deployment not suitable for the context of this thesis.

A cheap and scalable robotic platform that emulates echolocating bats has not yet been realised, and the Ro-BAT project aims at filling this gap.

3 | Signal Processing Methods

This Chapter will describe the full signal processing pipeline devoted to the echolocation capabilities of the Ro-BAT and the conditions in which the robot is intended to operate, together with some results regarding the implementation of Direction of Arrival estimation algorithms. The hardware details and their characterisation is the topic for Chapter 4, here only the necessary information for understanding the rationale behind the methodologies employed is presented.

3.1. Signal Processing Pipeline

The signal processing pipeline that was developed is largely inspired by the work of Steckel et al. [48] and further developments of the Cosys-Lab at the University of Antwerp [39]. Their idea of a sonar system based on a single emitter and multiple receivers is at the basis of the content of this thesis. There are a few differences regarding the objective with which the two systems are built, though: while their sensor is thought as a plug-and-play reliable alternative for autonomous robotics navigation, the robot proposed in this thesis is a first attempt to deploy active echolocation on a small mobile platform with the greater objective of studying how a swarm of agents which base their movement strategy on acoustic mechanisms can use active echolocation effectively to achieve a coordinated collective behaviour.

The geometry of the problem posed is assumed to be two-dimensional: the robot moves on a plane, so only two coordinates are necessary to establish the relative position of objects around it. These two coordinates are the distance from the robot to the objects, and the angle at which they are positioned with respect to the robot's longitudinal axis. Assuming, as explained above, that the objects around the robot reflect the emitted calls back to the robot's position, the distance from the robot to the objects can be computed by knowing the time difference between the instant in which the call is emitted, and the instant in which the echoes are received. Regarding the incident angle of the echo, Direction of Arrival estimation algorithms can be employed, after recording the received echoes by means of a microphone array. Among the many different array geometries, a

Uniform Linear Array (ULA) configuration has been selected for its simplicity, since it is fully capable of handling a two-dimensional localisation problem.

The field of view (FOV) of the robot is by necessity its front half-plane, since the audio source is pointed towards its front and the microphone array is linear, which introduces a front-to-back ambiguity solved by only considering one of the two half-planes with respect to the array orientation. The angular axis θ is defined as $\theta \in [-90^\circ; 90^\circ]$, where -90° corresponds to the robot's right hand side, 90° to its left hand side, and 0° is in correspondence of its longitudinal axis.

Another simplification introduced in the system is that the robot will focus only on the first echo received, meaning that at every instant only the objects placed on a semi-circle with the lowest radius with respect to the robot's position are located.

3.1.1. Call Emission and Array Recording

Since the Ro-BAT aims at emulating the echolocation strategies of Frequency Modulating (FM) bats, the type of signal that constitutes the emitted call will be a wide-band chirp, in particular of the linear type. This assures that all the frequencies of interest have the same weight in the emitted signal. An example of a call signal used by the robot can be seen in Fig. 3.1. After an emission, echoes will be generated by obstacles posed around the robot. Both the call and the echoes are recorded by the microphone array, which enables a spatial processing of the recorded audio.

The signal captured by the k-th microphone in the array after the emission of a call and the reception of the subsequent echoes can be modelled as:

$$y_k(n) = h_k(n) * \left(s_c(n) + \sum_{i=1}^{L} \alpha_i s_c(n - \tau_i - \tau_k) \right) + e_k(n) \quad \text{for } k = 1; \dots; M \tag{3.1}$$

where $h_k(n)$ is the impulse response of the single sensor, $s_c(n)$ is the emitted call, α_i and τ_i are respectively the attenuation coefficients and the propagation delays due to the reflection process, L is the number of echoes recorded, τ_k is the propagation delay due to the position of the sensor in the array, M is the number of sensors in the array and e_k is the noise present in each sensor.

Figure 3.1: Time trend and spectrogram of the emitted signal.

3.1.2. Filtering and Ranging

The first step in the audio data processing is constituted by a matched filter, described by:

$$y_k^{MF}(n) = \mathcal{F}^{-1}\{Y_k(j\omega)\cdot S_c^*(j\omega)\} \tag{3.2}$$

where $Y_k(j\omega)$ is the Discrete Fourier Transform (DFT) of the signal recorded by the i-th microphone, $S_c^*(j\omega)$ is the complex conjugate of the DFT of the call signal, and the $\mathcal{F}^{-1}$ operator denotes the Inverse Discrete Fourier Transform. The output of the filter will be a series of scaled and time-delayed autocorrelation functions of the emitted call, if we consider as negligible the spectral distortions introduced by the emission and reflection process. As a result, the call and the echoes onsets are easily individuated in the cross-correlations, since they are in correspondence of the local maxima of the autocorrelation functions. Other signals that do not match the template of the matched filter will instead be annihilated by the filter.

To better perform a peak-finding operation to reveal the position of the echoes, the envelopes of the cross-correlations are extracted by taking the absolute value of their Hilbert transform. Then, a different strategy is applied to isolate the call and the first echo, respectively.

- **Call detection:** a mean envelope is computed by averaging all the envelopes extracted from the cross-correlations. The call onset is then retrieved by a peak-finding operation executed on the mean envelope, to have a more reliable estimate. The first index returned by the algorithm represents the time instant wanted, since it is assumed that the call is the first relevant signal appearing in the analysed sequence.

- Echo detection: for each envelope a peak-finding operation is computed, starting the search from a tolerance window of samples after the peak representing the call, to avoid considering further oscillations belonging to the call autocorrelation function as reflections caused by very close objects. This introduces a "blind spot" in the robot's field of view. The tolerance window which worked best in terms of avoiding "ghost obstacles" versus locating obstacles as close as possible to the robot corresponds to a radius of $12.5\,\mathrm{cm}$. For each envelope the first index corresponding to a peak is returned, and the smallest among all of them is used as a reference to compute the distance from the nearest object.

Having retrieved the two onsets needed, the range from the nearest object is then computed according to:

$$d = \frac{1}{2} c \Delta t \tag{3.3}$$

where Δt is the time interval between the two peaks, c is the sound speed in air, considered to be $343\,\mathrm{m\,s^{-1}}$, and d is the estimated distance from the robot to the object. An example of the full filtering and ranging process is shown in Fig. 3.3, while the spectrogram of the same audio recording is reported in Fig. 3.2.

Figure 3.2: Spectrogram of the emitted call and first echo. The audio was extracted from the first channel of the array. The call and the echo are clearly distinguishable in the time-frequency representation of the recorded audio.

(a)

(b)

(c)

Figure 3.3: Example of the ranging process for locating an object placed at 50 cm in front of the robot. The three graphs refer to a single channel in the microphone array. It is easy to see how much the matched filter compresses in time the signals that match the template by comparing (a) and (b). The envelope reported in (c), then, improves the peak-finding algorithm accuracy, which needs to be tailored properly nonetheless to avoid locating the spurious peaks of the call's autocorrelation function as inexistent obstacles.

3.1.3. Direction of Arrival Estimation Algorithms

The following step in the localisation process is the Direction of Arrival estimation. Two of the possible strategies for this task are employing spatial filtering, or parametrizing the covariance structure of the array data to exploit the properties of one of its subspaces. I implemented and tested two algorithms belonging to the spatial filtering class, and a third one belonging to the parametric methods class. One relevant note is that all three of the algorithms implemented are effective on narrow-band signals, while the call emitted by the robot is inherently wide-band. To accommodate this requirement, each algorithm has been extended to a wide-band use case by means of an incoherent average [49, 56]: the input signal is divided into sub-bands which are processed individually. The results obtained by each sub-band are then averaged, and the DOA is computed on the basis of this average. This method is defined as incoherent because a series of power spectra are averaged, without maintaining any phase relationship between sub-bands.

The data fed to these algorithms is constituted of a rectangular window of length 1 ms centred on the cross-correlations portion corresponding to the echo identified in the previous processing step. The advantages of passing the filtered echo with respect to the raw audio data are that the filtered echo is much more isolated in the time history of the signal, reducing the bleeding-in of the emitted call even when the obstacle is very close to the robot, and also it contains the same amount of spectral information in a reduced number of samples, resulting in a faster computation time.

The DOA estimation algorithms implemented are:

- Delay and Sum (DAS): this algorithm focuses on passing undistorted the signals coming from a specific direction θ, while attenuating the signals coming from all the other directions. The array signal is assumed to be spatially white, meaning that the covariance matrix R is an identity matrix I of rank M. The formalisation of the equivalent optimisation problem is written as:

  $$h(\theta) = \arg\min_{h} h^H h \qquad \text{subject to } h^H a(\theta) = 1 \tag{3.4}$$

  The resulting filter coefficients are defined as:

  $$h(\theta) = \frac{a(\theta)}{M} \tag{3.5}$$

  where $a(\theta) = [1 \quad e^{-j\omega_s} \quad \dots \quad e^{-j(M-1)\omega_s}]^T$ is the propagation vector whose spatial frequency ω_s corresponds to the direction θ, and M is the number of microphones in the array. It is worth noting how the obtained coefficients are scaled complex exponentials, which correspond to pure delays: the output of the filter is the sum of the signals captured by the array, each one of them with an appropriate phase shift to create zones of constructive and destructive interference where it is desired. The name with which this procedure is usually found in literature reflects this very property.

The resulting pseudospectrum, which derives from the power of the filtered signal evaluated at all the angles of interest $\theta \in [-90^\circ; 90^\circ]$ is:

$$p(\theta) = \frac{a^H(\theta)\hat{R}a(\theta)}{M^2} \tag{3.6}$$

where $\hat{R}$ is the sample estimate of the covariance matrix of the array data. The DOAs of N sources are then estimated as the angles in correspondence of the N highest peaks of the function:

$$p_{avg}(\theta) = \frac{1}{B}\sum_{i=1}^{B} p_i(\theta) \tag{3.7}$$

where B is the number of processed sub-bands.

- Capon method: the second spatial filtering algorithm developed is the so-called Capon method. Contrary to the previous approach, this method puts effort into attenuating the actual signals which impinge on the array from unwanted directions, while passing undistorted the signals coming from the desired direction. This method is then data-dependent, since the corresponding optimisation problem takes the form:

$$h(\theta) = \arg\min_{h} h^H R h \qquad \text{subject to } h^H a(\theta) = 1 \tag{3.8}$$

The filter coefficients for every sub-bands assume the following form:

$$h(\theta) = \frac{R^{-1}a(\theta)}{a^H(\theta)R^{-1}a(\theta)} \tag{3.9}$$

where R is the covariance matrix of the array data. The resulting pseudospectrum for this spatial filter is:

$$p(\theta) = \frac{1}{a^H(\theta)\hat{R}^{-1}a(\theta)} \tag{3.10}$$

where once again the sample estimate of the covariance matrix $\hat{R}$ is used in place of the actual covariance matrix R.

Also in this case, the pseudospectra obtained from each sub-band elaboration were averaged before seeking the directions of arrival, which are defined as the angles corresponding to the N highest peaks of the averaged pseudospectrum as well.

The trade-off between DAS and Capon is that the Capon method has a better spatial resolution, meaning that it is able to discriminate between sources coming from two very close points in space much better than what DAS can usually do, at the cost of a more complex computation.

- Multiple Signal Classification (MUSIC): this algorithm belongs to the signal subspace methods class. The subspace exploited by this method is the one generated by the eigenvectors of the covariance matrix of the array data related to the sensors' noise. It can be proven that such eigenvectors and the propagation vectors associated to N distinct sources are orthogonal. So, once the eigenvector matrix related to noise $\hat{V}$ is estimated after an eigenvalue decomposition of $\hat{R}$, the DOAs of N wideband sources can be estimated in correspondence of the N highest peaks of the function:

$$p_{avg}(\ ) = \frac{1}{B}\sum_{i=1}^{B} p_i(\ ) \tag{3.11}$$

where the pseudospectrum for each sub-band is defined as:

$$p(\ ) = \frac{1}{a^H(\ )\hat{V}\hat{V}^H a(\ )} \tag{3.12}$$

One important difference between this parametric method and the two spatial filters mentioned before is that the expected number of sources to be detected with MUSIC must be known in advance. This translates well in the case in which only one obstacle at a time is considered, but it would scale poorly when an unknown number of obstacles which is greater than one would need to be treated by the robot.

3.1.4. DOA Algorithms Performance Comparison

Before deploying the full pipeline on the robot while it navigates, I tested and compared the performances of the three algorithms developed in terms of their computation time, spatial resolution, and accuracy for obstacles in various positions around the array. More information on the robotic platform employed are given in Chapter 4, here only the relevant tests results are displayed and explained.

Computation Time

This metric is extremely important for real time use cases of the DOA estimators, since it determines how responsive the system would be to an evolving environment. I ran each algorithm for 20 times on the robot keeping the sampling frequency of the system fixed at 176.4 kHz and feeding the same window of length 1 ms of data to the algorithms. I kept track of the difference in the time stamp between the instant in which the signal is presented to the DOA estimator and the instant in which the pseudospectrum is available for further analysis. The resulting mean computation times are shown in Tab. 3.1. The collected data show that DAS has the highest computation speed among the three, unsurprisingly: the required operations are relatively simple with respect to the other two methods. Capon requires the inversion of a matrix with a large rank, and MUSIC employs an eigenvalue decomposition of the same matrix followed by a sorting of its eigenvalues and corresponding eigenvectors.

DOA estimator	Delay and Sum	Capon method	MUSIC
Computation time [ms]	68	145	162

Table 3.1: Mean execution times for the DOA estimation algorithms implemented

Spatial Resolution

The spatial resolution refers to the minimum angular separation that two sources must have to be identifiable as separate by the DOA estimator. The metric that I chose to give an estimate of the spatial resolution of the algorithms is the Half-Power Beam Width (HPBW), a concept taken from the theory of phased array antennas. The HPBW is defined as the angle between the two points of the main lobe of the pseudospectrum that have a magnitude inferior by 3 dB with respect to its maximum. Empty beer bottles of the NRW (Nordrhein-Westfalen) model (height: 260 mm, base diameter: 64 mm) were used as reflectors, due to their availability and ideal geometrical and material properties in relation to the robot's emission. One bottle was placed at 50 cm of distance from the robot, aligned with its longitudinal axis. The robot emitted its call, and the echo caused by the object was recorded by the microphone array and filtered with a matched filter. The output of the matched filter was windowed with the usual rectangular window of length 1 ms centred on the autocorrelation function representing the echo, and this array signal was then fed to the three DOA estimators. The resulting pseudospectra were interpolated with a cubic spline by a factor of 25, to have a more accurate estimate of the main lobe width, and then normalised with respect to their maximum. The two −3 dB points were then seeked in the magnitude of the pseudospectra. The interpolated pseudospectra are reported in Fig. 3.4.

DOA estimator	Delay and Sum	Capon method	MUSIC
Half-Power Beam Width [ ]	19	4	4.6

Table 3.2: Half-Power Beam Width of the DOA estimators implemented. The cylindrical obstacle was placed at 0 and at 50 cm of distance from the robot

The obtained HPBW values are reported in Tab. 3.2, and they were obtained by averaging five distinct detections. Once again the results are not surprising: DAS has the widest main lobe among the three algorithms, while Capon and MUSIC have comparable performances that exceed by a fair margin the one offered by DAS. This very simple experiment is not intended to be an exhaustive analysis of the resolution capabilities of the three algorithms, since the effect of only one parameter was investigated and just for one single value of it. Instead, it is meant to be an indication of the general trend in the capabilities of the algorithms once they would be deployed on a moving robotic platform: if multiple sources need to be located with a good degree of separation, then more refined algorithms can provide the desired performance. If instead the interest is focused on the presence of obstacles of whatever nature in a general direction, then also a simpler algorithm like DAS could be applied.

(a)

Figure 3.4: Pseudospectra obtained by applying the DOA estimation algorithms to the echo generated by a cylindrical obstacle placed at a distance of 50 cm (part 1).

(b)

(c)

Figure 3.4: (continued) Pseudospectra obtained by applying the DOA estimation algorithms to the echo generated by a cylindrical obstacle placed at a distance of 50 cm. The call employed is a linear descending sine sweep from 60 kHz to 20 kHz, windowed with a Tukey window with roll-off factor of 0.3.

Obstacle positions

This last verification step takes into account how the position of the obstacles with respect to the robot's axis plays a role in the accuracy of their localisation. To perform this test, two bottles were placed at a distance of 50 cm with an angle separation of 30, starting from 90 and spanning the entire frontal half-plane. The two tallest peaks were then selected in the obtained pseudospectra as an indication of the position of the two objects.

What was observed for all the algorithms is a progressive decrease in the localisation accuracy of both objects the farther they are positioned with respect to the robot's axis: this is caused by how the acoustic power emitted by the robot is distributed in the frontal hemisphere. As it will be shown in Chapter 4, the radiation pattern of the source employed is concentrated along its axis, so echoes are stronger if they are generated by obstacles close to the robot's longitudinal axis. The performance of the DOA estimators depends strongly on the SNR of the array signal: to a low signal power corresponds a worse localisation performance in terms of accuracy. A comparison of the pseudospectra obtained for various objects configurations is shown in Fig. 3.5.

(a) Objects positions: 0 , 30

(b) Objects positions: 30 , 60

Figure 3.5: Pseudospectra obtained by echoes coming from pairs of objects placed in asymmetrical configurations around the robot (part 1).

(c) Objects positions: 60 , 90

Figure 3.5: (continued) Pseudospectra obtained by echoes coming from pairs of objects placed in asymmetrical configurations around the robot.

The pair of objects placed in configuration 1 (0 , 30) are located with good accuracy by all three algorithms, despite the echo at 0 producing a much more prominent peak in the pseudospectra than the one coming from the side. In configuration 2 (30 , 60) the object closest to the front of the robot can still be recognised, but the obstacle placed at 60 is completely invisible to the robot. In the third and final configuration (60 , 90) the localisation performance is poor for both obstacles. The overall level of the pseudospectra obtained can also give an indication about the received echoes strength: comparing the pseudospectra in 3.5a with the ones in 3.5c, there is an excursion of about 20 dB between their maxima.

3.2. Final Remarks on the Processing Pipeline

In the previous Section the full signal processing chain that gives the robot the ability to echolocate objects around itself has been detailed, and in particular great attention was dedicated to the Direction of Arrival estimation algorithms that have been tested as the mechanisms with which the robot would identify the directions from which echoes arrive to the microphone array. Since the application for such a pipeline is on a mobile platform whose focus is a real time reactive behaviour, in the end the decision to favour reaction speed over detection accuracy was taken: DAS was the only algorithm which was implemented and tested while the robot was performing an autonomous navigation of a controlled environment. The other two algorithms, i.e. Capon and MUSIC, are still valid as a more accurate alternative, when the computational power of the platform they are run on is enough to reduce their execution time to an acceptable level.

4 | Robot Design and Implementation

This Chapter is dedicated to explaining the architecture of the robot that I built and programmed, the active sensing Ro-BAT. I both present the details of every single component and give an overview on how the system is intended to work as a whole. In Section 4.1, I go over the hardware components that constitute the robot, starting from its base configuration and detailing the added components for the intended functionalities. In Section 4.2, I detail the acoustic measurements carried out to evaluate key aspects of the transducer and microphones employed, such as their frequency response and directivity, and how these quantities can influence the robot's capability to sense the world. In Section 4.3, finally, I explain how the signal processing pipeline described in Chapter 3 is integrated into the robot's behaviour.

4.1. Hardware

The robot design process was strongly influenced by the decision to use consumer-grade hardware wherever possible. This approach has a number of advantages: first, costs and development time for the realisation of a single robot are decreased. Secondarily, scalability is also improved, since once a satisfying prototype is established it can be steadily replicated. These two features are well suited for a swarm robotics scenario. The selection process for the single components and the design stage to reach a final working prototype are explained. A picture of the final robot and a block diagram highlighting the main components and their connections is shown in Fig. 4.1.

4.1.1. Robot Base Configuration

The starting point for the development of my Ro-BAT consists of a Thymio II wireless robot [37] by Mobsya [36] equipped with a Raspberry Pi Model 4B single-board computer (SBC) [29], powered by an external power bank through the Pi's USB-C socket. It is a