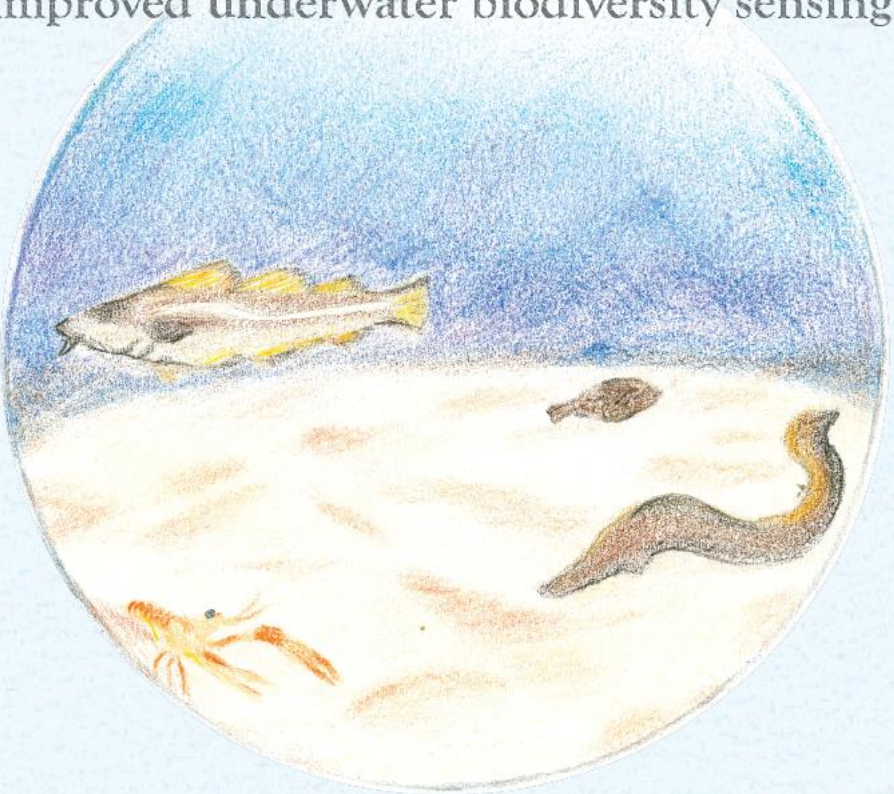
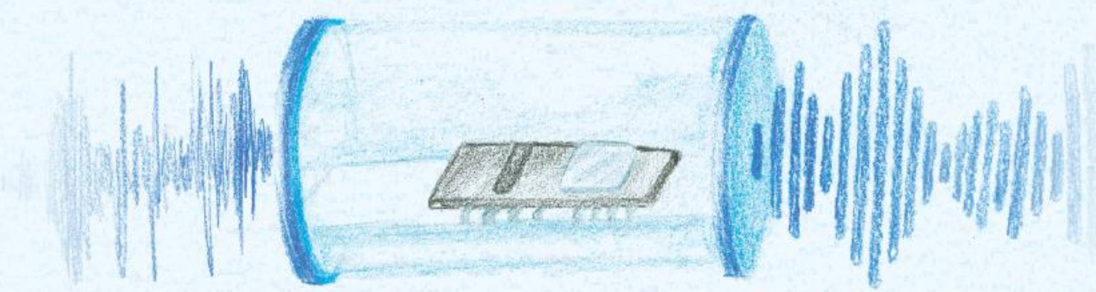
Revealing the sounds of the North Sea

An integrated passive acoustic monitoring system for improved underwater biodiversity sensing





Valentin Bordoux

Revealing the sounds of the North Sea

Valentin Bordoux

EngD Evaluation Committee

Primary supervisor
Prof. Dr T. Murk
Professor of Marine Animal Ecology
Wageningen University & Research

Other supervisors
Dr R. van der Ven, Wageningen University & Research
Dr R. Nijland, Wageningen University & Research

Other members
Dr D. Stowell, Tilburg University & Naturalis Biodiversity Centre Leiden
Dr JJ. Poos, Wageningen University & Research

Revealing the sounds of the North Sea

Valentin Bordoux

Thesis

Submitted in fulfilment of the requirements for the degree of Engineering Doctorate

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr C Kroeze,

in the presence of the

Evaluation Committee appointed by the Academic Board

to be defended in public

on Thursday 28 August 2025

at 2 p.m. in the Omnia Auditorium.

Valentin Bordoux
Revealing the sounds of the North Sea,
99 pages.

EngD thesis, Wageningen University, Wageningen, the Netherlands (2025)
With references, with summary in English
https://doi.org/10.18174/680244

Abstract

Marine biodiversity is increasingly threatened by human activities, necessitating robust monitoring to assess conservation and restoration efforts. Traditional methods, such as fisheries surveys and diver observations, are invasive, incomplete, and limited in scale. Deployable multi-sensor systems present a promising alternative for large-scale, long-term monitoring of marine animals, with passive acoustic monitoring (PAM) emerging as a particularly effective approach. Commonly applied to birds, PAM records animal sounds to monitor species presence. Since sound travels farther than light underwater and many marine species communicate acoustically, PAM is well-suited for marine environments. However, its application for fish and invertebrates remains restricted due to the lack of species-specific reference sounds and labour-intensive manual data analysis.

This EngD thesis aims to enhance the efficiency of PAM and support its integration into autonomous multi-sensor monitoring systems. The work is structured into four work packages (WPs): WP1 focuses on collecting underwater acoustic data for species identification and training deep learning models; WP2 involves designing adaptable sound detection models capable of performing across diverse environments with minimal training data; in WP3 an embedded, autonomous underwater recorder is developed with onboard processing capabilities; and WP4 introduces a dashboard for real-time visualisation and communication of acoustic data, supporting outreach activities such as stakeholder engagement and public participation.

Data were collected through field deployments, collaborations, and open-access repositories to train automatic sound detectors. An active learning workflow, Agile Modelling, was adapted to rapidly develop fish sound detectors with minimal manual effort, proving effective across various marine environments and sound types. In a

pilot study using North Sea recordings, a model trained for only one hour achieved a precision of 0.98 and a recall of 0.53 in detecting putative fish sounds. A prototype autonomous underwater sound and video recorder was also built and successfully deployed. A real-time dashboard was developed to display animal sound detections, using previously trained models. Further integration is needed in the audio-video recorder to enable onboard data processing, including support for deep learning models and connectivity to dashboards, facilitating use by biological researchers.

This work advances the feasibility of PAM for fish and invertebrates by addressing critical bottlenecks in data analysis and taking a step forward towards efficient collection of reference sounds. The developed tools and methods enhance marine biodiversity monitoring across broad spatial and temporal scales, supporting improved ecosystem assessments and conservation efforts.

Table of Contents

1 Introduction

1.1 Context

Marine biodiversity not only deserves protection for its own right, but it also plays crucial roles in the existence and sustainable development of human society, as recognised in the Sustainable Development Goals[1]. Marine biodiversity supports food security, generates local income through tourism, enhances water quality, strengthens coastal protection, sequesters carbon, and provides cultural services (Palumbi et al., 2009; Rodrigues et al., 2017). However, marine biodiversity is declining globally due to human activities such as destructive and over-fishing, shipping, urbanisation of coastal areas and subsequent pollution, and global warming (O'Hara et al., 2024). The study and protection of marine fauna received less conservation effort than terrestrial biodiversity, due to a lack of visibility and subsequently public awareness, historical research focus bias, and logistic challenges associated with surveying underwater (Caldwell et al., 2024). Enhancing public awareness and accurately monitoring marine ecosystems are essential for effective conservation and restoration efforts.

Unfortunately, monitoring marine environments presents significant challenges, primarily due to limited accessibility. Traditional methods, such as fisheries data collection and SCUBA-based visual surveys, are invasive, costly, dangerous, geographically restricted, incomplete, and difficult to scale (Suarez-Bregua et al., 2022). Developing new monitoring approaches is essential to overcoming these limitations. Sensor technologies offer promising solutions to monitor marine environments while minimising risks to researchers, reducing disturbances to marine ecosystems, providing access to remote or dangerous areas, and allowing for easy replication.

---

[1] https://sdgs.un.org/goals/goal14

Passive Acoustic Monitoring (PAM) involves listening and analysing ambient sounds, often using sound recorders, in various environments, including underwater. Sound travels about four times faster in water than air and typically farther than light, making it an effective communication medium for marine species. As a result, numerous marine organisms rely on sound for communication, navigation, predation, and reproduction (Erbe et al., 2016; Ladich, 2004; Solé et al., 2023). The study of these sounds is similar to the study of bird songs on land and is known as bioacoustics. Just as birds use vocalisations to communicate, marine species produce a wide range of sounds that provide valuable ecological insights (Mooney et al., 2020). Passive acoustic monitoring differs from active acoustic monitoring, which relies on sound emissions from tags or sonar. Instead, PAM uses microphones or hydrophones (underwater microphones) to passively capture ambient sounds, referred to as soundscapes. A soundscape is composed of multiple sounds that are typically categorised into three groups based on their source: biotic sounds (produced by non-human animals), abiotic sounds (resulting from environmental factors such as rain, waves, and wind), and anthropogenic sounds (made by human activities).

For analysis, sound recordings are often converted by acousticians into spectrograms, visual representations of the spectrum of frequencies varying with time, and then annotated (Figure 1). Annotations are indications of the presence of a sound event in a recording and appear as rectangular boxes on spectrograms; a label can be added to each annotation (Figure 1). When possible, labels correspond to the source producing the sound, e.g. a species name (Figure 1). Label naming is typically study-specific as no standardised practices exist yet.

By analysing biotic sounds with autonomous recorders, PAM enables minimally invasive biodiversity monitoring across broad temporal and spatial scales, including in dark and dangerous conditions. This approach has been widely applied to study various taxonomic groups, including birds, bats, cetaceans, and anurans. (Gibb et al., 2019; Mooney et al., 2020; Sugai et al., 2019). Yet, the application of PAM to monitor non-mammal marine animals, such as fish and invertebrates, remains in its early stages. Recent studies suggest PAM's untapped potential for monitoring fish in multiple contexts,

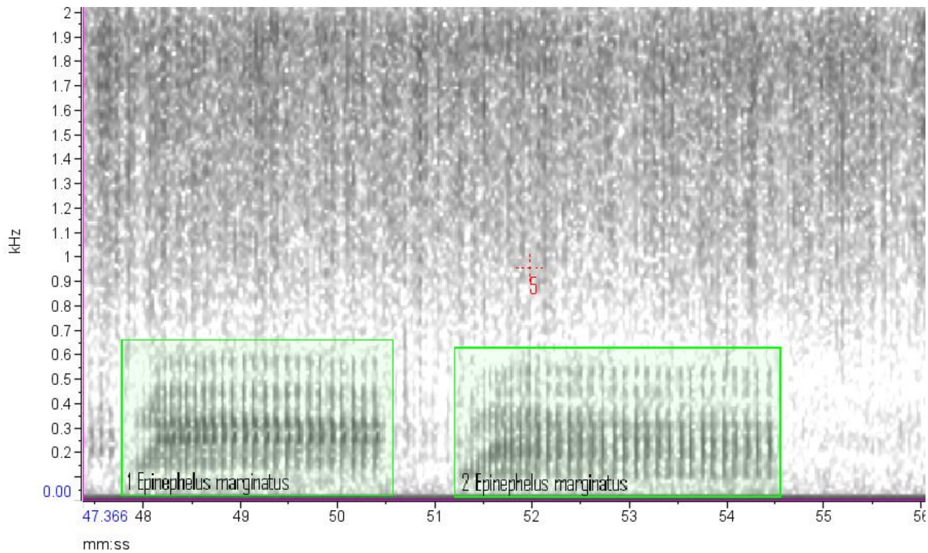


Figure 1: Example of a spectrogram using a greyscale colourmap including two annotations of sounds labelled as Epinephelus marginatus (dusky grouper)

including species presence assessment (Lindseth and Lobel, 2018), invasive species detection (Amorim et al., 2023), improved knowledge of red-listed species distribution (Bolgan et al., 2023), relative fish abundance in estuaries (Souza Jr et al., 2023), and the detection and characterisation of fish spawning sites (Chérubin et al., 2020; Wilson et al., 2019). A global inventory of underwater species based on their ability to produce sound suggests that PAM can effectively monitor a wide range of species (Figure 2). However, the effectiveness of PAM is currently limited by two major challenges: (i) the sheer volume of data generated by automatic recorders, which is impractical to analyse manually, and (ii) the scarcity of reference sound databases, which are essential for identifying sound sources.

(i) Data volume & analysis

Passive acoustic recordings can range from tens to thousands of hours, depending on deployment duration and the number of autonomous recorders used. Manually inspecting these recordings is time-consuming and therefore costly, limiting their use for long-term monitoring (Stowell, 2022). In terrestrial bioacoustics and marine mammal research, analysis has been successfully automated using

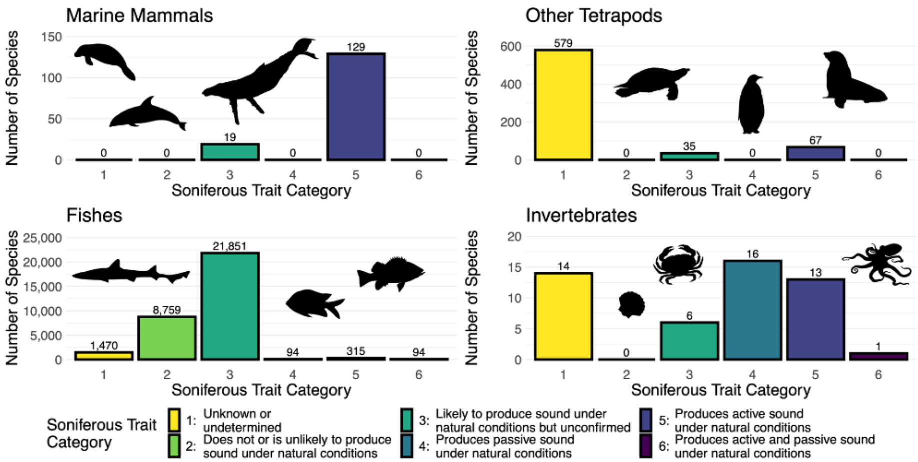


Figure 2: Inventory of underwater species based on known sonifery. For marine invertebrates, only the species studied for sound production are displayed. Sourced from Looby et al., (2023b).

supervised machine learning, for which detection models are trained on verified examples. More specifically, deep learning—neural network-based models—have outperformed traditional machine learning techniques in most bioacoustics applications (Stowell, 2022). However, supervised deep learning requires extensive amounts of annotated training data, particularly for applications across diverse environments and for different sound types. Currently, no large, annotated datasets exist for fish or marine invertebrate sounds, hindering the development of effective deep-learning models for automatic analysis. Although some fish sound detectors have been developed, these are limited to specific sounds or specific locations, reducing training data requirements, but limiting their applicability to other environments (Ibrahim et al., 2018; Laplante et al., 2022, 2021; Mouy et al., 2024; Urazghildiiev and Van Parijs, 2016; Waddell et al., 2021).

(ii) Limited reference sound databases

Passive acoustic monitoring for fish and marine invertebrates is impeded by the lack of reference sounds in reference databases. Phylogenetic analysis suggests that over 22,000 of the approximately 34,000 actinopterygian (ray-finned fish) species are known or presumed to produce sound (Looby et al., 2023a; Rice et al., 2022). However, a literature review found that active sound production has been directly studied in only about 4% of these fish species (Looby et al., 2022). Marine invertebrates also produce sounds (Solé et al., 2023), with reported active sound production in three groups: bivalves, echinoderms, and crustaceans. Among them, crustaceans have demonstrated evidence of using sound for communication (Solé et al., 2023). Sound production has been studied in approximately 50 marine invertebrate species, with 35 confirmed or likely to produce sound. Based on the 211,367 species listed in the World Register of Marine Species as of the date of writing, it is estimated that the count of marine invertebrate species largely exceeds 100,000 (Costello et al., 2013). Thus, significant knowledge gaps remain in the study of sound production in both fish and marine invertebrates.

Data collection initiatives emphasise the importance of gathering and classifying sounds from unidentified sources for future use (Looby et al., 2023a; Parsons et al., 2022). Sounds from unidentified sources can become valuable once linked to a reference sound and can be used to deduce the sound source species through cross-referencing (Vieira et al., 2024) or to estimate fish diversity and abundance in certain environments (Jarriel et al., 2024a). Recently, the efforts to fill reference databases have increased, including recording animals in aquaria (Almunia et al., 2024), localising sound sources with microphone arrays (Pyć et al., 2021), and integrating techniques such as portable audio-video arrays (Mouy et al., 2023).

The Marine Animal Ecology group at Wageningen University developed the Biodiversity Sensing Box (Figure 3), an autonomous monitoring system deployable in locations that are less applicable or accessible for other approaches. While currently focused on monitoring the North Sea, the system is adaptable for use elsewhere. Its modular design allows for easy replication with different sensor combinations, enhancing scalability and versatility. Currently, the Biodiversity Sensing Box integrates three monitoring techniques operating independently: sound and video recording, as well as environmental DNA (eDNA) collection (Yu et al., 2024). An autonomous sound recorder captures the soundscape continuously. Underwater cameras record video within a visibility range determined by water turbidity, allowing species identification under favourable conditions. Environmental DNA analysis detects species by comparing waterborne DNA fragments with a reference genome database. This method has been successfully applied to assess fish diversity, track temporal changes in fish communities, and monitor migration routes and has additional applications (Doorenspleet et al., 2025). However, despite rapid advancements, eDNA analysis still requires extensive laboratory work after sample retrieval and is susceptible to contamination in both field collection and analysis, limiting scalability and reliability.

Integrating eDNA, video, and acoustic sensing can improve

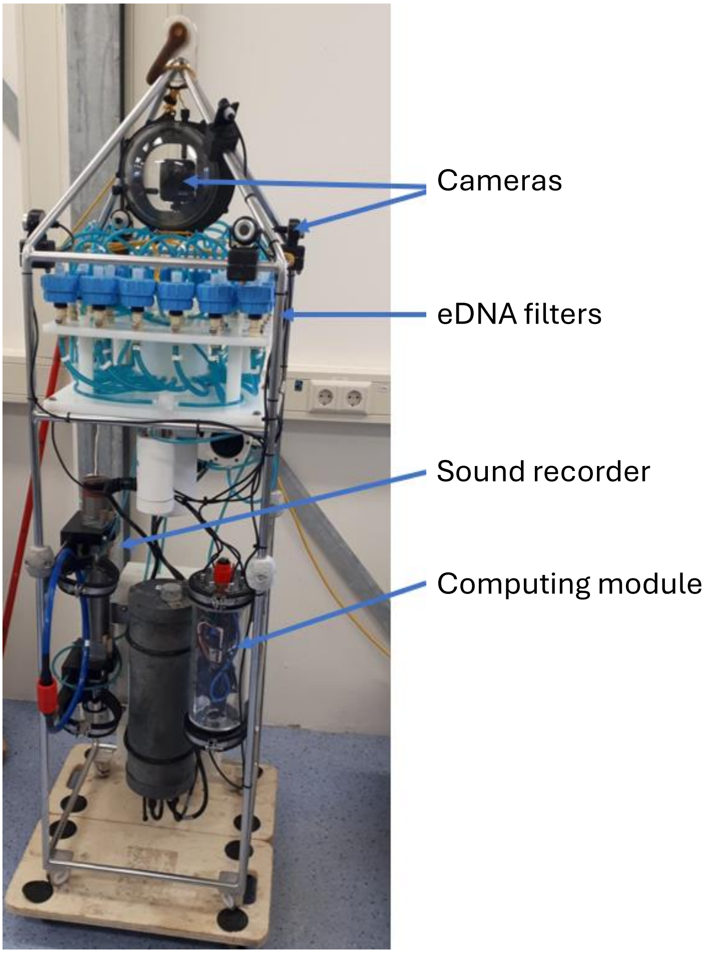


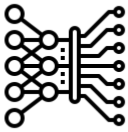
Figure 3: The Biodiversity Sensing Box

monitoring accuracy and coverage due to the complementarity of these techniques (Cabrito et al., 2024). In addition, combining these monitoring methods could enable synergistic applications, such as targeted eDNA sampling triggered by acoustic or visual detection of marine animals, thereby increasing the likelihood of capturing DNA fragments by confirming animal presence at the time of sampling. While the literature suggests combining eDNA and PAM in a single system, which would mutually benefit both techniques, such integration has yet to be implemented (Miksis-Olds and Watts, 2019). To achieve sound event triggered eDNA sampling, an onboard, real-time automatic analysis of acoustic signals is required. Such a smart sensor system, equipped with a computer running machine learning-based acoustic detection models, will improve monitoring capabilities by enabling real-time insights into local biodiversity and the synergistic integration of eDNA sampling, video recording, and bioacoustics monitoring.

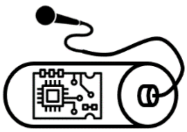
1.2 Aim and objectives of the project

This project aims to design and implement a novel, integrated PAM system designed for use within intelligent multi-sensor monitoring systems, such as the Biodiversity Sensing Box. The project will address key limitations currently hindering the broader application of PAM, specifically focusing on improving the accuracy, spatial and temporal coverage and system efficiency. To this end, a system capable of automated onboard processing and real-time communication will be developed. The project is structured around four interdependent work packages (WPs), each contributing to a specific aspect of system development and integration:

WP1: Collection of underwater bioacoustics data. Reference sounds are collected for species identification in recordings. In addition, datasets and annotations support the development of supervised deep-learning models. The data collection is designed for scalability and collaboration.



WP2: Development of machine learning bioacoustics detectors.
The models automatically detect sounds of interest in underwater recordings from files or data streamed by the hydrophone. The models provide, when possible, identification and confidence scoring. The method for developing detectors considers the lack of training data available and the possibility of being generalised to new sounds and environments.



WP3: Building an embedded system for real-time monitoring.
The embedded system includes a computer housed in a watertight enclosure and connected to a hydrophone for real-time data access and processing. The system runs machine learning models and has a dedicated power source.

WP4: Development of a graphical user interface for bioacoustic monitoring. A **dashboard (WP4a)** displays sound acquisition and automatic detection in real time and enables user interaction. Additionally, an **interface (WP4b)** facilitates rapid training and the use of machine-learning models for detecting sounds in underwater recording files.

Figure 4 shows a system analysis of the project, including the different work packages. The database supports the training of automatic detection and classification models. Trained models are integrated into the embedded system for real-time detection. Detections are displayed on the dashboard and can be combined with other monitoring techniques, e.g. triggering eDNA sampling (Figure 4). In addition to real-time onboard detection, models enable faster analysis of recordings compared to manual processing. The

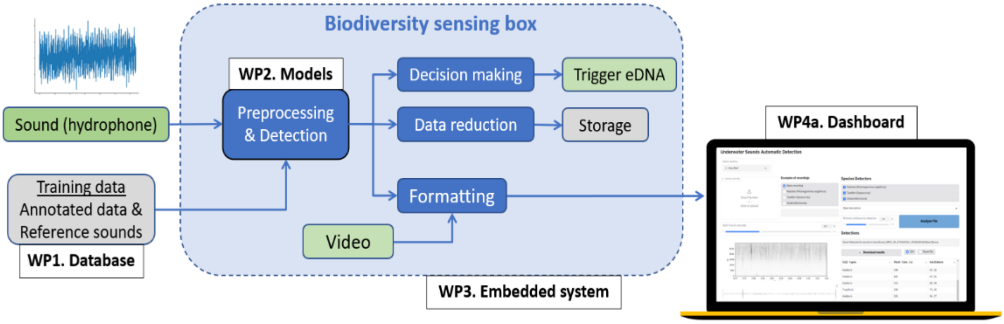


Figure 4: overview of the 4 interconnected work packages (WPs) to be developed in this project for developing an integrated PAM system for integration in a multi-sensor monitoring approach.

Interface (WP4b, not shown in Figure 4) allows stakeholders to use pre-trained models without programming, addressing common accessibility and reusability challenges in machine learning methods for bioacoustics (see Section 2.2 for details).

2 State-of-the-art and prior attempts

For each work package, an initial search for prior solutions was conducted at the beginning of the project, followed by regular updates to incorporate the latest developments. Analysing the state-of-the-art research is needed to build further on existing knowledge, identify key challenges, and select the most suitable, up-to-date methodologies. This section provides an overview of the relevant knowledge per WP, it does not aim to be an exhaustive literature review on bioacoustics monitoring approaches.

2.1 WP1. Availability of underwater bioacoustics data

The availability and accessibility of different types of underwater bioacoustics data were examined to develop a novel PAM system. Passive acoustic monitoring based on machine learning models requires two types of data to be most effective. First, *reference sounds* to identify the animals producing the recorded sounds. A recording featuring only the sounds of a specific species, when shared in an accessible database, is called a *reference sound*. Unidentified reference sounds can also be shared when the sound source is unknown, until the species name can replace a descriptor(examples provided in section 2.1.3). Second, datasets of recordings from the field, preferably with annotations. *Annotated recordings* contain labelled sound events, whereas recordings without annotations are referred to as *raw data*. If the sound source is unknown, descriptive labels are used until future identification and the data are then classified as *unidentified annotated data*, with the sounds termed *unidentified sounds*.

Annotations are essential for training deep learning models for automatic detection. For optimal performance, annotated data should be abundant and representative of the environment where the detector will be used. As the Biodiversity Sensing Box will primarily be deployed in the North Sea, this project will focus on recordings from the North Sea, while aiming to make the developed approach transferable to other regions in the future.

2.1.1 Existing reference sound databases

The current state of sound reference databases for fish and marine invertebrates was examined, as well as the infrastructure used in bird acoustics for comparison and inspiration. The field of bird acoustics is more established and developed than that of fish and marine invertebrates, providing valuable insights for future development.

The development and maintenance of reference databases is an active area of bioacoustics research. In a review listing existing underwater sound reference databases, (Jarriel et al., 2024b) found that of the 19 projects referenced, 14 focus on specific regions, such as the Ocean Biodiversity Listening Project in the western Pacific[2]The Australian Fish Chorus Catalogue[3], and the San Francisco Maritime National Park Historic Naval Sound archive[4](Jarriel et al., 2024b). One database, Animal Sound Archive[5] from The Museum für Naturkunde Berlin is dedicated to European regions, the database covers many taxonomic groups, including fish, but currently only contains recordings from two Actinopterygii species.

Five libraries have a global scope, e.g. FishSounds is dedicated to the collection of reference sounds of fish worldwide (Looby et al., 2023b). FishSounds.net is actively updated and compiles all reference fish sounds available in scientific literature and other sources (Cox et al., 2023). However, all reference databases only include a fraction of sound-producing species. For instance, of the more than 160 fish species reported to occur in the North Sea, 124 species are predicted

---

[2] https://sites.google.com/view/marine-ecoacoustics/projects/biodiversity-listening-project,
[3] https://doi.org/10.26198/qfj2-jj93
[4] https://maritime.org/sound/
[5] https://www.museumfuernaturkunde.berlin/en/science/animal-sound-archive

to actively produce sounds, while for only 16 species, recordings are available in reference databases (T. Maries, MSc thesis report, 2023).

Currently, there are no dedicated databases exclusively for marine invertebrates; instead, these species are included only in a few region-specific sound libraries. Each of the five databases containing marine invertebrate recordings features sound from only one to five species, limited to shrimp, sea urchins, crabs and lobsters.

In contrast, bird sound databases are already well-developed. The citizen-science platform Xeno-Canto enables global sharing and expert verification of bird recordings, currently covering 95% of all bird species (Vellinga and Planque, 2015). Recently, Xeno-Canto expanded to include sounds from grasshoppers, bats, and frogs. The database is widely used by researchers and companies to train machine learning models for biodiversity assessment. A key difference with FishSounds is the collaborative nature of Xeno-Canto, where citizens contribute recordings and species identifications, substantially accelerating data collection and expansion.

The Xeno-Canto approach, however, cannot be directly applied to underwater bioacoustics due to unique challenges. Recording underwater is constrained by the limited accessibility to the ecosystems, the need for specialised equipment (e.g., waterproof cameras, underwater sound recorders), and lower public awareness and engagement. Additionally, humans cannot perceive the direction of sound underwater, making species identification by visual confirmation unreliable. Due to these limitations, there is currently no large community recording and sharing underwater sounds. Currently, FishSounds does not permit user uploads; instead, administrators curate sounds exclusively from scientific publications. While this approach improves the reliability of species identification, it also slows down data sharing, including the sharing of unidentified recordings.

2.1.2 Availability of annotated underwater recordings

The use of annotated recordings is essential for training supervised deep-learning models. However, there is currently no preferred and widely adopted database dedicated to sharing large volumes of annotated underwater bioacoustics data, although several small initiatives exist. Jarriel et al. (2024b) catalogued thirty projects hosting underwater PAM data, detailing their status, data accessibility, and focus areas. The list includes repositories hosting raw or annotated datasets. Unfortunately, none of the reviewed databases contained datasets including annotations of fish sounds.

In addition to exploring data hosting initiatives, field studies that share their datasets, including annotations of fish sounds, were also examined. Unfortunately, as supported by a literature review of 100 publications concerning bioacoustics, in both terrestrial and marine studies, less than 20% of authors provide access to (part of) their raw data (Baker and Vincent, 2019). While several recent studies do discuss their manually annotated datasets of fish sounds, no access to the data is provided they do not provide access to the data (Monczak et al., 2019; Mouy et al., 2024; Picciulin et al., 2019; Watson et al., 2024). Personal exchanges with experienced researchers in underwater acoustics indicate that the substantial costs of field recording and manual data analysis discourage data sharing, favouring more restricted collaborations.

Sharing unidentified Annotated underwater recordings

Recent studies emphasise the ecological value of analysing unidentified sounds (Jarriel et al., 2024a; Mouy et al., 2024; Parcerisas et al., 2024). For example, the abundance of manually annotated unidentified sounds, assumed to be from fish, proved to be predictors for reef health indicators, such as fish abundance and coral cover, suggesting their potential use in assessing coral reef health (Jarriel et al., 2024a).

However, sharing unidentified sounds presents additional challenges due to the lack of dedicated infrastructure (Parsons et al., 2022). Moreover, inconsistent terminology in the literature, where descriptors such as "grunt," "pulse," "tonal," "knock," "boat whistle," "click," and "croak" are used interchangeably, hinders effective communication and comparison between studies (Looby et al., 2023a).

In the North Sea area, data-sharing efforts in underwater bioacoustics are primarily driven by individual initiatives rather than institution-wide databases. Recent studies that generated unidentified annotated data, including an overview of fish sounds in the Wadden Sea (Watson et al., 2024) and a soundscape analysis in the Belgian part of the North Sea (Parcerisas et al., 2023a, 2023b), have contributed to regional knowledge. However, these datasets remained inaccessible at that time, thereby limiting the possibility of benefiting from the latest research.

2.1.3 Availability of Raw Data Underwater Recordings

Raw acoustic data are more frequently shared, likely due to lower invested analytical costs. However, according to the latest inventories of openly accessible underwater acoustic data from the United Kingdom Acoustic Network[6] and a recent review (Jarriel et al., 2024b), no long-term acoustic datasets are currently available for the North Sea. Extending the geographical scope of the search, the SanctSound project, a three-year deployment of autonomous sound recording units across U.S. marine sanctuaries, has publicly released an extensive dataset (Hatch et al., 2024). Soundscapes recorded in the Grey Reef and Stellwagen Bank marine sanctuaries in the Northwest Atlantic are expected to share similarities with recordings

[6] List from the United Kingdom Acoustic Network: https://acoustics.ac.uk/open-access-underwater-acoustics-data/

from the North Sea and to contain fish sounds, although these remain unannotated.

2.2 WP2. State of the art of machine learning bioacoustics detectors

Existing initiatives, challenges, and potential solutions related to the automatic detection and classification of marine animals sounds were explored, with a particular focus on methods applicable to the North Sea region, when possible.

The advent of Deep Learning in Bioacoustics

Machine learning, signal processing, and data mining methods have been used for years to extract relevant ecological information from acoustic data faster. However, recent progress in deep learning, models based on deep neural networks, have demonstrated superior performance in bioacoustics applications compared to traditional machine learning methods (Mouy et al., 2024; Stowell, 2022). This improvement is largely attributed to the ability of deep learning models to identify complex patterns and representations from large datasets.

The application of deep learning in bioacoustics primarily relies on supervised learning, which necessitates substantial amounts of annotated training data. A key challenge in this domain is the generalisation issue, wherein models struggle to perform well on tasks or data distributions that differ from those encountered during training. Generalisation is particularly crucial in bioacoustics, as target sounds and environmental soundscapes exhibit significant variability, often leading to suboptimal model performance (Hamer et al., 2023; van Merriënboer et al., 2024). The extent to which a model generalises effectively is largely influenced by the quantity and diversity of its training data, yet obtaining or producing well-annotated datasets remains costly and time-consuming.

The most straightforward strategy to address generalisation is training models across diverse environments, increasing the likelihood of recognising familiar patterns in novel soundscapes. For instance, BirdNET, trained on extensive bird datasets from Xeno-Canto, has demonstrated reliable performance across multiple geographic locations (Kahl et al., 2021). In contrast, generalisation in aquatic bioacoustics remains particularly challenging due to the scarcity of annotated datasets covering diverse environments and sound types. The vast number of vocalising species and the complexity of underwater soundscapes further hinder the development of a globally applicable model.

Consequently, previous studies have largely focused on designing species-specific models tailored to particular locations (Guyot et al., 2021; Ibrahim et al., 2018; Waddell et al., 2021). These specialised models require only a reasonable amount of training data to achieve satisfactory performance. However, each new application in a different environment or for detecting a different sound requires additional annotated data and retraining of the model. This need for continual data collection and manual annotation poses a significant challenge in the widespread and quick adoption of bioacoustic deep learning models.

Navigating the fragmented landscape of computational bioacoustics tools

Deep learning methods have been successfully applied for efficient analysis of bioacoustics data. Numerous tools have been developed to facilitate model training and reuse (Napier et al., 2024). However, the proliferation of approaches has resulted in a fragmented landscape of independent initiatives, each with specific advantages but also contributing to the redundancy of development efforts (Darras et al., 2023). The number of existing tools hampers the identification, adoptability, and standardisation of suitable methods and tools, which becomes increasingly time-consuming and complex.

To mitigate these challenges, the website of the Global Library of Underwater Biological Sounds initiative hosts a catalogue of current sound-processing applications and their respective purposes for underwater bioacoustics, currently encompassing 92 references (Jarriel et al., 2024b). The resource needs to be continuously updated with the latest developments in automatic sound detectors for marine animals, e.g. FishSound-Finder (Mouy et al., 2024), Surfperch (Williams et al., 2024) and FADAR (Ibrahim et al., 2024). The list provides a centralised overview of existing tools for sound visualisation, signal processing, and automated detection. However, the short description provided for each tool is often insufficient to fully comprehend the capabilities of an application and determine its suitability for specific purposes compared to other tools.

Currently, 50 sound processing applications are reported to enable automatic sound classification or detection, according to Jarriel et al. (2024b). Most software does not allow for training custom models but instead provides pre-trained models for direct application to specific environments or species. Models trained on other taxonomic groups than fish or marine invertebrates will likely exhibit limited performance when transferred to such other applications or locations due to differences in sound types and background noise. Although some models have recently been developed for fish sound detection, their direct applicability also remains limited, as they were trained on specific species and within distinct environmental contexts (Ibrahim et al., 2024; Mouy et al., 2024; Williams et al., 2024). The most promising approach for application in the North Sea is to use software that allows training deep-learning models on personal data for dedicated applications[7](Bergler et al., 2022). Unfortunately, the

[7] https://arbimon.org/

limited availability of annotated fish and marine invertebrate sound data (as shown in Section 2.1) remains a major limitation for developing sound classification and detection models.

Machine learning methods in data-deficient context

Specific machine-learning methods have been developed to mitigate the challenge of limited annotated data. One of the most commonly applied methods in this context is transfer learning, where a model trained on one task is reused for another related task with minimal additional training, referred to as fine-tuning. Transfer learning has been shown to enhance model performance while significantly reducing the amount of training data required for various bioacoustic applications (Dufourq et al., 2022). It has proven effective in cross-taxa applications; for example, Ghani et al. demonstrated that knowledge gained from bird vocalisations improved model performance for other taxonomic groups such as frogs, bats or marine mammals (Ghani et al., 2023). A follow-up study successfully fine-tuned the bird detection model Perch to classify underwater sounds, including fish sounds, in tropical coral reefs (Williams et al., 2024). This promising approach has encouraged ongoing research into developing models capable of detecting and classifying sounds across multiple taxonomic groups and realms (Hagiwara, 2022; Nolasco et al., 2023; Robinson et al., 2024b, 2024a).

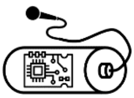
Beyond transfer learning, various other techniques can improve model performance while minimising the need for large training datasets. The following methods hold potential for bioacoustics applications:

- Simulation-to-reality (sim2real): Generates synthetic training data and enables models to learn from simulated environments (Li et al., 2020).
- Few-shot learning: Trains models using only a minimal number of annotated examples (Nolasco et al., 2023).

- Data augmentation: Expands training datasets artificially through transformations such as pitch shifting, noise addition, and time stretching, increasing data diversity (Park et al., 2019).
- Active learning: Involves a "human-in-the-loop" approach, where experts iteratively refine model predictions, reducing the need for extensive manual annotations (Qian et al., 2017).
- Self-supervised and semi-supervised learning: exploit partially annotated or even raw data to enhance model training, reducing dependence on fully annotated datasets (Baevski et al., 2020; Hagiwara, 2022; Moummad et al., 2023).

While many of these techniques have been successfully applied in terrestrial or marine mammal bioacoustics, their effectiveness in training models for other marine animal sound detection in temperate regions remains largely unexplored. Further research is needed to assess their potential in addressing data scarcity in this domain.

2.3 WP3. Components and prior solutions of an embedded system for real-time monitoring



This section examines the components and prior solutions for developing an autonomous sound recorder with real-time monitoring and onboard data processing capabilities.

Although a *hydrophone* is a sensor that converts underwater pressure levels into voltage, typically based on a piezoelectric component (Figure 5a), the term *hydrophone* is also commonly used to designate underwater sound recorders. Sound recorders are self-contained units comprising a pressure-sensing element, an analogue-to-digital converter, and a storage mechanism for the recorded signal (Figure 5b). When a recorder has its own power supply, it is called an autonomous recorder or autonomous recording unit. From this point on, a *(sound) recorder* refers to an autonomous sound recording unit, while a *hydrophone* is used to designate the pressure-sensing element without digitisation or storage capabilities.

Currently, the Biodiversity Sensing Box is equipped with a recorder, SoundTrap ST600 HF (Figure 5b) from Ocean Instruments (New Zealand). The SoundTrap ST600 is widely used in research due to its high autonomy, sensitivity, large frequency range, and robustness in a range of deployment conditions. SoundTrap ST600 sound recorder is contained in a watertight enclosure to withstand high pressure up to 50 bars (so up to 500m depth), therefore, the data from the SoundTrap ST600 recorder can only be accessed after deployment via retrieval of the SD cards.

a.

b.

Figure 5: a. Example of a hydrophone, A5 Hydrophone from Aquarian hydrophones (U.S.A). b. Example of a sound recorder, SoundTrap ST600 from National Instruments. The scale is different between the pictures.
Sources: https://www.aquarianaudio.com/a5-hydrophone.html; https://www.oceaninstruments.co.nz/product/soundtrap-st600-std-long-term-recorder/

To access and process the sound data in real-time, it therefore is necessary to create a custom autonomous recorder with onboard computing capabilities, including: a hydrophone, an analogue-to-digital signal converter, a computing unit capable of running deep-learning models, a power supply, and a watertight enclosure (Sousa-Lima et al., 2013). Two initiatives for self-made recorders were identified in the scientific literature. Caldas-Morgan et al. (2015) developed a custom sound recorder using a Raspberry Pi (a single-board computer), a self-made hydrophone, and a custom signal-conditioning board. Bagočius and Narščius (2021) alternatively integrated a professional audio digital recorder within a watertight enclosure, connected to a hydrophone, and successfully recorded sound at a depth of 50 meters.

For the desired application in the embedded system, a suitable hydrophone needed to be selected. Unfortunately, no recent review exists comparing the different hydrophones available on the market. There is, however, a personal blog describing a list of hydrophones, including some of their characteristics[8]. While not a verified and comprehensive source, the table gives a quick overview of the main hydrophone manufacturers and was used as an entry point into the topic for discussion with experts.

Following the selection of the hydrophone, the other technical choices can be made that will be described in the design process section of the WP3 (Section 3.3).

2.4 WP4. Design of a graphical user interface for bioacoustic monitoring

Graphical user interfaces enable system interaction and real-time functionality checks, facilitate the visualisation of data insights, including during deployments, and offer an engaging way to connect researchers, stakeholders, and the public with remote and unfamiliar environments.

Existing initiatives of sound analysis software that provide graphical user interfaces (GUIs) were first explored. In a general overview of

[8] https://zachpoff.com/resources/choosing-a-hydrophone-for-field-recording/

existing tools for sound analysis, Jarriel et al. (2024b) list 23 tools for data exploration (including listening, editing, and producing spectrograms) that offer GUIs. In addition to data exploration, these tools include additional functionalities for various stages of the bioacoustics workflow, such as tools for manual data annotation (e.g., Audacity, Raven Lite, Sonic Visualizer), automated detection and classification of acoustic events (e.g., Koe Bioacoustics, Luscinia, WASIS), or multiple functionalities (e.g., Avisoft-SASLab Pro, PAMGuard, RavenPro). Some initiatives are also developed to be used for sound analysis of specific taxonomic groups, such as AviaNZ, BatExplorer, Robots4Whales, or Sonobat.

Among the existing tools, RavenPro or PAMGuard already provide the possibility to visualise sound data in real time. However, adding new functionalities to such existing software can be challenging, time-consuming, and limited to open-source applications. In the future, the aim is to integrate video functionality into the dashboard and incorporate additional data from sensors that will be included in the Biodiversity Sensing Box. Creating a custom dashboard allows the provisioning of personalised data flow and different functionalities for user interaction, e.g. displaying highlights of automatic detection of unusual events. Given the specificity of the requirements of this project, programming libraries for the development of GUIs were investigated to create a custom solution.

Python-based initiatives were prioritised for compatibility with the solution chosen in the development of the embedded system (section 3.3). Having experience with the programming language also helps reduce development time. Prominent open-source libraries for GUI development include PyQt[9], Tkinter[10], Kivy[11], and WxPython[12]. PyQt is

[9] https://doc.qt.io/qtforpython-6/
[10] https://docs.python.org/3/library/tkinter.html
[11] https://kivy.org/
[12] https://wxpython.org/

a well-established, continuously maintained, open-source library that is supported by extensive documentation and an active developer community, facilitating efficient development.

More recently, a compelling alternative has emerged, Gradio[13], a library specifically designed for finding prebuilt components to create interfaces for machine learning applications. The library also features an integrated web server to share the new application with users and supports integration in Hugging Face, a widely used platform for sharing machine learning models and applications. Thus, enabling the creation of both deployable, dedicated interfaces and accessible, permanent web applications. Gradio has been used in prominent projects, including BirdNET Analyzer (Kahl et al., 2021), an open-source application for bird detection that shares several functionalities with the objectives of WP4, such as using a pre-trained model to analyse recordings, displaying recordings, and extracting results of automatic analysis of recordings.

[13] https://www.gradio.app/

3 Methodology

The methodology chapter presents, for each work package and based on the research of prior solutions in the previous chapter, the design process, the current solution and the main challenges encountered.

3.1 WP1. Collection of underwater bioacoustics data

This section focuses on the collection of annotated data for developing deep-learning models and reference sounds for species identification in recordings (Figure 6).

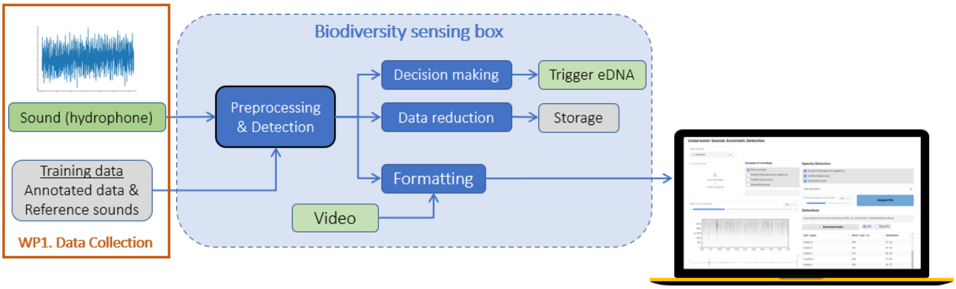


Figure 6: Overview of the development of an integrated passive acoustic monitoring system, for onboard processing of sound in a multi-sensor system, the Biodiversity Sensing Box, with a focus on work package 1, the data collection.

3.1.1 Design Process of the underwater bioacoustics data collection

Data were sourced from initiatives allowing the sharing and reusing of their datasets and reference sounds. These were then aggregated into a local repository for model development. The aim was to gather underwater sound recordings, preferably annotated and preferably from the North Sea. To achieve this, data were collected through openly accessible repositories, prior scientific studies, data-sharing collaborations, and in situ hydrophone deployments.

Search in repositories and prior studies

The SanctSound project (Hatch et al., 2024) served as an initial source of long-duration recordings containing marine animal sounds, given the lack of North Sea recordings in repositories and prior studies at that time. The SanctSound cloud repository hosts thousands of hours of raw audio data from multiple locations, with three sound recorders per site. These data were used to get familiarised with acoustic datasets, including sound visualisation and manual annotations. While annotation methods have been described in previous studies, practical experience is essential for improving efficiency and accuracy. The annotated data provided valuable hands-on experience in processing, analysis, and machine learning applications. Ultimately, SanctSound served as a preliminary resource for skill development and was incorporated into a local data repository to support model training.

Reference sounds were obtained from the comprehensive FishSounds.net database. Although the platform does not yet offer a simple querying mechanism for its reference database, it allows for the manual download of reference sound species by species.

Collaboration and scientific meetings

To facilitate data sharing and strengthen collaborations, contact with experts in the field of bioacoustics through various scientific events was initiated, including the ARISE Day at Naturalis, the annual Bioacoustics day, the 5th World Ecoacoustic Congress, and the interdisciplinary art & science project Voices of the North Sea. These events provided opportunities to directly and indirectly establish connections, exchange knowledge, and explore potential data-sharing agreements.

A key outcome was the formation of a temporary working group on underwater bioacoustics during the Bioacoustics Day 2023. Formed in collaborations with researchers from the Vlaams Instituut voor de Zee (VLIZ), the University of Groningen, and the Norwegian Institute of Marine Research, the group aims to address challenges in sharing underwater sound data. Its primary focus is on standardising the classification of unidentified fish sounds based on measurable acoustic characteristics and improving data-sharing infrastructure. To advance these goals, two dedicated workshops were held. An additional agreement was settled with VLIZ, granting access to a subset of their data in exchange for involvement in model development, analytical insights, and annotations.

Unfortunately, some meetings did not result in collaborations due to differences in research priorities, time and budget constraints, or hesitancy toward data sharing. Nonetheless, these scientific meetings played a crucial role in identifying collaboration opportunities, understanding challenges in data sharing, and fostering discussions on standardisation in underwater bioacoustics.

Sound recorder deployments

To collect underwater recordings of the soundscape in North Sea environments, multiple deployments were conducted at various locations, using the SoundTrap ST600 recorder alone or connected to the Biodiversity Sensing Box. Several challenges were encountered, as the recording of marine animal sounds underwater remains a complex field with a lack of standardised protocols. Insights gained from these experiences were used to enhance data collection through improvements in both software and hardware usage (Table 1). Unfortunately, deployments are costly, particularly offshore, limiting the number of attempts.

Table 1: Overview of experiences from deployments of the underwater sound recorder

Deployments	Status	Issues	Learnings
Artificial reef	Data was accidentally deleted during extraction	Mistake in software use	Reading the documentation more carefully
Artificial wave breakers blocks	Only 30-second file recordings	Software misfunctioning	Temporary fix with functional settings Developer contacted for permanent solution
Wind farm	Boat motor masks all the sounds	Sound recorder positioned under the boat	The recorder should be deployed further from the boat, or the boat engine turned off.
Bruine Bank	Corrupted SD card, data inaccessible	Incompatible hardware, only tried on a short test	Only use SD cards recommended by the manufacturer.
Bruine Bank 2	Acoustic recordings from a 360 camera	Limited duration due to the camera's autonomy	Video soundtrack can also be used to gather recordings (uncalibrated, relatively short battery life, so short recording duration)
Haringvliet dam	Masking noise from waves and buoy	Sound recorder positioning	Sound recorders should be deep (>1m) and not attached to a surface buoy if there are a lot of waves
Texel-Oudeschild	Successful 1-week recording		Harbour can be suitable for recordings. Hanging from the floating structure set up, not too many boats
Side projects			
Coral reef in Kenya	Successful day-long recordings		Annotation is difficult in an environment with a high sound event density
Costal area in Krk (Croatia)	Successful months-long recordings		The underwater buoy and bottom-anchored setup were validated

3.1.2 Current status of the underwater bioacoustics data collection

A local data repository has been established to aggregate datasets, annotations, metadata, and reference sounds, serving as the central resource for model development. This repository is designed to be scalable and shared in the future, facilitating the collection, storage, and reuse of data.

Given the infrastructure required to host large datasets of recordings (and annotations), existing International and National initiatives for data hosting and sharing were explored. The most promising solutions currently include the WorldWide Soundscape Project (Darras et al., 2024), the OPUS portal[14] by the International Quiet Ocean Experiment working group (Boyd et al., 2011), and the Global Library of Underwater Biological Sounds (Parsons et al., 2024). As these initiatives remain under development, their progress is being actively monitored, and opportunities to contribute the collected data have been explored.

Reference fish sounds are sourced from FishSounds.net (Cox et al., 2023), though identification of newly recorded fish sounds remains dependent on expert acousticians and the availability of verified reference datasets. New fish sounds yielded by the project can be added to FishSounds.net as well.

For the exploration and annotation of recorded data, Raven Pro 1.6 was used, a widely adopted bioacoustics software developed by the Cornell Lab of Ornithology.

During the Bioacoustics Day 2023, an initiative was launched to create a shared database for unidentified underwater sounds. Hosted within the EUDAT Collaborative Data Infrastructure[15], this database facilitates data exchange by categorising sounds based on measurable acoustic characteristics. To ensure consistency in

---

[14] https://opus.aq/index.html
[15] https://b2drop.eudat.eu/

naming unidentified fish sounds, a decision tree was developed by a temporary working group during a dedicated workshop. The outcomes of this initiative will be published in an upcoming paper, currently in preparation.

Given the success of several collaborative efforts, this will be further followed up by engaging in scientific discussions and participating in region-specific events and mutual site visits. Collaborations are crucial in overcoming challenges related to data collection, data sharing and standardisation in underwater bioacoustics.

Building on previous deployment experiences, the collection of additional underwater recordings is planned across various North Sea environments. Optimal placement strategies involve positioning recorders above the seafloor or attaching them to floating structures, depending on environmental conditions. To reduce interference from anthropogenic noise, locations with minimal human activity will be prioritised unless anthropogenic noise is to be qualified and quantified as well. Promising sites include offshore shipwrecks and artificial structures, which provide biodiverse, sheltered environments well-suited for bioacoustic monitoring.

3.1.3 Design challenges of the underwater bioacoustics data collection

Establishing new collaborations is time-consuming and takes time, especially when it involves building a network in a field unfamiliar to the research group. Still, collaborations are crucial in bioacoustics, given the scarcity of available resources and the cost involved in obtaining recordings, particularly in marine environments. Dependency on data to share, that were lacking at the start of the project, posed an additional difficulty in successfully starting up collaborations.

Also, generating annotated datasets proved to be difficult. As demonstrated in a study on marine mammal sounds, annotating underwater bioacoustic data is a complex process prone to inconsistency and subjectivity (Nguyen Hong Duc et al., 2021). The decision to annotate a sound often requires interpretation, particularly when the signal is faint, barely visible, or similar to another sound. As a result, the annotation process may have weak reproducibility.

Additionally, manual annotation is time-intensive and resource-demanding, making it impractical and costly to have multiple experts validate each annotation. While inconsistencies can be partially mitigated by establishing clear annotation guidelines, as proposed by Parcerisas et al. (2023b), challenges remain. For example, a common approach is to require the annotator to be able to both see and hear the sound event for it to be annotated positively. However, visualisation depends on specific parameters used for the spectrogram computation and display, and sound detection ability varies between annotators, further complicating standardisation.

Currently, there is no consensus within the scientific community on standardised annotation methods for bioacoustics, as approaches must be tailored to different recording conditions and research objectives. Further research and collaboration is needed to establish standardised, reproducible, and comparable annotation procedures in underwater bioacoustics.

3.2 WP2. Development of machine learning bioacoustics detectors

The aim of WP2 was to develop a model capable of automatic sound detection across diverse environments and sound types in the North Sea (Figure 7).

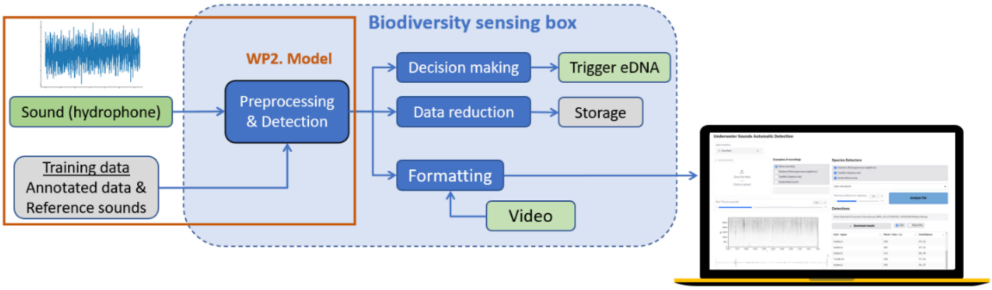


Figure 7: Overview of the development of an integrated passive acoustic monitoring system, for onboard processing of sound in a multi-sensor system, the Biodiversity Sensing Box, with a focus on the work package 2, the machine learning model for automatic sound detection.

An overview of existing methods (see Section 2.2) revealed numerous potentially suitable approaches, each with distinct advantages. But the fundamental challenge in automatic marine animal sound detection remains the scarcity of annotated data required to train models. To address this limited availability of environment- and species-specific annotated datasets, the most recent machine-learning techniques designed for data-scarce scenarios were explored.

3.2.1 Design Process of machine learning bioacoustics detectors

Initially, it was hypothesised that artificially generating more training data could compensate for the lack of annotated data, and the sim2real approach was considered. To increase the amount of annotated data available, various data augmentation techniques were also explored (Figure 8; Hwang et al., 2020; Park et al., 2019). Given the substantial time and effort required to implement and validate ideas, especially with limited experience in machine learning, advice was sought from researchers experienced in applying machine learning to bioacoustics before conducting any experiments.

Experienced data scientists stressed that high-quality training data are irreplaceable for model training. While generating annotated data can enhance performance, these improvements are generally minor. Artificially generated data often fail to fully represent real-world data, leading to limited performance in practical applications.

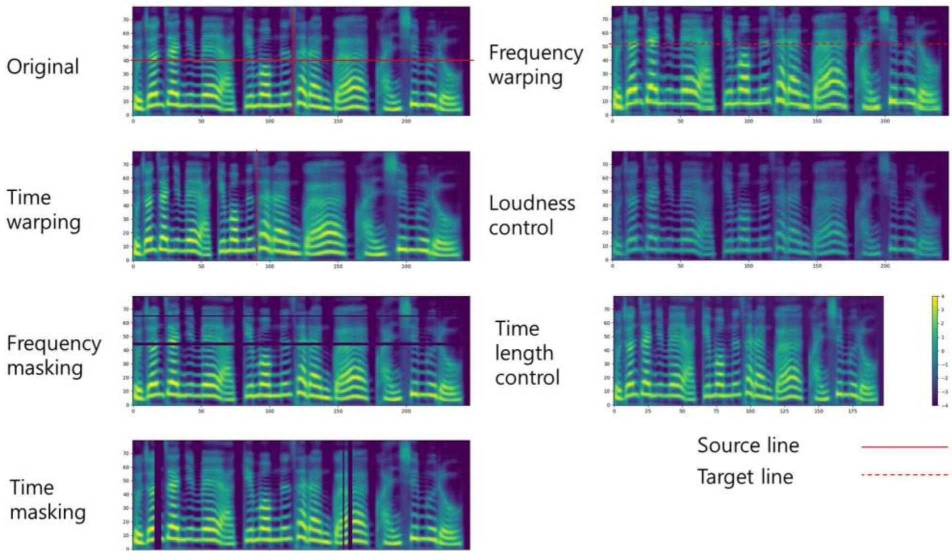


Figure 8: Examples of the effect of different data augmentation techniques on a mel-spectrogram. On the left, the original mel-spectrogram and augmentation techniques proposed in SpecAugment (Park et al., 2019); on the right, augmentations techniques proposed in (Hwang et al., 2020), from which the image is sourced as well.

Published models that exhibit state-of-the-art performance on test datasets sometimes remain largely ineffective on real-world data, due to specificities not captured in curated or limited training data (Schall et al., 2024). Experts' recommendations were to focus on acquiring real data to fine-tune large models, which are pre-trained with very large datasets containing multiple environments and types of sound data and extensive computational power.

At the start of the project, transfer learning had not been applied to fish sounds, as this taxonomic group is understudied in bioacoustics. Using the AVES model (Hagiwara, 2022), experiments were done using transfer learning on fish sounds from the Grey Reef environment, sourced from the SanctSound project. The experiment demonstrated that transfer learning could be effective for detecting some fish sounds. However, performance dropped significantly when focusing on the detection of types of sounds containing only a few examples in the training set. Additionally, preliminary tests to detect similar sounds in different locations showed poor performance. Extracts from the results using AVES and a convolutional neural network to classify fish sounds are provided in Appendix A.

Studies on bird sound detection and some pilot experiments suggest that automatic detection across different locations and target sounds is feasible if the training data adequately represent the model's operating conditions, including similar calls and background noise. This implies that models must be trained with annotations specific to both the call type and the environment to achieve high performance. However, the choice of machine learning method impacts the required number of annotations, directly influencing the amount of manual effort needed.

Few-shot learning enables models to perform automatic detection or classification using only a limited number of examples. In bioacoustics, researchers have adapted this paradigm from computer vision approaches to detect animal sounds with as few as five examples (Nolasco et al., 2023). Figure 9 illustrates the few-shot learning workflow for bioacoustic detection. Few-shot learning for bioacoustics has the potential to facilitate sound detection with

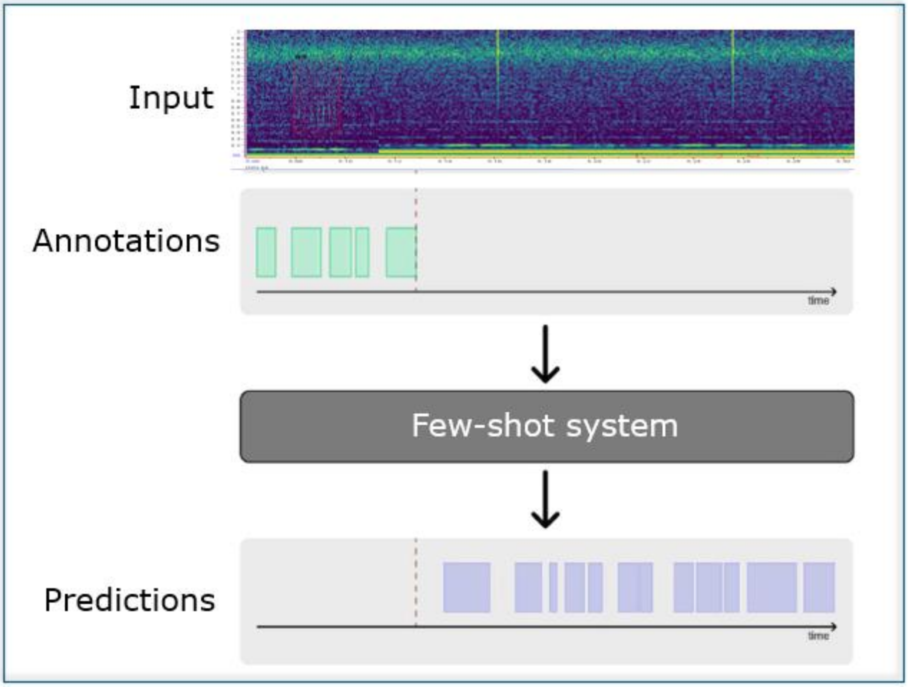


Figure 9: Few-shot sound event detection. The first 5 sound events are given as examples, in standard supervised learning they would be considered the training set, and the remaining sounds must then be detected. Adapted from (Nolasco et al., 2023).

minimal annotation effort, but there is no consensus yet on the best-performing approach (Morfi et al., 2021). A task dedicated to the development of few-shot learning approaches for various animal sounds was introduced in the yearly Detection and Classification of Acoustic Scenes and Events (DCASE) challenge (Mesaros et al., 2017). While every year the challenge includes sounds from diverse taxonomic groups, it has never featured fish or marine invertebrate sounds.

To assess the applicability of few-shot learning for fish sounds, a transfer learning approach based on the AVES model was developed and applied in the 2024 edition of the DCASE challenge, task 5 (Liang et al., 2024). The results were promising, demonstrating that few-shot learning could accelerate the discovery of new sound occurrences by a factor of 2 to 8, depending on the location and associated challenges. However, the detector's recall score, measuring the proportion of detected sounds, often fell below 50%, meaning that more than half of the calls in the recordings were missed with this automated approach. While this approach could speed up the collection of annotated data, its performance was insufficient for use as a reliable detector in most monitoring contexts. Furthermore, each deployment in a new environment or for a different sound type would still require a time-consuming cycle of few-shot detection followed by manual verification to train high-quality detectors.

Few-shot learning showed promising results in detecting rare sound events faster, but with limited precision and recall. Therefore, developing a reliable detector or multiple detectors suitable for various locations in the North Sea would remain prohibitively time-consuming due to the verification step needed. During the project, an alternative approach called Agile Modeling was published (Williams et al., 2024). This method employs an active learning (human-in-the-loop) strategy to quickly train reliable detectors for conditions where training data are limited. The method was considered promising for the purpose of this project, with the potential to improve performance, reduce manual effort and be adaptable to any environment.

3.2.2 Current solution: Agile Modelling

Agile Modeling, further mentioned as Agile Modelling, combines transfer learning and active learning and was first developed for bird acoustics in December 2023[16]. The approach was later adapted to detect biotic and anthropogenic sounds in multiple coral reef locations (Williams et al., 2024). Agile Modelling can quickly train models for the classification of specific types of sound in a given environment, with an iterative loop involving user feedback (Figure 10). While the method applies to training multiclass classifiers, the focus here was on binary classification, distinguishing one target sound (positive class) from all other sounds and background noise (negative class). While this method can be used for training multiclass classifiers, where the goal is to predict the correct class among multiple ones, this work focuses on binary classification, identifying one target sound (positive class) against all other sounds and background noise (negative class). The approach is reproducible with different target sounds, enabling detection of various sound types. Given the expected low diversity and abundance of biotic sounds in North Sea environments, the primary challenge is not sound classification but the detection of sound events within potentially long recordings. From this point, binary classifiers trained using Agile Modelling will now be referred to as detectors.

[16] https://nips.cc/virtual/2023/76893

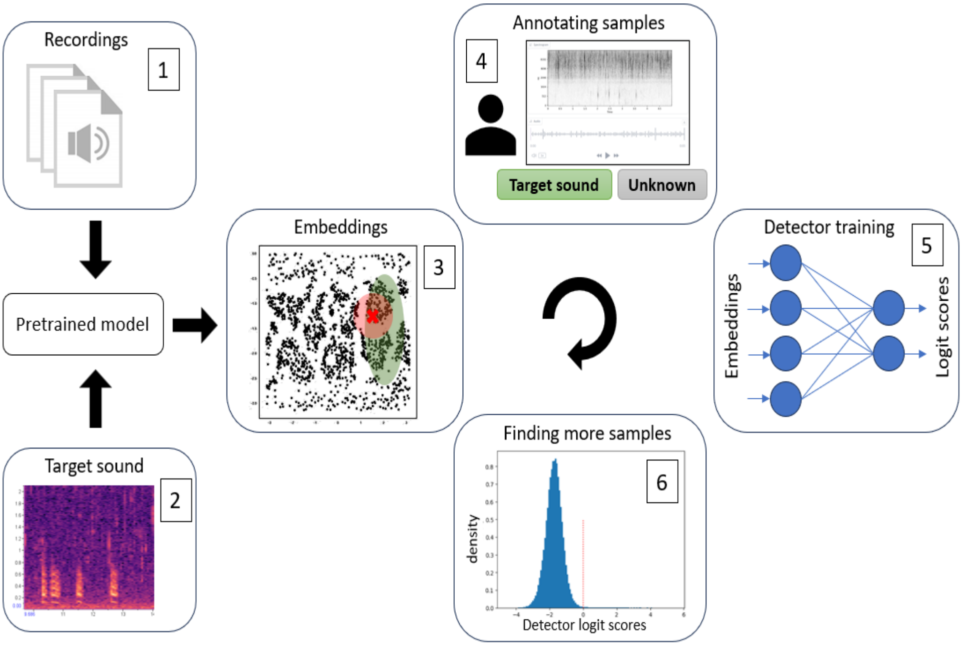


Figure 10: Agile Modeling pipeline. 1. Recordings are cut in 5-second windows and projected in a high-dimensional embedding space by a pre-trained model (Surfperch ;Williams et al., 2024). 2. The target sound is projected in the embedding space (red cross). 3. The closest samples to the target sound in the embedding space (red circle area) are shown to the user for annotation. 4. After annotating some samples, a detector constituted of a single-layer neural network can be trained (the dimension of the input layer is reduced to four nodes for visualisation). 5. The detector attributes a prediction score to all points of the embedding space. 6. The user can request more samples around the logit score of their choice to annotate. The process is iterated to increase the number of annotated examples therefore improving the detector until reaching satisfactory performance. The final goal is to train a detector able to separate the target sound from the unknown sound samples in the embedding space (green area in 3).

The method can be divided into 6 steps as presented in Figure 10:

1. Recordings from a given environment are divided into 5-second segments, which is the sample size required for the pre-trained model, currently SurfPerch (Williams et al., 2024). The pre-trained model generates a mathematical representation (high-dimensional vector) of each sample, referred to as embeddings.

2. The user provides one example of the target sound, which is also converted into an embedding (red cross in step 3).
3. A similarity search is first used to select samples resembling the target sound in the embedding space (red area in step 3).
4. Selected samples are suggested to the user for annotating as “target sound” or “Unknown” (sound).
5. Annotated data are used to train a one-layer linear binary classifier, the detector. The goal of this approach is to train the detector to separate the target sounds from the unknown sounds in the embedding space (green area on step 3, assuming all target sound samples are within the green area, which is a simplified representation).
6. The user can request more samples around a specific logit score given to them by the detector (red line on step 6). The score can be interpreted as a confidence level: if the score is high, the detector is confident that it is a target sound (which is not necessarily true), and a very low score indicates confidence that the sample is not the target sound. The user can continue to provide more annotated data to improve the detector by repeating steps 4, 5 and 6 until satisfactory performance (see Section 3.2.3 for details on performance measures).

After training, the model can be used to inspect all recordings for target sound occurrences and create a table of detections, with indication of confidence, filename and starting time of each detection.

Agile Modelling enables users to train a detector from a single example of a target sound, utilising already trained multi-species models through transfer learning. The model is iteratively improved as the user annotates additional samples selected based on confidence scores from the current version of the detector, following an active learning approach. This process reduces the amount of training data required and facilitates finding additional target sounds, an especially time-consuming task when target sounds are rare, thereby accelerating the development of deep learning-based detectors.

3.2.3 Pilot Experiment on Agile Modelling

To confirm the applicability of detectors created using Agile Modelling, the performance of a model was evaluated on data from a deployment in the Harbour of Texel, Oudeschild.

During the active learning process, the model iteratively assessed its performance using a subset of the user-provided annotations. A training and validation split was generated, with the detector trained on the former and evaluated on the latter. The original protocol recommends stopping the iterative training once a predefined performance threshold on the validation set is reached, with the threshold value based on prior experiments. However, there is a risk that performance metrics derived solely from the validation set do not reliably represent the quality of the detector. Preliminary experiments indicated that the limited number of samples in early iterations tended to result in an overestimation of detector performance. In addition, the evaluation, done on samples annotated by the user but chosen based on the results of the model, may introduce a biased representation of the selected sounds from the overall dataset. Therefore, a subsequent pilot experiment was conducted to assess the reliability of measuring performance on the validation set and to evaluate both the performance of a detector trained with Agile Modelling as well as the associated time efficiency in finding target sound occurrences compared to manual annotation.

A dataset from a one-week recording from the harbour of Texel-Oudeschild was used, divided into a training pool and a test set (Figure 11). The test set was manually annotated by visual and aural inspection to be used as ground truth for evaluating the performance of the model. Due to the time-consuming nature of annotation, the test set was limited to 6% of the total dataset, representing 10 hours of recordings, randomly sampled in 60 ten-minute periods across the dataset. The active learning training was conducted on the remaining 94% of the data, called the training pool, consisting of 156 hours of recording.

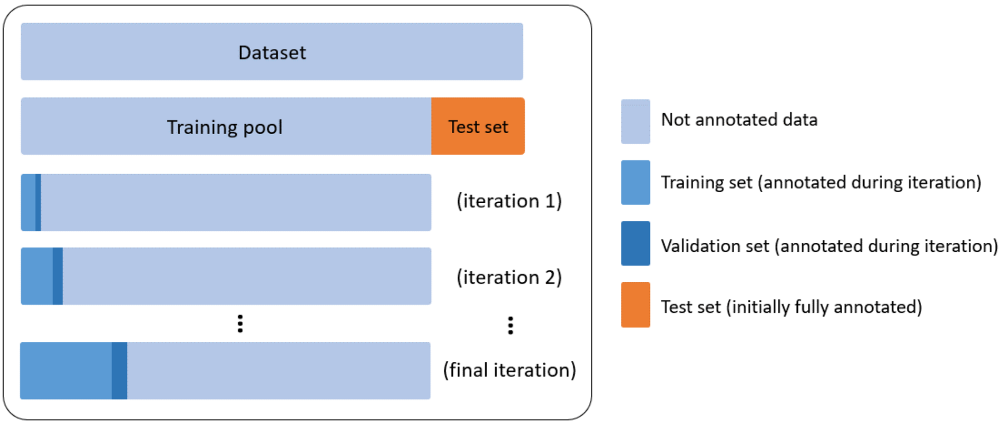


Figure 11: Data splits for evaluation and during the iterative training. To facilitate visualisation, the size of the splits is not proportional to their real size. The test set is 6% of the dataset, fully manually annotated and only used for evaluation of the model at each iteration. The rest of the dataset, the training pool, is used for training a detector using Agile Modelling. The training set and validation set are samples annotated by the user during the iterative process. The training\validation split size is 80\20. The user can decide when to stop annotating new samples.

Appendix B provides details on the performance metrics, experimental protocol, and parameter settings used. In each Agile Modelling iteration, newly annotated samples were added to progressively train and assess the detector (Figure 11). Performance was evaluated using Precision, Recall, F1-score, and AUC-ROC. Precision evaluates the proportion of correct detections, and Recall measures the proportion of target sounds missed by the detector. The F1-score balances Precision and Recall, offering a single measure that accounts for both false positives and false negatives. The previous metrics depend on the minimal confidence score value for which the model considers an example positive or negative, usually referred to as the decision threshold. The best decision threshold is generally specific to the purpose of the detector and the specific task. The AUC-ROC score reflects the model's ability to distinguish between classes across all decision thresholds, making it independent of any specific decision cutoff. To ensure a reliable assessment, results were averaged over a 5-fold cross-validation process, where the data was divided into five parts, each used once as a validation set while the others trained the model (Figure 1 of Appendix B).

As anticipated, initial iterations revealed an overestimation of detector performance on the validation set across all evaluated metrics compared to the test set (Figure 12). However, with increasing annotated samples, performance metrics on both datasets demonstrated a trend toward convergence. The AUC-ROC exhibited the most consistent values between the validation and test sets (Figure 12D). Despite the general trend of increasing test set performance across most metrics (except the Recall) throughout the iterations, validation set performance mainly declined (Figure 12C). This divergence underscores the unreliability of validation set metrics as accurate indicators of the detector's true quality. Consequently, employing the validation set AUC-ROC score as a training termination criterion, as proposed by the original protocol (achieving a threshold of 0.95), may lead to misleading interpretations.

At the final iteration, the detectors achieved an average Precision of 0.978 (Figure 12A) and a Recall of 0.532 (Figure 12B). These directly interpretable metrics indicate that while the final detector exhibited an almost perfect rate of correct positive detections, it failed to identify approximately half of the target sound events within the test set. The substantial disparity between Precision and Recall suggests that the performance of the detectors could be potentially improved

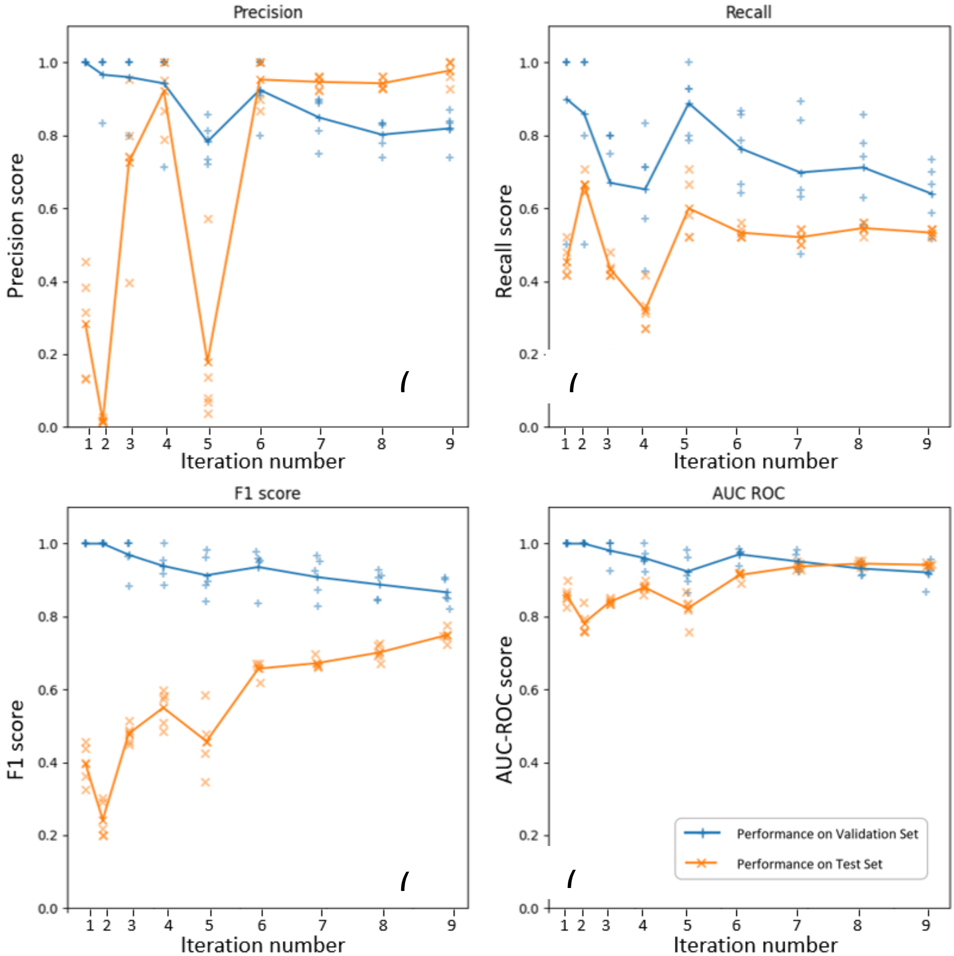


Figure 12: Performances on different metrics of the detector on the validation set and the test set after each iteration during training with Agile Modelling: Precision (A), Recall (B), F1 score (C) and AUC ROC score (D). Performance scores for the validation set are represented in blue and for the test set in orange. The markers in transparency indicate the scores of each model during the 5-fold cross-validation. The plain markers connected by the line are the average of the five models. The different spacing between the iteration numbers is proportional to the number of annotated samples added at each iteration (450 samples were annotated at the end).

by employing a different decision threshold (see Appendix B).

Adjusting this threshold may lead to a significant increase in Recall with a minimal reduction in Precision. However, the optimal threshold value is context-dependent (e.g. class distribution, detector's intended application, difference in soundscapes per environment), necessitating user adjustment of an additional parameter. The F1-score for the test set reached its maximum at the final iteration with a score of 0.689 (Figure 12C). Similarly, the best AUC-ROC score on the test set was 0.92 at the final iteration (Figure 12D).

The complete iterative process required 56 minutes by a user proficient in the data, Agile Modelling, and the associated script. During the training phase, 148 instances of the target sounds were annotated, corresponding to a detection rate of 2.64 sounds per minute. In contrast, six hours of manual annotation of the test set resulted in 40 annotations of target sounds, corresponding to a detection rate of 0.11 sounds per minute. Therefore, the application of Agile Modelling resulted in the detection of occurrences 24 times faster than manual annotation, considering only the iterative training phase.

Following model training, automatic detection was performed on the 166 hours of the dataset using the best-trained model. This analysis resulted in 485 detections in 90 minutes using GPU acceleration on an NVIDIA GeForce RTX 2080 Ti graphics card with 12GB of RAM. Given the model's precision score of 0.978 on the independent test set, most of these detections are likely to represent true occurrences of the target sounds. Unfortunately, time constraints prevented manual verification of these results during this project. Considering that target sounds constitute less than 1% of the total dataset duration in the raw data, the identification of novel occurrences could be significantly increased if most of the detector's pre-detections are indeed accurate. While this pre-detection approach increases data anal valuable insights, it is important to recognise that, like any detection methodology, it may introduce some bias in the detected sounds. The recall score of 0.54 (Figure 12B) indicates that nearly half of the target sounds in the test set were undetected. This could suggest a bias in the detector’s performance, for example, if sounds with short duration or low signal-to-noise ratio were systematically missed. Further analysis is needed to confirm whether this reflects a true sampling bias or limitations in the detection methodology itself.

Nevertheless, the findings of this pilot experiment indicate a substantial gain in time efficiency of the proposed method in identifying novel sound occurrences compared to manual annotation, particularly for infrequent sounds. The results also suggest that performance evaluation based only on the validation set is insufficient to estimate the quality of the detector (precision, recall or F1) reliably, which poses a challenge for the straightforward application of Agile Modelling. As illustrated in Figures 12B, C, and D, increasing the training sample size seems to improve the reliability of performance estimation on the validation set, with results progressively converging toward test set performance. While the final detector's performance is not yet sufficient to entirely replace manual annotation of recordings, it could serve as a valuable tool for initial data exploration, rapid assessment of sound or species presence, and efficient collection of target sound instances for training more extensive models.

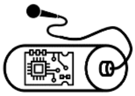
Further investigation, with the inclusion of more annotated samples, is necessary to verify the potential for convergence between validation and test set performance across all metrics. Moreover, the generalisability of these findings requires additional experiments conducted on diverse environments and acoustic datasets. The observed results may reflect the unique characteristics and inherent complexities of this experiment, particularly the presence of multiple anthropogenic sound sources in the dataset. To assess the method's generalisability, future work should evaluate the pre-trained detector on recordings from different geographic locations. Research is in progress in collaboration with De Vlaamze Instituut voor de Zee (VLIZ, Belgium) to further understand the advantages and limitations of Agile Modelling.

3.2.4 Design challenges of machine learning bioacoustics detectors

The primary challenge in the design process of developing automated sound models was the vast array of existing initiatives and unstandardised methods. The field is evolving exceptionally fast, necessitating continuous adaptation of the approach in response to frequent updates and the publication of new relevant methods. For instance, Agile Modelling applied to rare bird sounds with the model Perch (Ghani et al., 2023) was first communicated in a scientific meeting in December 2023 and was updated soon after, using a new pre-trained model, called SurfPerch, on sounds from Coral reefs and birds (Williams et al., 2024).

Limited initial knowledge and experience with machine learning and bioacoustics further complicated the estimation of each method's potential and rendered experimentation to compare them all particularly time-consuming. Time was therefore invested in building a network of researchers working with various experiences in the same field, aiming for cooperation, and relying on the advice from these experts while developing skills and intuition through practice.

3.3 WP3. Building an embedded system for real-time monitoring



An embedded system for real-time monitoring of underwater sound was designed in collaboration with partners engaged in the parallel development of an underwater video monitoring system with an interest in including an acoustic part. The contribution of this project focused on the sound acquisition and onboard processing, while also contributing general engineering support for the development of the embedded audio-video monitoring system (Figure 13).

3.3.1 Design Process of the embedded system for real-time monitoring

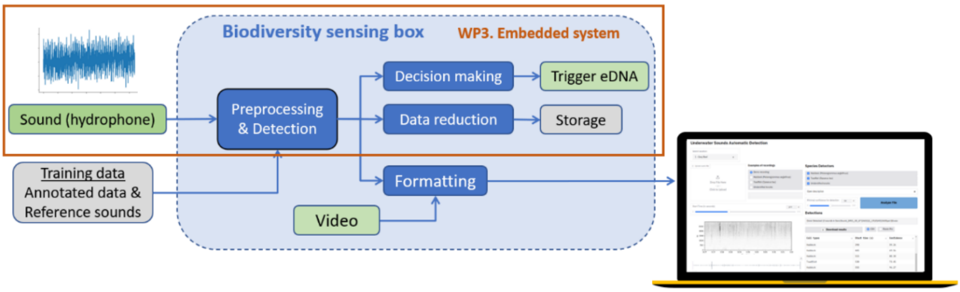


Figure 13: Overview of the development of an integrated passive acoustic monitoring system, for onboard processing of sound in a multi-sensor system, the Biodiversity Sensing Box, with a focus on work package 3, the embedded system supporting data acquisition and processing.

Participation in the design of a combined audio-video monitoring embedded system was not initially part of the project's proposal, but became necessary early in the project. The acoustic recorder used in the Biodiversity Sensing Box did not provide real-time data access, an essential feature for the onboard computing work planned for this project.

Maria Sokolova from the Agricultural Biosystems Engineering group (Wageningen University) led the development, while Rick Hendriksen from Wageningen Technical Solutions (Wageningen University &

Research) was responsible for designing, prototyping, and manufacturing the embedded video system that would now be extended with acoustic monitoring. Regular meetings were held with the team and the Biodiversity sensing box project leader, Reindert Nijland, to schedule important milestones such as tests and deployments.

Prior to involvement in the project, several design decisions were made, including the selection of onboard computing hardware. An NVIDIA Jetson Nano was chosen for its integrated GPU, making it suitable for embedded machine learning applications. Additionally, a Raspberry Pi, a low-cost single-board computer frequently used for deployment purposes, was selected specifically for video data acquisition.

The self-made recorder solutions described in prior studies were not replicated due to their development time requirements and lack of suitability. The recorder developed by Caldas-Morgan et al. (2015) included a custom-made signal-conditioning board but lacked design plans, making reproduction excessively time-consuming. Consequently, commercially available solutions, as used by Bagočius and Narščius (2021), were considered. However, because their recorder lacked a computer for real-time data access and processing, an essential requirement for the objectives of this project, a different solution was ultimately chosen.

Considering factors such as price, compactness, and recommendations from acoustics experts, the A5 hydrophone (Aquarian Audio[17], Anacortes, WA, USA) and the AMS-22 audio interface (Zoom Corporation[18], Tokyo, Japan) were chosen for the underwater sound acquisition. The A5 hydrophone is suitable for targeting low-frequency sounds, with a linear frequency response

---

[17] https://www.aquarianaudio.com/
[18] https://www.zoom-europe.com/

from 20 Hz to 10 kHz and omnidirectional directivity for sounds below 20 kHz. It is functional at depths up to 100 meters, sufficient to cover the North Sea basin region. The AMS-22 is the most compact pre-made audio interface on the market and includes phantom power, a specific type of power supply that improves the sensitivity of microphones. A great advantage was that the AMS-22 had already been tested for compatibility with the A5 hydrophone by the technicians at Aquarian.

Next, the audio interface, hydrophone, computers, and watertight enclosure needed to be integrated. The watertight enclosure is a transparent acrylic cylinder of 50 cm length and 12 cm diameter (Blue Robotics[19], Torrance, California). Software was required for acquiring signals from the audio interface and subsequently processing them on the computer. Python was chosen for software development because of its widespread use in machine learning and in computational bioacoustics research, where it also features robust libraries for sound acquisition and processing. Solutions considered had to be compatible with the Linux operating system used in the embedded system. Although Python is typically less performant than C in embedded systems, it offers faster development and access to well-established libraries for machine learning and audio processing.

The initial design included both computers (the Raspberry Pi and the Jetson Nano), a video acquisition interface board (Shield), and the audio acquisition interface (AMS-22) inside the watertight enclosure (Figure 14). The 4 cameras and hydrophone outside the watertight enclosure are connected to the Shield and the AMS-22, respectively, via watertight cable passages. The power source was also connected to the watertight enclosure using a specialised underwater cable that includes Ethernet transmission and power. This cable can be plugged

[19] https://bluerobotics.com/

into a battery for autonomous deployment or into a socket and a computer for locations with access to the power grid.

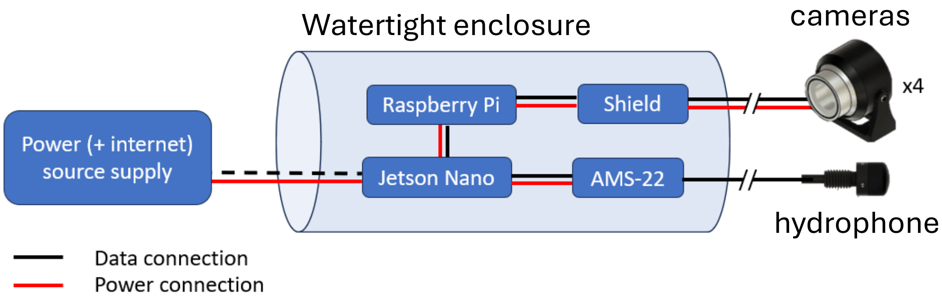


Figure 14: Connections of the components of the embedded system for real time monitoring.

The Raspberry Pi was tasked with acquiring camera feeds via the Shield board and transferring the data to the Jetson Nano. The Jetson Nano was expected to handle the acquisition and storage of the hydrophone data and the onboard processing of both types of data, audio, and video.

Based on this design, a Python application was developed to retrieve data from the AMS-22 on the Jetson Nano, save the data in files, and visualise them in a simple interface. This interface relied on a direct connection from the Jetson Nano to the internet source.

Initially, the Raspberry Pi was intended to be powered directly by the Jetson Nano. However, the USB-C port intended for this use on the Jetson Nano was unable to supply the required current, as specified in the manufacturer's datasheet for the computer. Consequently, a DC/DC converter was needed in the watertight enclosure to transform the 18V power from the external supply to the 5V power required for the Raspberry Pi. Additionally, the storage capacity of the computers was insufficient, necessitating the inclusion of an external SSD storage. Due to the unanticipated need to integrate a power converter and SSD storage, all the components could no longer fit within the original watertight enclosure. Obtaining a larger replacement enclosure was not feasible due to budget constraints, particularly because of the costs required to modify the cylinder to be suitable for the project.

External application of Jetson Nano

To address the space constraints in the watertight enclosure, the development leader decided to remove the Jetson Nano. By connecting the AMS-22 to the Raspberry Pi, the system could still function as an autonomous recorder of sound and video. However, this solution removed most of the onboard processing capabilities. Figure 15 shows the hardware configuration before adding the DC/DC power converter and before integration in the watertight enclosure.

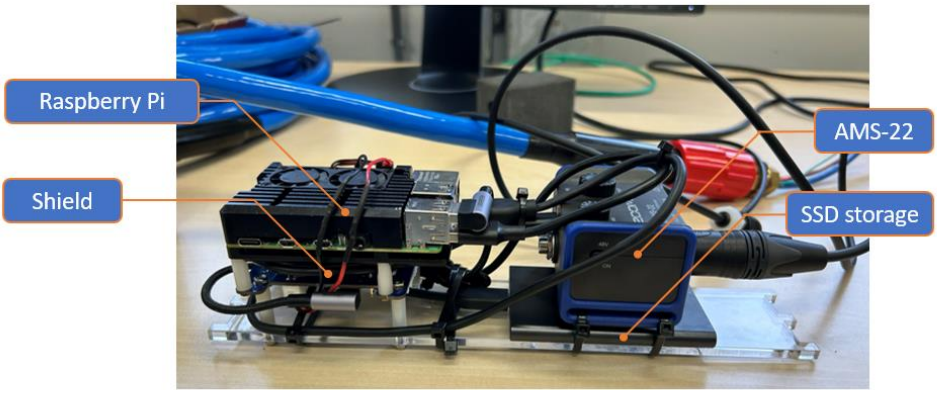


Figure 15. Hardware for autonomous recording of sound and video in the embedded system before integration in the watertight enclosure.

The removal of the Jetson Nano from the watertight enclosure made it necessary to access data differently. Data from the Raspberry Pi were now accessed externally using the Secure Shell (SSH) protocol through the Jetson Nano, which has been relocated above water, outside the watertight enclosure and connected via an underwater cable (Figure 16). This modification required partial redevelopment of the sound acquisition software. Notably, real-time data visualisation on the interface was no longer possible without a direct connection to the onboard computer.

Adapting the interface to run on the Raspberry Pi was estimated to require several days of work due to the difference in the way the data are accessed. Since the plan remains to use the Jetson Nano because its computational power is required for onboard data processing with machine learning models, the interface was not adapted for the Raspberry Pi. This decision was made to avoid additional development effort that would soon become obsolete.

Pilot study on temperature

An additional aspect of the embedded system's design was verifying the internal temperature of the watertight enclosure, as the lack of airflow could cause it to overheat during operation. The Raspberry Pi is natively equipped with a temperature sensor on its CPU. The temperature was observed during an hour of operation in the watertight enclosure, placed outside water, and showed a stabilisation after approximately 10 minutes, suggesting a state of thermal equilibrium. In ambient air, heat from the watertight enclosure dissipates more slowly than underwater due to the air's lower thermal conductivity. Additionally, North Sea underwater temperatures typically range from 10–20°C, which is lower than the air temperatures during testing (around 22°C). As a result, the functioning was validated under more demanding conditions than those expected in actual use, ensuring safe operation.

3.3.2 Current solution of the embedded system for real-time monitoring

The current iteration of the embedded system incorporates the elements shown in Figure 15 (the Rasberry Pi, the Shield, the SSD storage and the AMS-22) plus a power converter and is now housed within the cylindrical watertight enclosure (Figure 16). This system is equipped with a hydrophone and four industrial-grade cameras and can be mounted on the Biodiversity Sensing Box frame for deployment. The system operates autonomously when powered by a battery pack, but it can also use a socket as a power source when available. Underwater sound pressure levels are measured using an A5 Hydrophone, connected to the Raspberry Pi via an AMS-22 audio interface, which amplifies and digitises the acoustic signal. A Python application on the Raspberry Pi records audio at a sampling rate of 44 kHz, storing the data in files of adjustable duration. Simultaneously, the system captures and stores video streams from the four cameras. All acoustic and video data are securely saved to an external SSD housed within the watertight enclosure.

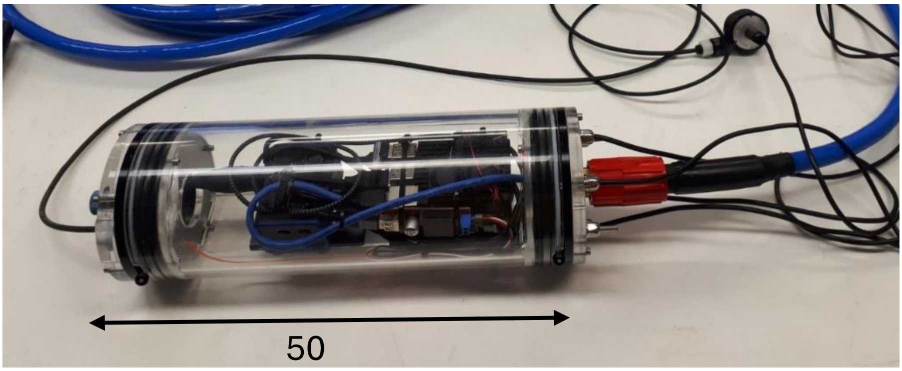


Figure 16: The embedded system. An autonomous sound and video recorder for underwater environments.

As a proof of concept, the embedded system was deployed both in a freshwater location using power from a socket and in the sea using a battery pack.

3.3.3 Design challenges of the embedded system for real-time monitoring

The design process of the embedded system presented significant challenges due to the parallel development of the video and sound components, which were handled by different individuals with sometimes diverging priorities.

Late-stage design changes often have substantial consequences, particularly when earlier development work is rendered incompatible with new decisions. In this project, critical factors such as space constraints in the watertight enclosure and power limitations were recognised early as key considerations. However, time pressure, including shipping time and the annual budget deadline, forced accelerated decision-making. The lack of thorough research and planning on these critical characteristics ultimately led to issues that could have been anticipated in an ideal situation.

The discussion that resulted in the decision to exclude the Jetson Nano from the watertight enclosure highlights divergent design priorities and strategies within the team. One approach prioritised iterative development through functional proof-of-concept stages, minimising risk and allowing for easier stakeholder demonstrations. The other emphasised aligning the design with long-term goals during development. Although hierarchical decision-making within the team facilitated quick resolutions, it often prioritised short-term convenience over balanced, collaborative solutions. In this case, a competitive or accommodative approach was taken, favouring the lead developer’s perspective, rather than a compromising or collaborative strategy. This is a consequence of joining an already started project.

Within more regular project developments, such challenges can be mitigated through thorough risk analyses when making critical design decisions, particularly for prototypes and during early-stage development. Effective communication and collaborative decision-making processes should then be prioritised, ensuring that the interests of all team members are considered. In this case, the acoustic aspects were considered of minor importance. In situations where interests are more balanced, involving the project leader can help facilitate decisions that serve the overall best interest of the project.

3.4 WP4. Design of a graphical user interface for bioacoustic monitoring

The development of a graphical user interface (GUI) aimed to create an interactive dashboard for real-time visualisation of sound data, with plans to incorporate video in the future, while enabling interaction with the embedded system (Figure 17). Interfaces provide users with an intuitive way to interact with embedded systems and serve as tools for visualising progress, benefiting stakeholders and fostering public engagement. This is particularly valuable for raising awareness about lesser-known environments, such as the underwater ecosystems of the North Sea.

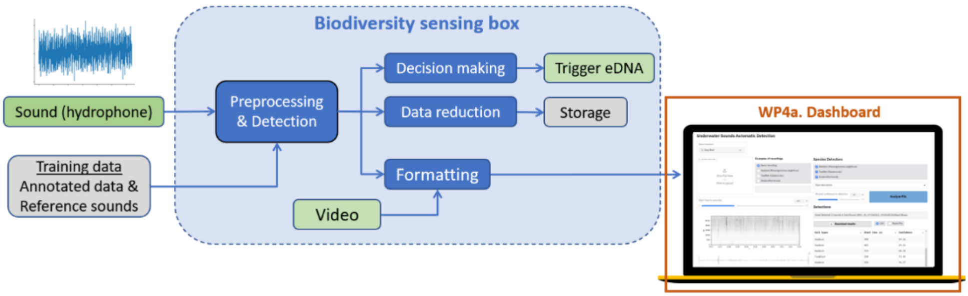


Figure 17: Overview of the development of an integrated passive acoustic monitoring system, for onboard processing of sound in a multi-sensor system, the Biodiversity Sensing Box, with a focus on work package 4, the dashboard for displaying sound and detection in real time.

3.4.1 Design Process of the graphical user interface for acoustic monitoring

Developing a dashboard for an embedded system often provides greater flexibility for functionalities than existing applications. Custom GUI development enables the integration of specialised data and allows for personalisation to address both project-specific needs and stakeholder expectations. In this project, the future wish to integrate diverse data types, including video and DNA, with the acoustics, along with potential future user interaction capabilities, justified the development of a custom graphical interface.

The first interface was designed to display real-time sound data from the Jetson Nano inside the embedded system and was developed using the Python PyQt library. The interface offered limited functionalities: selecting the duration of the files containing the recording, visualising the input acoustic signal as a waveform, displaying status messages, and starting and stopping to record (Figure 18).

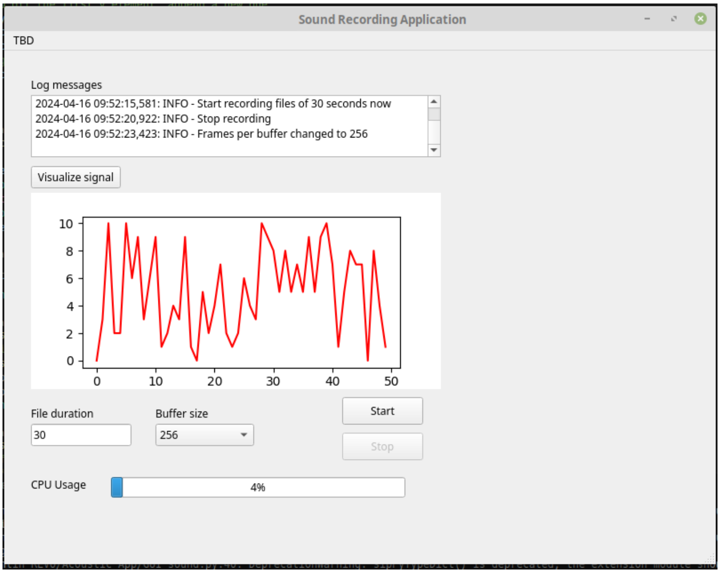


Figure 18: First iteration of an interface to record and display sound acquisition in real time.

As outlined in section 3.3.1, the development of this interface was discontinued due to changes in the embedded system design, which restricted direct data access and, consequently, visualisation of the recorded sound. Nonetheless, this initial development provided valuable insights into the challenges and opportunities of sound visualisation and laid the groundwork for future advancements.

Meeting developers from the BirdNET team during an eco-acoustics conference provided an excellent opportunity to discuss interface solutions. Gradio, a library dedicated to creating interfaces for machine learning applications, was identified as a promising solution. The development of an application based on Gradio was initiated, following a design process inspired by agile software development methodologies. ("Manifesto for Agile Software Development," 2001).

Agile-inspired methods have become predominant in software development as they are suggested to improve development processes. These approaches are particularly well-suited for software development with rapid iteration, low-stakes decision-making, and enhanced stakeholder engagement and adaptability. A review of agile software development practices concluded that a development method is called agile when it is incremental, cooperative, straightforward, and adaptive (Abrahamsson et al., 2002). One of the practices in agile software development is to produce a minimum viable product quickly and improve it iteratively based on regular discussions with users or customers, rather than adhering to a fixed set of requirements. In this framework, tasks are typically divided among team members, who use tracking tools to manage progress effectively.

While the project did not involve collaborative development, an iterative design approach driven by user feedback was pursued. In total, 11 researchers and users with diverse profiles were observed and interviewed while testing the prototype of the interface, enabling improvements informed by their suggestions. Observations were particularly valuable for enhancing the intuitive use of the interface. Through iterative refinement, the interface quality improved to reach a usable and distributable version. Figure 19 illustrates two iterations of a specific interface component during the development process.

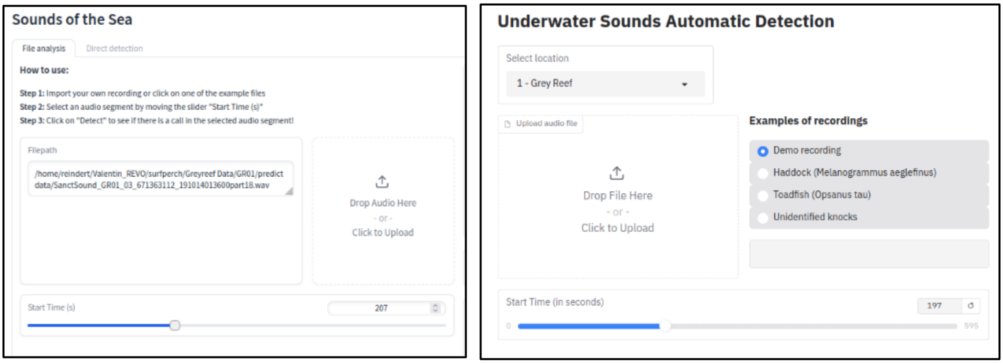


Figure 19: Partial visualisation of the interface at different time of the development. The version on the left is older than the one on the right.

Initially, the plan was to include all functionalities in a single GUI. The existing interface supported the use of pre-trained models for detecting specific sounds across various environments, displaying spectrograms, and playing back both recordings and sounds automatically detected. The future additions will include real-time sound display from an acquisition device connected to the computer hosting the dashboard, and later, integrating a video stream and enabling user interaction with the embedded system.

Two distinct use cases for the GUI, with different future development trajectories, were identified:

1. A deployment-focused dashboard: As originally envisioned, this version would display real-time sound detections from the embedded system during field deployments. Designed to work

exclusively with the embedded system, it will later incorporate the video stream and enable interactions with the embedded system.

2. A Standalone analysis tool: Aimed at a broader audience, this application enables offline analysis of recordings using pre-trained models and does not require a connection to the embedded system. Future enhancements should expand its capabilities to include custom model training and support for additional environments.

Therefore, to streamline development and usability, the GUI was split into two separate applications:

- The Dashboard (WP4a): Tailored for real-time sound and detection monitoring from the WP3 embedded system, with planned support for video and eDNA data.
- The Interface (WP4b): Designed for users analysing recordings with pre-trained models, independent of embedded hardware.

This separation reduced complexity, simplifying both development and end-user interaction.

3.4.2 Current Solutions: Dashboard and Interface

Two GUIs were developed, the Dashboard (WP4a) and the Interface (WP4b).

The Dashboard

The Dashboard displays in real-time the recorded sound as a spectrogram, and a history of the detections in a table (Figure 20). Automatic detection, done by pre-trained models, runs every 5 seconds during recording. Every 10 minutes, the recorded sound data and the detection are saved to local storage. The dashboard is designed to be coupled with the embedded system (WP3).

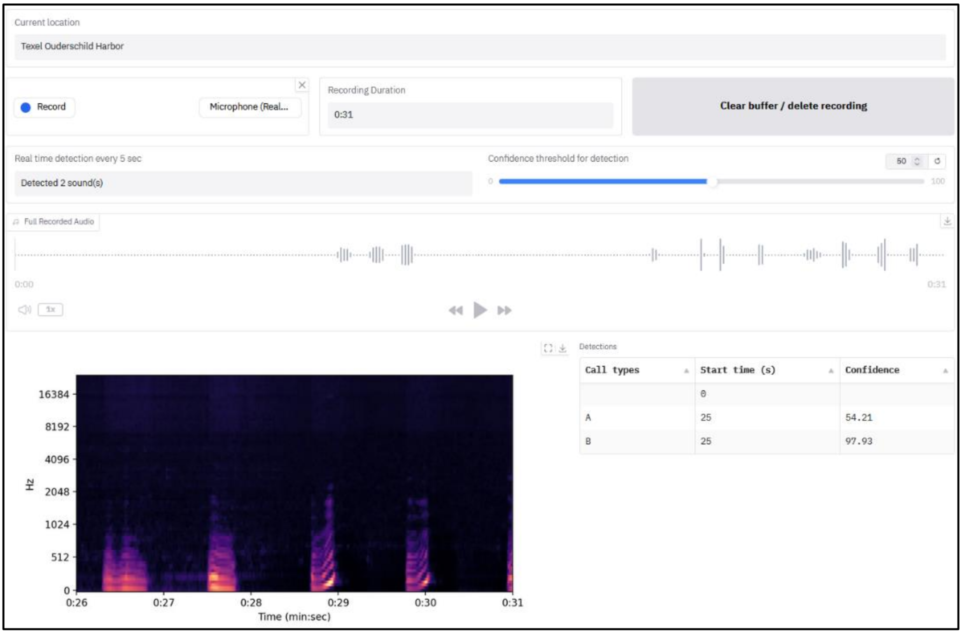


Figure 20: Dashboard for real time monitoring and detection of sounds

The Interface

The Interface is meant for researchers and bio-acousticians to automatically detect sound events in underwater recordings, thanks to pre-trained deep-learning models (WP2). Pretrained detectors for different sound types and different locations can be used to analyse recordings provided by the user. The results can be downloaded as annotation files in a format compatible with the software Raven Pro 1.6 or as a table. The most recent version of the Interface is displayed in Figure 21.

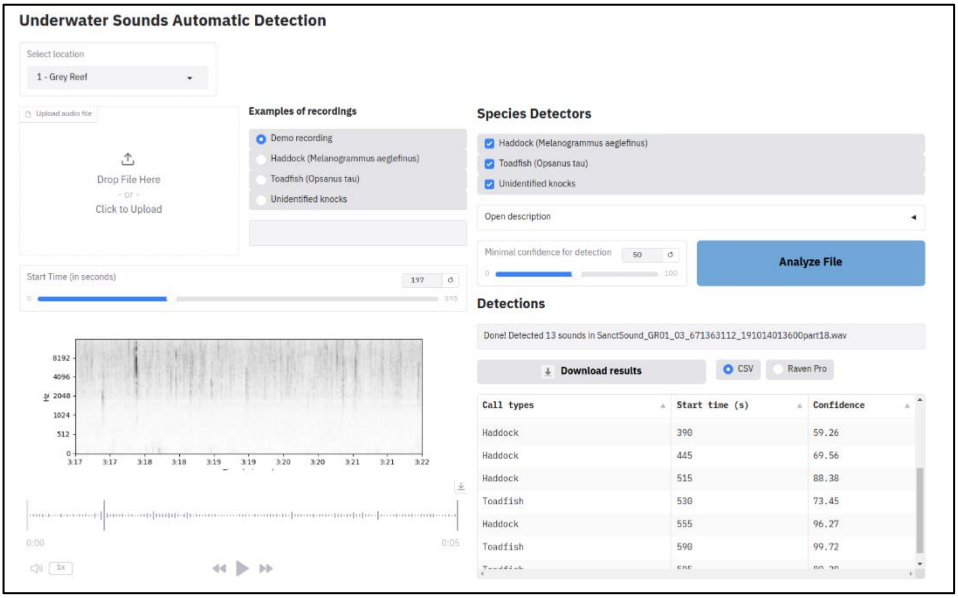


Figure 21: The Interface for automatic sound detection using pre-trained deep learning models (version December 2024).

3.4.3 Design challenges of the graphical user interface for acoustic monitoring

Implementing an iterative development method based on user feedback was more challenging than anticipated for several reasons:

- The first time a user tries the interface is the most informative regarding intuitiveness, but this can only be tested once per person; after that, a new user is required.

- While it was easy to see what users did not understand, they were often not able to suggest what could be improved and how, or what they would prefer.
- Some requests for improvement were technically infeasible due to the limitations of the library (high-level programming) or would be too time-consuming. In such cases, a workaround solution was looked for to satisfy the requests.
- Suggesting different ideas was unexpectedly difficult. Some ideas can be difficult for a user to visualise based on simple explanations, making it difficult to gauge their interest. Doing a test implementation is often more effective in obtaining feedback, but it can be time-consuming.

Overall, the iterative process facilitated rapid and flexible development, proving more effective than striving for a final product from the outset. However, the frequent release of minor updates (several versions per week) posed challenges for systematic evaluation. Moreover, qualitative attributes of the interface, such as intuitiveness and aesthetics, are inherently difficult to quantify. While subjective user assessments can provide valuable insights, conducting such evaluations for every iteration is tedious and time-consuming. Given the objective of producing a functional prototype, the development of core functionalities and the qualitative analysis of user feedback and satisfaction were prioritised within the time constraints of the project. Future feedback from external users will offer valuable insights to guide further development after sharing the project in open-access platforms such as GitHub[20] and Hugging Face[21].

[20] https://github.com/

[21] https://huggingface.co/

4 Summary of deliverables

This project aimed to design and implement an innovative, efficient passive acoustic monitoring (PAM) system for integration into intelligent multi-sensor systems. The developed system aims to enable real-time communication and automated onboard processing of acoustic data, based on four work packages (WPs): WP1 focused on bioacoustics data collection and annotation to support species identification and model training; WP2 advanced automatic underwater animal sounds detection, particularly when training data are limited; WP3 contributed to the design of an embedded system for real-time sound acquisition and processing; and WP4 initially focused on developing a user interface for real-time monitoring with visualisation of automatic sound detections, which later evolved into two separate interfaces: one for integration with the embedded system and a standalone tool for data analysis. The following sections detail the key deliverables achieved within each WP.

4.1 WP1. Collection of underwater bioacoustics data

Recordings from eight locations, including two in the North Sea, one of which originated from a deployment within this project, were successfully collected and stored in a local data repository. Due to the lack of annotated data available, several hours of recordings from different datasets were manually annotated with the assistance of two students. These annotations support model validation and ecological analysis and are also included in the repository.

As part of this effort, a protocol for annotating underwater sounds was developed based on the work of a student supervised during the project. This protocol includes the creation of a standardised sound dictionary, improving the consistency of the annotation process, and facilitating data analysis. Having a dictionary of sounds also enables the comparison of sounds between locations (Vieira et al., 2024).

For future monitoring purposes, a list of soniferous fish species known in the Wadden Sea and for which reference sounds are available was made (Appendix D). This list was created by crossing the references of species detected in the Wadden Sea by any monitoring method[22] with those catalogued on the FishSounds database (Cox et al., 2023). The list was created to identify the fish species responsible for the sounds recorded in Texel-Ouderschild Harbour (see Section 3.2). However, due to the lack of acoustic references for most fish species, reliably determining the sound-producing species was not possible. Since the Wadden Sea is part of the North Sea region, this list can be easily expanded to cover the broader North Sea by including species that occur there but are absent in the Wadden Sea.

Additionally, the database of unidentified sounds, shared with researchers from IMR, VLIZ, and RUG, currently contains 12 described unidentified sounds from the North Sea that later should become identified. This collaborative database continues to grow as research advances, contributing to the broader understanding of underwater soundscapes.

In addition to the above-mentioned data collections, WP1 also yielded valuable methodological insights. Considerable expertise was gained in using the SoundTrap ST600 sound recorder and in deploying underwater recording units. These advancements bolster the capacity of the Marine Animal Ecology group for future acoustic research.

Finally, the project facilitated new network connections and strengthened existing ones for the Marine Animal Ecology group. Collaboration with researchers in bioacoustics, including the

---

[22] https://swimway.waddensea-worldheritage.org/fish-species

acoustic team from VLIZ, has expanded the group's connections and opportunities for future collaborative research.

4.2 WP2. Development of machine learning bioacoustics detectors

In Work Package 2, the aim was to address the challenge of marine animals' sound automatic detection under data-scarce conditions. The development of few-shot learning for bioacoustics using an innovative method was promising, and as a contribution submitted to the DCASE Challenge 2024 task 5 (Liang et al., 2024). A detailed technical report[23] and the source code[24] are publicly available from the DCASE Challenge website, facilitating further research and development.

Although the DCASE Challenge has been running for several years, few-shot learning methods for detection in bioacoustics had not yet been applied to real-world data. The applicability of the method submitted to the DCASE challenge was demonstrated by improving the speed of gathering fish sound annotation five times over the classical manual annotation method. These findings were shared with the bioacoustics research community through a presentation at the 5th World Ecoacoustic Congress[25] In Madrid.

Using Agile Modelling, seven models were trained to identify putative sounds of different species across three distinct environments: a coral reef, an offshore wind farm, and a harbour. Models for detecting toadfish (*Opsanus tau*) sounds, presumed haddock *(Melanogrammus aeglefinus*) sounds and for different unidentified sounds were developed. The method was also applied successfully for detecting a

---

23 https://dcase.community/documents/challenge2024/technical_reports/DCASE2024_Bordoux_66_5.pdf

24 https://github.com/vbordoux/dcase_task5_Bordoux_WUR

25 https://ecoacoustics2024.org/

sound assumed to be from anthropogenic source based on its characteristic, suggesting the applicability of Agile Modelling to sound from different type of sources.

In a pilot study, a model was trained to investigate sounds, presumed to be emitted by fish based on temporal and spectral characteristics, recorded at Texel Harbour. An example of application of this model is shown in Figure 22, which displays the count of automated detections of sound per two-hour intervals, time of day, and the average value of the relative sound pressure levels (mostly influenced by boat noise). This study suggests the potential of the method for the analysis of extensive acoustic recordings that would otherwise demand

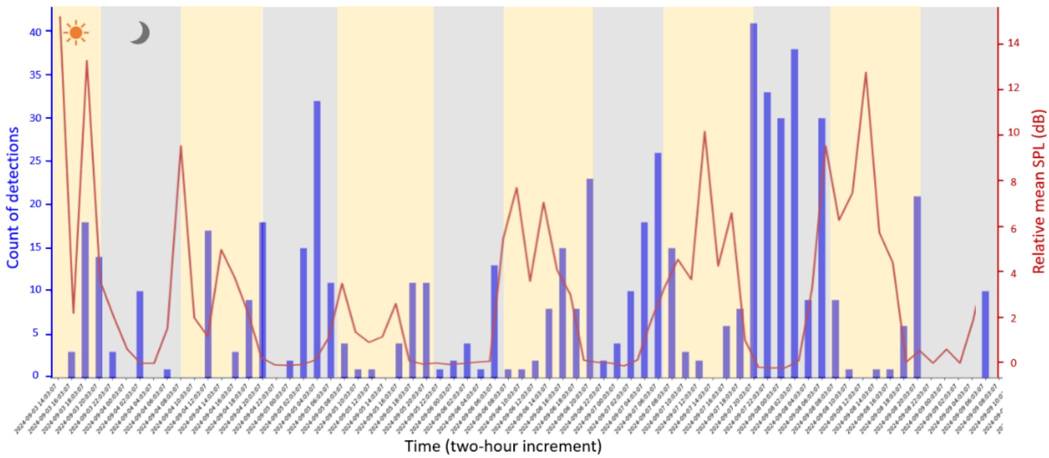
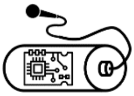


Figure 22: Number of automatic detections of fish sounds per 2-hour recording from Texel-Oudeschild harbour. The grey and yellow areas represent night and daytime, respectively, and the red curve represents the average value of the relative sound pressure level per 2 hours.

prohibitive effort or present logistical challenges (in this investigation, 166 hours of data were processed in 90 minutes). Continued development and validation of this approach is warranted.

4.3 WP3. Building an embedded system for real-time monitoring



Sound and video data were successfully collected by the embedded system running on a socket power supply in a pond on the Wageningen University campus (Figure 23, left). Unfortunately, shortly before its scheduled offshore deployment, the Raspberry Pi was accidentally formatted, causing the loss of the sound recording application. Due to time constraints, this human error could not be corrected before deployment, preventing testing of sound collection by the system on autonomous power (Figure 23, right). Now fixed, the embedded system is currently available to function as an autonomous recorder for both sound and video, ready for future deployments within the biodiversity sensing

Figure 23: Deployment of the embedded system in the pond on the Wageningen University campus with a power connection supply from the building (left). Biodiversity Sensing Box ready to be deployed at sea with the embedded system and its power supply mounted on it (right).

box.

During the deployment of the Biodiversity Sensing Box at sea, the system was observed to topple sideways. In response, Rick Hendricksen proposed a modified design to prevent tilting. A stability analysis was performed to evaluate the design's effectiveness under the strong underwater currents expected in the North Sea (up to 2 knots). The analysis revealed that the Biodiversity Sensing Box remained unstable even with the design modifications. The proposed modified design of the Biodiversity Sensing Box and calculations for the analysis are shown in Appendix C.

4.4 WP4. Design of a graphical user interface for bioacoustic monitoring

The Dashboard for real-time monitoring and detection is ready for deployment once the Jetson Nano is reintegrated into the embedded system. Although the system is not yet fully operational, a proof-of-concept test was conducted by playing underwater recordings through a speaker and using a laptop microphone as a substitute for the hydrophone connected to the embedded computer (Figure 24).

During this test, real-time detection of the target sound was successful. However, some instances of the target sounds were missed, likely due to distortions introduced by the speaker and the automatic signal processing applied to the laptop microphone input.

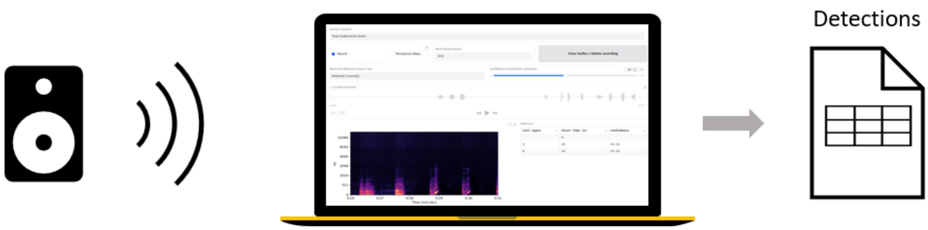


Figure 24: Testing real-time monitoring and detection of fish sounds with a mock system. The speaker plays a soundscape recorded in an underwater environment. The laptop's microphone plays the role of the hydrophone, and the laptop runs the model doing automatic detection, which would normally run on the onboard computer.

Upon full hardware integration, the dashboard will operate on the embedded computer and will be accessible via a local or web server to visualise the system functioning underwater.

The interface for offline file analysis is fully functional. It currently integrates the seven trained sound-detection models and supports analysis across three environments. Users can run file analyses and extract the resulting data efficiently.

5 Discussion

This chapter discusses the contributions of this project to the development and improvement of PAM techniques for marine biodiversity, in the context of multi-sensor monitoring system integration. This project was separated into four packages: WP1. Collection of underwater bioacoustics data, WP2. Development of machine learning bioacoustics detectors, WP3. Building an embedded system for real-time monitoring, and WP4. Development of a graphical user interface for bioacoustic monitoring. This discussion establishes the research context for the design of this EngD project, then situates each work package's contribution within this context, and concludes with future development prospects.

Passive Acoustic Monitoring for fish and marine invertebrates

With increasing attention and emerging cost-effective technologies, the development of PAM as a monitoring technique for fish is undergoing quick progress, which will subsequently enhance the monitoring of marine invertebrates (Mooney et al., 2020; Parsons et al., 2024; Stowell, 2022). Based on insights from the available literature and the work realised in this project, two positive feedback loops and the main bottlenecks in the development of PAM for fish and marine invertebrates are underlined in Figure 25 and below.

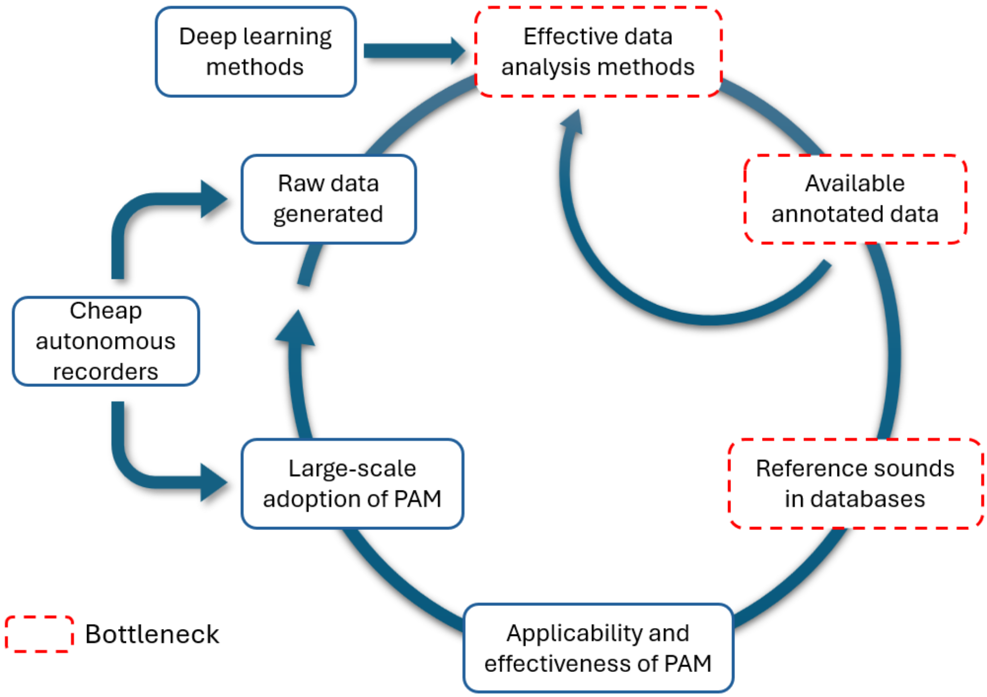


Figure 25: Positive feedback loops in PAM advancement for fish and marine invertebrates: key drivers and main bottlenecks (in dotted red squares).

The development of autonomous acoustic recorders has recently led to a massive increase in the amount of available raw data, making it nearly impossible to analyse manually. However, accessible and effective data analysis methods, such as Agile Modelling, enable the processing and annotation of larger datasets. Such progress allows for unprecedented processing of hours of recordings into annotated data, previously unmanageable in terms of costs and duration of analyses (Stowell, 2022). If a proportion of this annotated data is made available for other studies, this will, in turn, support the development of better methods for data analysis and feed into a positive feedback loop, accelerating the process even more. Furthermore, the availability of annotated data and the improvement of data analysis methods will facilitate the acquisition of new reference sounds, e.g. when a sound is matched with in situ visual fish identification (Dantzker et al., 2024; Mouy et al., 2023; Vieira et al., 2024). The availability of databases with more reference sounds will increase PAM applicability to study more specific species for various purposes, e.g. species distribution knowledge for red listing, relative population abundance for fishery quotas, or detection of fish spawning aggregation sites to protect the area (Bolgan et al., 2023; Souza Jr et al., 2023; Wilson et al., 2019). By advancing progress and overcoming current limitations, PAM could emerge as a cost-effective, reliable solution for large-scale underwater monitoring, offering unique information that complements other monitoring methods (Cabrito et al., 2024).

To summarise, at the start of this EngD work, the application of PAM for marine animal monitoring is limited by three major bottlenecks (Figure 25):

- Effective data analysis methods: High cost associated with manual data analysis, and limited accessibility and reusability of existing automated data analysis.
- Available annotated data: Lack of openly accessible annotated data.
- Reference sounds in database: Scarcity and reliability of reference sound databases for identification of sound sources.

Through the different work packages, this EngD project made several contributions to advancing progress and overcoming bottlenecks for PAM for marine animals, as depicted in Figure 26.

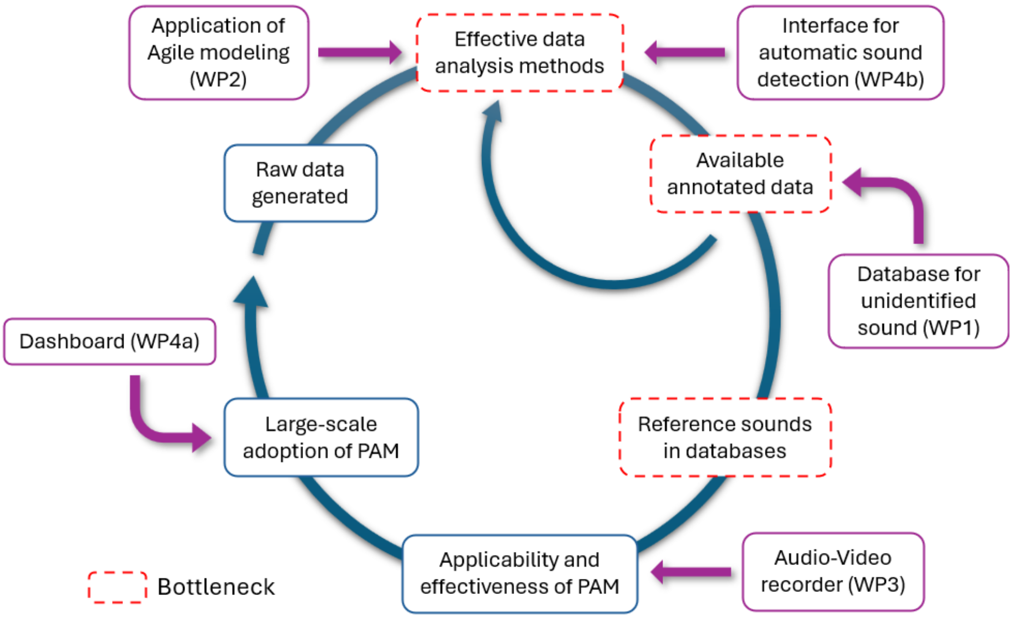


Figure 26: Contribution of this EngD, placing the outcome of the different work packages (in purple), in the context of the development of PAM for fish and marine invertebrates. The main bottlenecks are indicated in dotted red squares.

Toward automated data processing with deep learning

The outcomes of WP2 demonstrate that automatic detectors can be rapidly developed for marine animal sounds across temperate underwater environments. This builds on Agile Modelling, a method initially designed for rare bird sound detection (Williams et al., 2024), which was successfully adapted for the detection of fish sounds in different environments.

Automatic detection of fish sounds using deep learning models has already been achieved for specific fish species or in specific environments, e.g. brown meagre, Caribbean groupers, and unidentified sounds in the Strait of Georgia (Ibrahim et al., 2024; Laplante et al., 2022, 2021; Mouy et al., 2024). In these studies, detectors trained for specific utilisation generally achieve higher performance (based on F1 score) for fish sound detection or classification than the models developed in WP2. However, the model development in previous studies included a lot of fine-tuning that has been specifically made to suit each dataset, such as window length, model architecture, threshold selection, spectrogram generation parameters, or feature selection. In addition, each of these detectors required dozens of hours of manually annotated data to be trained and additional development time for finding the best parameters for the model and training. Unfortunately, specialised detectors as the one built in the reported studies have limited transferability, meaning that the extensive time spent doing manual annotations (not mentioned in the study but likely to be tens to hundreds of hours) must be spent anew for different environments. In this EngD project, using Agile Modelling and starting with only one annotated sound, a fish sound detector was developed. The detector reached a precision score of 1 and a recall score of 0.54 after less than an hour of training, including annotation time, without fine-tuning parameters (see Section 3.2.3). Such performances cannot be considered sufficient to be applied confidently to replace a human annotator. However, the method can be used to get initial insight into long deployment or to very efficiently gather training data to train a detector, saving huge amounts of time on manual annotation work, especially in environments with low sound event activity or for detecting rare sounds.

While many researchers in biology or ecology are interested and would benefit from using machine learning techniques in PAM, the accessibility of such methods is limited by the knowledge of computer science and programming skills needed. This EngD project demonstrates how interdisciplinary collaboration can bridge this gap effectively. Current research on fish sounds still largely relies on manual annotation methods, with multiple studies explicitly highlighting the need for automated detection systems to enable reproducible and scalable PAM investigations. (Dantzker et al., 2024; Jarriel et al., 2024a; Mouy et al., 2023). The method used in WP2 can help advance such studies, though significant work remains to reach progress comparable to terrestrial bioacoustics research.

Currently, terrestrial bioacoustics emphasises developing standardised, accessible methods, using graphical user interfaces and models capable of operating across diverse regions and species or even taxonomic groups.(Bergler et al., 2022; Hagiwara, 2022; Kahl et al., 2021; Robinson et al., 2024a, 2024b). In both realms, terrestrial and underwater, the creation of robust detectors for a range of different types of sounds and environments requires sufficient training data similar to the conditions where the detector will be used. The much more limited annotated data available for fish and marine invertebrates sounds, a bottleneck addressed above (Figure 26) and in WP1 and WP2, are not yet sufficient to build a robust detector that can be used in different environments. For this reason, training specialised models for detecting one target sound in a specific environment is currently more likely to result in high-performance detectors than developing a general model that is applicable everywhere.

Agile Modelling is, to this date, the fastest method to train specialised detectors of marine animal sounds in different temperate environments such as a harbour, an offshore wind farm, a rocky plateau, or a coral reef (Section 4.2). Due to time constraints, the method was not compared with the most novel approaches for quick annotation, e.g. the combination of active learning and label propagation to clusters from unsupervised learning (Napier et al., 2025; Parcerisas et al., 2024). Future work should also investigate the application of Agile Modelling for multiclass classification, already possible with the current version, which was not addressed in this EngD work due to time constraints.

The application of Agile Modelling is a step toward addressing two of the main bottlenecks in PAM for fish and marine invertebrates. First, directly contributing to providing the ability to analyse large volumes of data with effective and accessible state-of-the-art machine learning methods. Secondly, indirectly addressing the lack of available annotated data: by rapidly accelerating the detection of sounds and providing a method that is accessible to everyone, researchers will be able to process a lot more data than with manual annotation.

Future development of Agile Modelling should further enhance its accessibility. The current version, implemented in a Jupyter Notebook, enables users to execute code sections by section, guided by comments and instructions. While this approach improves accessibility, it still requires Python proficiency, particularly for adjusting the code to specific data requirements. Integrating Agile Modelling into a user-friendly interface (to be incorporated into the WP4b Interface) will further increase accessibility by removing the need for coding.

An additional point of improvement, Agile Modelling relies on a general classification model such as BirdNET, Perch or SurfPerch for training specialised detectors. Development of new general classification models that outperform the best bird detection models, such as NatureLM-audio, should be included and tried to facilitate the training of better detectors (Robinson et al., 2024a).

Methods for reference sound collection

Reference sounds are required for sound-based identification of species in field recordings. An additional challenge of PAM for marine animals remains with the lack of reference sounds available. Historically, reference fish sounds have been collected in aquarium studies. However, these reference sounds may not be representative of sounds made in situ because the fish produced sounds prompted by unnatural stressors, e.g. electroshocks or manual stimulation (Clark and Dunn, 2022). In addition, sound production in fish is often associated with social behaviour (Amorim et al., 2015) or in response to environmental cues, which can be difficult to recreate in tank conditions. Recent developments in simultaneous video and acoustic monitoring have proven successful in identifying fish sounds in situ (Dantzker et al., 2024; Mouy et al., 2023). In WP3, a system was developed that can record sound and video simultaneously for real-time monitoring, specifically, the embedded system is also usable to associate sounds with images underwater. Localising the sound source is necessary to reliably confirm the identity of the marine animal producing the sound, e.g. with echolocation using an array of underwater sound recorders (Dantzker et al., 2024; Mouy et al., 2023; Pyć et al., 2021). Therefore, future efforts should focus on integrating multiple hydrophones combined with one or several cameras to reconstruct the position of the sound source.

Another promising approach to rapidly gather reference sounds is cross-referencing unidentified sounds recorded in different locations against species lists (Vieira et al., 2024). The authors also utilised environmental conditions, such as time of day, depth, and habitat type, to narrow down the list of potential species or families producing a given sound, although they were unable yet to confirm the species. The study suggests that larger databases of unidentified sounds can provide information about the local acoustic community, help to better understand soniferous species, and serve as a proxy for biodiversity.

In WP1, a collaborative database was established in partnership with VLIZ, RUG, and IMR to facilitate the sharing of unidentified sounds among North Sea researchers, addressing the current lack of data-sharing initiatives (Parsons et al., 2024). As knowledge expands in the different research teams, this collaborative database will provide access to a growing number of unidentified sounds. By comparing sounds recorded in different locations, alongside environmental data and local species list, unidentified sounds are expected to be matched to their corresponding species, as suggested by Vieira et al. (2024).

An additional source of recordings for fish and marine invertebrate sounds has potentially been overlooked. No research has yet systematically examined fish and marine invertebrate sounds in existing publicly available datasets originally published for studying marine mammals or released without annotations (Darras et al., 2024; Hatch et al., 2024). Although direct ground-truthing is not possible anymore, these unrelated datasets likely contain fish and marine invertebrate sounds that could be exploited for cross-referencing in acoustic studies, removing the cost needed for additional data collection. Currently, identifying and analysing these sounds is highly time-consuming. However, with the advancement of automated detection and analysis tools, this process could become significantly more efficient shortly.

Over time, reference sounds for most of the 22,000 fish species expected to produce sound will be collected, expanding the applications of PAM. Because of the restrictions of the current manual approaches, the current rate of discovery is slow (Looby et al., 2022; Rice et al., 2022). Both traditional methods, such as aquaria studies, and novel approaches, like deployable audio-video arrays, continue to steadily increase the number of fish species recognised as actively soniferous while also providing essential reference sounds (Figure 27). However, these methods inherently depend on the detection and annotation of animal sounds, making the process labour-intensive and time-consuming (Dantzker et al., 2024; Mouy et al., 2023; Vieira et al., 2024. Automating fish sound detection and annotation would significantly improve the efficiency of these approaches. Thus, the development of accessible automation tools for fish sound detection, such as Agile Modelling (WP2), also accelerates the expansion of reference sound databases.

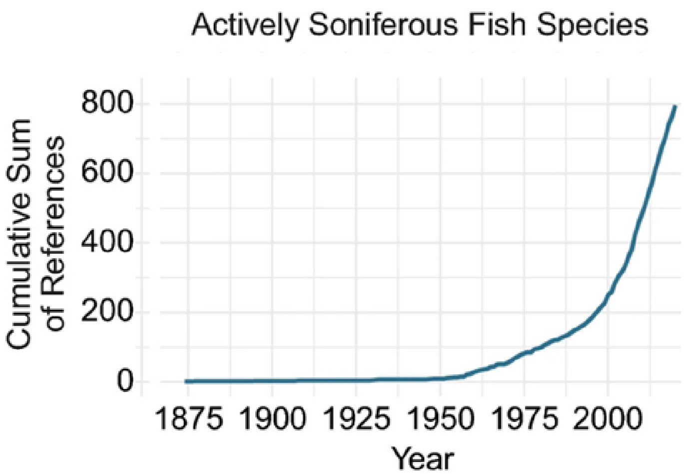


Figure 27: Cumulative references of actively soniferous fish species. Source: Looby et al., 2022.

A key application of the automated onboard sound detection system in the Biodiversity Sensing Box was to create a novel method for collecting reference sounds. This innovative approach would use a decision-making algorithm (Figure 28) to trigger targeted eDNA sampling upon detecting sound of interest, potentially allowing DNA-based identification of sound-producing species.

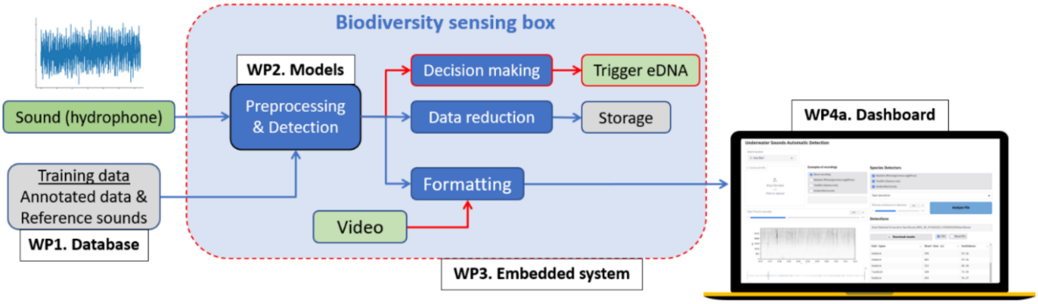


Figure 28: Future development steps for the Biodiversity Sensing Box: links and elements that remain to be developed are highlighted in red.

Triggering eDNA sampling based on sound detection

Project findings, discussions with eDNA researchers, and expertise developed in computational bioacoustics revealed several unresolved challenges that must be addressed before implementing an efficient audio-eDNA combination for reference sound collection. Confirming the sound source based on eDNA sampling is unlikely to be effective because the eDNA could originate from any animals upstream, not necessarily the ones making the sound. Sound can travel in all directions, eDNA only with the current, so the distance or position of a sound source relative to water currents can compromise eDNA detection (Rishan et al., 2023). Also, eDNA samples often contain DNA from multiple species, so they cannot reliably certify the species that made the sound (Rishan et al., 2023). Finally, current state-of-the-art automatic fish or marine invertebrate sound detection is largely limited to familiar sounds. Detecting unexpected or rare sounds not included in the training of a model is a research field known as open-set classification, which has been studied in birds but appears to be unexplored in fish (Tavares, 2022). Therefore, existing automated sound detection methods are not expected to be advanced enough yet to enable the effective integration of PAM with eDNA techniques.

Nonetheless, the method could be used in niche applications as eDNA-based monitoring has demonstrated applications beyond biodiversity assessment, including population abundance estimation (Spear et al., 2021) and population genetic analyses (Parsons et al., 2024). Therefore, triggering eDNA sampling based on sound detection could enable targeted monitoring of specific species with known vocalisations, such as haddock or cod, or under specific conditions, such as river migrations (e.g., triggering sampling in response to Alosa fallax mating sounds (Langkau et al., 2016)).

Challenges and opportunities for PAM in the North Sea

This EngD thesis suggests that PAM has potential to become can become one of the standardmethods for monitoring fish in the North Sea, with this work representing a step toward that goal. Initially, concerns were expressed about the feasibility of recording fish sounds in this region due to their low propagation distance (De Jong et al., 2007; Ladich, 2004) and the lower fish diversity compared to less deteriorated tropical ecosystems (Froese and Pauly, 2002). Additionally, doubts arose regarding the possibility of training machine learning models due to the limited availability of annotated data. Nevertheless, in this project, fish sounds in the North Sea were successfully recorded, manually annotated, and automatically detected. During the project, new studies have also confirmed the feasibility of recording fish sounds in the North Sea region (Watson et al., 2024) and have combined unsupervised machine-learning techniques with active learning for efficient data annotation (Parcerisas et al., 2024). When combined with automated data processing, PAM provides significant advantages for ecological monitoring. Unlike eDNA analysis, video surveys, or diver-based visual assessments, PAM enables continuous spatial and temporal monitoring under conditions where other methods would be unapplicable, and at potentially lower costs.

While PAM is a powerful tool, it is inherently limited to detecting soniferous species. Studies have proposed using fish sounds as environmental proxies for habitat biodiversity monitoring (Di Iorio et al., 2018) and further research is required to validate this approach in different marine environments. To enhance flexible monitoring capabilities, future efforts should focus on deploying multi-sensor systems with complementary applications (Cabrito et al., 2024) at strategic locations such as artificial reefs, migratory corridors, biodiversity hotspots, and marine protected areas.

Future perspectives

Development of passive acoustic monitoring

With the main bottlenecks being addressed, the development of PAM for fish and marine invertebrates sounds is nearing an acceleration point, whereafter application is going to increase drastically. PAM will then become a major asset in monitoring marine biodiversity, especially when combined with other monitoring techniques in multi-sensor systems. Future development pathways are outlined in Figure 29 and discussed below, except for improvements to Agile Modelling, which were covered previously.

Multi-sensors integration

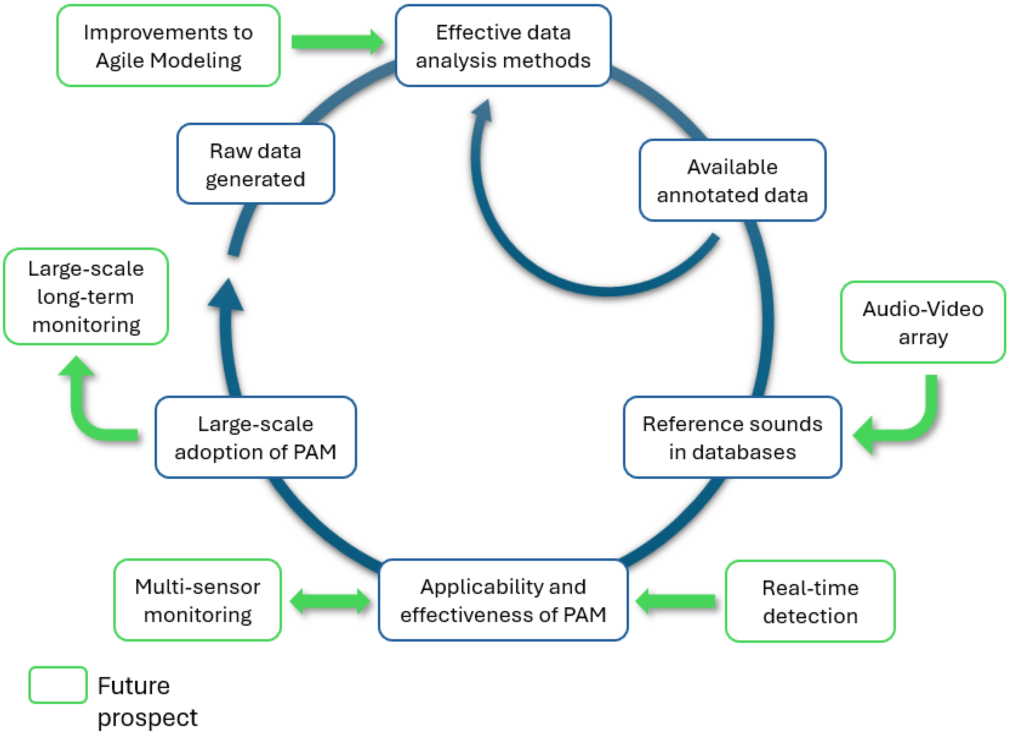


Figure 29: Future perspective for the development and application of PAM for fish and marine invertebrates

Insights from this project highlight the promise of combining audio and video recordings for ecological monitoring. While audio-video arrays have successfully identified new reference sounds in healthy tropical reefs, their effectiveness in temperate or degraded ecosystems, such as the North Sea, remains untested. These environments pose unique challenges, including lower marine animal diversity and abundance, and higher turbidity, which may limit observable events for sound identification and reference collection. Current methods rely heavily on cherry-picking high-quality events, where a sound of interest matches with a sufficiently clear visual capture, enabling identification. In temperate regions, such events are likely rarer, potentially complicating or compromising the use of the video-audio array method for reference sound collection. Further research is needed to adapt these approaches for broader ecological contexts. A key limitation to the number of events recorded with audio-video monitoring is that deployment duration is constrained by the battery life of cameras (Dantzker et al., 2024; Mouy et al., 2023). In contrast, sound monitoring is highly energy-efficient, with recorders capable of running for months or even years, depending on duty cycles, sampling rates, and battery capacity (D'Eu et al., 2012). To overcome this limitation, the development of the embedded system (WP3) with real-time onboard processing could enable cameras to be activated based on real-time sound detection. This event-triggered approach would significantly reduce camera usage, extending battery life and allowing for longer deployments. As a result, more marine animal sounds could be captured in a single deployment, enhancing the chances of capturing high-quality events.

Monitoring applications and implications

The necessity for long-term ecological monitoring in the North Sea presents an opportunity where PAM can play a crucial role. Offshore wind farm construction is re-introducing artificial hard structures that are expected to enhance biodiversity. This development, however, still needs to be followed over several years before conclusions can be drawn about the ecological impact. The ability of PAM to operate over long temporal scales makes it a particularly suitable method for long-term studies. The integration of PAM with complementary methods, such as energy-efficient triggered video recording and, in the future, targeted eDNA sampling, enables multimodal data collection through systems like the *Biodiversity Sensing Box*. When enhanced with automated data processing, such replicable systems can support robust and scalable ecological assessments of marine environments. This approach will be developed and applied within the context of the NWA-funded initiative, Hybrid Labs[26], to assess the impact of floating offshore energy-producing structures on biodiversity and the potential multi-use of the area. Long-term monitoring of these installations can provide valuable insights for designing nature-inclusive infrastructure, improving understanding of species ecology, and informing conservation strategies and stakeholders of the developments underwater.

Beyond ecological monitoring, the data collected through PAM and multi-sensor systems can be utilised to raise awareness and communicate conservation efforts. Marine ecosystems play a vital role in supporting human society and global ecological stability, yet they often receive less attention than terrestrial environments due to their inaccessibility. To address this, the real-time monitoring dashboard (WP4a) is ready for implementation and will display sound and video data once the Jetson Nano is integrated into the embedded system (WP3). This autonomous system will provide real-time visual and acoustic insights into underwater ecosystems, observable from a dashboard (WP4a), making them more accessible to the public and policymakers. By increasing visibility and understanding of marine environments, such monitoring initiatives can help bridge the knowledge gaps and facilitate communication among stakeholders

[26] https://hybridlabs.tudelft.nl/

about the pros and cons of offshore floating energy structures and opportunities for nature-inclusive design.

6 Conclusion

The objective of this project was to improve the efficiency, accuracy and applicability of passive acoustic monitoring (PAM) techniques for underwater biodiversity monitoring and for the integration of PAM into a multi-sensor deployable system.

During the project, underwater acoustic data were collected through deployments and collaborations, and a regional data-sharing initiative was started. The collected data were organised into a local repository to support the development of automatic marine animal sound detectors using an innovative approach, tailored to address the scarcity of annotated data. The development of an autonomous recorder for sound and video was undertaken, and the system is expected to be enhanced with onboard processing capabilities. Additionally, a dashboard was developed to display the acquired acoustic data in real time, along with automatically detected sounds of interest. Integrating the design in the context of the current state of research, two positive feedback loops were identified that drive PAM advancements for marine fish and invertebrates. These positive feedback loops were constrained by three main bottlenecks: the high cost of manual data analysis and limited access to automated methods, the lack of openly available annotated datasets, and the scarcity of reference sounds.

This work demonstrates how novel deep-learning methods can rapidly train automatic detectors for various sounds in diverse marine environments. Developed detectors streamline and accelerate the collection of annotated data, supporting continuous improvement of deep learning models. Moreover, Agile Modelling only requires one annotated sound to initiate training and demands minimal programming and machine learning expertise. Consequently, this approach significantly improves accessibility for bioacousticians to efficient data analysis methods. Future efforts will aim to further enhance accessibility for non-programmers by developing a graphical user interface.

In addition to data analysis methods and available annotated data, addressing gaps in reference sound databases is critical for advancing PAM. Two novel and promising methods exist to gather rapidly reference sounds: cross-referencing unidentified sounds from multiple locations and deploying integrated acoustic-video arrays in the field. In this project, a step was made toward using both methods in the North Sea region for the first time: a collaborative database was initiated to share unidentified sounds focused on the North Sea, and an autonomous recorder for sound plus video was developed. Real-time detection of fish sounds from playback recordings was achieved in this project to test the approach; additional efforts are required for integration of the detectors in the sound recorder for application in situ under water. More sounds need to be gathered in different locations of the North Sea, and the applicability of acoustic-video arrays should be tested in deteriorated as well as intact temperate environments. Both cross-referencing and acoustic-video array approaches will benefit from improved automatic detection, as sound annotation remains a prerequisite for their application.

Combining sound and activity-triggered video sensors is promising for extending deployment time and improving monitoring quality by providing complementary data in different environmental conditions. In contrast, the state of the art in both eDNA monitoring and automated sound detection for marine animal techniques is concluded not yet suitable to be synergistically combined for most applications.

By addressing key bottlenecks in PAM, this project lays the foundation for more effective and widespread adoption of the technique and facilitates its integration into multi-sensor systems to advance marine ecosystem monitoring. In the future, multi-sensor systems will undoubtedly play a key role in monitoring marine ecosystems due to their ability to provide complementary data from multiple sources, enhance detection accuracy, extend monitoring capabilities across various environmental conditions, and the possibility to replicate them to increase temporal and spatial monitoring scale. Beyond advancing marine science knowledge, monitoring systems with real-time observatories can increase public awareness about and interest in hard-to-visit environments. Ultimately, improved ecological data and greater visibility will enhance support for conservation and restoration efforts and sustainable management of marine biodiversity in relation to human activities in the ocean.

References

Almunia, J., Fernández-Maquieira, M., Flores, M., 2024. Aquariums as Research Platforms: Characterizing Fish Sounds in Controlled Settings with Preliminary Insights from the Blackbar Soldierfish Myripristis jacobus. J. Zool. Bot. Gard. 5, 630–640. https://doi.org/10.3390/jzbg5040042

Amorim, M.C.P., Vasconcelos, R.O., Fonseca, P.J., 2015. Fish Sounds and Mate Choice, in: Ladich, F. (Ed.), Sound Communication in Fishes. Springer, Vienna, pp. 1–33. https://doi.org/10.1007/978-3-7091-1846-7_1

Amorim, M.C.P., Wanjala, J.A., Vieira, M., Bolgan, M., Connaughton, M.A., Pereira, B.P., Fonseca, P.J., Ribeiro, F., 2023. Detection of invasive fish species with passive acoustics: Discriminating between native and non-indigenous sciaenids. Mar. Environ. Res. 188, 106017. https://doi.org/10.1016/j.marenvres.2023.106017

Baevski, A., Zhou, Y., Mohamed, A., Auli, M., 2020. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations, in: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 12449–12460.

Baker, E., Vincent, S., 2019. A deafening silence: a lack of data and reproducibility in published bioacoustics research? Biodivers. Data J. 7, e36783. https://doi.org/10.3897/BDJ.7.e36783

Bergler, C., Smeele, S.Q., Tyndel, S.A., Barnhill, A., Ortiz, S.T., Kalan, A.K., Cheng, R.X., Brinkløv, S., Osiecka, A.N., Tougaard, J., Jakobsen, F., Wahlberg, M., Nöth, E., Maier, A., Klump, B.C., 2022. ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. Sci. Rep. 12, 21966. https://doi.org/10.1038/s41598-022-26429-y

Bolgan, M., Parmentier, E., Picciulin, M., Hadjioannou, L., Di Iorio, L., 2023. Use of passive acoustic monitoring to fill knowledge gaps of fish global conservation status. Aquat. Conserv. Mar. Freshw. Ecosyst. 33, 1580–1589. https://doi.org/10.1002/aqc.4020

Cabrito, A., Maynou, F., Simide, R., Mouillot, D., Lossent, J., de Juan, S., 2024. Non-extractive fish diversity assessment in Mediterranean rhodolith beds. Aquat. Conserv. Mar. Freshw. Ecosyst. 34, e4212. https://doi.org/10.1002/aqc.4212

Caldwell, I.R., Hobbs, J.-P.A., Bowen, B.W., Cowman, P.F., DiBattista, J.D., Whitney, J.L., Ahti, P.A., Belderok, R., Canfield, S., Coleman, R.R., Iacchei, M., Johnston, E.C., Knapp, I., Nalley, E.M., Staeudle, T.M., Láruson, Á.J., 2024. Global trends and biases in biodiversity conservation research. Cell Rep. Sustain. 1. https://doi.org/10.1016/j.crsus.2024.100082

Chérubin, L.M., Dalgleish, F., Ibrahim, A.K., Schärer-Umpierre, M., Nemeth, R.S., Matthews, A., Appeldoorn, R., 2020. Fish Spawning Aggregations Dynamics as Inferred From a Novel, Persistent Presence Robotic Approach. Front. Mar. Sci. 6.

Clark, F.E., Dunn, J.C., 2022. From Soundwave to Soundscape: A Guide to Acoustic Research in Captive Animal Environments. Front. Vet. Sci. 9. https://doi.org/10.3389/fvets.2022.889117

Cox, K.D., Looby, A., Vela, S., Riera, A., Bravo, S., Davies, H.L., Rountree, R., Spriel, B., Reynolds, L.K., Martin, C.W., Matwin, S., Juanes, F., 2023. FishSounds Version 1.1: Data Archive, User Experience, and Online Resources, in: Popper, A.N., Sisneros, J., Hawkins, A.D., Thomsen, F. (Eds.), The Effects of Noise on Aquatic Life. Springer International Publishing, Cham, pp. 1–12. https://doi.org/10.1007/978-3-031-10417-6_35-1

Dantzker, M.S., Duggan, M.T., Berlik, E., Delikaris-Manias, S., Bountourakis, V., Pulkki, V., Rice, A.N., 2024. Who said that? Deciphering Complex Coral Reef Soundscapes with Spatial Audio and 360° Video. https://doi.org/10.1101/2024.12.16.628659

Darras, K.F.A., Pérez, N., -, M., Dilong, L., Hanf-Dressler, T., Markolf, M., Wanger, T.C., 2023. ecoSound-web: an open-source, online platform for ecoacoustics. F1000Research 9, 1224. https://doi.org/10.12688/f1000research.26369.2

De Jong, K., Bouton, N., Slabbekoorn, H., 2007. Azorean rock-pool blennies produce size-dependent calls in a courtship context. Anim. Behav. 74, 1285–1292. https://doi.org/10.1016/j.anbehav.2007.02.023

Di Iorio, L., Raick, X., Parmentier, E., Boissery, P., Valentini-Poirier, C.-A., Gervaise, C., 2018. ‘Posidonia meadows calling’: a ubiquitous fish sound with monitoring potential. Remote Sens. Ecol. Conserv. 4, 248–263. https://doi.org/10.1002/rse2.72

Doorenspleet, K., Jansen, L., Oosterbroek, S., Kamermans, P., Bos, O., Wurz, E., Murk, A., Nijland, R., 2025. The Long and the Short of It: Nanopore-Based eDNA Metabarcoding of Marine Vertebrates Works; Sensitivity and Species-Level Assignment Depend on Amplicon Lengths. Mol. Ecol. Resour. 25, e14079. https://doi.org/10.1111/1755-0998.14079

Dufourq, E., Batist, C., Foquet, R., Durbach, I., 2022. Passive acoustic monitoring of animal populations with transfer learning. Ecol. Inform. 70, 101688. https://doi.org/10.1016/j.ecoinf.2022.101688

Erbe, C., Reichmuth, C., Cunningham, K., Lucke, K., Dooling, R., 2016. Communication masking in marine mammals: A review and research strategy. Mar. Pollut. Bull. 103, 15–38. https://doi.org/10.1016/j.marpolbul.2015.12.007

Froese, R., Pauly, D., 2002. FishBase: A Global Information System on Fishes [WWW Document]. World Wide Web Electron. Publ. Wwwfishbaseorg. URL https://www.fishbase.in/home.htm (accessed 4.20.23).

Ghani, B., Denton, T., Kahl, S., Klinck, H., 2023. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Sci. Rep. 13, 22876. https://doi.org/10.1038/s41598-023-49989-z

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185. https://doi.org/10.1111/2041-210X.13101

Guyot, P., Alix, F., Guerin, T., Lambeaux, E., Rotureau, A., 2021. Fish migration monitoring from audio detection with CNNs, in: Proceedings of the 16th International Audio Mostly Conference, AM ’21. Association for Computing Machinery, New York, NY, USA, pp. 244–247. https://doi.org/10.1145/3478384.3478393

Hagiwara, M., 2022. AVES: Animal Vocalization Encoder based on Self-Supervision.

Hamer, J., Triantafillou, E., Merriënboer, B. van, Kahl, S., Klinck, H., Denton, T., Dumoulin, V., 2023. BIRB: A Generalization Benchmark for Information Retrieval in Bioacoustics. https://doi.org/10.48550/arXiv.2312.07439

Hatch, L., Barkowski, J., Baumann-Pickering, S., Bell, C.W., Best, B., Brown, J., DeAngelis, A., Joseph, J., Kim, E., Kitchen, D., Kok, A.C.M., Kügler, A., Kumar, A., Lammers, M.O., Margolina, T., McKenna, M., Merkens, K., Reeves, L.P., Rowell, T.J., Shoemaker, M., Stanley, J.A., Stimpert, A., Southall, B., Stone, B., Tartt, M., Van Parijs, S., Zang, E.J., 2024. SanctSound: Building Data Systems for Sound Decisions, in: Popper, A.N., Sisneros, J.A., Hawkins, A.D., Thomsen, F. (Eds.), The Effects of Noise on Aquatic Life. Springer International Publishing, Cham, pp. 1547–1557. https://doi.org/10.1007/978-3-031-50256-9_58

Hwang, Y., Cho, H., Yang, H., Won, D.-O., Oh, I., Lee, S.-W., 2020. Mel-spectrogram augmentation for sequence to sequence voice conversion. https://doi.org/10.48550/arXiv.2001.01401

Ibrahim, A.K., Chérubin, L.M., Zhuang, H., Schärer Umpierre, M.T., Dalgleish, F., Erdol, N., Ouyang, B., Dalgleish, A., 2018. An approach for automatic classification of grouper vocalizations with passive acoustic monitoring. J. Acoust. Soc. Am. 143, 666–676. https://doi.org/10.1121/1.5022281

Ibrahim, A.K., Zhuang, H., Schärer-Umpierre, M., Woodward, C., Erdol, N., Chérubin, L.M., 2024. Fish Acoustic Detection Algorithm Research: a deep learning app for Caribbean grouper calls detection and call types classification. Front. Mar. Sci. 11. https://doi.org/10.3389/fmars.2024.1378159

Jarriel, S., Formel, N., Ferguson, S., Jensen, F., Apprill, A., Mooney, A., 2024a. Unidentified fish sounds as indicators of coral reef health and comparison to other acoustic methods. Front. Remote Sens. 5. https://doi.org/10.3389/frsen.2024.1338586

Jarriel, S., Looby, A., Larsen, K., Munger, J., Parsons, M., 2024b. Compilation of existing underwater PAM repositories, libraries, and applications for sound processing. https://doi.org/10.5281/zenodo.12096525

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Inform. 61, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236
Ladich, F., 2004. Sound Production and Acoustic Communication, in: von der Emde, G., Mogdans, J., Kapoor, B.G. (Eds.), The Senses of Fish: Adaptations for the Reception of Natural Stimuli. Springer Netherlands, Dordrecht, pp. 210–230. https://doi.org/10.1007/978-94-007-1060-3_10
Laplante, J.-F., Akhloufi, M.A., Gervaise, C., 2022. Deep Learning for Marine Bioacoustics and Fish Classification Using Underwater Sounds, in: 2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). Presented at the 2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 288–293. https://doi.org/10.1109/CCECE49351.2022.9918242
Laplante, J.-F., Akhloufi, M.A., Gervaise, C., 2021. Fish recognition in underwater environments using deep learning and audio data, in: Ocean Sensing and Monitoring XIII. Presented at the Ocean Sensing and Monitoring XIII, SPIE, pp. 97–102. https://doi.org/10.1117/12.2585991
Li, P., Liu, X., Palmer, K.J., Fleishman, E., Gillespie, D., Nosal, E.-M., Shiu, Y., Klinck, H., Cholewiak, D., Helble, T., Roch, M.A., 2020. Learning Deep Models from Synthetic Data for Extracting Dolphin Whistle Contours, in: 2020 International Joint Conference on Neural Networks (IJCNN). Presented at the 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–10. https://doi.org/10.1109/IJCNN48605.2020.9206992
Liang, J., Nolasco, I., Ghani, B., Phan, H., Benetos, E., Stowell, D., 2024. Mind the Domain Gap: a Systematic Analysis on Bioacoustic Sound Event Detection. https://doi.org/10.48550/arXiv.2403.18638
Lindseth, A.V., Lobel, P.S., 2018. Underwater Soundscape Monitoring and Fish Bioacoustics: A Review. Fishes 3, 36. https://doi.org/10.3390/fishes3030036
Looby, A., Cox, K., Bravo, S., Rountree, R., Juanes, F., Reynolds, L.K., Martin, C.W., 2022. A quantitative inventory of global

soniferous fish diversity. Rev. Fish Biol. Fish. 32, 581–595. https://doi.org/10.1007/s11160-022-09702-1
Looby, A., Cox, K., Bravo, S., Rountree, R., Juanes, F., Riera, A., Vela, S., Davies, H.L., Reynolds, L.K., Martin, C.W., 2023a. Fish Sound Production Research: Historical Practices and Ongoing Challenges, in: Popper, A.N., Sisneros, J., Hawkins, A.D., Thomsen, F. (Eds.), The Effects of Noise on Aquatic Life. Springer International Publishing, Cham, pp. 1–20. https://doi.org/10.1007/978-3-031-10417-6_92-1
Looby, A., Vela, S., Cox, K., Riera, A., Bravo, S., Davies, H.L., Rountree, R., Reynolds, L.K., Martin, C.W., Matwin, S., Juanes, F., 2023b. FishSounds Version 1.0: A website for the compilation of fish sound production information and recordings. Ecol. Inform. 74, 101953. https://doi.org/10.1016/j.ecoinf.2022.101953
Manifesto for Agile Software Development, n.d.
Mesaros, A., Heittola, T., Diment, A., Elizalde, B., Shah, A., Vincent, E., Raj, B., Virtanen, T., 2017. DCASE 2017 Challenge setup: Tasks, datasets and baseline system. Presented at the DCASE 2017 - Workshop on Detection and Classification of Acoustic Scenes and Events.
Miksis-Olds, J.L., Watts, A., 2019. Better together—Combining acoustics and environmental DNA to understand ecosystems. J. Acoust. Soc. Am. 145, 1654–1654. https://doi.org/10.1121/1.5101078
Monczak, A., Ji, Y., Soueidan, J., Montie, E.W., 2019. Automatic detection, classification, and quantification of sciaenid fish calls in an estuarine soundscape in the Southeast United States. PLOS ONE 14, e0209914. https://doi.org/10.1371/journal.pone.0209914
Mooney, T.A., Di Iorio, L., Lammers, M., Lin, T.-H., Nedelec, S.L., Parsons, M., Radford, C., Urban, E., Stanley, J., 2020. Listening forward: approaching marine biodiversity assessments using acoustic methods. R. Soc. Open Sci. 7, 201287. https://doi.org/10.1098/rsos.201287
Morfi, V., Nolasco, I., Lostanlen, V., Singh, S., Strandburg-Peshkin, A., Gill, L., Pamula, H., Benvent, D., Stowell, D., 2021. Few-shot

bioacoustic event detection : A new task at the DCASE 2021 challenge.
Moummad, I., Farrugia, N., Serizel, R., n.d. SELF-SUPERVISED LEARNING FOR FEW-SHOT BIRD SOUND CLASSIFICATION.
Mouy, X., Archer, S.K., Dosso, S., Dudas, S., English, P., Foord, C., Halliday, W., Juanes, F., Lancaster, D., Van Parijs, S., Haggarty, D., 2024. Automatic detection of unidentified fish sounds: a comparison of traditional machine learning with deep learning. Front. Remote Sens. 5. https://doi.org/10.3389/frsen.2024.1439995
Mouy, X., Black, M., Cox, K., Qualley, J., Dosso, S., Juanes, F., 2023. Identification of fish sounds in the wild using a set of portable audio-video arrays. Methods Ecol. Evol. 14, 2165–2186. https://doi.org/10.1111/2041-210X.14095
Napier, T., Ahn, E., Allen-Ankins, S., Schwarzkopf, L., Lee, I., 2025. LEAVES: An open-source web-based tool for the scalable annotation and visualisation of large-scale ecoacoustic datasets using cluster analysis. Ecol. Inform. 103026. https://doi.org/10.1016/j.ecoinf.2025.103026
Napier, T., Ahn, E., Allen-Ankins, S., Schwarzkopf, L., Lee, I., 2024. Advancements in preprocessing, detection and classification techniques for ecoacoustic data: A comprehensive review for large-scale passive acoustic monitoring. Expert Syst. Appl. 124220. https://doi.org/10.1016/j.eswa.2024.124220
Nolasco, I., Singh, S., Morfi, V., Lostanlen, V., Strandburg-Peshkin, A., Vidaña-Vila, E., Gill, L., Pamuła, H., Whitehead, H., Kiskin, I., Jensen, F.H., Morford, J., Emmerson, M.G., Versace, E., Grout, E., Liu, H., Ghani, B., Stowell, D., 2023. Learning to detect an animal sound from five examples. Ecol. Inform. 77, 102258. https://doi.org/10.1016/j.ecoinf.2023.102258
O'Hara, C.C., Frazier, M., Valle, M., Butt, N., Kaschner, K., Klein, C., Halpern, B.S., 2024. Cumulative human impacts on global marine fauna highlight risk to biological and functional diversity. PLOS ONE 19, e0309788. https://doi.org/10.1371/journal.pone.0309788
Palumbi, S.R., Sandifer, P.A., Allan, J.D., Beck, M.W., Fautin, D.G., Fogarty, M.J., Halpern, B.S., Incze, L.S., Leong, J.-A., Norse, E., Stachowicz, J.J., Wall, D.H., 2009. Managing for ocean

biodiversity to sustain marine ecosystem services. Front. Ecol. Environ. 7, 204–211. https://doi.org/10.1890/070135

Parcerisas, C., Botteldooren, D., Devos, P., Hamard, Q., Debusschere, E., 2023a. Studying the Soundscape of Shallow and Heavy Used Marine Areas: Belgian Part of the North Sea, in: Popper, A.N., Sisneros, J., Hawkins, A.D., Thomsen, F. (Eds.), The Effects of Noise on Aquatic Life: Principles and Practical Considerations. Springer International Publishing, Cham, pp. 1–27. https://doi.org/10.1007/978-3-031-10417-6_122-1

Parcerisas, C., Roca, I.T., Botteldooren, D., Devos, P., Debusschere, E., 2023b. Categorizing Shallow Marine Soundscapes Using Explained Clusters. J. Mar. Sci. Eng. 11, 550. https://doi.org/10.3390/jmse11030550

Parcerisas, C., Schall, E., te Velde, K., Botteldooren, D., Devos, P., Debusschere, E., 2024. Machine learning for efficient segregation and labeling of potential biological sounds in long-term underwater recordings. Front. Remote Sens. 5. https://doi.org/10.3389/frsen.2024.1390687

Park, D.S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E.D., Le, Q.V., 2019. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition, in: Interspeech 2019. pp. 2613–2617. https://doi.org/10.21437/Interspeech.2019-2680

Parsons, M.J.G., Lin, T.-H., Mooney, T.A., Erbe, C., Juanes, F., Lammers, M., Li, S., Linke, S., Looby, A., Nedelec, S.L., Van Opzeeland, I., Radford, C., Rice, A.N., Sayigh, L., Stanley, J., Urban, E., Di Iorio, L., 2022. Sounding the Call for a Global Library of Underwater Biological Sounds. Front. Ecol. Evol. 10.

Parsons, M.J.G., Looby, A., Chanda, K., Di Iorio, L., Erbe, C., Frazao, F., Havlik, M., Juanes, F., Lammers, M.O., Li, S., Liffers, M., Lin, T.-H., Linke, S., Mooney, T.A., Radford, C., Rice, A.N., Rountree, R., Sayigh, L.S., Sousa-Lima, R.S., Stanley, J.A., Thomisch, K., Urban, E.R., van Zeeland, L., Vela, S., Zuffi, S., Nedelec, S.L., 2024. A Global Library of Underwater Biological Sounds (GLUBS): An Online Platform with Multiple Passive Acoustic Monitoring Applications, in: Popper, A.N., Sisneros, J.A., Hawkins, A.D., Thomsen, F. (Eds.), The Effects of Noise on

Aquatic Life: Principles and Practical Considerations. Springer International Publishing, Cham, pp. 2149–2173. https://doi.org/10.1007/978-3-031-50256-9_123
Picciulin, M., Kéver, L., Parmentier, E., Bolgan, M., 2019. Listening to the unseen: Passive acoustic monitoring reveals the presence of a cryptic fish species. Aquat. Conserv. Mar. Freshw. Ecosyst. 29, 202–210. https://doi.org/10.1002/aqc.2973
Pyć, C.D., Vallarta, J., Rice, A.N., Zeddies, D.G., Maxner, E.E., Denes, S.L., 2021. Vocal behavior of the endangered splendid toadfish and potential masking by anthropogenic noise. Conserv. Sci. Pract. 3, e352. https://doi.org/10.1111/csp2.352
Qian, K., Zhang, Z., Baird, A., Schuller, B., 2017. Active learning for bird sound classification via a kernel-based extreme learning machine. J. Acoust. Soc. Am. 142, 1796–1804. https://doi.org/10.1121/1.5004570
Rice, A.N., Farina, S.C., Makowski, A.J., Kaatz, I.M., Lobel, P.S., Bemis, W.E., Bass, A.H., 2022. Evolutionary Patterns in Sound Production across Fishes. Ichthyol. Herpetol. 110, 1–12. https://doi.org/10.1643/i2020172
Rishan, S.T., Kline, R.J., Rahman, M.S., 2023. Applications of environmental DNA (eDNA) to detect subterranean and aquatic invasive species: A critical review on the challenges and limitations of eDNA metabarcoding. Environ. Adv. 12, 100370. https://doi.org/10.1016/j.envadv.2023.100370
Robinson, D., Miron, M., Hagiwara, M., Pietquin, O., 2024a. NatureLM-audio: an Audio-Language Foundation Model for Bioacoustics.
Robinson, D., Robinson, A., Akrapongpisak, L., 2024b. Transferable Models for Bioacoustics with Human Language Supervision, in: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1316–1320. https://doi.org/10.1109/ICASSP48485.2024.10447250
Rodrigues, J.G., Conides, A., Rodriguez, S.R., Raicevich, S., Pita, P., Kleisner, K., Pita, C., Lopes, P., Roldán, V.A., Ramos, S., Klaoudatos, D., Outeiro, L., Armstrong, C., Teneva, L., Stefanski, S., Böhnke-Henrichs, A., Kruse, M., Lillebø, A.,

Bennett, E., Belgrano, A., Murillas, A., Pinto, I.S., Burkhard, B., Villasante, S., 2017. Marine and Coastal Cultural Ecosystem Services: knowledge gaps and research priorities. One Ecosyst. 2, e12290. https://doi.org/10.3897/oneeco.2.e12290

Schall, E., Kaya, I.I., Debusschere, E., Devos, P., Parcerisas, C., 2024. Deep learning in marine bioacoustics: a benchmark for baleen whale detection. Remote Sens. Ecol. Conserv. 10, 642–654. https://doi.org/10.1002/rse2.392

Solé, M., Kaifu, K., Mooney, T.A., Nedelec, S.L., Olivier, F., Radford, A.N., Vazzana, M., Wale, M.A., Semmens, J.M., Simpson, S.D., Buscaino, G., Hawkins, A., Aguilar de Soto, N., Akamatsu, T., Chauvaud, L., Day, R.D., Fitzgibbon, Q., McCauley, R.D., André, M., 2023. Marine invertebrates and noise. Front. Mar. Sci. 10. https://doi.org/10.3389/fmars.2023.1129057

Souza Jr, P.M., Olsen, Z., Brandl, S.J., 2023. Paired passive acoustic and gillnet sampling reveal the utility of bioacoustics for monitoring fish populations in a turbid estuary. ICES J. Mar. Sci. 80, 1240–1255. https://doi.org/10.1093/icesjms/fsad085

Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10, e13152. https://doi.org/10.7717/peerj.13152

Suarez-Bregua, P., Álvarez-González, M., Parsons, K.M., Rotllant, J., Pierce, G.J., Saavedra, C., 2022. Environmental DNA (eDNA) for monitoring marine mammals: Challenges and opportunities. Front. Mar. Sci. 9. https://doi.org/10.3389/fmars.2022.987774

Sugai, L.S.M., Silva, T.S.F., Ribeiro, J.W., Jr, Llusia, D., 2019. Terrestrial Passive Acoustic Monitoring: Review and Perspectives. BioScience 69, 15–25. https://doi.org/10.1093/biosci/biy147

Tavares, T.F., 2022. Open-set classification approaches to automatic bird song identification: towards non-invasive wildlife monitoring in Brazilian fauna. IEEE Lat. Am. Trans. 20, 2388–2394. https://doi.org/10.1109/TLA.2022.9904764

Urazghildiiev, I.R., Van Parijs, S.M., 2016. Automatic grunt detector and recognizer for Atlantic cod (Gadus morhua). J. Acoust. Soc. Am. 139, 2532–2540. https://doi.org/10.1121/1.4948569

van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., Denton, T., 2024. Birds, bats and beyond: evaluating generalization in bioacoustics models. Front. Bird Sci. 3. https://doi.org/10.3389/fbirs.2024.1369756
Vellinga, W.-P., Planque, R., n.d. The Xeno-canto collection and its relation to sound recognition and classification.
Vieira, M., Ríos, N., Muñoz-Duque, S., Pereira, J., Carriço, R., Fernandez, M., Monteiro, J.G., Pais, M.P., Quintella, B.R., Silva, G., Silva, R.P., Fonseca, P.J., Amorim, M.C.P., 2024. Cross-referencing unidentified fish sound data sets to unravel sound sources: a case study from the Temperate Northern Atlantic. Front. Remote Sens. 5. https://doi.org/10.3389/frsen.2024.1377206
Waddell, E.E., Rasmussen, J.H., Širović, A., 2021. Applying Artificial Intelligence Methods to Detect and Classify Fish Calls from the Northern Gulf of Mexico. J. Mar. Sci. Eng. 9, 1128. https://doi.org/10.3390/jmse9101128
Watson, M.S., Kok, A.C.M., Opzeeland, I.V., Eriksson, B.K., 2024. The biological soundscape of temperate reefs: characterizing natural and artificial reefs in the Dutch Wadden Sea. https://doi.org/10.21203/rs.3.rs-4593657/v1
Williams, B., van Merriënboer, B., Dumoulin, V., Hamer, J., Triantafillou, E., Fleishman, A.B., McKown, M., Munger, J.E., Rice, A.N., Lillis, A., White, C.E., Hobbs, C.A.D., Razak, T.B., Jones, K.E., Denton, T., n.d. Leveraging tropical reef, bird and unrelated sounds for superior transfer learning in marine bioacoustics.
Wilson, K.C., Semmens, B.X., Gittings, S.R., Pattengill-Semmens, C., Širović, A., 2019. Development and evaluation of a passive acoustic localization method to monitor fish spawning aggregations and measure source levels, in: OCEANS 2019 MTS/IEEE SEATTLE. Presented at the OCEANS 2019 MTS/IEEE SEATTLE, pp. 1–7. https://doi.org/10.23919/OCEANS40490.2019.8962663
Yu, H., Amador, G.J., Cribellier, A., Klaassen, M., de Knegt, H.J., Naguib, M., Nijland, R., Nowak, L., Prins, H.H.T., Snijders, L., Tyson, C., Muijres, F.T., 2024. Edge computing in wildlife

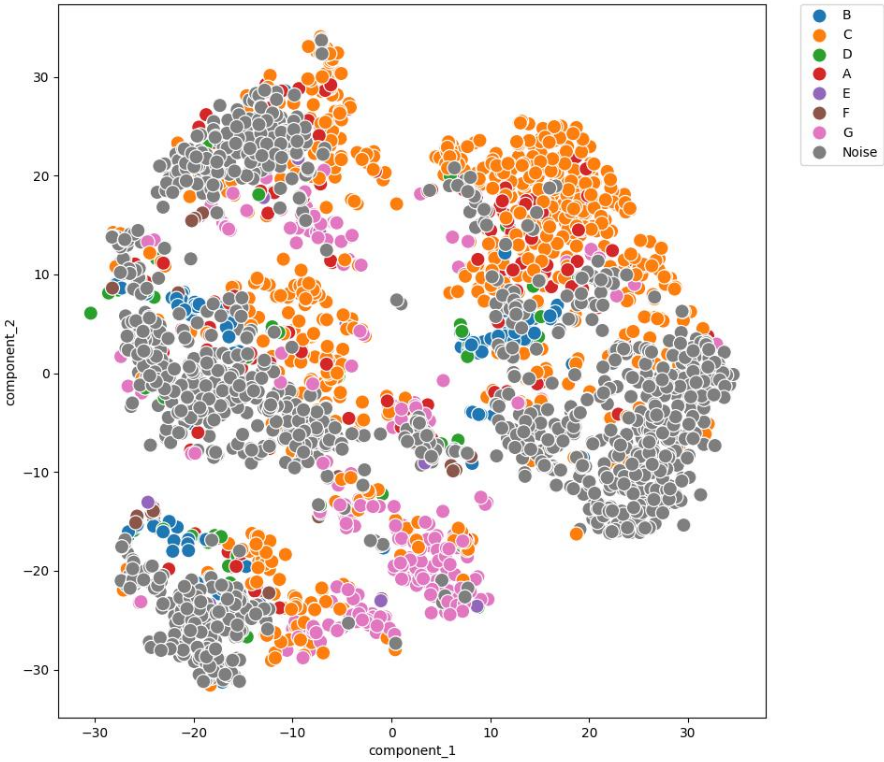
behavior and ecology. Trends Ecol. Evol. 39, 128–130. https://doi.org/10.1016/j.tree.2023.11.014

Appendices

Appendix A. Preliminary results on classification of fish sounds

This appendix shows some preliminary results from experiments using different models for the detection of fish sounds in recordings from Grey Reef, recorded within the SanctSound project (Hatch et al., 2024).

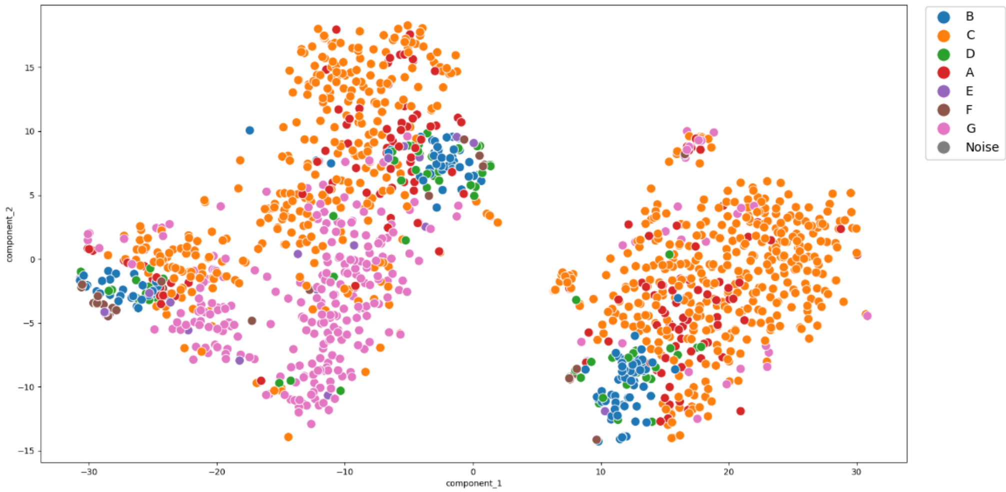
Separability of the data based on a dimensional reduction of the embeddings from AVES using TSNE



Appendix A Figure 1: TSNE of the embeddings from AVES. The letters represent different types of fish sounds.

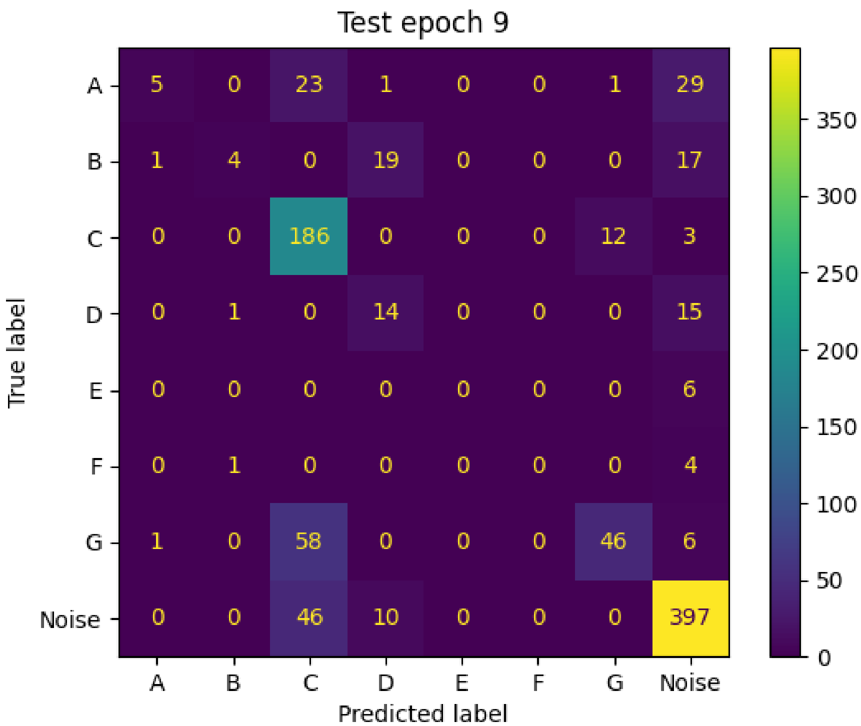
While some clusters can be observed, the separation is not nearly perfect, indicating that some noises are not distinguishable from calls by the model and vice versa.

Removing the noise, a confusion between different types of calls can also be observed. While the call type B is well grouped, A and C overlap mostly, indicating that the model does not make a difference between the two types of calls. This could be due to the similarity of the calls but could also to confusion of the annotator between the



Appendix A Figure 1: TSNE of the embeddings of fish calls. The letters represent different types of fish sounds.

Confusion matrix of the results of AVES model to classify different type of fish sounds.

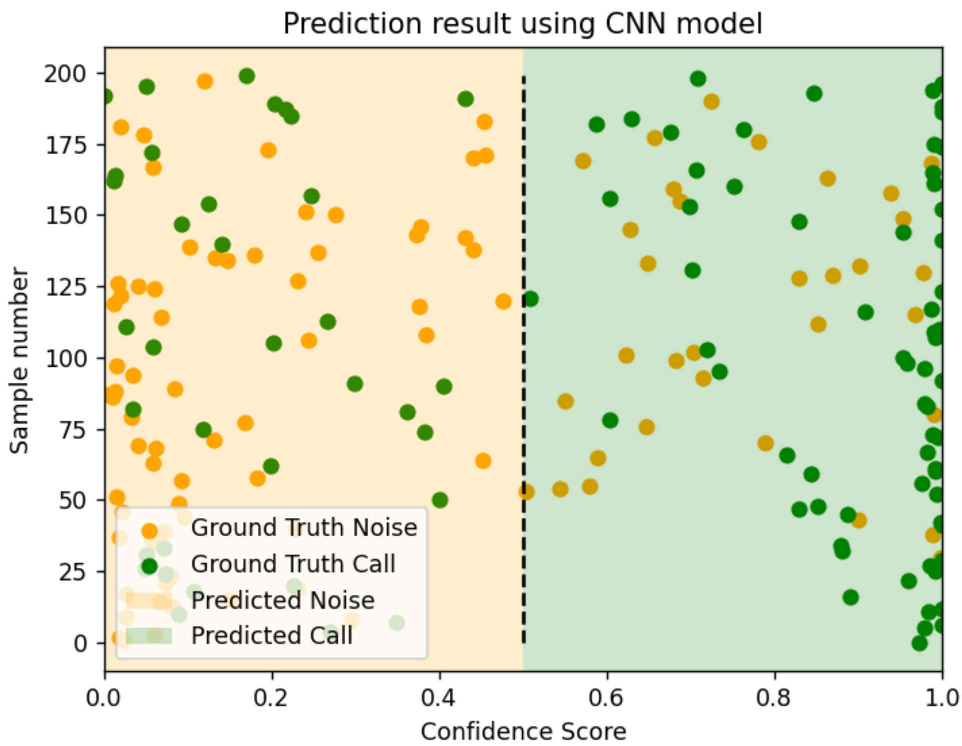


Appendix A Figure 2: Results of a classification on a validation set using AVES to separate different types of fish sounds

As expected from the visualisation of the embeddings above, the classifier displays some confusion between call and noise and between noise and call. Numerous call type A are classified as C. The classifier displays good performance for classifying the type C sound, which is most abundant in the training set, but also confuses other sounds, such as the G and Noise types, as C. This is likely the result of the imbalance in the dataset used for training, which overrepresents the more abundant type C sounds.

Binary classification using a Convolutional Neural Network (CNN)

Figure 4 from Appendix 1 shows that most sound and noise snippets were successfully classified with high confidence (green markers with a score close to 1 and orange markers with a score close to 0). However, many samples were also misclassified. After a qualitative



Appendix A Figure 3: Result of a binary detection of fish sounds and noise using a CNN model

verification, some of the misclassified samples (both noise and calls) were found to be errors in annotations, where the annotator did not provide the correct label to the sample by mistake.

This experiment was one of the first to familiarise me with machine learning libraries, and the training of deep learning models was using a simplified case where the test set originated from the same deployment as the training set, hence the relatively good performance.

Appendix B: Metrics and protocol of the experiment on Agile Modelling

Definitions

In binary classification, metrics are often expressed in terms of **True Positive (TP)**, **True Negative (TN)**, **False Positive (FP)** and **False Negative (FN)**. Table 1 of Appendix B shows the definition of each of these terms for the prediction and the real label.

Appendix B Table 2: Binary confusion matrix

		Predicted label (model prediction)	
		Positive	Negative
Real label (manual annotation)	Positive	TP	FN
	Negative	FP	TN

From these definitions, **Precision** and **Recall** scores are calculated as such:

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

The **F1 score** combines Precision and Recall in one score using the harmonic mean:

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FN + FP}$$

These metrics are sensitive to the threshold used for the detectors.

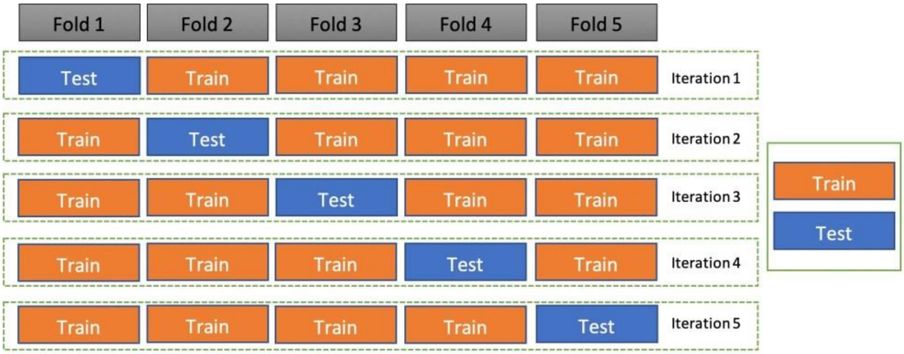
A **detector threshold** is the limit score above which a detector considers a sample as positive. Output scores are typically normalised between 0 and 1. But the threshold can be selected by a user, impacting the value of the Precision, Recall and therefore F1 scores. For example, if the prediction score of a true negative sample is 0.6, the sample would be counted as a false positive with a threshold of 0.9 but as a true negative with a score of 0.5. The value of a threshold is typically application dependent and a trade-off between Recall and Precision.

The Area Under the Curve of the Receiver Operating Characteristic Curve (AUC-ROC) score is the value of the surface area under the Receiver Operating Characteristic (ROC) curve. The ROC curve is the plot of the Recall against Specificity for every detector threshold (from 0 to 1 if normalised).

$$Specificity = \frac{TN}{TN + FP}$$

The AUC-ROC has the advantage of being independent of the threshold of the detector but is not directly interpretable compared to Precision or Recall.

K-fold cross-validation is a common method for evaluating classifiers. It consists of splitting the data into K equally sized parts and training K models with different training data and evaluating each model on a different split of the data. A 5-fold cross-validation protocol is shown in Figure 2 of Appendix 2. For each iteration, the performance of the model is computed, and the average performance is given as the result. This approach limits the variability of scores due to specificity in part of the training or validation data, and so, it is estimated to be a more reliable evaluation method. K-fold cross-validation is called stratified when the proportion of positive and negative samples is the same in the different folds.

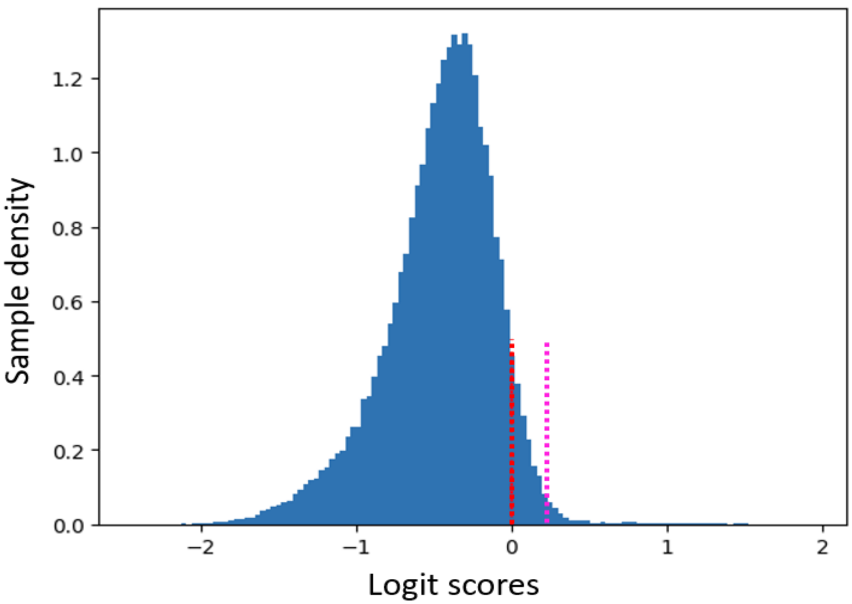


Appendix B Figure 1. Protocol for 5-fold cross-validation.

Experimental protocol

After manual annotation, the test set was generated using a sliding window approach with a window length of 5 seconds on recordings with a sampling rate of 32000Hz, corresponding to the format expected by the model SurfPerch; the hop size was set to 5 seconds. A window was counted as positive if 20% of the duration of the window is overlapping with a manual annotation or if more than 50% of one or several annotations are contained in the window. Otherwise, the window was labelled negative.

The Agile Modelling approach started with one example of the target sound. Using a similarity search with the highest similarity to the initial example of the target, 20 samples were annotated. These 20 samples were used to train the first detector. By iteration, more samples were added. The samples were selected around a subjective logit score, where it was assumed that some samples would be positive (i.e. predicted to contain a target sound by the detector) and some negative (pink dashed line on Figure 2 of Appendix 2). The number of samples annotated at each iteration started with 20 and increased by 10 for every iteration, up to 90 (8 iterations), reproducing the scenario a user could use, annotating more and more samples as the model becomes more reliable. At each iteration, 5 models were trained using a stratified 5-fold cross-validation method and their performance was evaluated on the validation set and on the test set using Precision, Recall, F1 and AUC-ROC scores. The iterative training



Appendix B Figure 2. Distribution of logit scores from a detector applied to the samples from the deployment in the Harbour of Texel-Oudeschild. The pink dashed line represents an example of a score used to find more samples to be annotated by the user. The red dashed line is the threshold of the detectors.

process was stopped after one hour. At the end of the training, 148 samples were labelled positively, and 291 samples were labelled negatively. The total number of samples labelled differs from the sum of samples labelled during the iterations because duplicates can be shown during the training and re-annotated, which would count for only one sample at the end.

Detectors used were one-layer linear classifiers, trained with a batch size of 12, a learning rate of 0.001 for 128 epochs, and a detector threshold of 0 (in logit score, red dashed line in Figure 2 of Appendix 2). These parameters were kept constant for the whole iterative training process.

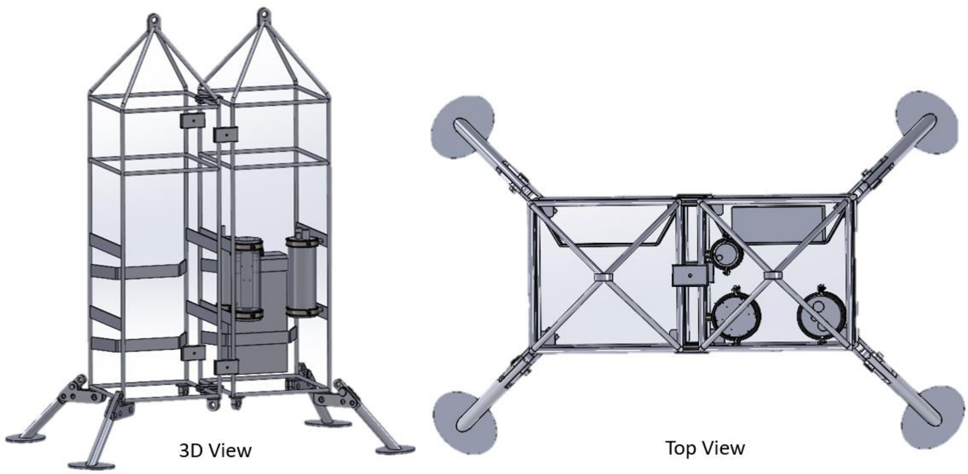
Additional parameters for the computation of the spectrograms and of the embeddings, the display of the spectrograms, and the training of the models were kept unchanged from the original script available on the Github of the SurfPerch project[27].

It is important to notice that Agile Modelling is an intrinsically very unrepeatable process because it builds on iteration where samples are selected using a semi-stochastic approach: to accelerate the research of similarity sounds on the whole Training pool, the search is done on a subset of the all data and an early stop condition is activated after some time of not finding better samples. In practice, it was observed that the similarity search of the first samples produces non-repeatable results, e.g. the top 5 samples would not necessarily figure within the output of a request for the top 10 samples. Therefore, the results obtained on the same dataset, using the same reference sounds, are likely to vary if the iterative training is repeated.

[27] https://github.com/BenUCL/surfperch/tree/surfperch

Appendix C. Theoretical study on the Biodiversity Sensing Box stability

In the first deployment, the Biodiversity Sensing Box unfortunately fell sideways on the seafloor. Based on a theoretical analysis of the stability of the box's frame under high underwater currents, Rick Hendriksen suggested a new design to prevent tilting (Figure 1 of Appendix 3). The stability of the design was estimated accounting for the expected force of the currents in the North Sea, the shape and the mass of the box. Details of the calculation are available in Figure 2 of Appendix 3.



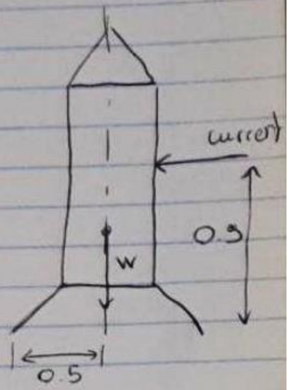
Appendix C Figure 1: 3D model of the solution suggested to improve the stability of the Biodiversity Sensing Box during deployment.

1) Drag Force

$\rho = 1025 \text{ kg.m}^{-3}$
$V = 1 \text{ m.s}^{-1}$ (2 knots)
$C_d = 1.05$ (cube)
$A = 1 \times 1.5 \text{ m}^2$

$F = \frac{1}{2} \rho v^2 C_d \times A$

$= \frac{1}{2} 1025 \times 1^2 \times 1.05 \times 1.5$
$= 807 \text{ N}$



2) Drag Torque

$T_{current} = F \times 0.9 = \boxed{726 \text{ Nm}}$

3) Weight Torque

$W_{frame + equip} \simeq 120 \text{ kg}$

underwater: $W \times 0.80 \simeq 100 \text{ kg}$

$T_{weight} = 100 \times 9.81 \times 0.5$
$= \boxed{490 \text{ N.m}}$

4) Stability analysis

Stable if $T_{weight} > T_{current}$

$490 \ll 726$

Unstable under this conditions

Appendix C Figure 2: Theoretical study of the stability of the Biodiversity Sensing Box under strong water current conditions.

Appendix D. List of fish species in the Wadden Sea, with sonifery knowledge and available recordings online

Appendix D Table 1: Fish species list from the Wadden Sea[28] with reported sonifery and number or recordings available (according to Fishsounds.net).

Index	Latin name	Common name	Known as actively soniferous	Recording available
1	*Abramis brama*	Carp bream	Yes	0
2	*Acipenser sturio*	Sturgeon	Yes	0
	Acipenser sp.			
3	*Agonus cataphractus*	Hooknose		
4	*Alburnus alburnus*	Bleak		
5	*Alosa fallax*	Twaite shad		
6	*Ammodytes marinus*	Lesser sandeel		
7	*Ammodytes tobianus*	Small sandeel		
8	*Anguilla anguilla*	Eel	Yes	0
9	*Aphia minuta*	Transparent goby		
10	*Arnoglossus laterna*	Scaldfish		
11	*Aspius aspius*	Asp		
12	*Atherina boyeri*	Big-scale sand smelt		
13	*Atherina presbyter*	Sand-smelt		
	Atherina sp.			
14	*Barbus barbus*	Barbel		
15	*Belone belone*	Garfish		
16	*Blicca bjoerkna*	White bream		
17	*Buglossidium luteum*	Solenette		
18	*Callionymus lyra*	Dragonet		
19	*Callionymus maculatus*	Spotted dragonet		
20	*Callionymus reticulatus*	Reticulated dragonet		
21	*Carassius carassius*	Crucian carp	Yes	0
22	*Carassius gibelo*	Gibel carp		
23	*Chelidonichthys lucerna*	Tub gurnard	Yes	0
24	*Chelon aurata*	Golden grey mullet	Yes	0
25	*Chelon labrosus*	Thick-lipped grey mullet		

[28] https://swimway.waddensea-worldheritage.org/fish-species

26	*Chelon ramada*	Thin-lipped grey mullet		
	Chelon sp.			
27	*Chirolophis ascanii*	Yarrell's blenny		
28	*Chondrostoma nasus*	Common nase		
29	*Ciliata mustela*	Five-bearded rockling		
30	*Clupea harengus*	Herring	Yes	2
31	*Conger conger*	Conger eel	Yes	0
32	*Conodon nobilis*	Barred grunt		
33	*Coregonus oxyrinchus*	Houting		
34	*Coregonus albula*	European cisco		
	Coregonus sp.			
35	*Cottus perifretum*			
36	*Ctenolabrus rupestris*	Goldsinny		
37	*Cyclopterus lumpus*	Lumpsucker		
38	*Cyprinus carpio*	Carp		
39	*Dicentrarchus labrax*	Sea bass		
40	*Echiichthys vipera*	Lesser weever		
41	*Enchelyopus cimbrius*	Four-bearded rockling		
42	*Engraulis encrasicolus*	Anchovy	Yes	0
43	*Entelurus aequoreus*	Snake pipefish		
44	*Esox lucius*	Pike	Yes	0
45	*Eutrigla gurnardus*	Grey gurnard	Yes	2
46	*Gadus morhua*	Cod	Yes	8
47	*Gaidropsarus vulgaris*	Three-bearded rockling		
48	*Galeorhinus galeus*	Topeshark		
49	*Gasterosteus aculeatus*	Three-spined stickleback	Yes	1
50	*Gobiosoma bosc*	Naked goby		
51	*Gobius niger*	Black goby	Yes	1
52	*Gymnocephalus cernua*	Ruffe		
53	*Helicolenus dactylopterus*	Bluemouth redfish	Yes	0
54	*Hippocampus hippocampus*	Sea-horse	Yes	0
55	*Hippoglossoides platessoides*	American plaice		
56	*Hyperoplus immaculatus*	Greater sandeel		
57	*Hyperoplus lanceolatus*	Great sandeel		
58	*Lampetra fluviatilis*	River lamprey		

59	*Leucaspius delineatus*	Sunbleak		
60	*Leuciscus aspius*	Asp		
61	*Leuciscus idus*	Ide		
62	*Leuciscus leuciscus*	Common dace		
63	*Limanda limanda*	Dab		
64	*Liparis liparis*	Sea-snail		
65	*Liparis montagui*	Montagu's Sea-snail		
66	*Lota lota*	Burbot	Yes	0
67	*Maurolicus muelleri*	Pearlsides		
68	*Merlangius merlangus*	Whiting		
69	*Merluccius merluccius*	European hake	Yes	0
70	*Micropogonias undulates*	Atlantic croaker	Yes	6
71	*Microstomus kitt*	Lemon sole		
72	*Misgurnus fossilis*	Weatherfish		
73	*Molva molva*	Ling	Yes	0
74	*Mullus surmelutus*	Surmullet		
75	*Mustelus asterias*	Starry smooth-hound		
76	*Myoxocephalus scorpius*	Bullrout	Yes	1
77	*Neogobius melanostomus*	Round goby	Yes	0
78	*Neogobius fluviatilis*	Monkey goby		
79	*Oncorhynchus mykiss*	Rainbow trout	Yes	1
80	*Osmerus eperlanus*	Smelt		
81	*Pagellus bogaraveo*	Blackspotted seabream	Yes	0
82	*Parablennius gattorugine*	Tompot blenny		
83	*Perca fluviatilis*	European perch	Yes	4
84	*Petromyzon marinus*	Sea lamprey	Yes	0
85	*Pholis gunellus*	Butterfish		
86	*Platichthys flesus*	Flounder		
87	*Pleuronectes platessa*	Plaice		
88	*Pollachius pollachius*	Pollack	Yes	1
89	*Pollachius virens*	Saithe	Yes	2
90	*Pomatoschistus lozanoi*	Lozano's goby		
91	*Pomatoschistus microps*	Common goby		
92	*Pomatoschistus minutus*	Sand goby	Yes	0
93	*Pomatoschistus pictus*	Painted goby	Yes	2
94	*Ponticola kessleri*	Kessler's goby		

95	*Proterorhinus semilunaris*	Western tubenose goby		
96	*Pungitius pungitius*	Ninespine stickleback		
97	*Raja clavata*	Thornback		
98	*Rhodeus amarus*	European bitterling		
99	*Rutilus rutilus*	Roach		
100	*Salmo salar*	Salmon	Yes	1
101	*Salmo trutta*	Sea trout	Yes	1
102	*Sander lucioperca*	Pike perch	Yes	4
103	*Sardina pilchardus*	Pilchard		
104	*Scardinius erythrophthalmus*	Common rudd		
105	*Scomber scombrus*	Mackerel		
106	*Scophthalmus maximus*	Turbot		
107	*Scophthalmus rhombus*	Brill		
108	*Scyliorhinus canicula*	Lesser spotted dogfish		
109	*Silurus glanis*	Wels catfish	Yes	3
110	*Solea solea*	Sole		
111	*Spinachia spinachia*	Sea stickleback		
112	*Spondyliosoma cantharus*	Black seabream		
113	*Sprattus sprattus*	Sprat		
114	*Squalius cephalus*	Common chub		
115	*Squalus acanthias*	Spurdog		
116	*Symphodus melops*	Corkwing	Yes	0
117	*Syngnathus acus*	Great pipefish		
118	*Syngnathus rostellatus*	Nilsson's pipefish		
119	*Taurulus bubalis*	Long-spined sea scorpion		
120	*Tinca tinca*	Tench	Yes	0
121	*Trachurus trachurus*	Horse mackerel	Yes	0
122	*Trisopterus luscus*	Bib		
123	*Trisopterus minutus*	Poor cod		
124	*Zoarces viviparus*	Eelpout		
total		**Number of species**	**soniferous species**	**recorded species**
		123	**37**	**16**

Appendix E. Summary of the thesis

Marine biodiversity is facing significant global threats due to human activities such as overfishing, habitat destruction, pollution, and climate change. Monitoring marine ecosystems is essential to be able to understand and mitigate the decline in biodiversity and habitats as a first step for effective conservation and restoration measures. Traditional monitoring methods based on fishery or SCUBA diving are invasive, costly, incomplete and limited to certain areas. Listening to animal sounds, a technique known as Passive Acoustic Monitoring (PAM), is a potentially broadly applicable and cost-effective alternative for current marine biodiversity monitoring practices.

The underwater world was once assumed to be silent, but it is now clear that sound plays a vital role for marine animals. From whales and dolphins to fish and invertebrates, many marine species depend on sound for communication, much like birds do on land. Listening to these communications provides a non-invasive way to study biodiversity and animal behaviour. Passive acoustic monitoring is a well-established technique in marine mammal research, but remains underused for fish and invertebrates. Two key challenges limit its broader application. First, the scarcity of known and accessible reference sounds limits species identification from acoustic recordings. Second, PAM generates vast amounts of data, often hundreds or thousands of hours, which are time-consuming to manually analyse and automated AI-based methods are not yet efficient and widely accessible for fish and marine invertebrate sounds. Overcoming these limitations could make PAM a reliable, cost-effective, and widely applicable monitoring method. Combined with other emerging monitoring techniques, it could significantly enhance marine biodiversity assessments using a data-driven approach.

This project aimed to improve the efficiency and usability of PAM while facilitating its integration into autonomous multi-sensor

monitoring systems. The effort was structured into four work packages (WPs). WP1: Underwater bioacoustics data collection - Acoustic recordings were gathered from various marine environments to capture a variety of marine animal sounds. Datasets, which include human-made annotations, are essential for training AI models to automate bioacoustics data processing, while reference sounds are critical for species identification. WP2: Development of machine learning bioacoustics detectors - Deep learning-based AI models were trained to automatically identify animal sounds in underwater audio recordings, focusing on techniques reducing the manual efforts necessary for training, including data annotation. WP3: Creation of an embedded system for real-time monitoring - A portable underwater recorder was developed, integrating audio and video capture along with onboard processing to support real-time sound detection from the AI models and communication with researchers. WP4: Design of a graphical user interface for bioacoustic monitoring - A dashboard was developed to display live recordings and automate the detection of animal sounds from the underwater recorder.

In Chapter 2, established methodologies and knowledge relevant to the design of each work package were reviewed. Most reference sounds of marine animals, aside from marine mammals, are missing and remain unknown, and no annotated datasets of fish sounds were available at the beginning of the project. This limited the data collection aimed for in WP1. Although species identification requires reference sounds, the detection of unidentified fish sounds has previously contributed to ecological understanding of biodiversity, and species identification can still become possible later. Existing detection models, developed for specific birds, fish species, or different taxa, did not apply to the objectives of WP2, as they were trained for different uses. Training an AI model for automatic species sound detection requires data that reflects real use conditions, but such training data were also unavailable. Creating training datasets involves manual annotation of recordings, a process that is both time-intensive and costly. Therefore, several machine learning approaches suitable for limited-data scenarios were identified. For WP3, existing autonomous sound recorder solutions lacked onboard processing capabilities and were therefore unsuitable. Appropriate hardware for an autonomous recorder was identified, and collaboration began with partners developing a similar recorder for video, with interest in adding sound capability. Multiple libraries were found to support the creation of graphical user interfaces as envisioned for developing in WP4.

Chapter 3 presents the design processes, implemented solutions, and associated challenges for each work package. In WP1, bioacoustics data were collected using underwater recordings from various environments, obtained through collaborations, open-access sources, and field deployments. Animal sounds were successfully recorded during a deployment in the North Sea. Reference sounds used for species identification were sourced from an open-access database compiling scientific knowledge on fish sounds. However, most fish species still lack known reference sounds, and only limited data exist for invertebrates. In WP2, a workflow based on active learning, a technique where a model works with a human to improve its performance iteratively, was adapted from bird bioacoustics and applied to train models for the detection of animals in various marine environments. WP3 saw the development and deployment of a prototype underwater audio-video recorder for data collection. Further work is needed to incorporate onboard data processing, as originally intended. In WP4, two graphical user interfaces were developed. One is a real-time detection dashboard for use with the embedded system. The other is a standalone software tool that enables users to analyse acoustic data after deployments, using AI models pre-trained in different environments.

Chapter 4 summarises the key outcomes of each work package, while Chapter 5 places them in the broader context of passive acoustic monitoring (PAM) research. WP1's data collection, sourced from deployments and collaborative efforts, supports the training of automatic detectors across diverse environments, gathers reference sounds for species identification. These will be extended and made available as initiatives for global bioacoustics data sharing are being developed. In WP2, the active learning workflow enabled successful training of detection models for diverse sounds from different locations, requiring minimal data and manual effort. The approach is both flexible and user-friendly, representing a significant step forward in making AI techniques more accessible to marine bioacoustics researchers. The WP3 audio-video system records video and sound simultaneously. Future enhancements, based on research in tropical reefs, will add multiple hydrophones for sound source localisation via triangulation. This will enable visual species confirmation and the collection of new reference sounds, an approach not yet tested in temperate or degraded environments. WP4's interfaces further increase the accessibility of AI in underwater PAM. The standalone interface allows for the application of pre-trained models, with plans to support new model training in future versions, following the development direction observed in terrestrial bioacoustics. The real-time dashboard interface will be further developed and applied in a larger-scale follow-up project aimed at public and stakeholder engagement for monitoring biodiversity in future offshore floating wind farms in the North Sea.

This project has contributed to advancing underwater PAM for fish and marine invertebrates by enhancing automated data analysis methods and taking a step towards the integration of PAM in autonomous multi-sensor monitoring systems. The expansion of underwater bioacoustics data collection will continue, allowing for the identification of more species and the training of better models. The method employed to train AI automatic sound detectors applies to any environment, providing bioacoustics researchers with an efficient and accessible data analysis tool. The audio-video recorder, already deployed for data collection, will be improved to facilitate reference sound collection and real-time communication, using the dashboard highlighting marine biodiversity. These advancements bring PAM closer to becoming a viable, cost-effective, and broadly applicable solution for underwater monitoring. Progress has also been made towards integrating PAM with other monitoring techniques, enabling the combination of acoustic and video data recording, onboard processing, and triggering water sampling for DNA analysis based on sound detection. The ongoing advancement of PAM and its integration into multi-sensor systems will enhance biodiversity assessment, ultimately improving marine life conservation and restoration efforts.

Appendix F. Acknowledgements

I would like to express my gratitude to all those who have significantly shaped my life during these years of pursuing my Engineering Doctorate and contributed to this project in one way or another.

Thank you to the innovation programme Next Level Animal Science from the Animal Science Group of Wageningen University and Wageningen University and Research for financially supporting this work.

Thank you to the people who made this journey possible. Ronald, thank you for offering me this incredible and unexpected opportunity, without which I wouldn't be here. And Tinka, thank you for that inspired spark of creativity and effort that led to the project proposal; it set everything in motion. Thank you also for your support throughout the project, especially for helping me find direction and meaning during moments of doubt at the beginning, and for sharing genuine enthusiasm for this work. I will carry forward your optimism, determination, and ambition to improve things as a source of inspiration.

A special and deeply felt thanks to Rosa, whose unfailing support and guidance made this journey both possible and meaningful. Even if it may not have always felt that way to you, I am truly grateful for your presence when it mattered most, for your availability, your constructive attitude, and your communicative enthusiasm, which helped carry me through many difficult moments. Among the many things I've learned from you, one of the most lasting is how to supervise with integrity and care, something you taught through your example. I hope to carry that forward and pass it on in my own way.

Big thanks also to the amazing people from MAE, and especially those with whom I shared the office and lunches on most days: Daniel, Saskia, Sophie, Joshua, Lara, Xantia, Jeroen, and Auriel. And thank you for being, in turns, great advisors, attentive supporters, celebration buddies, and even rubber ducks when needed. I’m also grateful to everyone else in the MAE group that I have seen less but who contributed to making such a welcoming and inspiring environment to work in.

Thank you to all the teachers, colleagues, and students from whom I’ve learned so much throughout this journey. In particular, I’m grateful to Karen for sharing your vast knowledge of fish acoustics with me, to Clea for your support in computer science, especially at the beginning, when I felt alone and knew so little, and to Erik for meticulously and with great passion, teaching me how to dive.

Thank you as well to everyone who contributed to building and developing the EngD program at WUR. It has been an incredibly enriching learning experience. In particular, I’m grateful to the EngD office and especially to Femke Brouwer, the program coordinator, for your tireless support and the energy you invested in both the program and its candidates.

I would like to thank the members of the committee who played an essential role by accepting to assess the quality of this work.

On a more personal note, this chapter of my life would not have been the same without my homes, the Achtertuin and the Voortuin, and all their homies. There were tears and laughter, a lot of mess and beauty, tough and easier days, but this is where life happened, and where it was worth living. Thank you to my parents, Gisele and Patrice, for their support and understanding in every decision I've made, and to my brothers, Lucas and Raphael, for always being there and listening with care. To Croc-mou and Sheena, my ever-reliable sources of support and stress relief (as long as you were well fed), you brought so much joy and fun into my life.

And finally, to Chloé, thank you for being my unwavering support. The many ways in which you have contributed to both this work and my

life, and continue to do so, mean more than I could ever clumsily express here.

The research described in this thesis was financially supported by the innovation program Next Level Animal Science from the Animal Science Group of Wageningen University and by Wageningen University and Research.

Cover design by Valentin Bordoux