The Assessment and Development of Methods in (Spatial) Sound Ecology

Becky E Heath[1,2]

(Rebecca Emma Heath)

Supervision:

Dr. Lorenzo Picinali[1-]

Prof. Rob M. Ewers[2]

Dr. C. David L. Orme[2]

[1] Dyson School of Design Engineering, Imperial College London

[2] Georgina Mace Centre for the Living Planet, Imperial College London

Submitted:

October 2022

Accepted:

February 2023

A thesis submitted for the degree of Doctor of Philosophy

ABSTRACT

As vital ecosystems across the globe enter unchartered pressure from climate change industrial land use, understanding the processes driving ecosystem viability has never been more critical. Nuanced ecosystem understanding comes from well-collected field data and a wealth of associated interpretations. In recent years the most popular methods of ecosystem monitoring have revolutionised from often damaging and labour-intensive manual data collection to automated methods of data collection and analysis. Sound ecology describes the school of research that uses information transmitted through sound to infer properties about an area's species, biodiversity, and health. In this thesis, we explore and develop state-of-the-art automated monitoring with sound, specifically relating to data storage practice and spatial acoustic recording and data analysis.

In the first chapter, we explore the necessity and methods of ecosystem monitoring, focusing on acoustic monitoring, later exploring how and why sound is recorded and the current state-of-the-art in acoustic monitoring. Chapter one concludes with us setting out the aims and overall content of the following chapters. We begin the second chapter by exploring methods used to mitigate data storage expense, a widespread issue as automated methods quickly amass vast amounts of data which can be expensive and impractical to manage. Importantly I explain how these data management practices are often used without known consequence, something I then address. Specifically, I present evidence that the most used data reduction methods (namely compression and temporal subsetting) have a surprisingly small impact on the information content of recorded sound compared to the method of analysis. This work also adds to the increasing evidence that deep learning-based methods of environmental sound quantification are more powerful and robust to experimental variation than more traditional acoustic indices.

In the latter chapters, I focus on using multichannel acoustic recording for sound-source localisation. Knowing where a sound originated has a range of ecological uses, including counting individuals, locating threats, and monitoring habitat use. While an exciting application of acoustic technology, spatial acoustics has had minimal uptake owing to the expense, impracticality and inaccessibility of equipment. In my third chapter, I introduce MAARU (Multichannel Acoustic Autonomous Recording Unit), a low-cost, easy-to-use and accessible solution to this problem. I explain the software and hardware necessary for spatial recording and show how MAARU can be used to localise the direction of a sound to within ±10˚ accurately. In the fourth chapter, I explore how MAARU devices deployed in the field can be used for enhanced ecosystem monitoring by spatially clustering individuals by calling directions for more accurate abundance approximations and crude species-specific habitat usage monitoring. Most literature on spatial acoustics cites the need for many accurately synced recording devices over an area. This chapter provides the first evidence of advances made with just one recorder.

Finally, I conclude this thesis by restating my aims and discussing my success in achieving them. Specifically, in the thesis' conclusion, I reiterate the contributions made to the field as a direct result of this work and outline some possible development avenues.

Copyright

Declaration of Originality

This work was funded through the National Environmental Research Council (NERC) as part of the Quantitative Methods in Ecology and Evolution (QMEE) Centre for Doctoral Training under grant number NE/P012345/1. This thesis is all my work except where it is explicitly stated below or in the text.

General:

The writing of this thesis was completed entirely by me. Valuable comments were provided by Lorenzo Picinali and Robert Ewers pre this thesis submission and by Kate Jones and Thrishantha Nanayakkara during the assessment viva.

Chapter 1:

The literature review and writing of this chapter were completed entirely by myself. Lorenzo Picinali and Robert Ewers provided valuable comments on an earlier draft of this chapter.

Chapter 2:

The idea to investigate compression was mine that I considered while reviewing the scripts used in Sarab Sethi's research. I developed this into a complete study with the invaluable contributions of Sarab Sethi, Lorenzo Picinali, Robert Ewers, and David Orme. Robert Ewers, Lorenzo Picinali, and David Orme provided me with continued guidance and supervision through this study's data collection, curation, and formal analysis. I collected the data and completed all formal analysis and data visualisation except for the final Figure (Figure 2.5) and whole-dataset model, which David Orme created. Lorenzo Picinali, David Orme, Sarab Sethi, and Robert Ewers provided valuable comments on this chapter up to its submission (and later acceptance) to a journal. This chapter was also developed with "minor corrections" from two anonymous reviewers.

Chapter 3:

The idea for this chapter was suggested by Robert Ewers, Lorenzo Picinali and David Orme, who submitted the idea of "3D acoustics for ecosystem monitoring" as the basis of my PhD grant. I developed the MAARU device's hardware and software, with advice on hardware from Lorenzo Picinali and Sarab Sethi and general supervision from Lorenzo Picinali, Robert Ewers, and David Orme. Lorenzo Picinali also advised beamforming, data visualisation, and theory in spatial acoustics. Reiji Suzuki provided me with support and bug fixes while installing HARKBird and created the array configuration file vital for all spatial analyses. Neel Le Penru helped quality-check the installation instructions for the proposed device. Lorenzo Picinali, David Orme, Sarab Sethi, Robert Ewers and Neel Le Penru provided valuable comments on an earlier draft of this chapter.

The field deployment was organised for this research by me. Recorders and solar panels were tethered in the trees by Jack Bath, Mark Griffiths, Oscar Housego, and Aiden Prall from I Am Lumberjack LTD, who also helped me design the final tethering strategy used in this research.

Chapter 4:

Robert Ewers, Lorenzo Picinali and David Orme suggested applying spatial acoustics to ecosystem monitoring. My idea was to orient the MAARU devices vertically to monitor vertical stratification and cluster signals by direction for abundance monitoring. I developed the data extraction pipelines in this chapter; however, the data used in this chapter and the HARKBird configuration were the same as in chapter 4, so the duplicate acknowledgements apply here. Robert Ewers, Lorenzo Picinali and David Orme helped me develop the formal analysis. Robert Ewers and Lorenzo Picinali made valuable comments on an earlier draft of this chapter.

Chapter 5:

This review chapter was conceptualised and written by me, with comments from Lorenzo Picinali.

Acknowledgements

First, I would like to extend thanks to my supervisors, Dr Lorenzo Picinali, Professor Rob Ewers and Dr David Orme, who have provided me with consistent academic and personal support over these last four years. The lab groups they have curated have provided an exciting and inspiring environment for me to grow as a scientist that will forever leave its mark on me and my work. Similarly, the assessors of my PhD, Professor Kate Jones and Professor Thrishantha Nanayakkara, have provided me with incredibly thoughtful and useful discussions on this work, resulting in many valuable improvements.

On a similar note, I would like to thank some of the many other scientists and researchers I have had the pleasure of working alongside who have made my time working at Imperial so enjoyable: Mike Boyle, Dom Brass, Ella Brassingdon, Ian Chard, Marco Comunità, Liv Daniels, Rapolas Daugintis, Isaac Engel, Hollie Folkard-Tapp, Adam Fraser, Matt Garrod, Ross Gray, Ali Green, Niche Gregory, Craig Henry, Aiden Hogg, Ben Ingledow, Neel Le-Penru, Emily Lewis-Brown, Valentina Marconi, Steph Martin, Ciar Noble, Danielle Norman, Fredie Poznanski, Aaron Prairie, Terhi Riutta, Sarab Sethi, Isaac Stoppard, Kevin Sum, Thibault Vicente and so many others. From fellow turret dwellers to friends made in the forests of Borneo, I have met many fantastic people thanks to this PhD that will stay in my life beyond my time at Imperial.

I would also like to extend thanks to Sophie Hughes from Imperial Wellbeing, who accidentally or otherwise provided me with mantras that kept me going when I was not sure I would ever actually make it to this point. Beyond Dyson and Silwood Park, I would also like to thank Renee Boling and Rebecca Stewart from Imperial Outreach, who have given me the platform to share my work with 100s of school children, which have been a constant source of enrichment and inspiration.

Much of this work has relied on data I collected with invaluable help and contribution from many others. Particularly I would like to thank Professor Henry Bernard for permitting this research to take place a SAFE, Jani Sleutel, Nursyamin (Syamin) Zulkifli, and the many other research staff (several of whom mentioned above) for their advice and assistance in the field. I will always remember my time at SAFE as a time of happiness, wonder, blood (credit to leeches), sweat, tears, and elephants – although the less said about the latter, the better.

I would also like to thank Mark Taylor, Patrick English, and Len Pegler-Vincent at Chalk Wood, and Peter Coles from the Woodland Trust for allowing me to research their sites here in the

UK. I would also like to extend additional thanks to Peter Coles for providing me with extensive information on the site's history. This UK fieldwork could not have taken place without Jack Bath, Mark Griffiths, Oscar Housego, and Aiden Prall from I Am Lumberjack LTD – who have been so enthusiastic and generous since first joining the project.

So many people outside of academia have given me so much in the last few years that I cannot possibly thank them all here. First, I would like to extend thanks to everyone who lent me a cat, a house, or both: Rob and Cris, Matt and Laura, Kate, and of course, Kizzy and Dan. Beyond that, I would like to thank my incredible friends, family, and everyone else who has remotely feigned interest in a graph I've been excited to show them: Morris, Elspeth, Hawkett, Sarah, Adam, Georgie, Ben, Becca (and baby Max), Mark, Rosie, Jake, Bethan, friends from Newstead, Dartford, Exeter, the Dodgers, and the Translation Inn.

And finally, finally, I would like to *especially* thank Kizzy for all her support, my wonderful mum Steph, my little brother Billy and Big Tone. It's not nearly as much as you deserve, but this is for you.

Publications and Contributions

PAPERS:

How index selection, compression, and recording schedule impact the description of ecological soundscapes

B. E. Heath, S.S. Sethi, C. D. L. Orme, R. M. Ewers, L. Picinali. *Ecology and Evolution* 2021. https://doi.org/10.1002/ece3.8042

Contribution: Led conceptualisation, fieldwork, running formal analysis and writing the manuscript.

Appearance in Thesis: Chapter 2

Autonomous Rainforest Soundscape Identification: Quantifying the Impacts of Data Lossy Compression, Recording Length, and Index Selection

B. E. Heath, S. S. Sethi, C. D. L. Orme, R. M. Ewers, L. Picinali. E-*Forum Acousticum* 2020. https://doi.org/ 10.48465/fa.2020.0365 (Conference Paper)

Contribution: Led Conceptualisation, fieldwork, running formal analysis and writing the manuscript.

Appearance in Thesis: Chapter 2

OPEN DATA:

Ecoacoustic Study Design Variation: Raw Files

B. E. Heath, S. S. Sethi, C. D. L. Orme, R. M. Ewers, L. Picinali. *The SAFE Project Zenodo* 2021 [Data Set]. https://doi.org/10.5281/zenodo.5159914

Contribution: Led data collection, processing, and curation

Appearance in Thesis: Chapter 2

Ecoacoustic Study Design Variation: Impact on Acoustic Indices and AudioSet Fingerprints

B. E. Heath, S. S. Sethi, C. D. L. Orme, R. M. Ewers, L. Picinali. *The SAFE Project Zenodo* 2021 [Data Set]. https://doi.org/10.5281/zenodo.5153193

Contribution: Led data collection, processing, and curation

Appearance in Thesis: Chapter 2

TALKS:

Spatial Acoustics (Soundscape Recording Session) +

Compression in Ecoacoustics (Soundscape Analysis Session)

B. E. Heath *UKAN Soundscapes SIG.* 2022. London, UK

Appearance in Thesis: None

(Spatial) Autonomous Soundscape Recording

B. E. Heath *TR@NSNET: Passive Acoustic Monitoring (Current Practice and Future Solutions).* 2022. Madrid, Spain (Reorganised to Online due to the Pandemic)

Appearance in Thesis: Chapter 3

Autonomous Rainforest Soundscape Classification and the Impact of Lossy Compression

B. E. Heath*, (e-)Forum Acousticum.* 2020. Lyon, France (Reorganised to online)

Appearance in Thesis: Chapter 2

Impact of Lossy Compression on the Usability of Audio in Soundscape Level Ecoacoustics

B. E. Heath, *International Bioacoustics Council (IBAC)* 2019**b**

Appearance in Thesis: Chapter 2

OPEN GITHUB REPOSITORIES:

www.github.com/BeckyHeath...

/Experimental-Variation-Ecoacoustics-Analysis-Scripts

Scripts for all the inter-study methods comparisons

B. E. Heath, C. D. L. Orme

Appearance in Theses: Chapter 2

/multi-channel-rpi-eco-monitoring

Software for automating on-device data collection, processing, and transmitting multichannel audio data. (Adaptation of https://github.com/sarabsethi/rpi-eco-monitoring)

B. E. Heath, S. S. Sethi, C. D. L. Orme

Appearance in Thesis: Chapter 3

/multi-mic-recorder-analysis

All the scripts used to evaluate the power consumption, accuracy, and signal degradation of the MAARU recorders

B. E. Heath

Appearance in Thesis: Chapter 3

/DOA_HeightTracking

Scripts used to collate HARKBird and BirdNET outputs, assess abundance through optimal clustering and explore vertical space usage.

B. E. Heath

Appearance in Thesis: Chapter 4

Glossary of Terms

As with many developing fields, the terms used to describe concepts in sound ecology are undergoing constant development and occasional changes in their meaning. Further, several of these terms are used in other fields to explain slightly different things. The terms used in this thesis and their intended definitions are listed below.

Term	Definition in this Thesis
Ecoacoustics	The study of the whole *soundscape* to infer ecological phenomena. Ecoacoustics includes the use of acoustic indices and machine-learning soundscape descriptors. Note: Ecoacoustics is elsewhere used as an umbrella term to describe any instance where sound is used to describe ecological phenomena (i.e., including bioacoustics)
Bioacoustics	The study of individual sounds to study ecological phenomena. This includes the production, transmission, and detection of sounds by animals. This includes identifying species/ individuals from a soundscape, studying song learning and evolutionary histories and studying individual responses to sounds.
Sound Ecology	The umbrella term used to describe any instance where sound is used to explore ecological phenomena. Therefore, including both ecoacoustics and bioacoustics.
PAM	Passive Acoustic Monitoring The practice of using an acoustic recorder which does not require a researcher present. Note: PAM recorders can be left in the field for a long time, but this does not make them autonomous unless they require no external influence at all.
ARU	Autonomous Recording Unit Completely autonomous PAM. ARU recorders are self-charging and can transfer data autonomously. The defining difference between PAM recorders and ARUs is that PAM recorders can be left in the field for a long but fixed time, while ARUs can be left in the field indefinitely.
Audible Sounds/ Audible Range	Refers to the human audible range (20 – 20,000Hz)
Acoustic Indices	Abstract, numeric soundscape descriptors. Acoustic Indices reduce a soundscape to an individual or a set of numbers so they can be easily analysed.
Analytical Indices	"Traditional" Acoustic Indices refer to straight mathematical calculations on a soundscape to output descriptive characteristics.
AudioSet Fingerprint	A specific acoustic index derived from the VGGish CNN, which was pre-trained on an audio data ontology (AudioSet). The AudioSet Fingerprint has 128 values that can describe a soundscape.

Contents

1. CHAPTER 1: SOUND ECOLOGY

1.1. CHAPTER OVERVIEW

Ecosystem monitoring is revolutionising in terms of methods for data collection and analysis. This thesis appraises some of the emerging state-of-the-art methods in a subset of ecosystem monitoring which focuses on sound. This cross-disciplinary field is built on insight from computer science, engineering, physics, biology, and ecology. This introductory chapter aims to briefly highlight critical contextual information across these fields to ensure readers from any background can engage with the content.

I will first give context to the necessity of environmental conservation and research (section 2), moving to how ecosystems are currently monitored (section 3) with a later focus on sound (section 3.4). Beyond introducing the biological argument for sound as a species monitoring tool (section 4), I will introduce the physical properties of sound, with a specific focus on how sound is recorded in field monitoring (section 4.3). The following section will focus on recorded sound analysis (section 5), exploring both traditional and deep learning methods on a single-species (section 5.2) or whole ecosystem level (section 5.3). I will then introduce some recent advances in the field and thoughts on where the field could be heading in the near and distant future (section 6). Finally, this chapter concludes by summarising this information (section 7), explaining the whole thesis objectives (section 8), and overviewing the research documented in the following chapters (section 9).

1.2. ENVIRONMENTAL RESEARCH

1.2.1. The Need for Nature

The importance of conserving ecosystems goes beyond an emotional attachment to a loss of charismatic fauna or a sense of duty and respect for the world we are part of. The environment provides us with irreplicable services necessary for humanity's sustainable continuation. Researchers have worked to comprehensively describe the benefits ecosystems provide humans, often attributing financial costs or worrying outcomes of loss. The formal discussion of these "Ecosystem Services" was popularised by the Millennium Ecosystem Assessment (MA) (Millenium Ecosystem Assessment , 2005), and has undergone continuous development and reframing. We now consider these as Nature Based Solutions, the use of existing natural processes to mitigate threats to development goals (explored further in 1.2.3).

Table 1.1 outlines the main Ecosystem Services given by the Millennium Ecosystem Assessment. These services broadly cover the benefits humans receive and depend on from the environment.

Regulating Services, e.g.,	**Cultural Services, e.g.,**
Purification of water and air Pollination Climate regulation and carbon sequestration Waste decomposition Natural disturbance regulation: resistance to floods/ strong wind Predator/ prey regulation Disease Regulation	Food Water Materials (timber, fertiliser, organic matter) Genetic diversity (e.g., for crop stability) Medicinal resources (pharmaceuticals, study/ test organisms) Energy
Supporting Services, e.g.,	**Provisioning Services, e.g.,**
Nutrient cycling Soil formation Primary production (photosynthesis) Habitat provisions	Cultural motifs (symbols, folklore, media) Spiritual and Historical Recreation, healing, and Eco-therapy Science and discovery Education

As a brief overview, the MA looked to create a global human-centred appraisal of interactions between people and ecosystems, intending to guide "decision makers" to make decisions that benefit both the people and the planet. The MA is a policy document that assimilated science from a wealth of sources to draw general conclusions and provide guidance. The MA defined ecosystem services as "the benefits people obtain from ecosystems", explaining the role of ecosystems in regulatory, supporting, cultural and provisioning services (Table 1.1). It describes how these ecosystem services contribute to people's security, health, and beneficial social interactions—ultimately tying ecosystem services and biodiversity to people's well-being and fundamental freedom of choice and action (Millenium Ecosystem Assessment , 2005).

The overall conclusions synthesised and popularised by the MA were that:

(1) Humans have caused irreversible damage to ecosystems as the demand for natural resources has increased.
(2) Humans have enormously benefitted from the extraction of natural resources; however, the extraction rate is unsustainable and driving some groups into extreme poverty, with returns on extractions potentially diminishing.
(3) The harm done to ecosystems may stop us from achieving the millennium development goals.
(4) The MA outlines several methods that are necessary to reverse the damage already done and ensure the sustainability of people on the planet. These methods require significant changes in policies and practices beyond those that existed when the MA was written.

Here intertwines human welfare with healthy environments, ecosystems and wildlife. Despite its publication nearly 20 years ago, anthropogenic expansion has threatened the environment and the delicate balance in which all life is a part.

1.2.2. Threats to the Environment

The environment is currently under countless stressors, such as the climate approaching tipping points (Mckay et al., 2022), widespread pollution (Cousins et al., 2022), habitat destruction (Gatti et al., 2021), and global declines in biodiversity (Ceballos et al., 2015).

A climate tipping point is an event in a climate-change-driven changing earth system. The systems with tipping points have a feedback relationship with the climate, which will cause further warming once reached, making the tipping point impossible to return from. As tipping points are seemingly approaching, they have increasingly become a source of public concern (Mckay et al., 2022). Further, if one tipping point is reached, it is speculated that this could cause a "tipping cascade", which could amplify global warming effects and potentially result in a global climate tipping point (Mckay et al., 2022). Six climate tipping points have been defined, spanning: the extent of ice sheets (Antarctica and Greenland), the continuation of ocean currents (AMOC), monsoons (the Sahara and West Africa), the El Niño southern oscillation, and forest extents (the Amazon and Boreal Forests) (Mckay et al., 2022). The form of these tipping points, along with their expected onset, varies between locations; however, early warnings of destabilisation have already been seen in the Greenland Ice Sheet, Atlantic Meridional Overturning Circulation (ocean currents), and the Amazon rainforest (Boers, 2021; Boers & Rypdal, 2021; Boulton et al., 2022; Mckay et al., 2022).

In particular, the Amazon, previously a reliable sink of ~150-200Gt of Carbon, now acts as a net carbon source (Gatti et al., 2021). The drivers behind this are thought to be climate change-induced drying, droughts and human-driven degradation (~17% since the 1970s (Mckay et al., 2022) (SPA (Scientific Panel for the Amazon), 2021)). Climate and rainfall-driven dieback of the Amazon was not expected until 3 to 4°C of warming or ~40% deforestation (Mckay et al., 2022; Nobre et al., 2016), but recent research, which includes tipping cascades into models, has indicated this may happen much sooner (20-25% deforestation)(Lovejoy & Nobre, 2018; Mckay et al., 2022)).

Deforestation, climate change, and species overexploitation have been credited with driving the planet's sixth mass extinction, even by highly conservative measures (Brooke et al., 2008; Butchart et al., 2006; Ceballos et al., 2015; Hoffmann et al., 2010). The natural world is in constant flux, where extinction is expected as desirable traits shift when the planet and its various biomes gradually change over the course of millennia. This pre-existing rate of extinction is known as "background extinction" and occurs at a rate of 2/MSY (Million Species per Year. AKA 2 extinctions, per 100 years, per 10,000 species) (Ceballos et al., 2015). However, a recent metanalysis found that actual species losses varied between 8 and 100 times that amount (Ceballos et al., 2015). The paper describes how under background extinction, it would be expected that just nine vertebrate species should theoretically have gone extinct since 1900; however, even under a conservative estimate, this number is actually at least 468 (69 mammals, 80 birds, 24 reptiles, 146 amphibians and 158 fish) (Ceballos et al., 2015). Biodiversity is pivotal in sustaining ecosystem services, especially crop pollination and water purification. Beyond this, biodiversity has extreme sentimental and cultural importance for people across the planet. Earth is not a planet for humans that other creatures take up space in; instead, we must consider Earth as a natural system of which we are just a part.

Our unsustainable exploitation of the environment is potentially irreversibly damaging natural systems, and the time we have left to mitigate this damage is rapidly diminishing. In this precarious time, it is necessary to act from a comprehensive understanding of ecosystems to protect, mitigate and make space for them as we enter unknowns in climate, extinctions, and anthropogenically reduced natural habitat extents.

1.2.3. Nature as a Solution

Since the MA and the subsequent popularisation of the term ecosystem services, there have been vital developments in this school of thought, namely through the introduction of "Nature-Based Solutions". The theory of Nature Based Solutions NbS is built upon the understanding, popularised by the MA, that many vital processes can be moderated by healthy ecosystems (carbon sequestration, clean air, medicines, genetic resources, and stabilising shorelines) (Millenium Ecosystem Assessment , 2005). NbS involve addressing societal challenges purely by enhancing natural processes. NbS covers ecosystem-based adaptation (EbA), ecosystem-based mitigation, eco-disaster risk reduction and green infrastructure (Griscom et al., 2017).

NbS have been used as solutions to many practical problems. From agroforestry in Panama (where farmers grow multiple species, including trees, shrubs, crops and livestock in the same area), which increases carbon sequestration, local biodiversity and economic return (Paul et al., 2017). To the Gulf of Mexico, where "living shorelines" have been introduced that aid the recruitment of oyster reefs, stabilising the shoreline and reducing the effect of waves and erosion while increasing the abundance and diversity of economically important species (Scyphers et al., 2011). And finally, in cities, green spaces and increased canopy cover can reduce air temperature, which is especially vital on hot days (Bowler et al., 2010).

Literature on ecosystem services and NbS allow us to quantify and discuss how the conservation of nature is vital for the continuation of humankind and how employing nature-positive practices may also be the solution to some of the most significant problems we currently face. Moreover, while NbS provide tremendous potential, their uptake has been limited as there are substantial challenges involved with measuring and predicting their effectiveness, a range of flawed financial models, and inflexible and highly sectorised government (Seddon et al., 2020).

Generally, a greater area and variety of ecosystems will lead to a greater number and stability of species and ecosystem services. NbS balance this with socioeconomic trade-offs to better people's lives while minimising or restoring natural processes. Optimising outcomes for people and the planet requires a more nuanced understanding of ecosystem functioning (Dietze et al., 2018).

1.2.3. Good Research Informs Good Practice

To best optimise land management practices, we must develop an understanding of ecosystem function that is as comprehensive as possible. A recent review highlighted that good monitoring practice is vital to (1) compare and monitor communities temporally across the globe, (2) provide data to models to anticipate system responses better, (3) determine potential early warning signals prior to substantial ecosystem changes and (4) examine the stressors that are causing the most impact/ species that are most at risk (Besson et al., 2022).

Environmental research has already informed the theory and success of NbS. For example, research quantifying the vitality of habitat connectivity, extent and diversity has inspired developments to be mindful of biodiversity hotspots and crossings such as hedgerows (Baudry et al., 2000) and protected forests along waterways (riparian reserves - Mitchell et al., 2018). By monitoring the success of these projects, we can inspire other landowners to spare land of plant environmentally important species voluntarily. Conversely, if a typically successful practice does not work, there could be unknown drivers, such as competition from invasives, disease, pollution or illegal behaviour, that may only be uncovered through ecosystem monitoring. The theory of the niches has inspired the idea that animal biodiversity can be conserved by diversifying orchards with mixed fruit trees (Round et al., 2006) and agroforestry systems (Harvey & González Villalobos, 2007). Agroforestry involves growing trees and shrubs within land designated for crops and pasture. Trees planted in agroforestry systems can be harvested for timber, fruits, and nuts, while the whole system results in a more complex community assemblage than monoculture (Harvey & González Villalobos, 2007).

Ecosystem monitoring is necessary to inform decision-makers on mitigating environmental and economic decline. Specifically, continuous monitoring is needed to understand what, when, where and how much change is occurring, to understand what causes change and use that to inform practice (Sparrow, et al., 2020). Specifically, the right type of data needs to be collected in sufficient detail at appropriate spatiotemporal scales. As the extent and variety of monitoring equipment develops, we are entering unchartered territory in the scale of ecosystem appraisals. This kind of work uncovers surprising and vital previously unknown processes governing the success of ecosystems and will ultimately help us to make smarter decisions about how to protect them.

1.3. (AUTONOMOUS) ECOSYSTEM MONITORING

Understanding, monitoring, and providing evidence of why and how sustainable management practice works is necessary for optimisation and uptake. Conservation practice relies on evidence-based science, which relies on good data from the field. Having sufficient data and asking ecological questions with solid statistical backing is pivotal for evaluating trends and drivers of population change and collapse, designing and assessing conservation plans, and biodiversity policy commitments (Honrado et al., 2016). Later work cites the importance of high-resolution, multidimensional, and standardised data (Farley et al., 2018).

While many manual data-collection practices are still in widespread use today, there has been an explosion in the availability of machine-based data-collection that can be automated. This automated data collection quickly generates enormous amounts of data that surpass human inference. A recent review of automated ecological data collection described how recorders fall into one of three categories: (1) Acoustic Wave Recorders (microphones, hydrophones, geophones, and sonars), (2) Chemical Recorders (environmental sample processors, DNA, eDNA), and (3) Electromagnetic wave recorders (cameras, LiDAR, and Radar) (Besson, et al., 2022)

Beyond expanding data collection tools, scientists with ever-increasing computer and statistical literacy can draw from an increasingly broad toolbox of machine learning and deep learning-based analysis methods. Machine learning has revolutionised data analysis and consistently identifies features and patterns accurately and reliably in ecological datasets (Christin et al., 2019; Stowell, 2022). Data can now be collected and analysed autonomously, vastly reducing work hours and allowing researchers to ask a previously impossible range of questions in short time frames.

1.3.3. Traditional Methods

Traditional ecosystem monitoring is based on researchers regularly entering their study sites and manually collecting observations, data and specimens. These observations and collections could take the form of species counts (identified through sight and/or sound) (Huff et al., 2000), species capture for body condition testing, marking, tagging and/or tracking, sample collections and testing, or in-situ field experiments such as exclusion assays. While these methods have led to formative conclusions in natural history and ecology, they can be expensive, labour-intensive, and subject to experimental bias (Costello et al., 2016; Fitzpatrick et al., 2009). Not to mention that these assays require researchers to enter and potentially damage sensitive environments repeatedly. Additionally, many of these types of data collection do not retain "proof". An observation or behaviour may be observed in the field, but without any recorded evidence, it is impossible to prove that these recordings are not a result of an error. Even expert field identifiers get things wrong sometimes, and without a permanent record, important information could be erroneous or missed.

1.3.4. Chemical Methods

All organisms are in constant flux with their environment through respiration, photosynthesis, digestion, decomposition, excrement or various hormones or bodily fluids. These processes leave traces of information in the environment about what organisms and processes are at play. Carbon and Nitrogen content in the soil and air have long been studied as proxies for forest respiration and carbon sequestration. Collecting mixtures of chemical compounds from a forest can also be used as a fast and cheap method of metabarcoding, which can be used to look for indicators of certain species. While theoretically interesting, chemical metabarcoding requires extensive reference libraries and is not yet suited to large-scale field deployment (and references therein). DNA, or more specifically environmental DNA (eDNA), is more widely used to determine species presence and community compositions, using samples from the soil and air (Besson et al., 2022; Boussarie et al., 2018; Clare et al., 2022; Li et al., 2021). Recently portable and autonomous Environmental Sample Processors (ESPs) have been developed to perform DNA amplification and storage without human intervention (Yamahara et al., 2019). Fully automated eDNA collection and sequencing is not yet possible, but recent work indicates it may be in the future (Besson et al., 2022; Huo et al., 2021)

1.3.5. EM Wave Methods

EM wave sensors can either be active (where data is recorded after a signal is emitted – LiDAR and radar) or passive (digital cameras/ camera traps) (Besson, et al., 2022).

1.3.5.1. Camera Traps and Photographic Sensors

Ground-level camera traps (including thermal and inferred cameras) are among the most popular autonomous species monitoring methods due to their relatively low cost and easy deployment. Camera traps are self-contained, often battery-powered units triggered to take pictures (sometimes thermal and IR) by movement. The output of camera traps is usually 100s or 1000s of pictures of passing fauna (or falling leaves), which are frequently analysed manually or through deep learning algorithms. The success of camera traps may also be because, as visual animals, people are often more confident identifying an animal by its appearance than by how it sounds. As such, manual exploration and annotation of photographic data is perhaps more appealing. Alongside this, deep learning algorithms designed around image analysis are more developed than their acoustic counterparts. Some neural networks designed for acoustic scene classification are based on image analysis networks (Hershey et al., 2017)

Machine learning methods that take data from camera traps as input and have been trained for species tracking, counting, measuring, and monitoring behaviours (Besson et al., 2022 and references therein). Labelled camera trap training databases for many fauna and flora are already considerably well documented.

1.3.5.2. Remote Sensing (Hyperspectral Cameras and LiDAR)

Remote Sensing describes EM sensors usually mounted on drones or aerial vehicles. They generally fly overhead an area of interest and capture images of the research site. Hyperspectral cameras take multiple narrowband images, which can be used to extract indices from captured images to infer data such as vegetation intensity or type. Comparing these indices over time scales makes it possible to observe growth rates. A notable downside of hyperspectral imaging is the low resolution, which can be somewhat countered by the inclusion of LiDAR and/ or digital photography (Besson et al., 2022, and references therein). LiDAR (Light Detection and Ranging) is a form of laser imaging that emits pulsed laser light and records the scatter (much like sonar). LiDAR images can reconstruct high-resolution 3D representations of targets and can be used to study an environment's vertical structure and 3D complexity beyond just its canopy cover. LiDAR is known as Terrestrial Laser Scanning

(TLS) when employed at ground level. Satellite remote sensing and LiDAR can be combined for vegetation monitoring over huge-spatiotemporal scales (Pettorelli et al., 2016).

Other EM wave sensing methods involve flow cyclometers and electromagnetic ray sensing devices but are less commonly used (See Besson et al., 2022). Individuals can also be collared/ tagged and tracked with GPS and radio sensors, although this is invasive and not autonomous.

1.3.6. Acoustic Methods

Acoustic wave methods, explored in more detail in section 4, cover the use of devices which record waves of pressure through a medium. Specifically, that is the use of microphones (pressure waves in air), hydrophones (pressure waves in water) and geophones (pressure waves through the ground).

Sonar is an "active" alternative to the above, which uses sound-emitting recorders that use scattered returned echoes as measurements of organisms and the environment (much like LiDAR) (Besson et al., 2022). Sonar is almost entirely carried out in aquatic environments and can be used to pick up species as small as krill (Bernard & Steinberg, 2013) or fish and squid at depths of 800m (Dunlop et al., 2018). Recent machine learning developments in the use of sonar have also shown that sonar outputs can be used to distinguish between species (Porto Marques et al., 2021).

1.4. SOUND ECOLOGY: THEORY, RECORDING, AND PROCESSING

Sound has played an essential role in the development of humanity. While this is most notably using sound for communication, we have also used sounds to build community through music and dance, to hunt and avoid being hunted, to find water in the wilderness, and to cross busy city streets. While sometimes overlooked as a resource, the sonic environment can provide a wealth of information that we, and all species, can and do exploit daily. In this section, I explore how sound is used by animals, theories as to how the usage of audio communication evolved, introduce the physical properties of soundwaves, and how they are commonly recorded and analysed in the field.

1.4.3. The Soundscape and Sound Ecology

The soundscape was first defined in 1977 as "any portion of the sonic environment regarded as a field for study" (Schafer, 1977). Sound ecology refers to the acoustic relationship between organisms and the environment. This definition was taken further by Bernie Krause, who coined the term "biophony" for all sounds made by living organisms in 1998 (Krause, 1998). Stuart Gage, working with Krause at the time, later included geophony: non-biological sources of sound such as wind, rain, and running water (Krause, 2008). The anthropophony was defined later by Krause, Gage, Bryan Pijanowski and others as human-generated sounds such as speech, machinery, traffic, and music (Pijanowski et al., 2011). This paper also coins the phrase "Soundscape Ecology", which defines a framework of interactions between climate, humans (policies, activities, and values), landscape structure, natural species dynamics (life history, population and community dynamics, and geophysical motion), and ultimately patterns in the soundscape (Pijanowski et al., 2011).

Sound is a rich, polluted, and potentially understudied resource in ecosystem monitoring and global change (Buxton et al., 2017). The following sections detail how geophony, anthropophony and biophony come together to form the soundscape (Figure 1.1).

1.4.3.1. Geophony

Geophony describes all the natural (non-animal) sounds in an ecosystem. Geophonic sounds fall into four categories: wind, water, weather and geophysical forces (Krause, 2008). Often it is not these processes directly that are being heard but rather their effects, be that the sound of wind shaking trees and rustling through grasses or the sound of water crashing into rocks in waterfalls or splashing onto leaves as they fall from the sky. Geophony can shake the whole acoustic environment. A heavy thunderstorm, a thick blanket of snowfall, the eruption of a volcano or a gentle stream trickling near its source on a mountain created vastly different conditions in an acoustic environment that can vary enormously on a daily scale. Geophony often dominates a soundscape, and efforts have been made to identify and remove geophony from recordings (Metcalf et al., 2020).

1.4.3.2. Anthropophony

Anthrophony describes all the human-made noise in an ecosystem. Human-made noise can be separated into four categories: electromechanical, physiological, controlled sound and incidental (Krause, 2008). It has been speculated that industrialisation romanticised mechanical sounds as signs of authority and prosperity (Schafer, 1977). However, the expanse of the anthropophony is now having an adverse effect on vocal communication in terrestrial and marine environments (Derryberry et al., 2020; McCormick et al., 2018; Jérôme Sueur et al., 2019a).

1.4.3.3. Biophony

Most social species maintain relationships through preening, grooming, hugging and so on; however, close contact is not always possible, and acoustic communication can be used to maintain social connections at a distance (Chereskin et al., 2022). Further, sound can travel through substances and for distances that light cannot. Sound is often far more helpful than visual or tactile communication in densely vegetated, dark, or situations where calling individuals are far apart. This makes the sonic domain a precious resource for animals in dense, dark, or sparse habitats.

Animals use sound in courtship (Elise et al., 2019), bonding (Chereskin et al., 2022), threat alerting (McCormick et al., 2018), establishing hierarchies (Vanden Hole et al., 2014), establishing territories (Laiolo & Tella, 2006), echolocation (Jones & Teeling, 2006), parent-offspring identification (Warren et al., 2006), and even as primitive languages (e.g. Bruck et al., 2022; King & Janik, 2013). Animal noise emission is not just vocalisation; other properties such as stridulation, striking, wing clapping, chest-beating, and rattling are also used as non-vocal communication. Beyond just intentional sonic emissions, species also unintentionally create noise through breathing, moving, and feeding (Bradbury & Vehrenkamp, 1998). The amount of sonic information given by just a single individual over the course of a day is often overlooked but incredibly significant.

Every year, more studies are emerging that previously thought unvocal species or unvocal genders do sing and with more complicated repertoires than previously thought (primitive language in rodents - Dent et al., 2018; "purring" spiders - Sweger & Uetz, 2016).

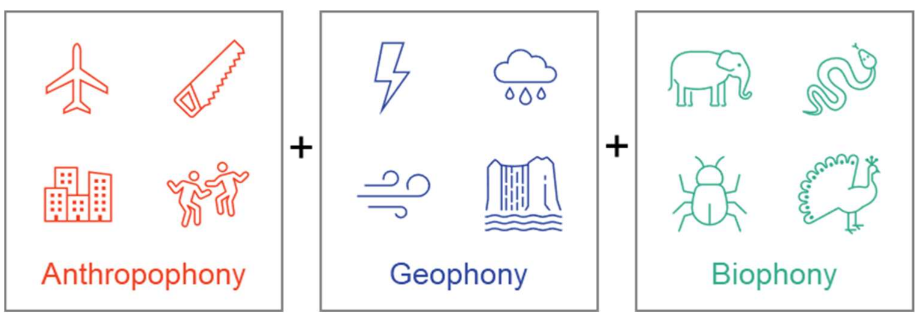


Figure 1.1 details the components of the soundscape: Anthropophony (human-made sound), Geophony (non-biotic environmental sound) and Biophony (biotic sound).

1.4.4. *The Evolution of Song*

Darwin speculated in the origin of species that complex birdsong had evolved as sexual selection relating to female "standards of beauty" (Darwin, 1861). In songbirds, it was generally thought that females do not sing alone, and any functionality of female song has largely been ignored or written up to hormonal imbalances. However, recent research has begun exploring the prevalence and role of singing females. Conversely, female Song has been found in 71% of examined songbirds (passerines), with analysis finding a 91.9% likelihood that ancestral female passerines were songbirds (Odom et al., 2014). Evidencing that the evolutionary basis for song and other means of vocal communication is still not widely understood.

A recent review (Alcocer et al., 2022) explored two conflicting arguments for the use of acoustic space in vocal animals: the Acoustic Niche Hypothesis (ANH – Hödl, 1977) and the Acoustic Adaptation Hypothesis (AAH – Morton, 1975). The acoustic niche hypothesis (ANH) centres around the idea that acoustic space (in frequency, temporal, and perhaps spatial domain) is a finite resource that organisms have competed for millennia. It is proposed that evolution drives species to structure their signals in relation to each other, partitioning the soundscape and avoiding masking via signal overlap (Alcocer et al., 2022; Hödl, 1977; Krause, 2008).

This theory logically entails that as species richness increases in an area, the acoustic space will become more saturated, which is the foundational principle of acoustic indices (explored later). AAH, an alternate theory, works on the principle that songs from multiple species may converge to maximise signal transmission (Morton, 1975). It is already known that different types of signals propagate and degrade at different rates through different habitats, so it is possible that in optimising their transmission, calls from vocal species converge on similar properties. Multi-species convergence would result in a less partitioned soundscape and may suggest that direct soundscape to species richness relationships may be underestimating diversity. Evidence supporting either AAH or ANH is mixed, with a meta-analysis showing an overall weak effect size for both (Alcocer et al., 2022 and references therein). Until the drivers of acoustic niche occupancy are determined, considering soundscape partitioning as a direct link to species diversity may be misleading (Alcocer et al., 2022).

1.4.5. Sound Recording and Processing

1.4.5.1. The Physical Properties of Sound

Sound is perceived from oscillating waves of pressure travelling through air or another medium that are received/ recorded by humans, animals, or microphones. Sound is produced by sudden movements that physically compress air locally and cause the propagation of pressure waves outwards. Commonly used properties and descriptors of sound are given in Table 1.2

Table 1.2 gives commonly used properties of sound, sound waves, sound recording and wave propagation.

Property	Commonly used Unit	Definition
Speed	ms^{-1}	The speed the wavefront of the sound wave travels in a medium. Speed depends on the medium's physical properties and can be affected by temperature/ humidity/ pressure. It is generally $343ms^{-1}$ in air and $1,480ms^{-1}$ in water (higher particle density).
Frequency	(k)Hz	The number of complete wave oscillations per second. The more oscillations, the higher the perceived frequency (or pitch). The lower bound of human hearing is 20Hz, and the higher bound of human hearing is 20,000Hz. Infrasound is a descriptor of sound too low frequency to be heard by people; ultrasound defines frequencies too high for human perception.
Wavelength	m	How long the complete cycle of a wave is spatially. In air, 20 Hz is 17m, and 20,000 Hz is about 1.7cm long.
Amplitude	dB	dB is a logarithmically scaled measure of power. 10dB = x10, 20dB = x 100, 30dB = x1000 and so on. As dB as a measure of power, it needs to be attached to a defined value to have any magnitude.
	dB SPL	dB SPL (Sound Pressure Level) is the magnitude of pressure oscillations relative to the quietest sound a human can hear ($2x10^{-5}$ Pa). The dB SPL is proportionate to the wave's amplitude and can be considered as perceived loudness.
	dBFS	dB FS (Full Scale) is the amplitude level compared to the maximum amplitude recordable before clipping occurs in digital recording.
Sonic Attenuation	-	The loss of signal as a sound wave propagates through a medium. Sound waves are attenuated more aggressively the closer the obstruction is in size compared to the signal's

		wavelength. High-frequency (short-wavelength) sounds are therefore attenuated more easily (Russ, 2013).
Sampling Rate	Hz	The number of samples (air pressure, voltage etc.) taken every second by a digital recorder
Nyquist Frequency (Recording Frequency)	Hz	Exactly one-half of the sampling frequency. The Nyquist frequency represents the maximum frequency oscillation recordable by the microphone. If the target frequency is less than half the size of the sample rate, the signal will become distorted through a process known as aliasing
(Temporal) Aliasing		A type of distortion that occurs when a signal's full extent cannot be captured. Aliasing happens in both visual and temporal mediums. An example of aliasing is when the spinning blades of a helicopter recorded digitally appear to be stationary (or even move backwards) as the shutter speed is not fast enough to capture more than a single frame per rotation. The same happens with acoustic recordings, where whole wave oscillations can happen within the time between samples. Waves can be reconstructed/ smoothed between sample rates, but there is no way of inserting a whole wave oscillation between samples. By ensuring that twice as many samples are recorded than oscillations, we can ensure waves can be reconstructed correctly. Aliased acoustic recordings sound lower pitch and distorted compared to the raw signal. Anti-Aliasing filters can be used to stop the reconstruction of waves with higher frequencies than half of the Nyquist frequency to reduce distortion.
Bit-depth	bit	The number of binary bits used to store each pressure sample—1-bit results in 2 levels (0 and 1), 2-bits in 4 levels and so on. Bit depth is a discrete levelling used to rebuild the soundwave in playback. Larger bit depths take up more storage but can recreate the soundwave to a much higher resolution. Bit Depth affects the dynamic range of the recording.
Frequency Modulation		Where the frequency of a signal or call changes during the call, often this happens very quickly. Frequency Modulation is perhaps more widely known for its use in radio transmission (e.g. FM radio), whereby changes in the frequency of a transmitted signal can be used to encode information.
Dynamic range	dB SPL/ FS	The difference between the loudest and quietest sounds, either: recordable by a device, of a signal, or in a soundscape.

In digital sound recording, a recorder takes air pressure samples (SPL) at a fixed rate (sample rate, typically 44100Hz) and stores those reading by a defined number of binary digits (bit depth, typically 16bit). Sampling involves transducing the sound pressure wave to an electrical signal and then digitally sampling the electrical signal at regular intervals. Specifically, a transducer converts the mechanical pressure energy to electrical energy in the circuit. Then an Analog to Digital Converter (ADC) or a Digital Analog Converter (DAC) is used to periodically sample the electrical impulse and store that information in the digital domain.

1.4.5.2. Autonomous Acoustic Monitoring

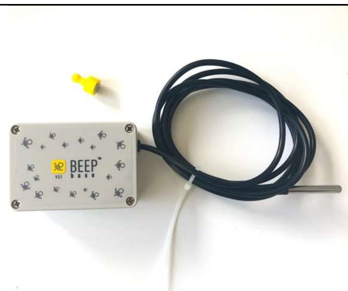
A major subset of acoustic monitoring is Passive Acoustic Monitoring (PAM). PAM centres around small computers that automatically process acoustic data recording and storage. These recorders can take many forms and are often designed to be left in the field for extended periods. These can take the form of Semi-Autonomous or Autonomous PAM recorders, where the former still require external interference for changing batteries/data collection while the latter have autonomous powering and data transfer.

Truly Autonomous Recording Units (ARUs) can mediate their powering and data transfer without external interference. Powering can be autonomised through renewable power sources or connection to the grid (Saito et al., 2015; Sethi et al., 2018). Data transfer can be autonomised through connection to the cloud via mobile networks (Sethi et al., 2018) or to central hubs via data transfer nodes (Bruggemann et al., 2021)

PAM has been employed in marine (Sousa-Lima et al., 2013), terrestrial (Roe et al., 2021) and freshwater (Desjonquères et al., 2020) habitats across the globe. Taxa-specific studies have been completed on birds (Shonfield & Bayne, 2017), bats (Gallacher et al., 2021), other terrestrial mammals (Spillmann et al., 2015), marine mammals (Zimmer, 2011), and insects (Jain & Balakrishnan, 2012). These autonomous and semi-autonomous soundscape recorders are being deployed as tools for continuous ecosystem monitoring across habitat and even continent-scale (Roe et al., 2021; Sethi et al., 2021; Sethi, Ewers, Jones, Signorelli et al., 2020).

Despite a recent deceleration of uptake, primarily credited to an international microchip shortage (Shead, 2021), all PAM devices have begun to explode in their uptake, with acoustics consistently considered a solution in modern ecosystem monitoring (Table 1.3).

Table 1.3 explains some of the emerging and commonly used commercial methods of PAM (including semi-autonomous and autonomous PAM, denoted here as ARUs) in sound ecology. AudioMoth, Wildlife Acoustics, HaikuBox and Beep.nl are available for purchase at the time of writing, while the others require self-building or are not yet in large-scale production.

Device Name	Image	Importance/ Novelty	Reference
Beep.nl (PAM)		Records temperature, sound, and weight Designed to monitor beehives Runs Autonomously. Sends data over networks. Can run for one year on AAA batteries	(BEEP, 2022)
BUGG (ARU)		Records audio (20Hz-80kHz) An autonomous recorder that uploads data via 3G/4G networks (offline mode optional). Works with grid or off-grid powering. Uploads to a web-based platform that performs soundscape feature extraction, species, and anomaly detection.	(BUGG, 2022)
Swift *and SwiftOne* (PAM)		Records audio (maximum sampling rate 96kHz) Battery-powered and SD card storage. Lightweight and cost-effective	(Cornell Lab of Orthinology , 2022)

ecoPi: (ARU)		Autonomous solar-powered and networked modular devices based on Raspberry Pis with several specialisations. EcoPi:Bird – automated audible range recording EcoPi: Bird2D – stereo audible range recording with some positional determination EcoPi:Bug – visual recording EcoPi:Bleep – detects bats fitted with a radio transmitter within 50m range EcoPi:Boom – Plays attractive recordings and records	(OekoFor, 2022)
AudioMoth *and derivatives* (PAM)		Low-cost, easy to use, widely used and available. Battery-powered and SD card storage. AudioMoth - records up to 384kHz HydroMoth - optimised for underwater deployments Dev – exposed headers for users to customise µMoth – 32x24mm and weighs just 5g	(Open Acoustics Devices, 2022)

Wildlife Acoustics *and derivatives* (PAM)		Recording and analysis platform (kaleidoscope). Extensive guides and training. Battery-powered and SD card storage. Widely used SM4* – optional stereo recording up to 96Khz, 510hrs run time Mini* – smaller and cheaper. Optional stereo recording up to 96kHz, 210-1040hrs run time Micro – smaller and cheaper. Recording up to 96kHz, 180hrs run time EchoMeter Touch – smartphone adaptor for ultrasonic recording up to 384kHz * also have bat detector options which can record up to 500kHz	(Wildlife Acoustics , 2022)
RFCx (ARU)		Automated solar-powered and satellite networked audio recording up to 18kHz Real-time threat (poaching etc.) detection and reporting Dedicated app for biodiversity monitoring	(RFCx Guardian Platform, 2022)

HaikuBox (ARU)		Designed for home recording, connects to Wi-Fi and reports species detections to users through push notifications on smartphones	(Loggerhead Instruments, 2022)

Raw audio recordings can quickly generate large amounts of data that can become costly and difficult to manage. Many sound ecologists use audio compression to reduce their file sizes. Compression takes two forms either lossy (where "inaudible" data is irreversibly removed from the recording) or lossless (where audio is encoded to be a string of repeating patterns instead of samples, resulting in reduced file size without compromising the acoustic content). Both compression methods have benefits and losses, which are explored further in chapter 2.

1.4.5.3. Visualising Acoustic Signals

More data than can be feasibly listened to is quickly generated using PAM. Visualisation through spectrograms is used to quickly observe the whole recording and look for patterns.

A spectrogram visualises soundscapes with time on the x-axis and frequencies on the y-axis. The amount of energy (amplitude) in each frequency band for a particular timestamp is shown as a colour, usually darker for quieter amplitudes, approaching white as they get louder. This energy distribution is calculated through a Fast Fourier-Transform (FFT). FFT is a standard method of converting data from one domain (like time or space) to another scaled by abundance or the frequency of data occurring within a given range.

1.5. DATA ANALYSIS IN SOUND ECOLOGY

Autonomous data collection, especially if recording from multiple locations simultaneously, quickly results in more data than can feasibly be examined by a human observer, even with spectrograms. As such automated methods of analysis have been developed. In the following section, I will outline popular automated methods for analysing field recordings in both Bioacoustics and Ecoacoustics (defined below). This section will focus on analysis to determine species richness and other measures of biodiversity. Information on automated methods for determining call functionality and behavioural neurology is beyond the scope of this brief overview, but information can be found at (Mouterde et al., 2017).

Bioacoustics is the study of sound relating to the calls of individuals or specific species. Bioacoustics covers species-specific monitoring, studies of call functionality, and behavioural neuroscience involved with animal communication.

Ecoacoustics covers studies which consider the soundscape and all its components together. Conversely, ecoacoustics revolves around quantifying and classifying whole soundscapes using descriptive statistics and more abstract descriptors such as convolutional neural net (CNN) embeddings.

1.5.3. Machine Learning (ML) and Deep Learning (DL)

Machine and deep learning are processes in data analysis whereby computer algorithms can "learn" and better themselves over multiple iterations to describe data optimally.

Specifically, **Machine Learning (ML)** is a broad subset of statistics involving building models trained to represent data patterns. ML includes linear and logistic regression, (k-means) clustering, decision trees and random forests.

Deep Learning (DL) is a subset of ML**, including** models built from artificial Neural Networks (NN). Typically, NNs are built of node layers (or artificial neurons), each containing an input layer, at least one hidden layer and an output layer. The node layers interconnect and are associated with each other by different weights and thresholds. Nodes are activated, passing information to the next layer if the threshold of their output layer is reached. Unlike other ML, DL models can take in raw data (images, text, or audio).

ML and DL can be supervised, unsupervised or semi-supervised. Supervised learning requires labelled datasets and involves the model learning through a stepwise adjustment of weights and thresholds. After each adjustment, the model's error function (accuracy) is assessed, and the subsequent adjustment is made, hopefully moving towards a reduction in error. This optimisation process usually happens over a specified number of steps or until the model is deemed sufficiently accurate. Supervised methods can fall victim to overfitting, which occurs when a model has been pushed to match a training dataset too closely. Overfitting results in the model picking up artefacts in the training data that are not represented in the broader context, making the model less accurate on unseen data. NNs, Logistic and linear regression are examples of supervised learning.

Unsupervised methods do not require labelled data and are intended to pick up patterns and trends in the data and separate groups. This separation is often achieved by creating an artificial axis through the datasets, which creates the greatest separation between points – often revealing clusters in the datasets. Standard unsupervised methods are principal component analysis (PCA) and k-means clustering.

Finally, semi-supervised methods use a small subset of labelled data to describe an unlabelled winder dataset. When the models reach labelled data, it behaves similarly to a supervised method; for unlabelled data, it works to minimise the difference between the unlabelled point and other similar points. Including unlabelled points in semi-supervised learning makes the model more robust to variation in the data. Semi-supervised learning can be helpful when it becomes too costly or time-consuming to label the entire dataset.

1.5.4. Bioacoustics

Bioacoustics focuses on recognising species or individuals. Using bird calls to identify species is a traditional method, with notes on bird alarm calls and songs in almost every bird identification book. Retaining all target bird calls to memory requires many hours of training and skill and is not usually feasible. Fortunately, DL methods have been developed and are now widely used for detecting and classifying acoustic events.

Extracting species data from continuous soundscapes can be messy as soundscapes are usually noisy, and more than one species may be calling at any time. Therefore, automating bioacoustic analysis with DL is a two-step process: Detect and Classify (Stowell, 2022).

- **Detection** can occur as binary classification, sound event detection or image object detection (Morfi et al., 2021; Stowell, 2022; Stowell et al., 2019; Venkatesh et al., 2022). Binary classification returns yes/no across a time series of detections. Sound event detection involves the feature/ML-based detection and extraction of calls from a recording. Image object detection involves taking in spectrograms and using image boundary classifiers to separate individual calls (Stowell, 2022).
- **Classification** can be achieved by creating and training a new DL model or using an established setup. Training datasets can be created from personal libraries or open-source directories such as Xeno-Canto.org. Data augmentation, which involves artificially adding noise/ slight variations to labelled samples to increase training dataset size and variance, can also be employed (Stowell, 2022). DL classifiers are trained on a library of labelled calls from a target region. Once the classifier accurately predicts the known data, it can be used to classify unknown data.

DL classification models can be RAM intensive and take a while to train, but training needs only occur once; once trained, models are lightweight and deployable. Bioacoustics DL models are widely available as off-the-shelf programmes. BirdNET is one example with a lightweight download, user-friendly GUI, and can analyse unseen recordings in a fraction of their time (Kahl et al., 2021). BirdNET has an app, GUI and command line implementation for people with different skill levels and requirements. Bioacoustics-based species identification apps can be used to help twitchers, citizen scientists, and researchers identify and log the birds they see. These apps are easy to use and incredibly popular (BirdNet 1M+ downloads, Merlin Bird ID by Cornell Lab or Orthinology 1M+ downloads (Google Play, 2022)).

While showing promise in successfully identifying many taxa, bioacoustics depends on well-labelled training datasets, and under-represented species, especially insects, in those datasets often go undetected (Baker & Vincent, 2019). An alternative to bioacoustics that is less dependent on human-assigned value is an approach that considers the whole soundscape (ecoacoustics).

1.5.5. Ecoacoustics

Ecoacoustics considers the whole soundscape, combining biophony, geophony and anthropophony. In doing so, ecoacoustics gives space to less charismatic, unheard, or understudied species. Ecoacoustics can also monitor the anthrophony's manner, extent and perhaps impact. Ecoacoustics is broadly studied through two main routes: Acoustic Indices or DL-based soundscape embeddings.

- **Acoustic Indices** are statistical interpretations of a whole soundscape based around uncovering biophony. Specifically, they work by quantifying aspects of the soundscape related to increased biophony via the ANH, such as the range of frequencies in the signal and the spread of energy across different frequency bands. Over 60 acoustic indices have been developed since 2007 (Buxton, McKenna, et al., 2018). Examples: Acoustic Complexity Index (ACI) (Pieretti et al., 2011), Normalised Difference Soundscape Index (NDSI) (Kasten et al., 2012).

- Instead, **DL Embeddings** run the whole soundscape through a pre-trained acoustic classification algorithm, stopping the process just shy of the classification. What is left is a vector of numbers that can be used as an abstract soundscape descriptor (like a barcode or fingerprint). These abstract descriptors can then be used to train and build models for downstream classification. Examples: VGGish, wav2vec (Baevski et al., 2020; Hershey et al., 2017).

Acoustic indices have widespread use, primarily accredited to their ease of use and the wealth of guiding information on their usage (Bradfer-Lawrence et al., 2019; Browning et al., 2017; Gasc et al., 2013; Metcalf et al., 2021). Further, acoustic indices can be easily extracted through straightforward free packages in R: *seewave* (Jerome Sueur et al., 2008) and *soundecology* (Gasc et al., 2013; Villanueva-Rivera & Pijanowski, 2016). Their applications include monitoring species richness (Eldridge et al., 2018), community composition (Jérôme Sueur et al., 2008), the relative contributions of biophony, geophony and anthropophony to a soundscape (Kasten et al., 2012), approximating species abundance (Papin et al., 2019), as means of more intuitively visualising soundscapes (M. Towsey et al., 2018), or even as a means of mapping an area's biodiversity (Carruthers-Jones et al., 2019).

Acoustic indices have long yielded mixed results, which a recent meta-analysis aimed at exploring their efficacy in biodiversity representation (Alcocer et al., 2022). The meta-analysis stated that there were "significant shortcomings in the theoretical framework", referring to ambiguity in the evolution of the soundscape (ANH and AAH) (Alcocer et al., 2022). The meta-analysis also states a weak relationship and highly variable effect sizes between many of the most used acoustic indices and species diversity metrics (Alcocer et al., 2022). DL-based methods are briefly mentioned in the review, but their "complexity" limits their uptake in the field (Alcocer et al., 2022).

DL-based methods are being used as soundscape descriptors to a much smaller degree than acoustic indices. Models built from DL embeddings make for descriptive visualisations and are successful predictors of landscape, biomass, and species (Sethi, Jones, Fulcher, Picinali, Clink, et al., 2020). DL embeddings have been shown to outperform acoustic indices on landscape classification tasks and their robustness to experimental variation (Heath et al., 2021; Sethi, Jones, Fulcher, Picinali, Clink, et al., 2020).

These whole soundscape applications are lightweight, generalised models for exploring soundscapes but with unclear connections to ecological theory. No less because species vary in repertoire size, calling rate, calling behaviour, response to noise and detectability distance. Considering just whole community-level recordings does not take this variation into account.

1.6. THE FUTURE OF SOUND ECOLOGY

Bioacoustics and Ecoacoustics are exciting and developing fields, with studies increasing yearly (Alcocer et al., 2022). Development of hardware and software and increasing evidence of their applicability in the field of ecoacoustics are coming out regularly, with state-of-the-art changing at least every two years significantly (Stowell, 2022). A recent review of methods in ecosystem monitoring specifically cited that "technologies such as automatic recorders have not reached their full potential to support modern ecological monitoring" (Besson et al., 2022; Hampton et al., 2013; Tuia et al., 2022). In a maturing field, utilisation of the state-of-the-art requires a breadth of specialist research that may be off-putting for those unfamiliar (Besson et al., 2022). Updated guidelines in the field are regularly published; still, there is a lag between new applications emerging and this reaching common practice in the field. At the very least, it will be exciting to see how existing tools are applied on a broader scale as they become commonplace in the field.

1.6.3. Developments in DL

DL is already commonly used in bioacoustics through easy-to-use, low-CPU software like BirdNET (Kahl et al., 2021). As computer literacy increases with user-friendly GUIs and the inclusion of DL methods in widespread bioacoustics and ecoacoustics guidelines, so will their deployment in broader research.

Machine learning is a rapidly evolving field with new methodologies or applications coming out on an almost monthly basis. State of the art for computational bioacoustics was recently reviewed, highlighting possible avenues of development in the field, including Transformers, Active Learning and Few-Shot learning (Stowell, 2022 and references therein).

Transformers are a type of Semi-Supervised Neural Network designed for translation via natural language processing (NLP) (Vaswani et al., 2017). Transformers work well for NLP for the same reason they might apply well to bioacoustics: they are built to analyse sequential (such as time-series) data. Unlike RNNs, which also take in sequential data, they do not analyse it sequentially, meaning they lean well to parallelisation, allowing them to build sophisticated models from vast amounts of data quickly. Transformers benefit from positional encoding and self-attention, meaning that learning order information is built into the model, which can prioritise (give attention to) different aspects of data based on their importance training. Transformer models (wav2vec (Baevski et al., 2020)) have been applied to an enormous array of tasks and have been showing some promising early results in bioacoustics (Elliott et al., 2021; Stowell, 2022).

Few-shot learning is where a model is pre-trained across multiple similar tasks, such that when introduced to a new (but similar) task, it already has the infrastructure to build a classification model quickly. In bioacoustics, his new task could be introducing a new type of call to a model that has already learnt to classify several different calls successfully. (Stowell, 2022; Wolters et al., 2021).

Finally, Stowell suggests using **Active Learning,** which restructures the model to include more human-machine interaction (Stowell, 2022). The model shows some uncertain predictions to the user who feeds correct information into the system to assist optimisation. Active Learning requires fewer labelled data than a supervised model and makes better use of the researcher's time (Stowell, 2022 and references therein)

A very recent development in DL is AudioLM, a language model able to generate passable audio (Borsos et al., 2022). Using sound generation algorithms could create a broader range of augmented recordings for ML and DL model training – perhaps resulting in fewer artefacts. It may also be interesting to see whether synthetically generated "bird calls" can fool machine or even human listeners.

1.6.4. Determining Detection Ranges and Likelihood

While it sounds straightforward, the distance an acoustic signal will penetrate through an environment depends on several situational qualities, such as signal amplitude, frequency, attenuation profile, environmental masking, reverberant properties of the environment, and how direct the path is. Therefore, the audible range of a device depends heavily on where it is and what it is hoping to detect (Browning et al., 2017). Knowing a species' detection likelihood and distance is pivotal for extending bioacoustic monitoring beyond yes/no occupancy monitoring and towards abundance.

1.6.5. Developments in Hardware

The reduction in size and expense of micro-computers has been vital in the exploding tech available in sound ecosystem monitoring. As available technology develops further, we may see an influx in the capability of devices. I suspect the following developments will have significant impacts on the field in years to come: battery-free tech, low-latency satellite connection, terrestrial sonar, automated sensor deployment, biodegradable sensors, integrated sensors, and multichannel acoustic recording for sound-source localisation.

When regular computers and microcomputers are switched off, they go through a set of "shut-down" protocols that ensure the device enters the sleep phase safely. Immediately pulling the plug on battery-less devices causes them to lose power without running these protocols, sometimes corrupting the operating system and rendering the device useless. In ARUs powered by renewable (and often unpredictable) sources like sunlight, sudden power loss is a regular occurrence, meaning all devices using these power sources need to be equipped with clunky batteries and other power management hardware. These systems may only add to the issue as sometimes even sedentary batteries can drain power. A recent proposal for **Battery-Free Technology** (Lostanlen et al., 2021) explores a micro-computer that works “battery-free” and does not run into corruption in sudden power loss. Exploratory battery-free technology is promising and may be a valuable solution for all ecosystem monitoring practices requiring lightweight attachments, such as devices deployed by drones and devices used as animal tags or trackers.

ARUs may also benefit from connections **to widespread low-latency satellite communication networks** such as Starlink (Starlink, 2022). Acoustic data can reach large volumes quickly, and at present, ARU placement is limited to areas of solid, fast connectivity if the devices transmit data at least at the speed it is generated. With widespread low-latency connectivity, such as that provided by Starlink (20ms vs 600ms+ of standard network connectivity (Starlink, 2022)), it will be possible to deploy ARUs at a broader range of (remote) locations. Starlink currently covers large portions of Northern America, Australia and Western Europe but is hoping to expand to include the rest of the Americas, Europe, Africa, Australasia and some countries in North Asia and the Middle East in 2023 (Starlink, 2022).

Underwater sonar was mentioned in an earlier section, capable of identifying species of fish and arthropods (Bernard & Steinberg, 2013; Dunlop et al., 2018). However, research in **terrestrial applications of high-frequency sonar** remains understudied. Parallel to echolocation in bats, terrestrial sonar could potentially be used in 3D habitat reconstructions in terrestrial landscapes.

One drawback of even the most advanced ARUs currently available in ecosystem monitoring is that they require human interaction for deployment and removal. Recent speculative work has questioned whether **biodegradable sensors**, such as those used in the human body, could be used to collect data on enormous scales and pose that this would transform how natural ecosystems are managed and understood (Sethi et al., 2022). Outside of biomedical research, these sensors are in their speculative phase, but sensors of this type – deployed by drones, could lead to the automation of the entire fieldwork process.

Currently, most remote ecosystem monitoring research revolves around developing specialised equipment for acoustic monitoring OR visual monitoring OR chemical monitoring. However, getting a complete view of the species and ecosystem processes within an environment could benefit from devices with **Integrated Multimodal Sensors**. Multimodal sensing has been deployed in weather stations and agricultural IoT to report real-time multimedia data about ammonia, CO2, temperature, and sometimes sound and image (Lohchab et al., 2018). In ecosystem monitoring, this is much less explored but would give valuable insight into the holistic functioning of the system. Occasionally multiple sensor types are deployed simultaneously (Buxton, Lendrum, et al., 2018), but to my understanding, an integrated platform for ecosystem monitoring does not yet exist.

1.6.6. Linking Sensors to Real-Time Action

It was recently suggested that bioacoustics and ecoacoustics are going through the intermediate stages of an industrial revolution (Jones, 2022). The general four stages of the industrial revolution in any industry can be categorised as follows:

(1) Industry 1.0: "Mechanisation and steam power",

Akin to manual data collection, species traps etc.

(2) Industry 2.0: Mass production, assembly line, electricity

Electrical data collection and storage

(3) Industry 3.0: Computer and Automation

Automated data collection, transfer, and analysis pipelines

(4) Industry 4.0: Cyber-physical Systems and Networks

Automated intelligent (and potentially mediating) response to analytics

Ecosystem monitoring, particularly with sound, has surpassed manual mechanistic data collection (Industry 1.0) and has comfortably been collecting data through electronic recordings and storage for decades (Industry 2.0). Automated data collection and analysis (Industry 3.0) in the field is becoming more widespread but not yet ubiquitous. Finally, the use of systems that autonomously collect, analyse and respond to data (Industry 4.0) is only just emerging and is not widely used.

Integrated recording and response have widescale usage in agricultural IoT systems (Lohchab et al., 2018), including automated feeders, moisture control, feeding and ventilation. While not a fully integrated action station, an autonomous recording, detection, and *alert* system has been developed to report illegal behaviour in protected areas (RFCx Guardian Platform, 2022).

Wildlife Acoustics also recently announced a "Smart System", a networked bat detection system that enables wind turbine owners to respond to bat presence in real-time, hoping to reduce bat fatalities (Wildlife Acoustics , 2022). As PAM develops, we may see exciting new applications of real-time conservation responses as a direct result of ecoacoustic and bioacoustic triggers.

1.6.7. Enhanced Data Sharing

The scientific community already have a wealth of acoustic data across massive spatiotemporal scales, often belonging to just one lab or group of researchers. A comprehensive platform for acoustic data/ methods and research was recently suggested and is in progress (audioBLAST!, 2022), and would make such analysis possible.

1.6.8. Spatial Acoustics

Spatial Acoustics, explored in greater detail in chapters 3 and 4, describes an emerging methodology of enhanced recording and analysis that allows researchers to approximate the position of the origin of a sound. Knowing the position of an individual has many applications, including assessing the behaviour and movements of individuals, studying interactions, identifying individual identities, sub-setting (and beamforming) sounds for more detailed analysis, calculating species abundance more accurately, and inferring habitat usage/territory (Rhinehart et al., 2020).

Sounds can be localised when the difference in the time it takes for a signal to reach multiple known points in space is known. Specifically, if a signal is recorded at multiple fixed positions, the time difference between the signal reaching those positions can be used to triangulate the signal's position of origin. There are broadly two types of sound-source localisation: Hyperbolic and Direction of Arrival (DOA). Hyperbolic localisation occurs when microphones are widely spaced in an array across a study area; sounds that originate *within* the array can be approximately localised to a geographical position. DOA estimation instead centres around smaller microphone arrays, often on a single device, where external sound sources are localised to directions relative to the device rather than exact positions (Rhinehart et al., 2020). Multiple DOA recorders can also be used to determine positional information about a signal's location of origin. In both methods, localisation depends on ultra-fine scale temporal differences at each microphone. In ordinary (omnidirectional) recorders, the onboard clock of a device can drift slightly by a few seconds every day. While this does not have huge implications for omnidirectional recording, this can have a catastrophic effect on the ability of a multichannel recorder/ recording array to localise a signal.

Multiple established recorders can be used in acoustic localisation; however, this can become expensive (Wildlife Acoustics: £700+/unit, AudioMoth: £60+/unit, SAFE Acoustics: £200+/unit) and require additional hardware and software adaptations for necessary cross-microphone synchronisation. For this reason, most recorders used in localisation studies are built especially for a particular study or lab. Notable recently developed examples include CARACAL, VoxNet, and the Dev-Net/TAMAGO system used by the team behind HARKBird. The CARACAL (Conservation at Range through Audio Classification and Localisation) system consists of several units, each with four MEMS microphones and a GPS localiser and synchronisation (Wijers et al., 2021). The CARACAL system detects signals using thresholding; once detected and localised, signals are then beamformed and extracted to improve SNR. VoxNet, a development from Acoustic ENSBox (Girod et al., 2006), is a complete hardware and software system that centres around a wireless network of ARUs, each with four microphones, automated time synchronisation, localisation and network coordination (Allen et al., 2008). VoxNet was designed to be an “out of the box” tool for acoustic localisation, and the system houses internet connectivity for remote data transfer (Allen et al., 2008). The Dev-Audio/TAMAGO system used by the lab which developed HARKBird consists of single units, each with 7/8 microphones respectively (Suzuki et al., 2017), and the recording units are controlled by a laptop GUI (Suzuki et al., 2017).

While short-term experiments have proven themselves as interesting proofs-of-concept within the field of ecoacoustic localisation, extended recording periods are necessary to understand an ecosystem more comprehensively. In practice, this is difficult, with devices still requiring attention for data retrieval when SD cards fill up (Wildlife Acoustics Song Meters, e.g. - (Mennill et al., 2012)) when batteries run out (VoxNet), or both (CARACAL, Dev-Audio/TOMAGO). More recently, Crunchant et al., 2022 presented a solar-powered, GPS-synchronised device that they demonstrate using for wildlife sound localisation – the device requires manual data collection after SD cards become full. A recent study resolves this by using a completely Wireless Acoustic Sensor Network, which autonomously uploads data to a server through a gateway node (Bruggemann et al., 2021). Bruggemann et al., 2021 boast integrated real-time species ID and have expressed motivation to explore localisation in the future, but this has not yet been explored. This device also requires manual battery replacement.

Adding a spatial dimension to soundscape recording opens up a whole new dimension of available data and analyses that may prove incredibly fruitful in future studies. Specifically, rather than just knowing what and when things are happening in an ecosystem, we can instead ask what, when, and where.

1.7. CONCLUSION

In this overview of state-of-the-art sound ecology, I hope to have restated the importance of ecosystems by evidencing their necessity for the continuation of humanity while advocating for their continued monitoring. I briefly explain the most up-to-date methods in ecosystem monitoring, specifically through camera traps, canopy imaging, eDNA, and acoustics. After exploring the physical properties of soundwaves, I describe how the field of sound ecology has evolved from defining semantics and theory (including the ecological and evolutionary drivers of vocal communication) to large-scale automated monitoring. I introduce some of the most state-of-the-art methods in soundscape recording and analysis and speculate on how the field may develop in future. Specifically describing future and near future advancements in hardware development (specifically spatial acoustics), machine learning, cross-study standardisation, and automated action.

I have identified three key literature gaps within this literature review that this thesis is designed around answering. The first investigates an increasingly prominent challenge in long-term ecoacoustic monitoring: data storage. Acoustic datasets can become massive, and data reduction strategies are often used without known consequences. In this thesis, I identify some of the most common data management strategies and explore how and why these practices may affect how soundscapes are quantified. Secondly, I have introduced the fascinating new field of spatial acoustics, highlighting some of the vital applications of this tech whilst examining why uptake has been limited. In this survey, I found that cost, difficulty obtaining and using equipment and lack of autonomy were key drawbacks. The second aim of my thesis was to design a recorder that addressed this gap, with applications to increase the uptake of spatial acoustics and allow a broader range of users to utilise spatial acoustics. The third and final gap I uncovered was twofold: the scarceness of single-device spatial acoustics, and the lack of vertical axis spatial acoustics. Specifically, most spatial localisation requires three or more separate recorders, which can add up, even when the base recorders are low cost. It is underexplored whether useful ecological features can be derived from single-device spatial acoustics, so the third aim was to investigate what was achievable with just one of the devices I developed. Further, I noticed that spatial acoustics is almost exclusively considered on the horizontal axis, so I designed this study around exploring the vertical axis.

While this has been a largely high-level overview of the broad range of topics necessary for the development and usage of wildlife tech, I hope this has well contextualised the importance of the research area, the future landscape of the field, and ultimately the following chapters.

1.8. RESEARCH OBJECTIVES

This thesis aims to assess and develop methods in (spatial) sound ecology. I achieve that by exploring three research questions over the following three chapters.

(1) Investigate Whether Data Saving Practice in Ecoacoustics Affects the Quantification of Soundscapes

(2) Develop an Autonomous Multi-Channel Acoustic Recorder Capable of Sound-Source Localisation

(3) Explore Whether a Single Recorder Capable of DOA Estimation Can Be Used To Approximate Abundance and Monitor Behaviour

1.9. CHAPTER OVERVIEW AND OUTCOMES

(1) Investigate Whether Data Saving Practice in Ecoacoustics Affects the Quantification of Soundscapes

How Index Selection, Compression and Recording Schedule Impact the description of Ecological Soundscapes (Chapter 2)

I begin chapter two by explaining common variations in data collection and analysis in sound ecology, focusing on whole soundscape monitoring. Specifically, I introduce experimental protocols commonly used to reduce data storage and transmission costs (temporal subsetting and compression). In the chapter, I highlight how it is currently unknown whether compression (which works through the irreversible removal of data) impacts soundscape quantification. Further, I explain how despite the increase in studies calling for standardisation, there is still much variation in ecosystem sampling efforts.

I designed a study to investigate the impact of both compression and temporal subsetting on soundscape quantification. I highlighted analytical indices as the most used method of soundscape quantification and CNN-derived soundscape "fingerprinting" as a possible future solution. Both measures are investigated in this study. I used three days' worth of high-quality audio recorded across a gradient of land use. I collected the data for this study in Malaysian Borneo at the Sustainability of Altered Forest Ecosystems (SAFE) project. All acoustic data were collected with AudioMoth recorders.

The different experimental protocols were "simulated" post-fieldwork by creating augmentations of the collected audio with different recording lengths or were compressed to different rates. Acoustic Indices and the CNN-derived fingerprint were extracted from all raw and augmented acoustic datasets. I quantified the impact of compression by calculating the difference in the descriptive measures between each compressed recording and its raw (uncompressed) counterpart. Whether compression and the other experimental protocols affected soundscape quantification was assessed by building and testing the accuracy of a random forest classification model for all combinations of experimental protocols.

Overall, I found that both compression and recording length drove considerable variation in index values but that this had a more negligible effect on the model performance than whether Analytical Indices or the CNN-derived fingerprint was used. I found that the CNN-derived fingerprint was more robust to differences and worked better as a soundscape descriptor. I provide further evidence for the usability of CNN-derived soundscape fingerprinting but also provide evidence for its robustness which was previously only speculated. Finally, I use our findings to provide guidelines for future work, especially work which is restrained by data storage.

This finding is valuable as it means researchers on data volume restrictions from server costs or limited SD cards can confidently choose to sample data from a larger range by minimising the per/time data storage requirements with compression. At the time, the use of CNN-derived soundscape fingerprinting was not especially widespread and adding to the now much larger body of research advocating for its usage (especially on low-quality data) provides evidence of use cases in a broader context.

<u>Contributions:</u>

(1) An open-access article in *Ecology and Evolution* presents our research and provides recommendations for users limited by data constraints (Heath et al., 2021).
(2) Conference proceedings and presentation at Forum Acousticum 2020 (Heath et al., 2020)
(3) Talk at International Bioacoustics Congress (Heath, 2019)
(4) Two open datasets: raw audio (Heath B. E., Sethi, Orme, Ewers, & Picinali, 2021), and soundscape descriptors (Heath B. E., Sethi, Orme, Ewers, & Picinali, 2021).

(5) An open Github repository containing data analysis pipelines and data visualisation.

(2) Develop an Autonomous Multi-Channel Acoustic Recorder Capable of Sound-Source Localisation

Introducing MAARU – Multichannel Autonomous Recording Unit for Spatial Ecosystem Monitoring (Chapter 3)

I shifted focus in the third and fourth chapters to spatial acoustics, which is increasingly garnering attention in the field. Reviewing the literature revealed many use cases of spatial acoustics but highlighted that almost every study used recorders only suitable for short-term deployments that required highly specialised/ expensive equipment and/or extensive expertise. My first aim was to create a multichannel acoustic recording platform that was autonomous and suited to long-term field deployment, built from relatively low-cost commercially available equipment.

In my third chapter, I describe my process of developing MAARU, my low-cost, multichannel, autonomous recording unit. After considerable experimenting, I found that the best balance between ease of use, financial cost, and equipment availability came from the Seeed Studio ReSpeaker 6-Microphone Circular Array. I incorporated the sound card and different processing parameters into an established autonomised soundscape recording and data transfer platform. These new scripts and the necessary instructions are detailed in this chapter and on a GitHub repository (github.com/BeckyHeath/multi-mic-recorder-analysis). Weatherproofing and tethering used for this device required special consideration as for accurate sound-source localisation, the path between a signal and a receiver should be largely uninterrupted, voiding most usual methods of microphone cover. I developed the software, hardware and weatherproofing together and called this recorder MAARU (Multichannel Acoustic Autonomous Recording Unit). Getting a recorder that could record, process, and upload six audio channels simultaneously was only half of the challenge. The next was to determine whether this setup could (1) be used for sound-source localisation and (2) how well this setup and the localisation could persist in the field.

I explored several routes when determining how best to localise sound sources with my new device, landing on HARKBird. I worked with one of the developers of HARKBird (Reiji Suzuki, Nagoya University Japan) to create an install of HARKBird that was compatible with MAARU. I then tested how accurately audio recorded by MAARU could be localised with HARKBird. I tested the MAARU device's resistance to waterproofing and a six-month deployment in the field. In all cases, sound sources could be localised to within ±10˚ of their actual direction. Two of the four field devices suffered some terminal damage in the field because of animal damage and powering issues; However, from these results, I have been able to provide mitigating advice. Overall, I have successfully created a non-expensive, completely open, and accessible platform capable of accurate audible-range, sound-source localisation in field environments, the likes of which had not previously been introduced in this field.

The core work from this chapter is currently under review, but I hope this work will provide an accessible platform for others hoping to do spatial acoustics surveys without extensive funding or expertise.

Contributions:

- Article describing MAARU *in review.*
- Research Presented at an international conference (Heath B. E., (Spatial) Ecoacoustic Monitoring, 2022)
- I presented this work at two UK Acoustics Network workshops as a panellist (Heath B. E., Hardware Panel , 2022), and in two mini-talks (Heath B. E., Sound Recording and Sound Analysis, 2022)

- Two open GitHub repositories. The first contains all the scripts that record, process, and transfer data through the Raspberry Pi (MAARU) recorder. The second contains all the scripts used in accuracy processing and data visualisation in the chapter.

(3) Explore Whether a Single Recorder Capable of DOA Estimation Can Be Used to Approximate Abundance and Monitor Behaviour

<u>MAARU (Multichannel Acoustic Autonomous Recording Unit) in the Wild – Estimating Abundance and Exploring Vertical Stratification with Multichannel Acoustics (Chapter 4).</u>

The literature has begun to explore the applications and success rates of using multiple nearby recorders to determine the locations of calling individuals. I instead opted to explore what a single multichannel recording unit could do and whether it could add anything to ecosystem recording beyond that which was possible with a single channel recorder, even without determining the exact calling locations of individuals.

Therefore, the final aim of this thesis was to explore whether (and how) the recorder I developed in the previous chapter (MAARU) could be used to enhance ecosystem recording alone and what (if anything) we could learn from investigating the vertical axis. I determined two analysis routes here: firstly, to spatially cluster the direction of signals as enhanced abundance approximation and by using these directions as a coarse behavioural predictor for detected individuals. Over the chapter, I provide evidence for both use cases, showing reduced abundance approximations using DOA-approximated abundance and evidencing that different bird species call from different directions in the field test. This analysis relies on data from the MAARU recorder and an accurate classification algorithm. MAARU records at 16kHz, limiting the detected species to strictly below 8kHz. Excluding everything beyond that wipes many groups of species from the analysis and will reduce how generalisable these results are; however, this simultaneously provides an exciting standpoint for future development.

These kinds of analyses open a whole world of new questions that can be explored in future: whether morphology or diet influences vertical space usage, whether there is any link between functional traits and vertical height usage, do different forest management processes such as selective felling, agroforestry or rewilding influence the vertical space usage of species and could this perhaps be used to explain increases or relief of competition.

<u>Contributions:</u>

- Article describing the analysis pipeline *in progress.*
- An open Github repository containing all the scripts necessary to join raw HARKBird and BirdNET outputs, perform directional clustering, and some of the early visualisations of species- and location-specific DOAs.

2. CHAPTER 2: THE IMPACT OF DATA SAVING PRACTICE IN ECOACOUSTICS

2.2. CHAPTER OVERVIEW

Acoustic indices derived from environmental soundscape recordings are used to monitor ecosystem health and vocal animal biodiversity. Soundscape data can quickly become very expensive and difficult to manage, so data compression or temporal down-sampling are sometimes employed to reduce data storage and transmission costs. These parameters vary widely between experiments, with the consequences of this variation remaining mostly unknown.

I analyse field recordings from North-Eastern Borneo across a gradient of historical land use. I quantify the impact of experimental parameters (mp3 compression, recording length and temporal subsetting) on soundscape descriptors (Analytical Indices and a convolutional neural net-derived AudioSet Fingerprint). Both descriptor types were tested for their robustness to parameter alteration and their usability in a soundscape classification task.

I find that compression and recording length both drive considerable variation in calculated index values. However, I find that the effects of this variation and temporal subsetting on the performance of classification models is minor: performance is much more strongly determined by acoustic index choice, with AudioSet fingerprinting offering substantially greater (12%–16%) levels of classifier accuracy, precision, and recall.

I advise using the AudioSet Fingerprint in soundscape analysis, finding superior and consistent performance even on small pools of data. For audible range studies (excluding ultrasound) where data storage is a bottleneck, I recommend Variable Bit Rate encoded compression (quality = 0) to reduce file size to 23% file size without affecting most Analytical Index values.

The AudioSet Fingerprint can be compressed further to a Constant Bit Rate encoding of 64 kb/s (8% file size) without any detectable effect. These recommendations allow the efficient use of restricted data storage whilst permitting comparability of results between different studies. These results are, however, only relevant for soundscape level audible range monitoring, and further investigation should be taken if the study looks to investigate particular target species or include species that call outside of the audible range.

2.3. INTRODUCTION

Animal vocalisations come together with abiotic and human-made sounds to form soundscapes. These soundscapes can be recorded and quantified across large temporal and spatial dimensions to monitor species populations or infer community-level metrics such as biodiversity (Eldridge et al., 2018; Gómez et al., 2018; Roca & Proulx, 2016). Monitoring is crucial to effectively respond to threats such as disease, species loss and overlogging (Rapport, 1989; Rapport et al., 1998). Previously, the use of *in situ* expert listeners to monitor species presence and abundance was common (Huff et al., 2000), but: is costly and time-consuming; can damage habitats; and is prone to narrow focus and observer bias (Costello et al., 2016; Fitzpatrick et al., 2009). Advances in portable computing now permit remote recording of soundscapes but produce a volume of data that is very time-consuming to review manually, leading to the development of automated or semi-automated methods of analysis (Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020; M. W. Towsey et al., 2016).

Soundscape composition is primarily assessed using acoustic indices, which describe the soundscape in an abstract form. Analytical Indices are a type of acoustic index which are summary statistics that describe the distribution of acoustic energy within the recording (M. Towsey et al., 2014) –over 60 of which have been designed to capture aspects of biodiversity (Buxton, McKenna, et al., 2018; Jérôme Sueur et al., 2014). These are commonly used in combination to compare the occupancy of acoustic niches, temporal variation, and the general level of acoustic activity (Bradfer-Lawrence et al., 2019) across ecological gradients or in classification tasks (Gómez et al., 2018). These approaches have provided novel insight into ecosystems across the world (Fuller *et al.*, 2015; Buxton *et al.*, 2016; Eldridge *et al.*, 2018; Sueur, Krause and Farina, 2019) but are not without limitations and often have poor transferability (Bohnenstiehl et al., 2018; Mammides et al., 2017). This may result from a lack of standardisation: differing index selection, data storage methods, and recording protocols, which all lead to unassessed variation in experimental outputs (Araya-Salas et al., 2019; Bradfer-Lawrence et al., 2019; Sugai et al., 2019).

The output vector from the AudioSet convolutional neural net (CNN; Gemmeke *et al.*, 2017; Hershey *et al.*, 2017) is an attractive replacement for Analytical Indices. This pre-trained, general-purpose audio classification algorithm generates a multi-dimensional acoustic fingerprint (AudioSet Fingerprint) of a soundscape which can be used as a more effective suite of acoustic indices (Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020). The AudioSet CNN is trained on two million human-labelled anthropogenic and environmental audio samples, potentially giving it both greater transferability and discrimination than typical ecoacoustic training datasets. Unlike Analytical Indices, however, extra analysis (such as training classifiers/ predictive models) is necessary to relate the AudioSet Fingerprint to ecological processes and states.

In ecoacoustics, a continuous uncompressed or lossless recording is generally recommended (Browning et al., 2017; Villanueva-Rivera et al., 2011), but generates huge files. In this chapter, I considered two commonly used approaches to reducing storage requirements (M. Towsey, 2018). Firstly, mp3 compression, which is widely used in ecoacoustic studies (e.g. Saito *et al.*, 2015; Zhang *et al.*, 2016; Sethi *et al.*, 2018): this lossy encoding removes acoustic information inaudible to *human* listeners (Sterne, 2012). Mp3 compression will therefore remove all ultrasonic data akin to the calling frequencies of an increasingly diverse range of species of bats, small mammals and some birds, but is suspected of removing ecologically important data in the audible range too (e.g. Towsey, Truskinger and Roe, 2016; Sugai *et al.*, 2019). For this reason, Araya-Salas, Smith-Vidaurre and Webster (2019) have recently shown that audible range ecological information is lost under high compression from recordings of isolated animal calls. However, it is not known if this extends to recordings of noisier whole soundscapes.

Secondly, recording schedules also vary in ecoacoustic studies (Sugai et al., 2019). Bradfer-Lawrence *et al.* (2019) showed that longer and more continuous schedules give more stable Analytical Index values. However, ecoacoustic composition varies with the time of day (Bradfer-Lawrence et al., 2019; Fuller et al., 2015; Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020) and so reducing recording periods with temporal subsetting may reduce temporal variation and improve classification (Sugai et al., 2019) even with reduced data. Similarly, index calculation on longer recordings may average away anomalous calls and short-term patterns.

While clear standards are crucial for collaborative research in ecoacoustics, there is uncertainty in the literature on the impacts of the selection of index type, compression level and recording schedule on the quantification and classification of audible range ecological soundscapes. Here, we:

1) Investigated the impact of index selection on the accuracy of a random forest classifier.
2) Described the effects of compression, recording length and temporal subsetting on the values, variance, and classification performance of indices.

In describing how well ecological information is stored in acoustic data under different recording decisions, I identified stronger standards to improve classifier accuracy, precision and recall and provided a basis for comparison among studies.

2.4. METHODS

2.4.3. Study Area

Acoustic samples were collected in Sabah, Malaysian Borneo, at the Stability of Altered Forest Ecosystems (SAFE) project: a large-scale ecological experiment on habitat loss and fragmentation effects on tropical forests (Ewers et al., 2011), which included sites in the Kalabakan Forest Reserve (KFR). Historically, logging within KFR has been heterogeneous, reflecting habitat modifications in the wider area (Struebig et al., 2013), with higher than typical timber extraction rates. This is a diverse forest type from which we have recorded at least 175 species of bird and at least 50 species of amphibian from 26 sites (Sethi, Ewers, Jones, Picinali et al., 2020). Habitat ranges from areas of grass and low shrub to logged forest to almost undisturbed primary forest.

2.4.4. Soundscape Recording

Data were collected from three KFR sites representing a gradient in above-ground biomass (Figure 2.1a) (AGB: Pfeifer *et al.*, 2016): primary forest (AGB= 66.16 $t.ha^{-1}$), logged forest (AGB = 30.74 $t.ha^{-1}$), and cleared forest (AGB = 17.37 $t.ha^{-1}$) (Appendix A: Table 1). I recorded continuously from a single recorder for a mean of 72 hours at each site (range: 70 to 75) during February and March 2019 (Appendix A: Table 2.1). No rain fell during the recording period, so no recordings were excluded due to confounding geophony (Zhang et al., 2016). In all three sites, I placed individual omnidirectional (AudioMoth, Hill *et al.*, 2018) recorders, which were attached to trees (~ 50 cm diameter and 1-2 m above the ground) and recorded 20-minute samples with no break period and stored them as uncompressed files (‘raw’, .wav format) at 44.1kHz and 16 bits.

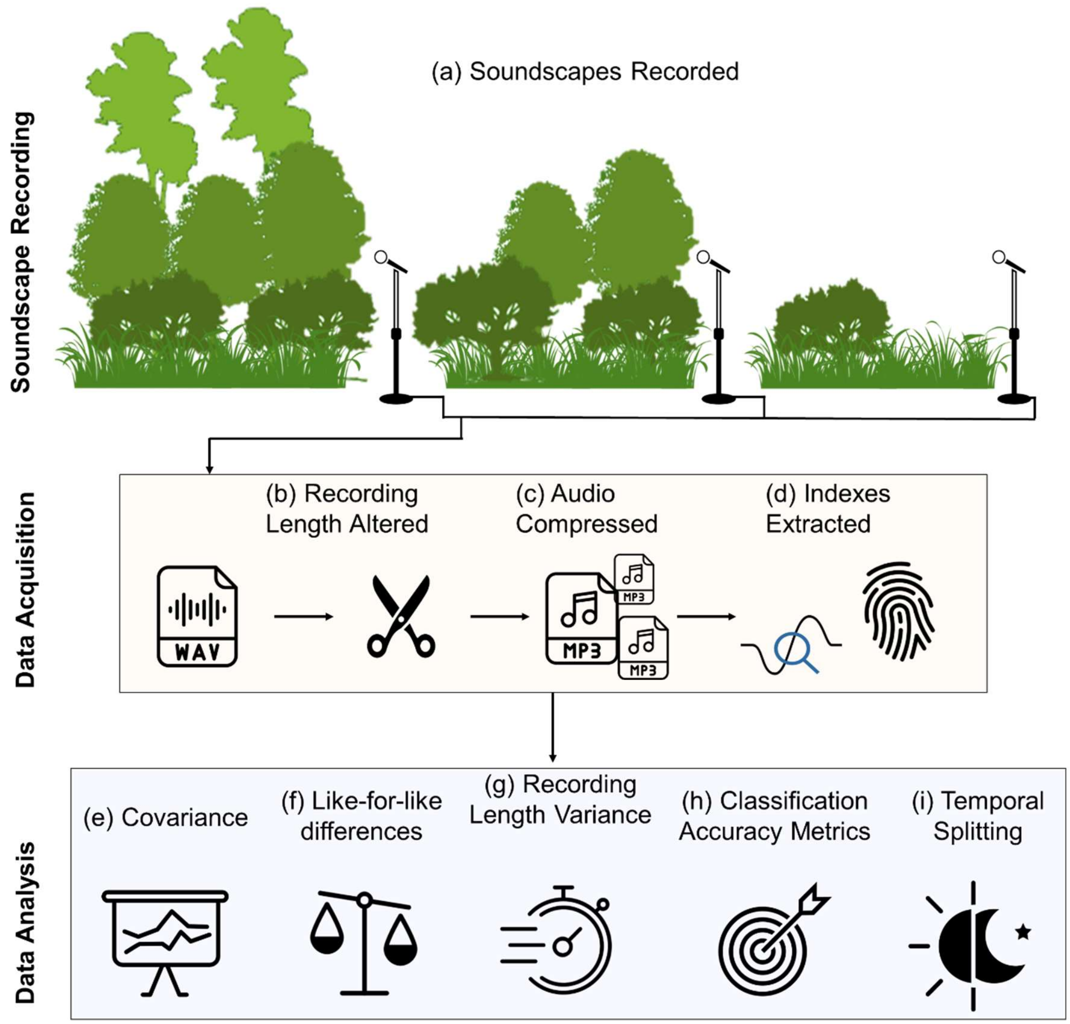


Figure 2.1: Experimental structure. Soundscape Recording: (a) Soundscapes from different forest structures in Malaysian Borneo are recorded. Data Acquisition: (b) Recording length is altered to 20-, 10-, 5-, and 2.5-min chunks; (c) all audio is compressed using nine lossy nine mp3 encoding techniques; (d) Analytical Indices and CNN Derived AudioSet Fingerprint are calculated from audio of all lengths and compressions. Data Analysis: (e) Index covariance is found per index type, and correlation with maximum frequency is found; (f) like-for-like differences of indices calculated from compressed versus uncompressed counterparts are found; (g) intragroup variance compared for the recording lengths; (h) the indices of both types, lengths, and compressions are tested with a supervised random forest classification task; (i) the dataset is split into temporal sections, and classification accuracy is found

2.4.5. Compressing and Re-sizing The Audio

Continuous 20-minute recordings were first split into recordings with a length of 2.5, 5.0 and 10.0 minutes, using the python package *pydub* (Webbie et al., 2018) (Figure 2.1b), resulting in 8, 4 and 2 times as many recordings, respectively. The audio was then converted to lossy mp3 format using the fre:ac LAME encoder (Kausch, 2019) under two standard LAME mp3 encoding techniques: constant bit rate (CBR) and variable bit rate (VBR) compression (Figure 2.1c). CBR reduces the file size to a specified number of kilobits per second; VBR varies bitrate per second depending on the analysis of the acoustic content and a quality setting (0, highest quality, larger bitrate; 9, lowest quality, smaller bitrate). Since bitrates are not directly comparable between VBR and CBR – and because storage savings are often the principal driver of compression choices – I used compressed file size as my measure of compression level. I used VBR0 and CBR320, CBR256, CBR128, CBR64, CBR32, CBR16 and CBR8, which resulted in file sizes ranging between 41.6% (CBR320) and 1.04% (CBR8) of the original raw file size and some reductions in Nyquist frequency (Table 2.1). I have not considered lossless compression here, as the storage capacity is much higher, and the files are obligatorily the same post-decompression. Previous studies have also found that lossless compressed audio is largely identical to raw audio (Linke & Deretic, 2020).

Table 2.1: Bitrate, percentage file size reduction and maximum encodable frequency for the experimental compression levels.

Compression Level	Bit storage/s	% File Size	Nyquist Frequency (kHz)
RAW	Constant: 768kb	100	22.05
VBR0	Variable: ~ 127 – 250kb	mean = 20.82 range = 32.64 – 16.63	22.05
CBR320	Constant: 320kb	41.6	22.05
CBR256	Constant: 256kb	33.35	22.05
CBR128	Constant: 128kb	16.67	22.05
CBR64	Constant: 64kb	8.33	22.05
CBR32	Constant: 32kb	4.16	11.025
CBR16	Constant: 16kb	2.08	8
CBR8	Constant: 8kb	1.04	4

2.4.6. Quantification of Soundscape Using Indices

2.4.6.1. Analytical Indices

I used the *seewave (version 2.1.6)* (Sueur, Aubin and Simonis, 2008) and *soundecology (version 1.3.3)* (Villanueva-Rivera and Pijanowski, 2016) packages in R (version 3.6.1; R Core Team, 2020) to extract 7 Analytical Indices (Figure 2.4d): Acoustic Complexity Index (ACI, calculated per minute and averaged), Acoustic Diversity Index (ADI), Acoustic Evenness (AEve), Bioacoustic Index (Bio), Acoustic Entropy (H), Median of the Amplitude Envelope (M), and Normalised Difference Soundscape Index (NDSI) (Appendix A: Table 3). These have been shown to capture diel phases, seasonality, and habitat type (Bradfer-Lawrence et al., 2019). These indices could not be calculated for all recordings due to file reading errors; however, this fault occurred in 0.3% of all recordings (Appendix A: Table 2.2).

2.4.6.2. AudioSet Fingerprint

The audio was converted to a log-scaled Mel-frequency spectrogram after 16kHz downsampling and then passed through the “VGG-ish” Convolutional Neural Network (CNN) trained on the AudioSet database (Gemmeke et al., 2017; Hershey et al., 2017) (Figure 2.1d). This generated a 128-dimensional embedding, and the 128 values in that embedding described the soundscape of a given recording in an abstracted form or fingerprint. Similarly, as in the Analytical Indices, some recordings could not be analysed by the AudioSet CNN; however, this was only in 0.2% of recordings (Appendix A: Table 2.2).

2.4.7. Data analysis

2.4.7.1. Impact of Index Selection: Auto-Correlation

Analytical Indices often summarise similar features of a soundscape (e.g. dominant frequency and frequency bin occupancy): this overlap may reduce the descriptive scope of the ensemble. I compared the degree of pairwise correlation between the individual Analytical Indices and between the individual values of the AudioSet Fingerprint. I also compared how well each index/feature correlated with the Nyquist frequency (Figure 2.1e).

2.4.7.2. Impact of Compression: Like-for-Like Differences

I used an adaption of Bland-Altman plots (Vesna, 2009, Araya-Salas, Smith-Vidaurre and Webster, 2019) to visualise the scaled difference (*D*) between raw (I_{raw}) and compressed (I_{com}) index values, as a percentage of the range of raw values R_{raw} (Figure 2.1f) :

$$D = \frac{I_{com} - I_{raw}}{R_{raw}} \times 100$$

D was not normally distributed (Appendix A.5.1.x), so median and interquartile ranges were reported. I determined that an index has been altered as a result of compression to be when: i) the interquartile range of *D* did not include zero difference or ii) median *D* was more than +/- 5% of the R_{aw}. I used Spearman rank correlation to test for a consistent trend in *D* with increasing compression. Reflecting their common use cases, *D* for Analytical Indices was calculated from the univariate values, while for AudioSet Fingerprints – which is intended as a multidimensional metric – D was calculated separately for each dimension and then given as a mean of all 128 values.

2.4.7.3. Impact of Recording Schedule: Recording Length

Recordings of longer length may have a reduced variance due to the smoothing of potentially important transient audio anomalies (such as nearby bird or cicada calls). I tested this by comparing the variance of the recording groups at different commonly used recording lengths. The index values are non-normally distributed, so I used Levene's test for homogeneity of variance (Figure 2.1g).

2.4.7.4. Impact of Parameter Alteration on Classification Task

I used random forest classification models to assess how well the soundscapes were represented by each index type under each different experimental parameter, using the *RandomForest (version 4.6-14)* (Liaw and Wiener 2002) package in R (Figure 2.1h). Models were trained on a 24 h period of data from each site and tested on the remaining 46+ h of audio. I used 2,000 decision trees to ensure accuracy had stabilised. The model was trained and tested separately for every combination of index type (Analytical Indices vs AudioSet Fingerprint), compression level and recording length. I determined the accuracy, precision and recall of each combination.

2.4.7.5. Impact of Temporal Subsetting

Soundscapes typically show considerable diel variation in both abiotic and biotic components. To assess the impact of this variance on model performance, I split my recordings into four 6-hour sections centred on the key periods of Dawn (06:00), Noon (12:00), Dusk (18:00) and Midnight (00:00) and then further subdivided these into 3 hour (8 sections) and 2 hour (12 sections) blocks to test how further reductions affected the model (Figure 2.1i). I trained and tested the random forest model again on each of the temporally subset recordings, with each section used to build models individually, and determined accuracy, precision and recall as before.

2.4.7.6. Modelling the Impact of Index Selection, Compression and Recording Length on the Accuracy Metrics

As the accuracy metrics are bound between 0 and 100%, I used a beta regression to model the relationship between each of the experimental parameters and performance metrics (Douma & Weedon, 2019). The model was built using the *betareg (version 3.1-3)* package in R (Cribari-Neto & Zeileis, 2010). To avoid fitting issues when performance measures are exactly 1, I rescaled all performance measures using m' = (m (n-1) + 0.5) / n, where n is the sample size (Smithson & Verkuilen, 2006). The model included pairwise interactions between file size, temporal subsetting, and recording length, and then all interactions of main effects and those pairwise terms with the index selection. I observed that variance in performance measures varied as an interaction of both index choice and a temporal subsetting (Appendix A.8a), so tested the inclusion of these terms in the precision component of the model. I first treated recording length and temporal subsetting as factors, but also tested a model considering these as continuous variables. I found the Akaike Information Criterion (AIC) was markedly lower in a beta regression model using factors and including the precision component (Appendix A: Figure 9.1).

2.5. RESULTS

Although Spearman pairwise correlations of Analytical Indices and Nyquist frequency were low on average (mean = 0.32, IQR = 0.22), I found some strongly correlated sets of indices (Figure 2.2). ADI, Bio and NDSI all showed strong similarities and were closely correlated with maximum recordable frequency; AEve and H were also strongly correlated (Figure 2.2). Some features of the AudioSet Fingerprint correlated with each other and maximum frequency, but in general, these features were more weakly correlated (mean = 0.14, IQR = 0.18, Figure in Appendix A: Figure 4.2).

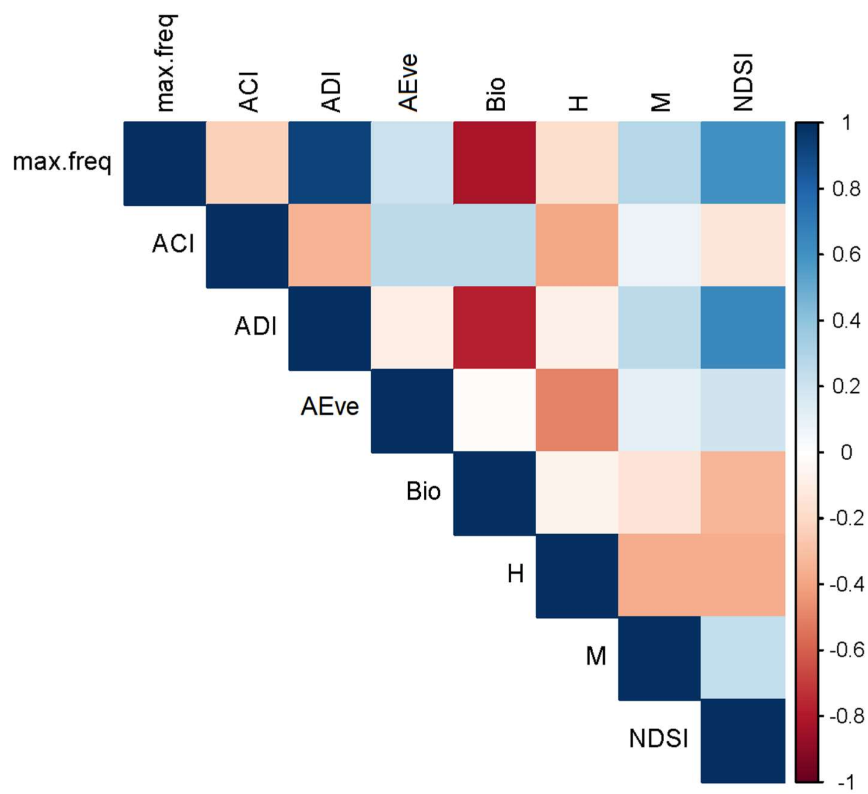


Figure 2.2: Pairwise Spearman correlation matrix for Analytical Indices (all recording lengths and all compressions) and maximum recordable frequency. The colour scale shows rho values.

2.5.3. Impact of Compression

2.5.3.1. Impact of Compression: Like-for-Like Differences

Both index types showed both differences under compression and clear trends with increasing compression (Figure 2.3) (confirmed with Spearman's rank correlation, all $p < 0.001$, Appendix A: Table 5.3). The mode of response showed three broad qualitative patterns, illustrated here using results from the 5-minute audio sample (other recording lengths in Appendix A: Figure 5.2.x). (1) Indices which were only affected above a threshold level of compression (AudioSet Fingerprint: CBR16; M: CBR32; and NDSI: CBR8). These indices typically showed low absolute *D* (median *D* typically <15%). (2) AEve and H showed the largest differences at an intermediate compression (CBR64) and relatively low absolute differences (median *D* typically < 30%). (3) The remaining indices showed a variety of responses: ADI showed a monotonic response above a threshold, ACI showed changes up to CBR64 and then stabilises, and Bio showed a stepped pattern of increase. However, all three showed increasing and large changes in absolute D (median *D* often > 75%) with increasing compression.

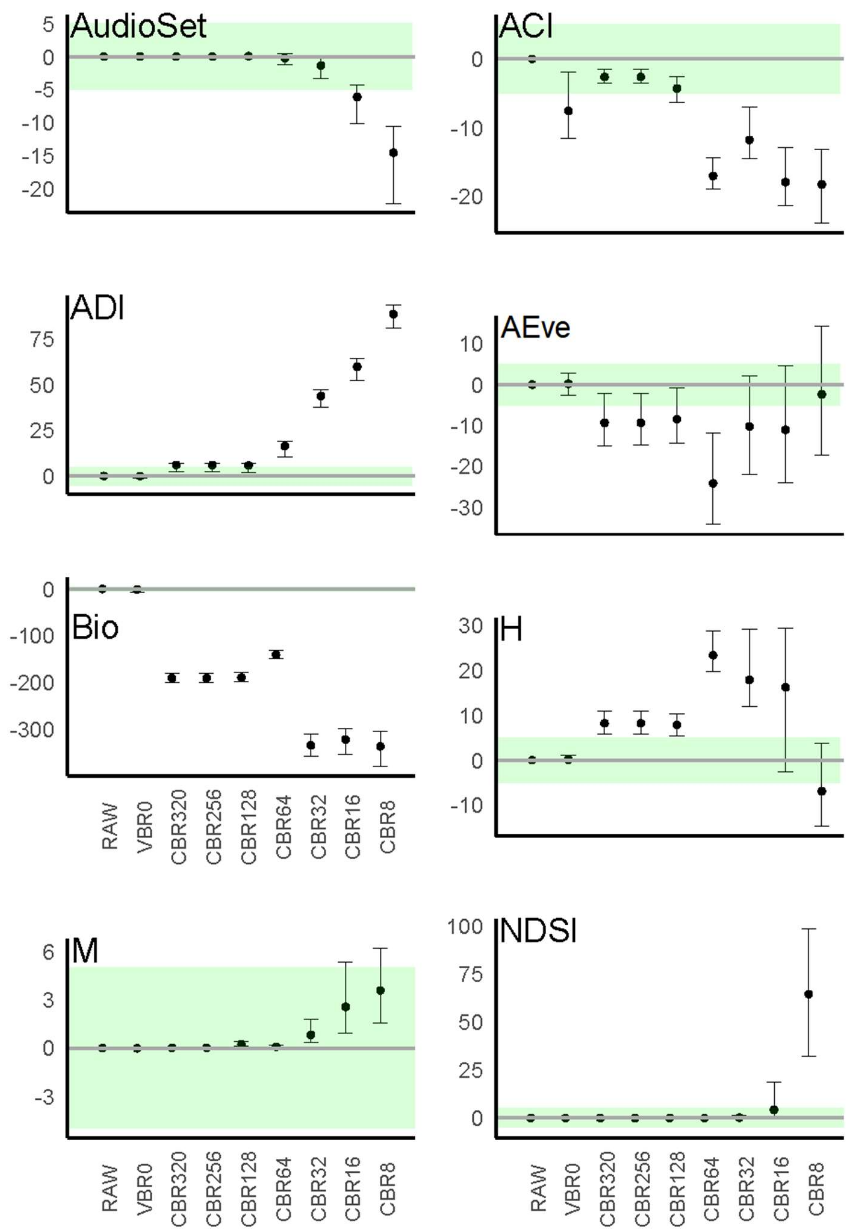


Figure 2.3: Scaled difference in acoustic indices from raw audio with increasing compression in 5-min audio samples (see Appendix A. Figure 5.2.x for 2.5- and 10- and 20-min examples). The horizontal green region shows the ±5% D. Dots and whiskers show the median and interquartile range of D from different indices under increasing levels of compression.

2.5.3.2. Impact of Recording Schedule: Recording Length

Three out of seven (43%) of the Analytical Indices (ADI, AEve and H) and a smaller proportion of the AudioSet Fingerprint values (46 out of 128; 36%) were found to have non-homogeneous variance in groups of different recording length ($p < 0.05$, Levene's test for homogeneity of variance, Appendix A: Table 6).

2.5.4. Impact of Index Selection

Confirming prior findings (Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020), I showed that habitat classifiers derived from 5-minute recordings using raw audio showed higher accuracy for AudioSet Fingerprint (93.8%) than Analytical Indices (80.9%, Table 2.2). This advantage was held across all recording lengths and performance metrics with performance gains of around 12-13% in accuracy, precision and recall (Appendix A: Table 7.1.2).

Compression decreased accuracy for both AudioSet Fingerprint (CBR8: 90.8%) and Analytical Indices (CBR8: 75.1%, Table 2.2). Classifiers trained on *compressed* AudioSet Fingerprint, however, still outperformed those trained on *uncompressed* Analytical Indices. For both index types, this reflected a decreased ability to differentiate between logged and primary forest. Interestingly, classifiers from both index types showed better discrimination between cleared land and logged forest under strong compression. These patterns were repeated across recording lengths (Appendix A: Table 7.1.1).

Table 2.2. Confusion matrices from random forest classifiers trained on AudioSet Fingerprint (a, c) and Analytical Indices (b, d) using uncompressed raw audio (a, b) and highly compressed CBR8 audio (c, d).

	AudioSet Fingerprint				Analytical Indices		
Observed		Predicted		Observed		Predicted	
a) Raw	Cleared	Logged	Primary	**b) Raw**	Cleared	Logged	Primary
Cleared	585	9	11	Cleared	484	67	49
Logged	11	508	44	Logged	97	421	46
Primary	17	14	521	Primary	9	61	486
c) CBR8	Cleared	Logged	Primary	**d) CBR8**	Cleared	Logged	Primary
Cleared	585	3	17	Cleared	484	23	98
Logged	2	488	73	Logged	9	379	175
Primary	11	53	488	Primary	9	115	428

2.5.4.1. Impact of Temporal Subsetting

Temporally subsetting poses a trade-off as when diel variation is reduced, so too are the recording hours available for analysis. Temporally subsetting the day into quarters (Figure 2.4) yielded a largely unpredictable effect on accuracy, precision and recall. There were clear differences in discrimination between pairs of sites. Notably, comparing cleared and primary forest had the highest precision across each temporal subset, index choice and compression (Figure 2.4 e,f), but the recall was not markedly different from other pairs (Figure 2.4 k,l). Temporal windows did not generally help discriminate between logged and primary forest (Table 2.2, Figure 2.4 g,h,m,n) and the performance difference between AudioSet Fingerprints and Analytical Indices was largely maintained.

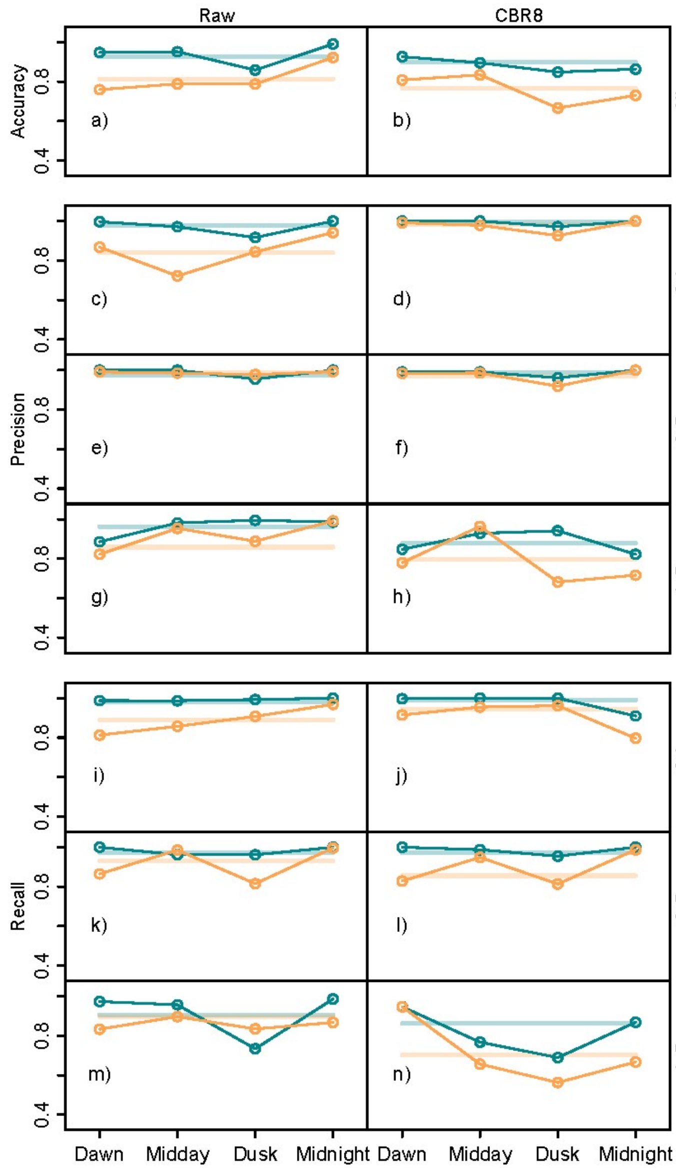


Figure 2.4: Classification model performance as a function of temporal sectioning (x-axis), compression (raw audio, left column; CBR8 compression, right column) and index choice (AudioSet Fingerprint: blue; Analytical Indices: orange). Pale horizontal lines show performance without temporal sectioning. Precision and recall are partitioned into pairwise performance by site (C, cleared forest; L, logged forest; P, primary forest

2.5.4.2. Combined Effects of Parameter Alteration on Classification Performance

Confirming prior findings (Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020), my model has demonstrated that performance measures were consistently higher when classifiers are trained on the AudioSet Fingerprint, rather than Analytical Indices (Accuracy: +16.9%

($z=10.38_{1799}$ $p<0.001$), Precision: +15.5% ($z = 9.717_{1799}$ $p<0.001$), Recall: +16.9% ($z=10.22_{1799}$ $p<0.001$), full model outputs Appendix A: section 9). Index type was by far the largest contributor to model accuracy (Table 2.3), although there was some effect of temporal subsetting, compression level and recording length. Despite the considerable impact of compression level on index values, it appeared to have a minor effect on model accuracy (Figure 2.5, Table 2.3). The effect of recording length appeared to increase as the days were cut into smaller temporal subsections. However, this effect was small compared to the contribution of index type (Figure 2.5). Temporal subsetting appeared to have minimal effect on the accuracy of the AudioSet Fingerprint classifier, which kept consistently high (70-100%, Figure 2.5). The classifier trained on Analytical Indices, however, became much more unpredictable when temporal subsetting was used (20-100%, Figure 2.5)

*Table 2.3. ANOVA table for the model terms in the beta regression model of the accuracy data. (Significance: *** p <0.001, ** p<0.01, * p<0.05). Equivalent tables for precision and recall are in Appendix A: section 9).*

	Df	**χ^2**
log10(File Size)	1	26.2128 ***
Temporal Subsetting	3	31.6818 ***
Frame Size	3	15.7820 **
Index Type	1	2985.9825 ***
log10(File Size): Temporal Subsetting	3	18.0278 ***
log10(File Size): Recording Length	3	2.9280
Temporal Subsetting: Recording Length	9	6.3156
log10(File Size): Index Type	1	59.0065 ***
Temporal Subsetting: Index Type	3	7.1061
File Size: Index Type	3	36.2699 ***
log10(File Size): Temporal Subsetting: Index Type	3	13.0715 **
log10(File Size): Recording Length: Index Type	3	0.8071
Temporal Subsetting: Recording Length: Index Type	9	7.1524

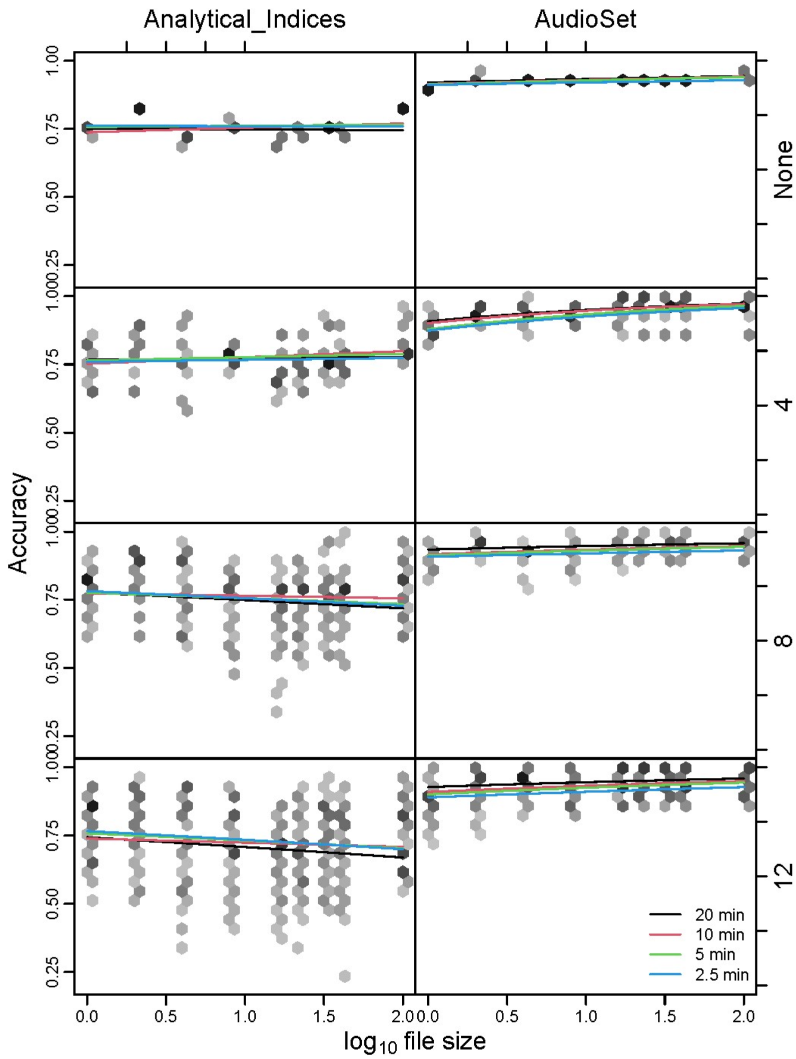


Figure 2.5: Classifier accuracy model predictions as a function of file size (x-axis), index type (columns), temporal subsetting (rows), and Recording Length (colours, see legend). Hexagon binning is used to show the distribution and density of the underlying data.

2.6. DISCUSSION

Ecoacoustics is a new and rapidly expanding field of ecology with great power to describe ecological systems (e.g. Sethi *et al.*, 2020), but methodological choices have proliferated that have poorly known impacts on ecoacoustic analysis. I have shown that the choice of acoustic index is key and confirm (Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020) that a multi-dimensional generalist classifier (AudioSet Fingerprint) outperforms more traditional Analytical Indices regardless of the levels of audio compression or recording schedule.

Analytical Indices have been constrained to a limited set of features within soundscapes, leading to high degrees of correlation. For example, ADI, AEve and H indices are all summaries of the evenness of frequency band occupancy (Jerome Sueur et al., 2008; Villanueva-Rivera et al., 2011). This non-independence can further decrease the dimensionality of suites of Analytical Indices, which are already typically small. Here, I use just the mean values of Analytical Indices, but other studies have incorporated both the mean and standard deviation (Bradfer-Lawrence et al., 2019), which provides further dimensionality. Although the AudioSet Fingerprint clearly benefits from a large number of relatively uncorrelated acoustic features, most Analytical Indices have the advantage of being designed to capture ecologically relevant aspects of the soundscape.

Compression affected the quantification of all indices in both index types (Figure 2.3), and – although the qualitative patterns were noisy – the groupings seen may reflect the underlying algorithms. The apparent threshold for AudioSet Fingerprint at CBR16 may be due to the obligatory loss in audio quality before samples pass to the CNN used to generate the AudioSet Fingerprint. The audio was down-sampled to 16kHz and then presented as a Mel-shifted spectrogram, which increases sensitivity in frequency ranges relevant to human hearing, akin to those frequencies favoured in commercial compression. Coupled with its variable quality training set (YouTube Videos), these factors may predispose AudioSet Fingerprint to perform as well with high-quality audio as with intermediate and low-quality mp3s.

The M and NDSI were also largely unaffected by compression until the frequency range was reduced. When mp3 audio is compressed below 32kb/s, the audio swaps from being encoded as MPEG-1 Audio Layer III (which supports a max frequency of 16-24kHz) to MPEG-2 Audio Layer III (max: 8-12kHz). This change in format results in the removal of signals beyond the cut-off frequency threshold. Further reduction is seen where at CBR8 when encoding changes again to MPEG-2.5 Audio Layer III (max: 4-6kHz). The M index is explicitly a measure of amplitude (Jérôme Sueur et al., 2014) and is largely unaffected until downsampling reduces amplitude. Similarly, NDSI measures the proportion of sound in biophonic vs anthropophonic frequency bands: as downsampling progressively eliminates sounds within the frequency range (2 – 11 kHz) containing most biophony, NDSI is known to increase (Kasten et al., 2012). The ADI index also shows a marked increase in the magnitude of the difference at higher rates of compression (CBR64). However, a small but significant difference can be observed from CBR256. The ADI index measures the spread of frequencies above a certain loudness threshold; the effect of compression on ADI may therefore suggest that certain high-frequency bands are dominant in this soundscape.

AEve and H, both of which describe the spread and evenness of amplitude over the full range of frequencies, showed a gradual increase in *D* that reversed when the Nyquist frequency reduced. The two measures differ in measuring dominance (AEve: Villanueva-Rivera *et al.*, 2011) and evenness (H: Sueur *et al.*, 2014) across bands but may share a common explanation. In both cases, compression preferentially removed amplitude from some bands, initially decreasing evenness, but downsampling removes bands entirely, possibly restoring a more even distribution.

ACI and Bio both shared a dependence on high-frequency or quieter sounds and were generally most severely affected by compression. ACI measures frequency band dependant changes in amplitude over time (Pieretti et al., 2011) and is reduced when there is minimal variation between time steps. Loss of "masked" sounds under low compression and then 16 – 24 kHz sound under CBR16 may reflect the loss of ecoacoustic temporal variation: this band includes the calling range of many invertebrates, birds, mammals and amphibians (Browning et al., 2017). The Bio index similarly quantifies the spread of frequencies in the range 2kHz-11kHz, all relative to the quietest 1kHz band (Boelman et al., 2007): loss of quiet frequency bands, therefore, make it uniquely sensitive to compression. Despite both of these indices incurring alterations 200% larger than the uncompressed range, the Analytical Indices classifier accuracy still showed robustness to compression, perhaps suggesting these indices are less important for classification than the others. Bradfer-Lawrence *et al.* (2019) have already shown that the Bio index contributes little additional power when classifying soundscapes but found that ACI was the strongest individual contributor in this suite of indices (Bradfer-Lawrence et al., 2019). my findings suggested this ranking may not be consistent across different levels of compression.

My findings reflect those of an earlier study that explored the effect of mp3 compression (VBR0 and CBR128) on indices describing specific bird calls (Araya-Salas et al., 2019). They found that compression did not cause a systemic deviation in all indices, but rather indices designed to capture extreme frequencies were less precise after compression, particularly with VBR-encoded files (Araya-Salas et al., 2019). While some of these principles are present in my findings, the use of a wider range of compressions has allowed us to develop a more complete description of the action of compression on soundscape indices in audible range recordings.

I found that even the highest rate of compression caused a comparatively small reduction in the overall accuracy of the classification task (5.8% and 3% for Analytical Indices and the AudioSet Fingerprint, respectively, for the 5-minute recordings without temporal subsetting). In both cases, the reduction in accuracy was explained by a higher degree of overlap between the primary and logged forest. When audio is compressed, the whole signal is altered, but higher frequencies and quieter sounds are more severely altered and reduced than others (Sterne, 2012). Higher and quieter frequencies (akin to specific animal vocalisations) may therefore be more important for separating logged and primary– but less so for discerning cleared from other forest types (which may be more dependent on overall amplitude). These proportionally small differences, while somewhat reassuring, should be considered with caution as they may be due to the large differences in habitat structure among my three habitat classes. Combining this with my relatively small sample size, I would like to emphasise that these findings may, therefore, not be generalisable to areas of more closely related forest.

Both Analytical Indices and AudioSet Fingerprint had similar changes in variance as a result of recording length. Transient vocalisers are, therefore, likely somewhat important in the determination of the AudioSet Fingerprint and variable importance in some Analytical Indices. The ACI index was not impacted by recording length despite specifically quantifying how the soundscape changes over time (Pieretti et al., 2011). The ADI, AEve and H all did incur an alteration in variance as recording length changed, interestingly these indices do not consider any temporal value but rather just the spread of frequency (Jérôme Sueur et al., 2008; Villanueva-Rivera et al., 2011), indicating that transient calls akin to short term anomalies in frequency are perhaps lost when recording windows are altered.

Finally, I found that subsetting audio data temporally and analysing them separately had an unpredictable impact on classification accuracy, with the AudioSet Fingerprint classifier staying consistently high while the Analytical Indices classifier was returning accuracies anywhere between 20 and 100%. Temporal subsetting can reduce the impact of diel variation on analyses but poses a trade-off as it reduces the amount of data used to train the classifier. Analytical Indices may perform better over longer recording periods as > 120 h of recordings are required for Analytical Indices to stabilise (Bradfer-Lawrence et al., 2019), yet in my study, I had just 70 – 75 h of recordings per site. Overall, I found that compression, recording length and temporal subsetting caused a small decrease in classifier accuracy, with the largest overall contributor being the choice of AudioSet Fingerprinting over Analytical Indices. The AudioSet Fingerprint classifier, temporally sectioned and trained on just 2 hours of data, was able to, on average, outperform the Analytical Indices classifier trained on the full 24h.

It is important to note that the AudioSet Fingerprint and several of the acoustic indices employ temporal downsampling that limits the dynamic range of a recording. This immediately excludes any ecologically relevant information in ultrasound frequencies from the analysis. Any studies wishing to include the rich array of species that call in ultrasonic frequencies should not use mp3 compression and should perform further investigation.

2.7. RECOMMENDATIONS AND CONCLUSIONS

This study was designed to compare distinct forest types in Malaysian Borneo and the recording periods used are relatively small. Based on the results of this study, I provide the following four recommendations in audible range monitoring; however, effort should be made to ensure they are generalisable to the desired area of deployment and not used when dealing with ultrasonic recordings:

1) I provide additional evidence for the viability and stability of AudioSet Fingerprinting rather than Analytical Indices when classifying soundscapes.
2) Lossless compression is always desirable, but if data storage/transmission becomes a bottleneck to a study, I advise using the VBR (quality = 0) mp3 encoder if using Analytical Indices, which will reduce the file size to roughly 23% of the original while having minimal impact on indices (other than ACI). The AudioSet Fingerprint, however, is more robust to compression and so can tolerate a minimum compression encoding of CBR64 (8% of the original file size) without significant effect.
3) If further compression is a necessity, use indices which describe the general energy of the system rather than those which are dependent on high-frequency or quieter sounds, such as ACI.
4) Temporal subsetting may be a useful alternative for capturing soundscape descriptors with AudioSet Fingerprinting when data storage costs are a bottleneck. However, temporal subsetting should be used with caution when using Analytical Indices owing to the variation in classification accuracy, precision and recall.

There exists a trade-off between the quality and volume of data that can be stored in ecoacoustics. I have investigated the impact of compression along a gradient of habitat disturbance, providing evidence that compressed audio can be used without severely affecting either of the index types. The ability to use compression may reduce experimental costs, remove bottlenecks in study design, and help remote ecoacoustic recorders reach true autonomy. Moreover, by providing a quantified description of how individual indices, and more broadly grouped index categories, respond to compression, I have enabled comparisons to be drawn between studies of compressed and non-compressed audio. Increasing comparability of studies will become progressively important as global ecoacoustic databases and recording sites grow and open up novel opportunities to explore datasets across huge temporal and geographic scales.

2.8. DATA ACCESSIBILITY

Acoustic Data: Available at 10.5281/zenodo.5159914

Analytical Indices/ AudioSet Fingerprint Data: Available at 10.5281/zenodo.5153193

Analysis Scripts: Available on Github at https://github.com/BeckyHeath/Experimental-Variation-Ecoacoustics-Analysis-Scripts

3. CHAPTER 3: INTRODUCING MAARU (MULTICHANNEL ACOUSTIC AUTONOMOUS RECORDING UNIT for SPATIAL ECOSYSTEM MONITORING

3.1. CHAPTER OVERVIEW

Acoustic localisation is the process of adding positional information to recorded audio with a variety of applications in ecosystem monitoring. Specifically, its uses span: counting individuals for better abundance, locating illegal behaviour such as logging/poaching, and monitoring organism behaviour such as habitat use or species interactions. Acoustic localisation depends on simultaneous recording from multiple microphones at multiple known positions. Several studies have shown the advantage of multichannel recording for acoustic localisation, but uptake remains limited as the equipment is often expensive, inaccessible, or only suitable for short-term deployments.

Here I present a low-cost, open-source device architecture built entirely from easily accessible commercially available equipment that is integrated into a solar-powered, networked system. my device, hereon MAARU (Multichannel Acoustic Autonomous Recording Unit), records and sends high-quality multichannel audio autonomously, removing the need for re-entry to the field entirely. Here I introduce MAARUs hardware and software, as well as the results of lab and field tests investigating localisation accuracy and durability.

I found that MAARU has no additional power demands compared to a similar omnidirectional device. MAARU can record across much of the audible range (sample rate: 16kHz). All devices remained watertight throughout a 6-month deployment period, and even fully weatherproofed devices were able to accurately localise signals to ±10°. In the field, two devices suffered rodent damage and system failure due to issues with solar panel attachment; however, I outline guiles to reduce the risk of reoccurrence in future work. Other on-board issues with energy distribution across microphones were also encountered. However, I designed a system of automatic error detection and correction.

Our proposed device is an exciting, accessible, low-cost option for those looking to explore spatial sound ecology accurately and easily. I show a successful port of this device to HARKBird, a sound source localisation platform, and outline other avenues of analysis. Ultimately the added directional element of multichannel recording may uncover patterns of behaviour not detectable with single-channel recording.

3.2. INTRODUCTION

Passive Acoustic Monitoring (PAM) describes the use of remotely deployed acoustic Autonomous Recording Units (ARUs) to study ecosystems in a way that is cheaper, less damaging, and more applicable to large-scale, high-resolution wildlife surveys than human field data collectors (Darras et al., 2019). The applications of PAM are broad and have been reviewed (Blumstein et al., 2011; Browning et al., 2017; Gibb et al., 2019; Mcloughlin et al., 2019), but include classifying individual calls to monitor species and biodiversity (Stowell 2022), building predictors of whole ecosystem health (Sethi, Jones, Fulcher, Picinali, Jane, et al., 2020), and building predictive models of acoustic species distributions (Villén- et al., 2022). Beyond the one-dimensional appraisal of soundscapes, spatial information can be determined from signals recorded from multiple fixed positions simultaneously. Studies have increasingly investigated spatial acoustics applications in often short-term deployments. Here I present a fully autonomous, networked PAM device which can record and transmit multichannel recordings in almost real time, providing spatial acoustic data over longer and more ecologically relevant time scales.

The principles of acoustic localisation work similarly to the human auditory system, whereby slight differences in acoustic signals at multiple known points (the two ears, in the case of a human) can be used to triangulate either a position (hyperbolic localisation) or direction compared to the receiver/s (DOA). In DOA estimation specifically, inter-channel delays can be systematically manipulated (beamformed) to find the time delay resulting in maximal correlation across channels and therefore estimate the calling direction of the signal (Mitianoudis, 2003). Once locations are known, the inter-channel delays can be manipulated to beamform to that location. This beamforming uses the principles of constructive and destructive interference to amplify calls from a desired direction, increasing the signal-to-noise ratio (SNR).

The applications of DOA estimation and acoustic localisation in an ecosystem monitoring context have been reviewed by (Blumstein et al., 2011; Rhinehart et al., 2020), with the latter identifying eight key purposes: Assessing the positions and movements of individuals, studying interactions, identifying individual identities, sub-setting (and beamforming) sounds for more detailed analysis, calculating species abundance more accurately, and inferring habitat usage/territory. Additionally, acoustic localisation has been used to locate illegal logging and poaching through chainsaw and gunshot sounds, respectively (Andrei, 2015; Wijers et al., 2019).

Several spatial ecoacoustic recorders have been developed that are showing promise. Currently, these devices largely suffer from issues that limit their easy adaptation to eco-acoustic studies, such as being non- or semi-autonomous, requiring specialist/discontinued equipment, lacking full 360° localisation, or lacking necessary hardware/ software for extended deployment in the field (Bruggemann et al., 2021; Smith et al., 2021; Suzuki et al., 2017; Wijers et al., 2019) (Appendix B: Table 1).

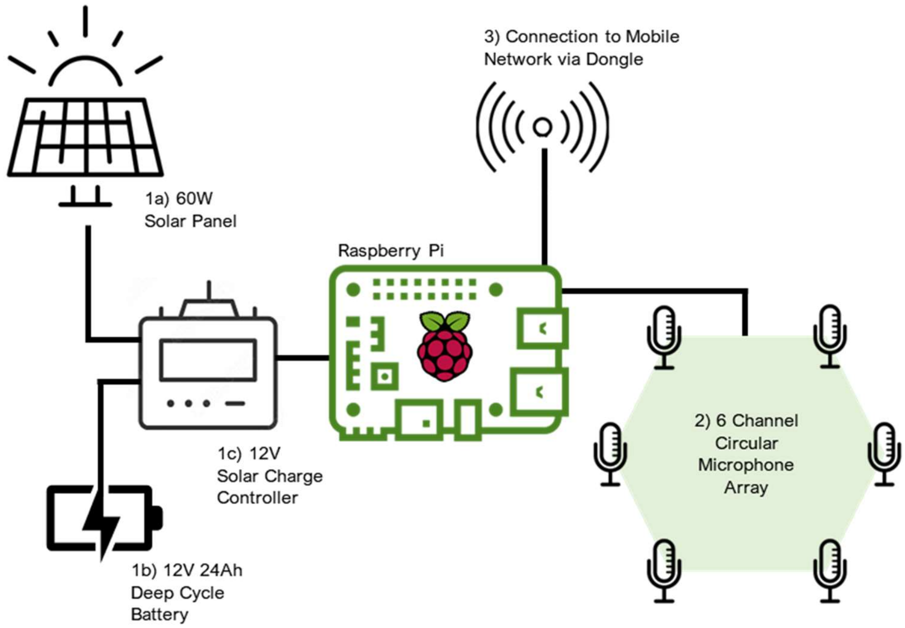
Here I present an autonomous, self-powering, networked device capable of DOA estimation built from low-cost, commercially available components. I demonstrate successful lab and field tests of MAARU and explain software and hardware adaptations necessary for multichannel recording. While there are still advancements to be made, my device demonstrates a low-cost and accessible new possibility in continuous multichannel recording.

3.3. DESCRIPTION

MAARU is based upon an omnidirectional (single-channel) autonomous acoustic recorder presented by Sethi et al., 2018, re-designed to localise signal DOA through multichannel recording. Changing the recording mode from omnidirectional to multichannel involves modifying: 1) the core device hardware, 2) the device weatherproofing, and 3) the onboard software. The open-source code and step-by-step setup instructions are available at https://github.com/BeckyHeath/multi-channel-rpi-eco-monitoring,

3.3.3. Core Device Hardware

This ARU centres around a Raspberry Pi, a multichannel microphone array (Circular 6-Microphone Seeed Studio ReSpeaker array), a (renewable) power source and a connection to cloud storage via a USB dongle (Figure 3.1).



. Figure 3.1 shows the schematic setup of MAARUs hardware. MAARU is centred around a Raspberry Pi with the following attachments: 1) a renewable power source, (a) a 60W solar panel, b) a 12V 24Ah deep cycle battery, and c) a 12V solar charge controller. 2) A ReSpeaker 6—Mic circular array connected via the ReSpeaker multichannel interface. And 3) a mobile dongle with a pre-paid mobile network sim card.

The microphones used in the ReSpeaker array (MSM321A3729H9CP) are small (3.76 x 2.95 x 1.1mm), omnidirectional, Micro-Electro-Mechanical System (MEMS), programmed to record at 16kHz. The six microphones are fixed in a circular array on a PCB, which attaches to the Raspberry Pi via a Seeed Studio Multichannel HAT Soundcard. Localisation/ DOA Estimation depends on small temporal differences between microphones, so recording must have highly accurate and consistent synchronisation (Mennill et al., 2012). In standard ARUs, microphone clock drift can cause delays from 1 to 10 seconds daily (Thode et al., 2006). To mediate clock drift, it is useful to have all recordings managed by the same internal clock (and sound card), which is most stable via a short cable connection on small DOA devices, such as this one. Other options include GPS (Crunchant et al., 2022), cable (Suzuki et al., 2017), network (Bruggemann et al., 2021) or through array self-organisation (Trifa et al., 2007).

Despite recording from multiple microphones, this device uses the same amount of power as the omnidirectional recorder (5W – Appendix B: Section 4). The powering system can be tailored to resources available at the study site (mains power/ exposure to wind/ hours of direct sunlight), provided 120Wh (5W x 24h) of energy is generated per day. At my woodland study site in the UK, solar panels were affixed in the forest canopy, where they were theoretically exposed to ~4h of direct sunlight daily. A 30W panel can generate a maximum of 120Wh in 4h of direct sunlight (120Wh ÷ 4h), and a 10Ah 12V power supply should store enough energy to charge MAARU for 24h without sunlight (120Wh = 10Ah x 12V). In practice, inefficiencies, fluctuations in sunlight, energy loss, and component degradation may make this setup unreliable, so instead, I use a 60W monocrystalline (higher efficiency) solar panel with a 24Ah 12V LiFEPO4 deep-cycle battery. A 12V solar charge controller protects the battery and regulates the Raspberry Pi's current supply and draw.

3.3.4. Weatherproofing

The microphone array must be positioned at the same orientation as the desired spatial separation (vertically for height or horizontally for direction) to localise sounds usefully. Ideally, the path between the sound source and the microphone must be uninterrupted. Other ARUs use the solar panel to protect from overhead rain (Sethi et al., 2018); however, this interrupts the above sound path and would interfere with localisation. In this deployment, I instead keep the power source and recording hardware separate: the solar panel is connected via a heavy-duty cable to a waterproof dry bag containing the battery and solar-charge-controller; an additional power cable then connects the solar-charge-controller to a separate container holding the Raspberry Pi, multichannel array, and dongle.

I found that modified 450ml waterproof Sistema plastic containers were a viable option for weatherproofing while minimising interference with the signal. Holes were drilled in and sanded down for: the microphones, raised components on the front of the microphone array, and cables to enter at the base. The microphone holes are covered by ePTFE Acoustic Vents, which have small pores that allow for the passage of air (sound) but are too small to permit water droplets (full details: Appendix B: Section 3).

The microphone array is pushed flush to the top of the container using polystyrene, and absorbent silica is included to keep the recording components of MAARU dry in case of condensation. The other holes in the microphone component (e.g., the cable holes) are flexibly weatherproofed using sniper tape and a silicone-based sealant.

3.3.5. Device Software

As in the omnidirectional device, the software used here records, optionally compresses, and then transmits acoustic data on a user-defined schedule, as well as receiving updates and sending log information daily. The core scripts are essentially the same as those used by (Sethi et al., 2018), but with the following alterations: installing the multichannel recording software, changing the compression mode from lossy to lossless, and altering the setup scripts for compatibility.

Using the Seeed multichannel software (Seed-voice card), MAARU records from all six channels of the ReSpeaker array simultaneously (16kHz, 16-bit per channel). The Seeed software installation requires a previous version of Raspbian (Buster, February 2020) and Python3 (Van Rosssum & Drake Jr, 1995). Several packages had to be altered or rolled back for compatibility.

Lossy compression codecs remove data irreversibly and non-uniformly from the audio to reduce file size. While mp3 compression has minimal impact on soundscape-level analysis and birdsong descriptors/detections (Araya-Salas et al., 2019; Heath et al., 2021; Stowell & Plumbley, 2014), the analysis needed in DOA estimation requires a much finer-scale comparison between signals. In place of lossy mp3 compression, I reduce file size using Free

Lossless Audio Compression (FLAC) (Xiph.org, 2011-2016), which is stable, easy to use, and upon decompression, is identical to the pre-compressed signal. FLAC compression across 6 channels takes longer than mp3 compression on a single channel, meaning (unlike in the omnidirectional device) compression occasionally takes longer than the audio recording. File locations were restructured to keep the "current sample", "finished raw samples", and "compressed files pending upload" separately and avoid tripping out the process.

Other, more cosmetic changes included: log file alterations (with additionally feedbacks information about the recorders' compression status), the inclusion of a kill time in recording (which allows MAARU to move on from faulty recordings), as well as the addition of the multi-mic config script, which enables MAARU to record from the sensors. Full details of the changes are traceable via GitHub.

3.4. PROOF OF CONCEPT

3.4.3. Lab Testing (Weatherproofing)

A test signal was played at various points around MAARU with and without weatherproofing (Figure 3.2). The test signal contained a 50-20000Hz sinus-logarithmic sweep signal (hereon, sweep) and five test tones (either pink noise or a Eurasian Wren song (Karel, H. 2020)) at 15-second intervals (Figure 3.2). The sweep and first test tones were played in position A, and the following four tones were played in order at each position until position E (Figure 3.2). Test tones were recorded from devices four times (twice with weatherproofing; twice without. Both recordings were conducted at 16kHz, 16-bit, all channels).

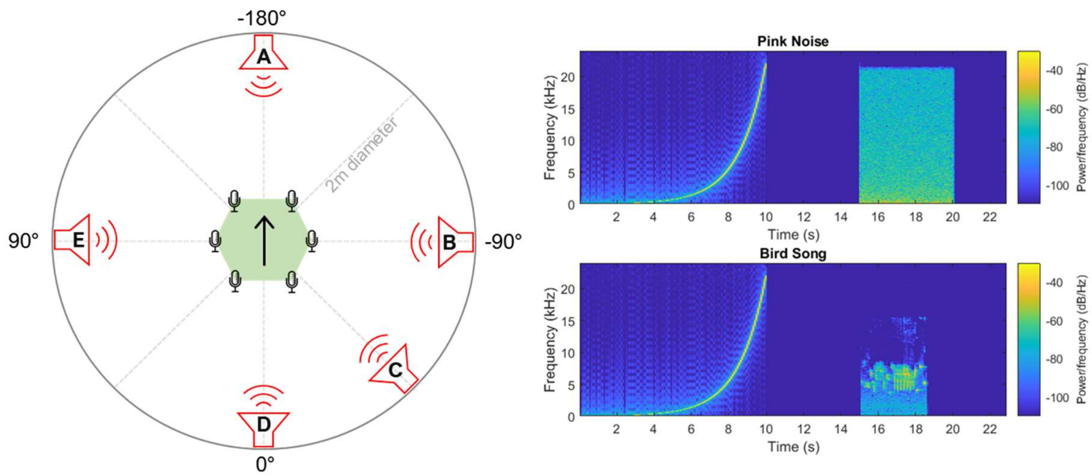


Figure 3.2 shows the setup of the lab tests. Devices were assessed (with and without weatherproofing) in the centre of the 2m test area (left). A 50-20000Hz sinus-logarithmic sweep signal (hereon, sweep) is played, followed by five test tones (original tones pink noise or Eurasian Wren, sweep and first test tone on right panels). The tones were played at 15 seconds intervals from positions (starting at A and continuing to E) 1m from MAARU. The power spectra of the sweep, as well as DOA estimation accuracy, were compared for devices with and without weatherproofing.

3.4.3.1. Impact of Weatherproofing on Power Spectra

Spectral data were analysed using MATLAB with the digital signal processing (DSP) toolbox (MATLAB, 2010). Weatherproofing caused a loss in the acoustic signal amplitude of 10-15dB between 1-4kHz, increasing to 15-20dB between 4.5kHz and 6kHz (Figure 3.3). The difference decreases at higher frequencies as the overall sensitivity to those frequencies also decreases (Figure 3.3). Other factors, such as test number and specific device used, also had an effect but to a much smaller degree (<3dB and <10dB, respectively, Appendix B: Figure 5.2.x).

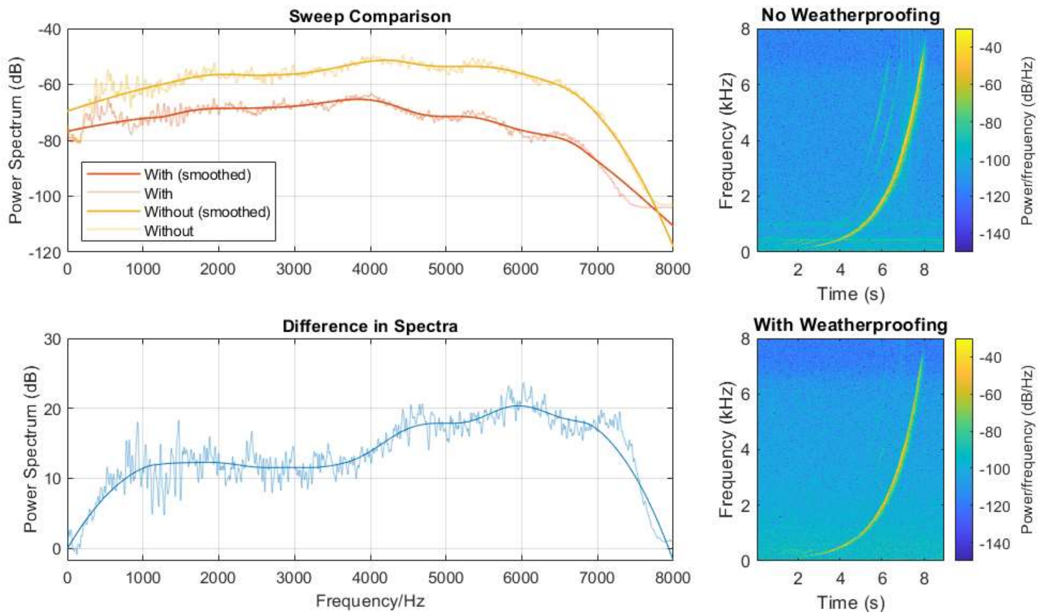


Figure 3.3 compares the power spectrum of sinu-logarithmic sweep signals (sweeps) recorded from devices with weatherproofing (yellow line) and without (orange line); the difference in these spectra (blue line); Spectrograms on the right show the resultant difference in an exemplar sweep from each group. Data is smoothed using robust locally weighted smoothing.

3.4.3.2. Impact of Weatherproofing on DOA Estimation

DOA was estimated using the open-source software HARKBird (Suzuki et al., 2017). HARKBird is the bio-acoustic adaptation of an open-source software platform designed for robot audition: HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) (Nakadai et al., 2017). HARK localises signals using MUSIC (Multiple Signal Classification) (Schmidt, 1979) which compares eigenvalues/ eigenvectors of covariance matrices across spectrograms from all 6-channels to determine how many signals are present and their locations. I run HARKBird in a dedicated Virtual Machine via VMWare with an Ubuntu 18.04 environment.

Analysing a signal using HARKBird requires the input of both the array-specific transfer function and a configuration file. The HARKBird algorithm was configured to localise to the nearest 5°. The only configuration difference between analysing signals from the weatherproofed vs un-weatherproofed signals was the power threshold (THRESH) used in signal tracking (24 for devices with weatherproofing, 29 for without) (Figure 3.4). THRESH refers to the MUSIC power. If THRESH is too small, unwanted noise is localised; if THRESH is too large, the signal will not be localised.

Signals recorded by un-weatherproofed devices were all accurately localised to within 5° (average error: 0.41°± 2.04° and 1.43°± 2.31° for pink noise and birdsong, respectively, Figure 3.4). Weatherproofing devices increased false detections (echoes) from 0 to 17% (Figure 3.4), possibly due to the decrease in the detection threshold. False positives were often at consistent angles for each test signal and were manually removed before analysis. Considering just the true detections, the weatherproofed device was able to localise all signals to within 10° (average error: 3.33°± 4.08° and 7°± 2.74° for pink noise and birdsong respectively, Figure 3.4, Full details including echoes in Appendix B: Figure 7.2.1).

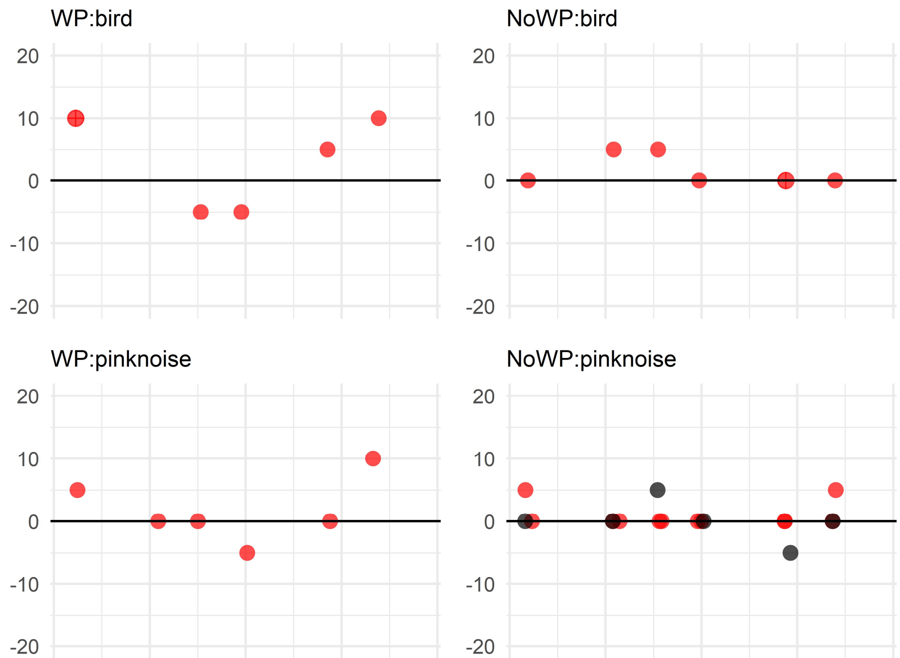


Figure 3.4 shows the predicted azimuth and error for all tested azimuths. WP/ No WP refers to whether the MAARU device being used has weatherproofing or not (respectively), and bird/pink noise refers to whether the test signal was birdsong or pink noise, respectively. Circle colours indicate different recorders. Excluding echoes, as in this figure, all detected signals were accurate to within ± 10°.

3.4.4. Field Testing

3.4.4.1. Study Site

Four acoustic recorders were deployed in two managed woodland areas with different forest structures in Southeast England: (1) Chalk Wood (London Borough of Bexley) is an area of Ancient Semi-Natural Woodland (ASNW), primarily used as a bridleway with relatively low footfall. (2) Joydens Wood (Woodland Trust, 2021) is an area of ancient woodland converted to Corsican pine by the Forestry Commission in 1956 (Plantation on Ancient Woodland, PAWS). The Woodlands Trust has been restoring the site since 1993, but large areas remain as conifer forest (Woodland Trust, 2021). Joydens Wood is a popular public woodland. The two sites are adjacent; however, dominant tree species vary locally.

Recorder locations were limited to areas with sufficient connectivity to mobile networks (5 Mbps, 10 Mbps if uncompressed audio streamed, assessed by checking upload speeds). Two recorders were placed on large oak trees in ASNW in Chalk Wood (ASNW1 and ASNW2), and two other recorders were placed on Corsican pine trees in PAWS in Joydens Wood (PAWS1 and PAWS2). All four recorders were deployed continuously between August 2021 and January 2022.

3.4.4.2. Recording Hours and Recorder Longevity

In pre-deployment trials, where the solar panels were completely exposed to sunlight, 24h recordings were consistently sent from the MAARUs. However, placement in field sites led to a drop-off in transmitted recording hours (Figure 3.5). As sunlight decreased when the UK entered the winter months, transmissions fell as low as one hour per day but rose again in the Spring (Figure 3.5). Despite reduced recording hours, I still acquired 1700+ hours (550+ GB) of data from the Woodlands.

All four units remained watertight for the whole deployment period. Other environmental factors caused damage of varying degrees to a particular MAARU's longevity and led to early device terminations (Figure 3.5). Two devices suffered early termination due to a partial panel fall (ASNW1), shorting the device and requiring a complete system reset, and severely rodent-chewed wires leading to an overload of the solar charge controller (PAWS2). PAWS2 returned to normal function when connected to a different power system post-deployment. ASNW2 had continuous onboard microphone issues leading to a mix of dead and unbalanced channel levels but remained powered for the whole deployment period. PAWS1 remained working well for the entire field deployment (Figure 3.5).

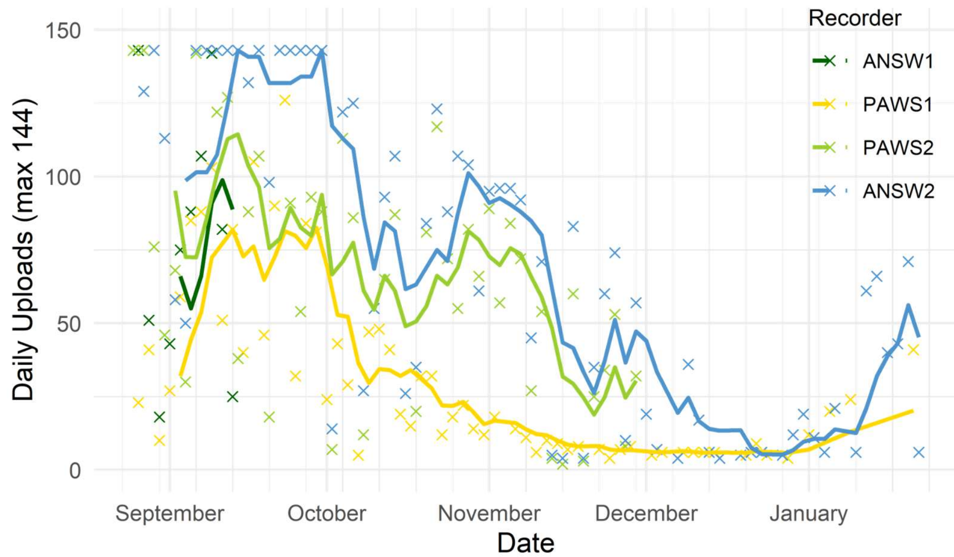


Figure 3.5 Device-specific longevity, two devices (ASNW1 and PAWS2) did not complete the deployment due to environmental factors. Solid lines represent the 5-day moving average.

3.4.4.3. Sustained Audio Quality

While sound-source localisation/ DOA estimation works primarily on slight temporal differences of a signal at each microphone, the longer-term overall spectra should be uniform for close-together microphones. I analysed microphone-specific spectra over 10-second windows in ten pre-deployment and ten mid-deployment recordings. During the pre-deployment period, the PAWS1, PAWS2, and ASNW1 devices had largely uniform spectra across their microphones. ASNW2 already showed considerable anomalous behaviour attributed to channels (1,3:6) failing and switching. After three weeks of deployment, all devices except PAWS1 suffered some degree of non-uniform microphone degradation resulting in a loss of amplitude in some channels (Appendix B: Figure 6.2).

I addressed the issues with amplitude by creating a script that autonomously detects and fixes issues with inter-channel discrepancy; it achieves the former by calculating the channel-specific maximum absolute level, using thresholding to determine activity in each channel and then processing the audio for analysis. If any channel in a recording has no signal, the recording is removed from analysis; if some channels have more amplitude than others, the difference in absolute peak amplitude is found, and the quieter channels are amplified by half of that difference; and recordings, where all channels are equivalent, are analysed as they are (full details: Appendix B: Figure 6.3).

Post-deployment, all MAARU devices were assessed for recording quality and DOA estimation accuracy as before (Figure 3.2). I show how adjusting amplitude can restore DOA estimation accuracy to levels equivalent to pre-deployment (Figure 3.6). PAWS1 remained consistent without adjustment, with all tones being detected to within 5° (average: -0.5° ± 2.84°, and 0°± 3.33° for bird song and pink noise, respectively). ASNW2 suffered considerable non-uniform microphone degradation, only detected 40% of signals in the DOA estimation task, and appeared to have random azimuth predictions, so it could not be used reliably. The amplitude adjustment algorithm automatically removed all ASNW2 recordings. Pre-amplitude adjustment, no signals were detected by PAWS2. However, post-amplitude adjustment, all signals were detected and accurate to within 10° (average: -2° ± 7.75° and -1° ± 6.87° for birdsong and pink noise, respectively). A notable difference in the pre-and post-deployment DOA estimation tests is the increase in false positives, ranging from 0-17% pre-deployment to 17-29% post-deployment.

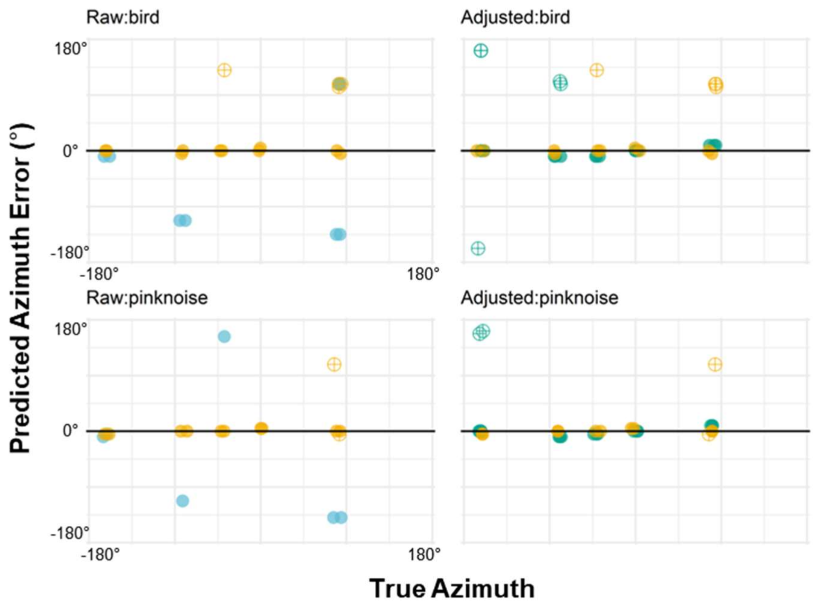


Figure 3.6 localisation tests completed on recorders post- 6-month field deployment. The predicted azimuth and error for all tested directions are shown. Raw/Adjusted refers to whether the localisation experiment was done pre- or post-adjustment (respectively), and bird/pink noise refers to whether the test signal was birdsong or pink noise, respectively. Filled circles are "detections", while empty circles are false detections (echoes). The adjusted amplitude algorithm removed faulty ASNW2 recordings and allowed for the inclusion of the previously uneven PAWS2. Post-amplitude-adjusting estimated directions across remaining recorders were accurate to within ±10° and ±5°for the PAWS2 and PAWS1 recorders, respectively.

The deployment also slightly impacted the spectral information of sweeps recorded by PAWS1 (Figure 3.6b – power spectra track strongly with differences of +/- 5dB). The PAWS2 incurred a slightly wider impact (broadly ±10dB), while ASNW2 incurred over 20dB difference (Full details: Appendix B: Section 8).

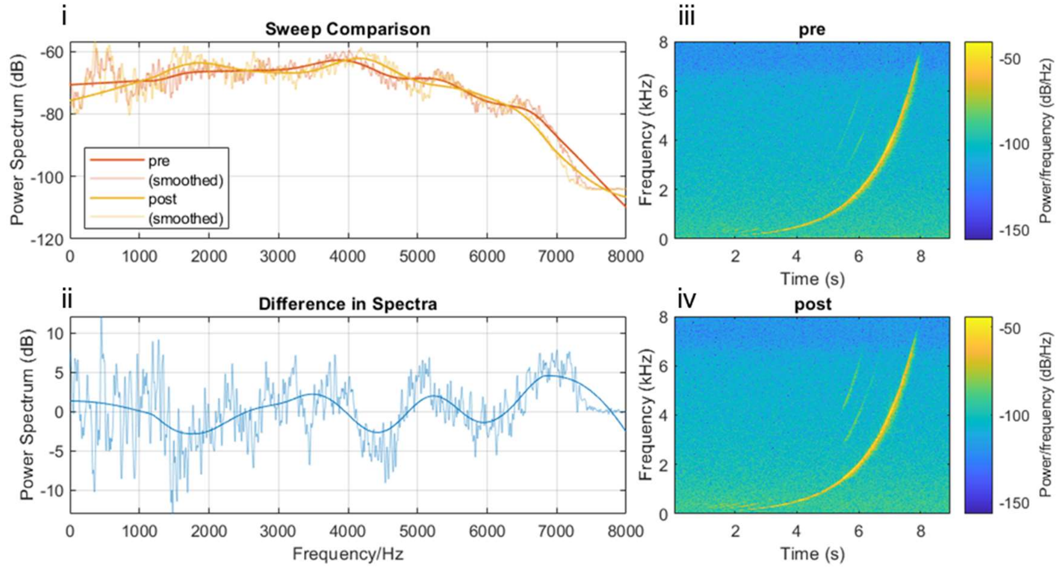


Figure 3.7 (i and ii) compare the spectra of a sweep recorded by the same device (PAWS1) pre- and post- 6-month deployment; the corresponding iii and iv show this as spectrograms.

Two devices suffered terminal damage in the field, one of which only affected the battery and was working well on reboot. Of the remaining two devices, one worked well throughout the deployment period, and the other suffered continuous onboard faults resulting in unusable data. After repowering and postprocessing, two of the four devices worked well and detected all lab signals to within 10°.

Table 3.1 summary of results from the 6-month field deployment

Device	Lasted Full Field Deployment?	Postprocessing	Sweep Response Difference	Post Deployment Precision	Post Deployment Recall
PAWS1	Yes	No	+/- 5dB	5°	100%
PAWS2	No (severely chewed wires)	Yes (Signal boost)	+/- 10dB*	10°*	100%*
ASNW1	No (partial panel fall resulting in terminal system reset)	NA	NA	NA	NA
ASNW2	Yes	Yes (removed)	20dB +	random	40%

* Refers to the post postprocessing results

3.5. DISCUSSION

I aimed to create a multichannel acoustic recording unit capable of DOA estimation made from readily available low-cost materials, which are robust and can survive extended field deployment. I evaluated MAARU for these properties in both lab and field tests. I report here the benefits and costs of using the proposed system, as well as considerations to be made for future deployments.

Low-cost and readily available components were key design aims of MAARU. All components were available to purchase online at the time of writing and required only the addition of a power drill (for microphone holes) and screwdriver to build. My device combines autonomous PAM with usable multichannel recording for £26 more than the omnidirectional recorder (SAFE Acoustics: £190/unit (Sethi et al., 2018), This device: £206 – both without powering, Appendix B: Table 2.1 and 2.2). Due to differences in microphone power requirements, both devices run at ~5W (one externally powered electret capacitor microphone vs six passive MEMS microphones). Providing power is an additional cost dependent on local resources such as grid connection or hours of direct sunlight. My proposed setup consists of a solar panel, deep cycle battery and solar charge controller, costing an additional £256 total.

An additional cost is in maintaining data transfer. Each device continuously generates and sends high-quality data (16.5GB/ day uncompressed, 7.2GB/day FLAC compressed), requiring a sufficient network upload speed (10/5 Mbps) and data plan. It is possible to run MAARU offline and manually collect data. However, this reduces autonomy and increases field revisit costs. I was able to cover my devices for £20/month/device using an unlimited data plan (SMARTY, 2022).

Our device costs £462 + £20/month, including data transfer and renewable power in a temperate climate. This price comes out mid-range compared to CARACAL (£150/unit, non-autonomous DOA) and other GPS synchronised single-channel devices whereby at least three recorders are needed (AudioMoth Dev £225 for three, SM2+ £2,100 for three, SAFE Acoustics £600 for three excluding powering, Full comparison Appendix B: Table 1). However, I believe that the advantage of MAARU being autonomous and stand-alone DOA justifies this additional cost.

The MAARU setup was also assessed for robustness to long-term field deployment. Despite the low-cost, accessible housing used, all devices remained watertight throughout the deployment. Two devices stopped recording data due to a solar panel partially falling from where it was tethered (ASNW1) and rodent damage to wires (PAWS2). While these factors caused the early termination of their devices, accounting for these issues with stronger/ more points of panel attachment and the use of armoured cable would reduce the risk of this reoccurring. Investigating how weatherproofing affected the sensitivity of MAARU, I found that weatherproofed devices had a loss of signal peaking at 20dB at 6kHz. This reduction in signal may be important, especially as it could make the device less responsive to species that call in that range.

The multichannel array chosen here (ReSpeaker 6-Mic) was selected as a low-cost (£42) and easy-to-use solution, sacrificing microphone quality, sensitivity, and recordable frequency range. ReSpeaker also has another 6-microphone array integrated with a Linux platform (Seeed Studio, ReSpeaker Core v2, £90) which, although not currently available, may have interesting applications in future. Specialist soundcards can also be used to perform multichannel recording with externally powered microphones which boast better signal quality and sensitivity (Audio Injector Octo, £122 – not currently available, Elk Audio, £324 – not currently available). Users looking specifically for quieter organisms or those that call at higher frequencies would benefit from a setup with more specialist equipment.

Beyond physical durability, I also investigated to what extent MAARU could accurately estimate DOA in the field. I found that weatherproofing resulted in an increase in maximum localisation error from ±5° to ±10°, with the mean detection error being slightly higher for birdsong than pink noise. The increased loss of accuracy in the birdsong tests is likely due to the decrease in signal around the frequency of birdsong call that was used (5-10kHz). Using raw audio post-deployment, I found that one of the MAARU devices was working as normal, localising all signals to within 5° (PAWS1). The second could not detect and localise any signals (PAWS2) reliably, and the fourth, which had considerable signal loss across all channels, detected only 40% with localisation errors up to 155°.

Spectral observations highlighted that some channels had lost amplitude. I was able to automate the detection and fixing of these issues with an algorithm that removes recordings with completely dead channels and fixes issues in recordings with uneven channel energy. Once amplitude was adjusted, the faulty recordings (ASNW2) were removed from the sample, and PAWS2 successfully localised all signals to within 10°.

Considerations should also be made to false detections (echoes), which increased from 0-17% pre-deployment to 17-29% post-deployment, as microphones may have degraded slightly in the field. While the rate of false positives in recordings post-deployment increased, they remained within the expected range of false positives for the MUSIC method used by HARKBird, which is 3-27% (Suzuki et al., 2017). I reduced the rate of false positives by more finely tuning the threshold case by case; however, this would become time-consuming with a large amount of data and is unscalable. Acoustic Echo Cancellation (AEC) algorithms may reduce false detections.

It is important to also consider that the sample rate used in MAARU recorders (16kHz) results in a somewhat limited dynamic range. While this will cover the majority of bird calls, this excludes all ultrasonic and some audible range sounds. While this has advantages, as it reduces the file size and necessary bandwidth, the target species should be considered carefully before MAARU devices are deployed.

While my lab localisation tests indicate device durability, it should be noted that all tests were in the direct nearfield, whereas field signals may be affected by other factors (attenuation, vegetation, variable topography, temperature and humidity gradients (Darras et al., 2016)). Most popular microphone arrays focus on determining an exact position on large-scale field deployments rather than a lab directionality test, so they cannot be directly compared here.

3.6. CONCLUSION

For the field of acoustic localisation for ecosystem monitoring to gain traction, equipment is needed that is “widely available, inexpensive, self-synchronising and low maintenance” (Rhinehart et al., 2020). Here I present a device that is an easily accessible, relatively low-cost, low-maintenance solution for users hoping to start with *autonomous,* audible range, DOA (spatial) ecoacoustics. MAARU doesn’t include self-synchronisation at present, as it was designed to be a stand-alone unit. Alterations could be made to MAARU to allow for GPS/ cable synchronisation and apply this device in a multi-ARU to perform DOA intersections and estimate caller positions.

I have demonstrated mixed efficacy of MAARU devices in a long-term field deployment showing good durability in cases that did not suffer external damage. I provide advice for future deployments to avoid these issues, specifically armoured cable and more tether points on solar panels in the canopy. As well as a successful port of audio from this device to HARKBird, I anticipate it would work well on other platforms such as HARKBird2 (Sumitani et al., 2019), PAMGuard or ODAS (Grondin et al., 2022). These platforms have their own additional features, such as echo/ noise reduction and t-sne signal clustering, which may reduce false positives and make data easier to appraise.

Localisation of signals in the field can describe patterns that single-channel acoustic ARUs alone cannot while having greater longevity and durability than human observers. Here I add a new dimension to acoustic studies allowing us to track more precisely the movement, interaction, and relation to the habitat of species in almost real-time. These considerations give us a complete 3D understanding of responses to environmental stressors and may help us more comprehensively understand the behaviour of at-risk organisms and environments.

3.7. DATA ACCESSIBILITY

GitHub: https://github.com/BeckyHeath/multi-mic-recorder-analysis
Acoustic data available on request

4. CHAPTER 4: MAARU IN THE WILD – APPROXIMATING ABUNDANCE AND EXPLORING VERTICAL STRATIFICATION WITH MULTICHANNEL ACOUSTICS

4.1. CHAPTER OVERVIEW

Calculating vital metrics such as abundance, density, and population number with acoustic Autonomous Recording Units (ARUs) is limited, as it is often impossible to determine the number of calling individuals in a recording. In the following chapter, I introduce a means of approximating bird abundance by optimally clustering calls by their direction recorded by MAARU (Multichannel Acoustic Autonomous Recording Unit). I also use this calling direction to perform some exploratory analysis of vertically stratified calling behaviour of bird species in Ancient Semi-Natural Woodland (ASNW) and Plantation on Ancient Woodland Site (PAWS), both in South-East England.

I introduce a novel pipeline whereby calls are recorded, localised, and classified using MAARU, HARKBird and BirdNET, respectively. Using spatial clustering, I could attribute the 5933 detected calls to 1000 individuals (331 in PAWS and 669 in ASNW). I also determined an estimate of the call rate of species in the area, finding that Blue Tits, Coal Tits, Long-Tailed Tits, Magpies, Rose-Ringed Parakeets and Wrens had the most frequently calling individuals. I found that the vertical calling angle was dependent on the species. However, there was no effect on whether they were found in PAWS or ASNW.

This novel application of spatial acoustics provides early insight into the abundance and calling behaviour of 19 species of British bird. In this chapter, I evidence that MAARU can be combined with an easily deployable and open-source analysis pipeline, providing an exciting stand-alone platform for approximating abundance alongside monitoring directional information.

4.2. INTRODUCTION

MAARU, explained in more detail in the previous chapter, is a Multichannel Acoustic Autonomous Recording Unit which records 6 channels of audible-range audio from a circular array continuously. Audio signals recorded from multiple channels simultaneously can be compared, through sound-source localisation algorithms, to be attributed positional/ directional information (also reviewed in greater detail in Chapter 3). In this chapter, I present some novel use cases of sound-source localisation in sound ecology, specifically using MAARU as a stand-alone DOA estimation platform for approximating abundance and exploring species behaviour.

4.2.3. Measuring Abundance

Understanding abundance is necessary for enhancing monitoring efforts and appraising species populations, risk, and the influence of conservation efforts (Farr et al., 2019). This is especially important when predicting the influence of climate change and other anthropogenic stressors (Furnas et al., 2019). In-person point counts from the field are connected to abundance through well-defined modes that incorporate detection probability and range (Johnston et al., 2014; Stowell, 2022). Specifically, imperfect field detections can be mitigated and modelled through distance sampling (Buckland, Anderson, Burnham, & Laake, 1993), and repeated counts (Royle, 2004). While this methodology has been shown to work well for small-scale wildlife monitoring wildlife (Zipkin et al., 2014), they do not scale well owing to the cost and field-hours involved in data collection. Acoustic data collected through PAM or ARUs seems an obvious solution, however the methodologies tying acoustic data to in-situ abundance are not clearly defined.

At present, known vocalisations can be identified in sound recordings and be used to build detection/non-detection occupancy models and infer species distributions (Furnas & Callas, 2015; MacKenzie et al., 2002). However, these methods are prone to false positives and do not give a clear reading of abundance. Several studies have looked to include the principles of distance sampling into acoustically detected counts using the “detection radius” of the recorder (Doser et al., 2021). While promising, calculating a detection radius is dependent on habitat type, background noise, vocalisation directionality and the target species (Pérez-Granados & Traba, 2021). While it may be possible to remotely model sound attenuation and signal-specific ARU detectability radii, it still remains impossible to determine whether two calls were made by the same individual with a single-channel recorder (Stevenson et al., 2015).

Localising individuals to measure abundance and density is cited as one of the main advantages of performing sound source localisation in ecosystem monitoring (Rhinehart et al., 2020). A limited number of studies have begun using spatial acoustics to population size in a way that echoes point counts based on the spatial human auditory system (Frommolt & Tauchert, 2014; R. W. Hedley et al., 2017; Spillmann et al., 2015; Wilson & Bayne, 2018). Hedley et al., 2017 used four microphones connected to two Wildlife Acoustics SM3s to test whether the location and number of simultaneous calls could be determined at varying distances from a recorder, finding that 95% of calls were detected and localised to within 12°. This setup, unfortunately, suffered some issues with inter-ARU clock drift and required regular (every 5 minutes) sonic re-synchronisation. To my knowledge, this setup has not been tested for long-term deployment in a more complex environment. Other studies have also had some success monitoring abundance with multichannel acoustics; however, these methods were applicable to just single species (Bitterns (Frommolt & Tauchert, 2014), Bornean Orangutans

(Spillmann et al., 2015), and Ovenbirds (Wilson & Bayne, 2018)). Individuals were detected in these studies using automated detection of a loud and distinctive long-call (Bornean Orangutan) or through visual inspection of spectrograms (Bitterns and Ovenbirds) which can be time-consuming.

The first aim of the following chapter is to provide a low-complexity, easily deployable use case of multichannel recording for unmanned signal clustering and ubiquitous bird species abundance approximation.

4.2.4. Vertical Stratification and the 3D Niche

Species distributions and spatial niches are traditionally considered in two spatial dimensions (2D) distributions across the Earth's surface and have long negated the vertical dimension (Chandler et al., 2020). Research has found, however, that a habitat's structural complexity (in three spatial dimensions 3D) is a better predictor of biodiversity than canopy cover alone (Davies & Asner, 2014). The concept of the 3D niche was recently defined as "an extension of the classic ecological niche concept which considers spatial, temporal and dietary specialisations in a complex, 3D habitat" (Gámez & Harris, 2022). The 3D niche considers a new range of explanatory opportunities for species and may even be necessary for explaining the coexistence and persistence of competing species (Oliveira & Scheffers, 2019). The 3D niche relies on the theory of vertical stratification, whereby the vertical layering of biotic and abiotic environmental conditions can lead to differences in biotic community assemblages.

Vertical stratification has been considered since the 40s (Allee, Park, Emerson, Park, & Schmidt, 1949) and has been evidenced in terrestrial, aquatic and subterranean environments (Thiel et al., 2021). Terrestrial forests are especially complex, sometimes with six vertical strata: soil, forest floor, herbaceous shrub, tree, and up to two emergent layers (Allee, Park, Emerson, Park, & Schmidt, 1949). The layering of vertical strata is due to gradient (light, temperature, humidity, pH, nutrient availability) and more discrete biotic and abiotic factors (substrate, habitat provisions, herbaceous structure). These variations can result in a vertical environmental turnover which occurs at a much faster rate than horizontal; in tropical forests, 20-30m vertical differences can result in climatic differences of 4°C and 4% humidity, equivalent to over 400m in altitude or kilometres in latitude (Scheffers et al., 2013).

Beyond just abiotic factors, vertical space provides variation in refugia for intra-/ inter-specific competition and predation pressure. Greater availability of niches in the vertical dimension could reduce the intensity of competition, which fundamentally shapes communities (Tilman, 2004). Species' ability to move vertically to buffered conditions within (micro) habitats may also allow them to tolerate a wider range of stressors, including temperatures in new climate conditions (Scheffers et al., 2013, 2017; Thiel et al., 2021). The combined effect of differential abiotic and biotic pressures and opportunities has resulted in the vertical stratification of many terrestrial animals' ranges, specifically documented in insects (Leahy et al., 2022), birds (Bradfer-Lawrence et al., 2018), and mammals (Chandler et al., 2020). A better understanding of how species interact with 3D, vertically stratified ecosystems may enable us to understand how ecosystem functions will perform under global change (Thiel et al., 2021).

Habitat threats such as selective logging, habitat loss, and fragmentation are driving species and biodiversity loss globally. When habitat is lost, species may respond by moving vertically or latitudinally, extending the extent of the threat. Practices such as mixed fruit orchards, keeping mixed strata of tree species within a plantation, agroforestry, canopy highways, and 3D structured corridors all maintain biodiversity by maintaining vertical stratification (Gámez & Harris, 2022; Harvey & González Villalobos, 2007; Round et al., 2006; Thiel et al., 2021). Tropical plantations, which are built with consideration of structural complexity, provide corridors and habitats for a larger number of species than those which do not (Harvey & González Villalobos, 2007; Round et al., 2006; Thiel et al., 2021). Without these kinds of considerations, unchecked deforestation pressures may disproportionately collapse 3D niches, lead to a higher degree of competition, and risk accelerating species loss (Levin, 1970).

Despite being considered in the 1950s, vertical stratification was not dedicatedly explored until canopy access became more widely available in the 1970s (Thiel et al., 2021). Since the 1970s, information about canopy-dwelling species has largely been derived through canopy access, cranes, and ground-to-canopy mist nets (Thiel et al., 2021). Even now, monitoring efforts are often conducted solely from the understory, which can result in an underestimation of community resilience as a larger number of more mobile species have been observed in the canopy (Bradfer-Lawrence et al., 2018). As the availability of small, affordable computers has exploded, so too has the means of collecting data from the field autonomously. Technologies capable of monitoring 3D niches were recently reviewed (Gámez & Harris, 2022), advocating for the use of high-resolution forest structure renderings (Hermosilla et al., 2014), multi-stratified camera traps (Moore et al., 2020), elevation-enabled GPS collars, and stratified eDNA measures via mosquitos (Gámez & Harris, 2022). Acoustic methods remain unmentioned in the review and, while underutilised in this area, provide some advantages (non-invasive, ubiquitous, low cost) that limit the above.

The second aim of this chapter is, therefore, to determine whether it is possible to explore vertical stratification with MAARU.

4.2.5. Chapter Aims

The overall aim of this work was to develop and appraise a means of autonomous multichannel acoustic recording for application to field research. In this chapter specifically, I explore whether data from MAARU (Chapter 3) can be used to:

(1) Use spatial localisation to give a more accurate indication of abundance and calculate relative biodiversity indices.

(2) Appraise vertical space usage of avian species in two different areas of managed forest

The study site consists of vertically stratified Ancient Semi-Natural Woodland (ASNW) and canopy-dominant Plantation on Ancient Woodland Site (PAWS) in Southeast England. This study presents an analysis of the first field data recorded by MAARU (Multichannel Acoustic Autonomous Recording Unit), a single multichannel field recorder capable of extracting the relative directions of acoustic signals (Chapter 3).

4.3. METHODS

4.3.3. Study Site

Two MAARU recorders were deployed in two managed woodland areas in Southeast England, Chalk Wood (Bexley Council, London) and Joydens Wood (The Woodland Trust). MAARU is a 6-channel (6 x 16kHz, 16-bit) autonomous acoustic recorder made the low-cost, ubiquitously available equipment (Chapter 3). The two field sites are geographically close; however, variations in the dominant tree species lead to clear forest structure differences.

(1) Chalk Wood is an Ancient Semi-Natural Woodland (ASNW) area primarily used as a bridleway with relatively low footfall. The ASNW site is made of mature broadleaf species, including sweet chestnut, oak, sycamore, ash, and field maple, with holly and yew comprising the shrub layer (Woodland Trust 2021). One MAARU recorder (ASNW1) was attached to the trunk of a large Oak Tree 15m from the ground.

(2) Joydens Wood is a popular public woodland mainly consisting of Plantation on Ancient Woodland (PAWS). Joydens wood was converted from natural oak woodland to Corsican Pine after the Forestry Commission gained ownership in 1956. In 1993, the Woodlands Trust bought the site, and despite felling efforts over the last 30 years, large areas of the woodland are still predominantly conifer forest (Woodland Trust, 2021). Currently, the PAWS sites in Joyden's Wood consist of a dense canopy of evergreen

Corsican Pine and a largely open understory with a few mixed ferns, several small Holly bushes, and some emerging saplings (Woodland Trust, 2021). The other recorder (PAWS1) was placed on the trunk Corsican Pine tree, also 15m from the ground.

Both recorders were MAARU recorders (Chapter 3), powered by solar panels attached by climbing slings in the canopy and programmed to transmit. FLAC compressed audio data continuously over mobile networks. As a result, the specific recorder locations were limited mainly to areas with sufficient connectivity to mobile networks (5 Mbps, 10 Mbps if uncompressed audio streamed - found via speed checker).

The deployment period was continuous between August 2021 and January 2022. However, inefficiencies in the power supply resulted in recordings that were not continuous over the experimental period. For standardisation, only days where over 16 hours of audio had been collected were used for testing. These recordings were manually screened to remove days where ongoing forest work occurred. The final study set consists of four 10-hour periods of continuous acoustic recording at each site (10 am-10 pm, 80 hours total).

4.3.4. Data Extraction

Beyond raw audio recording, the data used in the following analysis was extracted through two steps (Figures 4.1 and 4.2).

1- The DOA and species information was determined through HARKBird, a sound-source localisation programme (Suzuki et al., 2017), and BirdNET, a bird call classification platform (Kahl et al., 2021), respectively (Figure 4.1).

2- The localisation and species information is then used to approximate the species-specific calling rate and the number of calling individuals per recording (Figure 4.2).

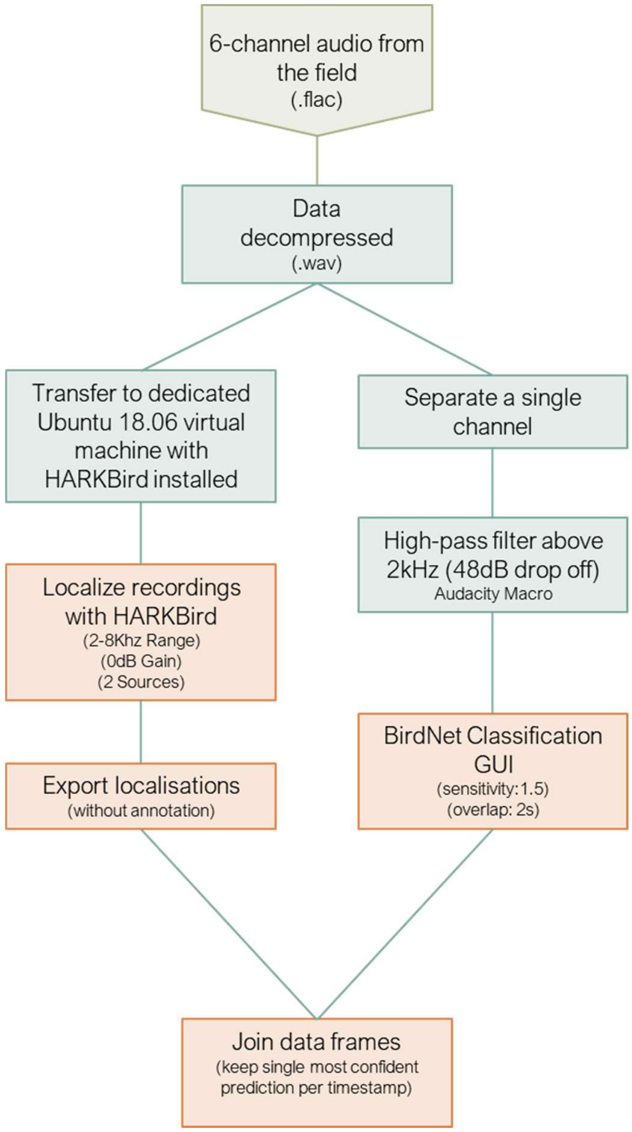


Figure 4.1 shows the steps in extracting localisation and species identifications from raw acoustic recordings, with essential software and settings for each step. Notably, as the detection range was limited to 2-8kHz in HARKBird, this was echoed by using a High-Pass filter prior to BirdNET classification.

4.3.4.1. Sound Source Localisation with HARKBird

HARKBird, explained in more detail in chapter 3, is a bioacoustic adaptation of software designed for machine listening: HARK (Honda research institute Japan Audition for Robots with Kyoto University (Nakadai et al., 2010, 2017)). HARK is an open-source platform which combines modules capable of speech detection (Missing Feature Theory), noise-robust localisation (MUSIC), and separation (GSS) and is designed to work on any microphone array in both quiet and noisy environments (Nakadai et al., 2010). HARKBird is a portable, customisable implementation of HARK with a simple GUI that allows users to record, localise, separate, visualise and annotate (environmental) sounds (Suzuki et al., 2017), with later implementations also performing T-SNE-based signal clustering (Sumitani et al., 2019).

HARKBird determines the calling azimuth (DOA) of a signal relative to a fixed array. MAARU recorders use a 6-channel 2D circular microphone array, so in my study, HARKBird localises recordings to the most likely 5° azimuth on a 360° circle. Prior lab tests, however, found that after waterproofing and extended deployment in the field, signals recorded by MAARU recorders could be localised to an accuracy of ±10° (Chapter 3).

HARKBird 2.0 was used in this study and was installed on a dedicated Ubuntu 18.06 Virtual Environment. As data was collected on-device and annotation will be performed by BirdNET, I used HARKBird solely as a localisation tool. When running localisations on HARKBird, the user should set several parameters, which I determined experimentally by manually inspecting inputs and outputs (Table 4.1). All below features were altered systematically until HARKBird localised only bird calls. Notably, the detection range is small (2-8kHz) because the sample rate of the recorder was 16kHz, and <2kHz was removed as this band was dominated by anthropophony from a nearby busy A-road and overhead air travel.

Table 4.1 shows parameters used for HARKBird in PAWS and ASNW

Parameter	ASNW1	PAWS1
Event Detection Threshold	24.75	25.25
Frequency Range	2-8kHz	2-8kHz
Predicted Source Number	2	2
Gain	0dB	0dB
Array	6-channel circular	6-channel circular

4.3.4.2. Automated Species ID with BirdNET

BirdNET is a dedicated deep neural network with 157 layers and more than 27 million trainable parameters that have been extensively trained on 226,078 labelled bird recordings (comprising 984 species from eBird, Xeno-Canto and the Macauley Library) and 15,000+ recordings of "non-bird noise" (found from AudioSet, Freefield1010, WarblR and the Macauley Library) (Kahl et al., 2021). Beyond this raw audio, augmented (warped, rolled, and artificially noisy) versions of the above were used to make the model more robust to differences in field recording protocols (Kahl et al., 2021).

The BirdNET DNN is based on a ResNet which takes spectrograms with enhanced temporal resolution (beyond human listening capacity) as input (Kahl et al., 2021). BirdNET boasts automated detection of over 70% of signals with accuracy reaching up to 100% (R. Hedley & Bayne, 2020). Its applications are widespread within academia, but BirdNET also lends itself to citizen science, with more than 1.1 million participants in 2020 (Wood et al., 2022) and over 1 Million downloads on Google Play. All recordings in this study were classified with the same BirdNET parameters (Longitude = 0.15, Latitude = 51.4, Week 34, Sensitivity = 1.5, and 2s overlap), with the outputs saved as .csv.

Only BirdNET classifications with a confidence level of over 0.5 are used in the following analysis. Further, species detected just once across the entire recording period were removed as potential false positives.

4.3.4.3. Abundance Approximation with DOA Estimation

In omnidirectional soundscape recording, it is often impossible to assign ownership to calls meaning, particularly for frequently calling individuals, that it can be impossible to attribute any measure of abundance. Here I define an experimental method for using DOA estimation to cluster calling azimuths and approximate the number of calling individuals from a soundscape (Figure 4.2).

The joined data frame containing the combined HARKBird and BirdNET outputs is split per recording and per species. This results in a list of DOAs of signals for a particular species over a (10-minute) time window. The DOAs are clustered optimally using the smallest number of k-means clusters, whereby the most extensive range of azimuths in each cluster is 20° (±10°) (Figure 4.2). The number of calls and average azimuth per group is then found and mapped from 2D (circular) azimuth to a pre-determined 1D height projection. These groups are used as an approximation of the number of calling individuals. Unless otherwise stated, all per-species analysis that is based on abundance has been estimated through this method (hereon DOA-approximated abundance).

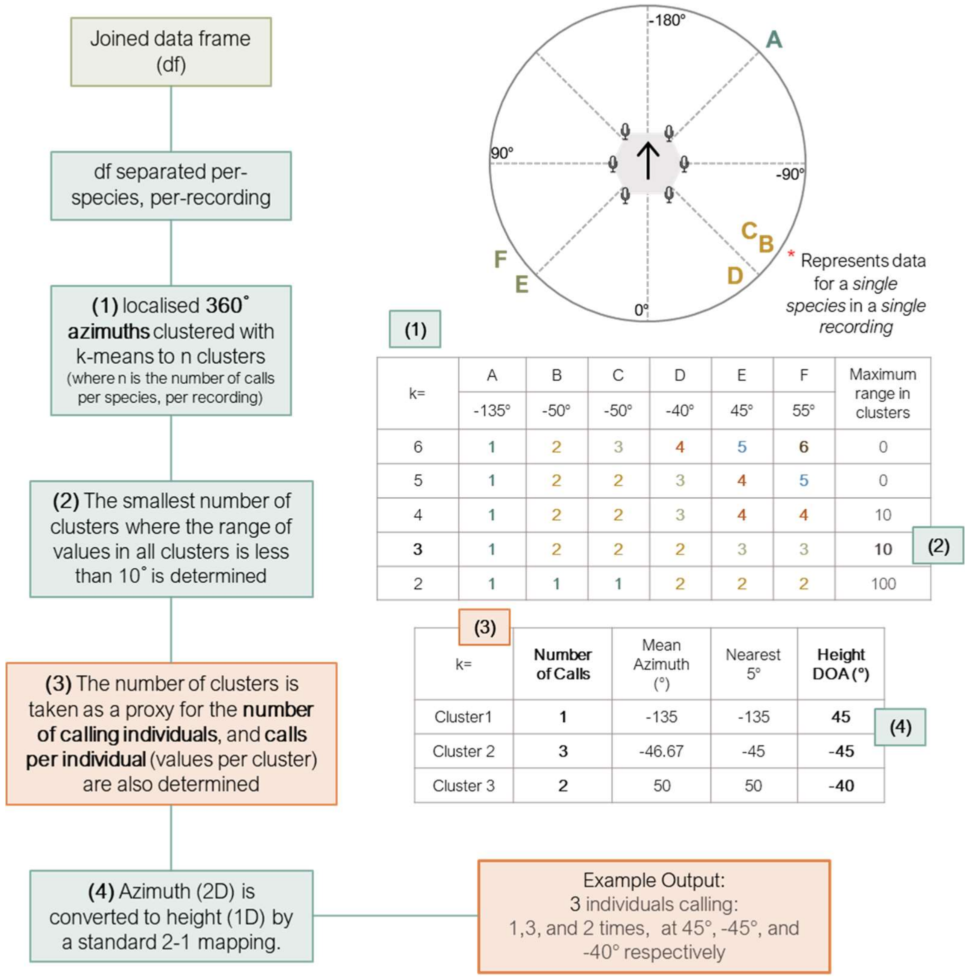


Figure 4.2 illustrates an example of the process of estimating abundance with spatial localisation.

4.3.5. Data Analysis

The following data analysis and visualisations are completed in R v.4.0.3 (R Core Team, 2022). These analyses work to highlight the advantages of using MAARU or other DOA recording units for enhanced environment appraisal. Here I highlight how DOA acoustics can be used to approximate species abundance and explore the calling locations of species.

4.3.5.1. Abundance Approximation

To demonstrate whether DOA estimation could be used to better approximate the abundance and calling rate of individuals, I compared the total number of detected calls per species per location with DOA-approximated abundance (Figure 4.1). The number of detections was compared with the DOA-approximated abundance through a Paired-Wilcoxon Signed-Rank test.

4.3.5.2. Diversity Indices

I use measures of site-specific species count (α diversity), site-specific unique species count (β diversity), and cross-site species count (γ diversity) as a first-point comparison across the sites.

I then calculate the Shannon Diversity at each site, comparing the use of DOA-approximated abundance and raw call volumes as proxies for species count. Shannon-Diversity measures the number of species and evenness of the population of each species in an area. I calculated Shannon Diversity with the diversity method in the “vegan” package (Oksanen, 2022). Shannon Diversity is calculated with samples determined hourly at each site. The difference in Shannon Diversity at each site, for each of the measures (DOA approximated abundance vs call volume), is compared pairwise through Welch’s T-Tests.

4.3.5.3. Calling Rate Approximation

The DOA- approximated abundance method records an estimate of the number of calling individuals but also the number of calls attributed to that individual. I show the median and quartiles of the number of calls per DOA-approximated individual per species. Using a one-sample Welch’s T-Test, I determine the species where individuals call more than once over a 10-minute period and therefore indicate the species most likely to be overestimated in just call-volume-related abundance approximations. I used a Bonferroni Correction of 19 species, making the p-value for 95% confidence $p<0.00263$.

4.3.5.4. Vertical Space Usage

I split the vertical space, depending on the DOA (θ), into the upper canopy (30° < θ ≤ 90°), mid-canopy (-30° < θ ≤ 30°) or understory (-90° < θ ≤ -30°) and describe species calling in each stratum. We also use a two-way ANOVA was used to test whether the DOA of a DOA—approximated individual can be predicted by which species it is, which location it's calling from or whether there is an interaction between the two.

4.4. RESULTS

4.4.3. Abundance Approximations

A total of 5933 calls were detected, localised, and classified over the 80-hour recording period (ASNW: 3240, PAWS: 2693). It is estimated that these calls came from 669 individuals in ASNW and 331 individuals in PAWS (1000 total). The PAWS recorder collected 45.4% of detections. However, it is estimated that this was from 33.1% of the total detected DOA—approximated individuals. Suggesting that PAWS may be made up of fewer, more frequently calling individuals

However, I found that the calculating abundance reduced the mean number of independent detections in ASNW (Wilcoxon Signed Rank Test: V = 120, p=0.0007) and PAWS (Wilcoxon Signed Rank Test: V = 91, p=0.0017) (Figure 4.3).

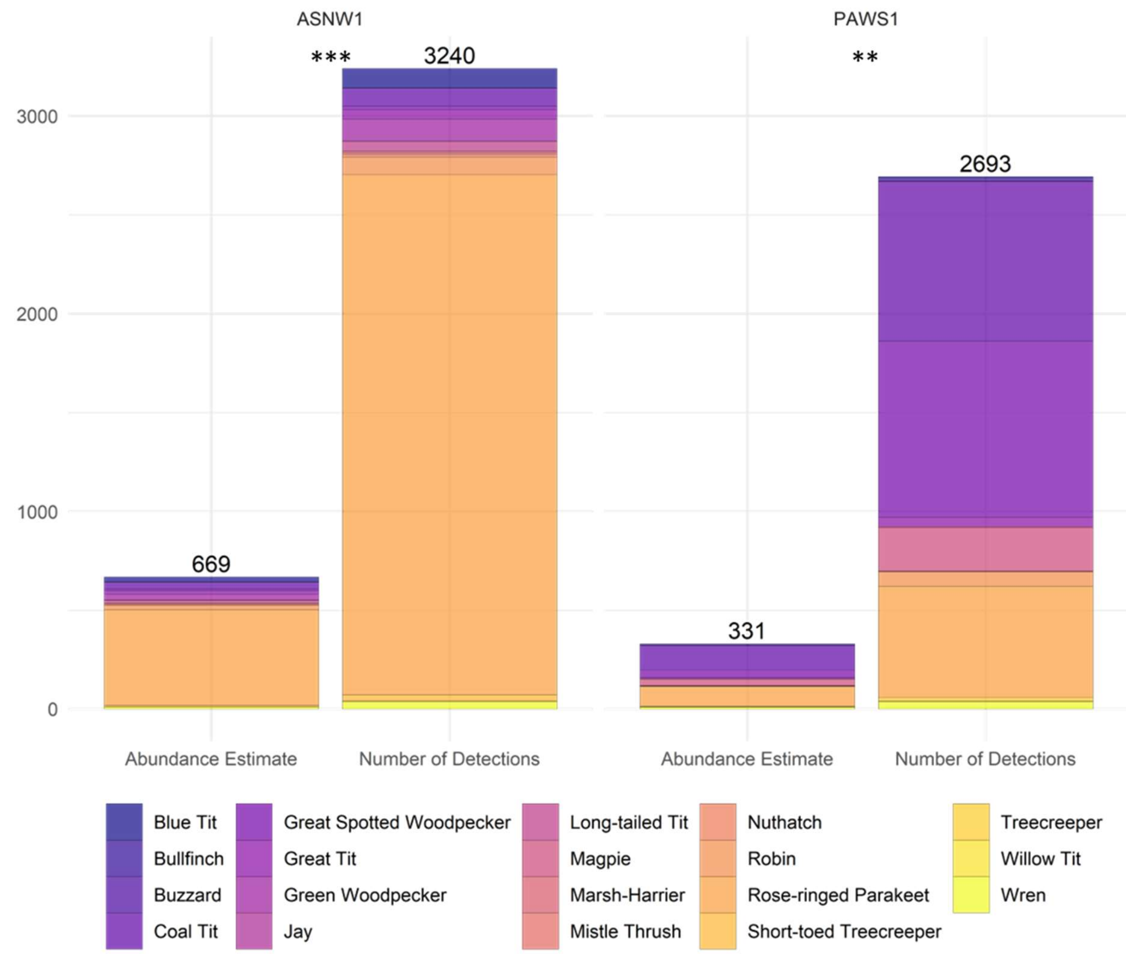


Figure 4.3 shows the overall number of detections and approximated abundance in both ASNW and PAWS. Abundance approximates are less than the number of detections in both sites.

4.4.4. Diversity Indices

ASNW had an α diversity of 16 (including three Red and Amber list species), while PAWS had an α diversity of 15 (including four Red and Amber list species). β diversity across the sites is seven species (four unique species in ASNW and three in PAWS). The γ diversity across both sites is 19 (Figure 4.3).

The Shannon Diversity in ASNW was larger than PAWS when both abundance (t(60.9)=-2.93, $p<0.0001$) and call (t(64.8)=3.51, $p<0.00081$) measures were used. The Shannon Diversity Index was also reduced when abundance was used in place of the number of calls in both ASNW (t(71.9)=-4.68, $p<0.0001$) and PAWS (t(59.5)= -2.94, $p<0.0048$) (Figure 4.4).

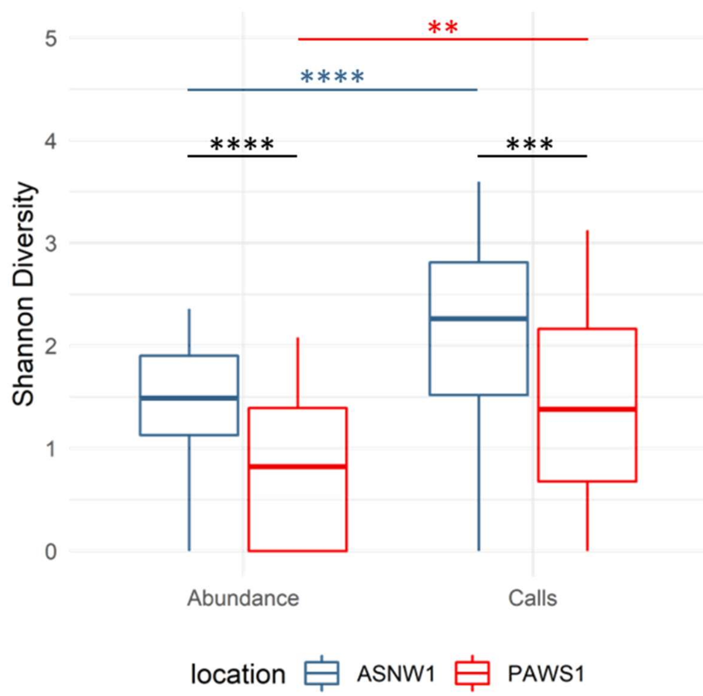


Figure 4.4 shows the Shannon Diversity Index (calculated hourly) at ASNW and PAWS, where either DOA-approximated abundance or call volume is used to count individuals. The Shannon Diversity is higher in ASNW than in PAWS, and using call numbers overestimates the Shannon Diversity when compared to DOA-approximated abundance.

4.4.5. Calling Rate Approximations

There was considerable variation in the number of calls made by individuals of different species. The largest number of calls made by a single individual over the course of 10 minutes was found in Great Spotted Woodpeckers (340 calls), Rose-Ringed Parakeets (63 Calls), and Robins (62 Calls). Considering detections across the entire experimental period, I found that six species call more than once (Table 4.2, Figure 4.5 – Full Details Appendix C: Table 3.1).

Table 4.2 shows the one-sample Welch's T-Test outputs comparing the calling rates of individuals across both recording locations (just significant findings).

Species	Mean	DF	T-Value	Raw P	Bonferroni Significance (p<0.00263)
Blue Tit	3.92	26	4.45	0.000145	*
Coal Tit	5.64	158	9.35	< 0.0001	*
Long-Tailed Tit	3	16	5.50	<0.0001	*
Magpie	6.36	35	4.94	<0.0001	*
Rose-Ringed Parakeet	5.50	580	17.13	<0.0001	*
Wren	3.57	20	4.25	0.000384	*

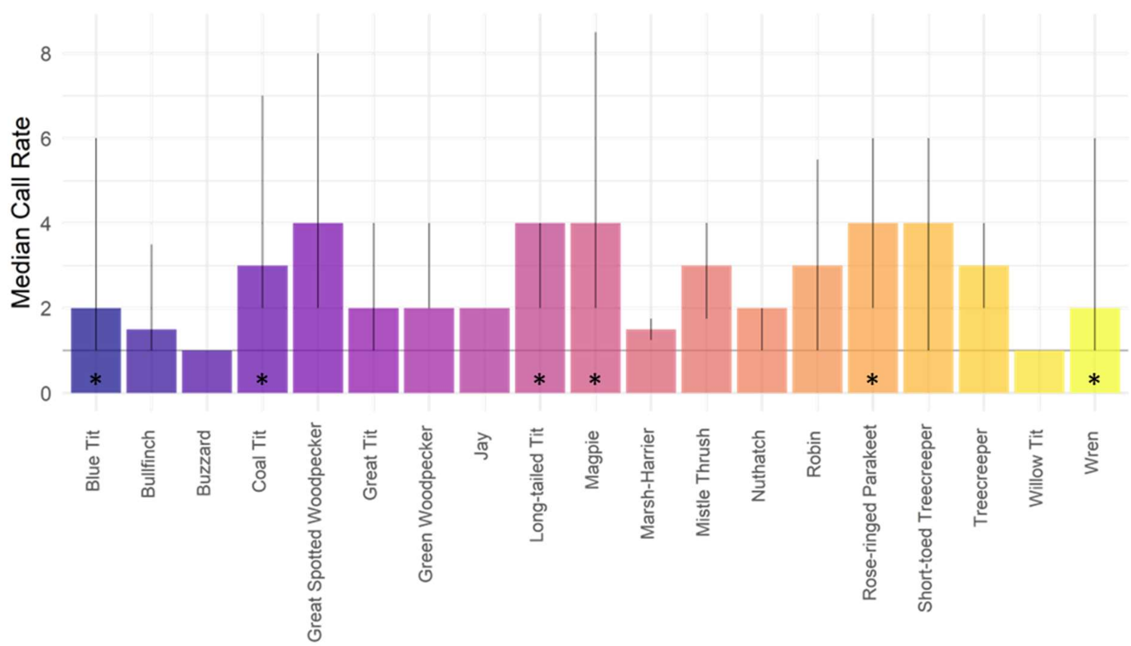


Figure 4.5 shows the per-species calling rate with median and inter-quartile ranges. The species found to call more than once (Bonferroni Corrected Welch's T-Test, 95% Confidence) are indicated by stars.

4.4.6. Vertical Space Usage

Of the 16 species in ASNW, six had median calling angles in the upper canopy (Blue Tit, Coal Tit, Jay, Magpie, Robin, and Treecreeper), one species had a median calling height from the understory (Great Spotted Woodpecker), and the rest from the mid-canopy. Of the upper canopy species, three were *only* detected in the upper canopy (Jays, Magpies and Treecreepers). Buzzards were only detected in the mid-canopy, and Nuthatches did not call from the understory, but the other 11 species were found to call from all three strata (Figure 4.6).

Conversely, all species in PAWS have median calling angles in the mid-canopy except for Nuthatches which were only detected in the upper canopy. Three species in PAWS were only detected in the mid-canopy (Buzzards, Long-tailed Tits and Willow Tits). Compared to the 11 in ASNW, only six species called from all three strata in PAWS. With less than half of the species (seven of fifteen) calling from the understory at all (Figure 4.6).

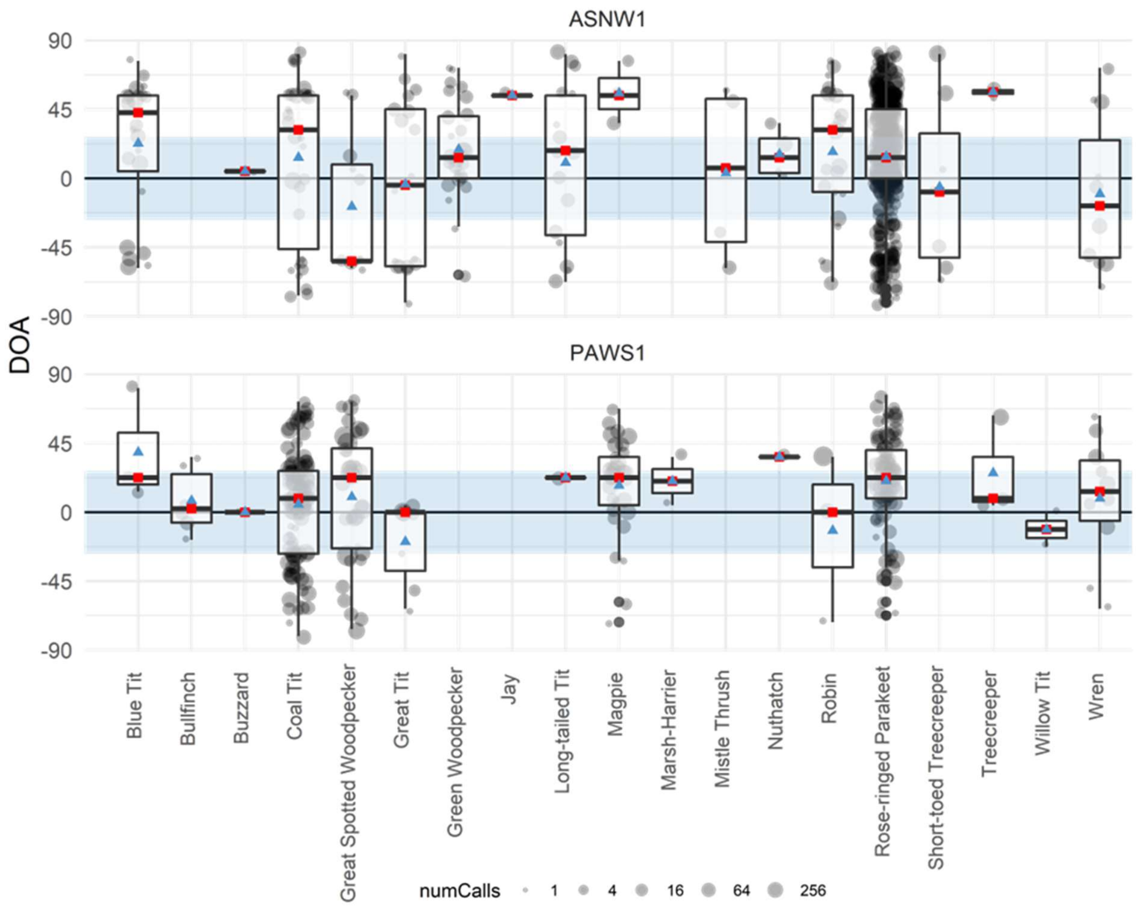


Figure 4.6 shows the calling angles of all detected DOA—approximated individuals in the experimental period. The number of calls made by each DOA-approximated individual is given by the point size. The shaded area between -30° and 30° represents the mid-canopy; above this is the upper canopy, and below is the understory. Boxplots indicate the range, outliers, and interquartile ranges of the calling angles of each species at each location. Means are indicated by red squares and medians by blue triangles.

Considering the overall dataset, a two-way ANOVA revealed that there was not a statistically significant interaction between species and location on the calling angle ($F(11) = 1.40$, $p=0.17$, Table 4.3). The analysis shows that the location alone did not have a statistically significant effect o the calling angle ($p=0.48$, Table 4.3), but that species did ($p=0.03$).

Table 4.3 Two-Way ANOVA output for DOA ~ Location Species*

	Df	Sum Squared	Mean Squared	F-Value	Pr(>F)
Location	1	896	895.8	0.498	0.4807
Species	**18**	**55866**	**3103.7**	**1.724**	**0.0303 ***
Location x Species	11	27690	2517.3	1.399	0.1677
Residuals	969	1744158	1800.0		

4.5. DISCUSSION

The pipeline I present details data extraction and provides the scripts necessary to combine raw outputs from MAARU, BirdNET and HARKBird to approximate abundance and explore the use of vertical space. Sound-source localisation in the field usually requires devices to be precisely time-synced over GPS, wireless networks, or cable networks, making localisation expensive, unreliable, and occasionally faulty (Rhinehart et al., 2020). This pipeline requires just one relatively low-cost device, all synced by the same onboard clock, removing the need for cross-recorder synchronisation. While using just one DOA recorder sacrifices the ability to locate the exact location of a calling individual, it should not be overlooked as a low-cost, easily deployable means of enhanced ecosystem recording.

I then explain using a simple k-means clustering optimisation process to group species-specific calls by DOA as a proxy for the number of calling individuals (the DOA-approximated abundance). Here I was able to provide insight into species-specific calling frequency and highlight individuals that may be under-/over-estimated in the population. Bird populations in the UK are fortunately very well documented, but this method can be extended to systems which are not so well understood. Call classification in terrestrial sound ecology is generally just used for occupancy/ range assessments, as attributing ownership to a signal is often impossible. Without knowing how many individuals are present, it is difficult to estimate population size, density or abundance. The number of individuals can be approximated from omnidirectional recording, but this is especially difficult for species without well-documented information on signal attenuation, signal directionality, or calling behaviour. My DOA-approximated abundance can be applied to any species, with the only requirement being that they can be classified with BirdNET (or similar). This DOA-approximated abundance estimate should not be taken as an absolute value of abundance but rather as a more accurate proxy. This abundance estimate is based on spatial differences over a fixed recording window; however, extending this into the temporal domain may develop this approximation to account for moving individuals and individual estimates across multiple recording windows. Further, the approximation also uses an experimentally derived localisation error of ±10° as a maximum range for clusters, which would under-estimate abundances for close by individuals. While helpful in separating individuals over a short time window, this approximation should be used cautiously and may be developed.

I found a similar number of species at each site (15 and 16 for ASNW and PAWS, respectively). It is possible that the species-specific detection radius at these sites differed, as sound often attenuates more quickly in more densely forested environments such as ASNW (Alcocer et al., 2022). Studies like this may benefit from determining the attenuation profile of the landscape before analysis. I found that for both PAWS and ASNW, using the call volume as an abundance measure caused an overestimate of Shannon Diversity than the DOA-approximated abundance. ASNW has a marginally higher Shannon Diversity, no matter which of the abundance measures was used. It should be noted that this refers only to successfully detected and classified audible-range bird calls. Any animals that vocalise outside of this frequency range or cannot be accurately detected by BirdNET will not be included in this analysis.

At both locations, Rose-ringed Parakeets drove un-even populations between species. In ASNW alone, 2630 Rose-ringed Parakeet calls were detected from over 480 individuals. Rose-ringed parakeets are an invasive species in the UK, which are found to have detrimental effects on native wildlife and socio-economic activities, such as crop damage (Heald et al., 2020). Rose-ringed Parakeets are famously noisy, with exposure to their loud and frequent calls driving a sharp decline in human tolerance of their invasion (Mori et al., 2020). Whether this domination of the acoustic space has an impact on the vocal behaviour of other species has been speculated but is unknown at this time (Peck, 2013). As just one recorder was deployed at each site, it may also be possible that this data may be skewed if the recorder was placed near a Rose-ringed parakeet (or other species) roost. The reduced dispersal of Rose-ringed Parakeets in PAWS, combined with the unique occurrence of Bullfinches, Marsh-Harriers, and Willow Tits, may further evidence the need for heterogeneity of landscape type for maximising species richness.

The data collection and processing pipeline had a few limitations regarding frequency range, species classification scope, and near-field DOA calculation. Due to a large amount of anthropogenic noise in the area, all data below 2kHz was removed. While this has benefits regarding privacy as this removes human speech (80-255 Hz), I also lose data from species that call in a lower range, such as Wood Pigeons and Collared Doves. This study and suggested analysis pipeline are also limited in terms of its species scope it just describes bird species classification and is by no means comprehensive. That being said, birds are vital indicator species owing to their sensitivity to environmental change and contributions to ecosystem services (Bradfer-Lawrence et al., 2018, and references therein). Should researchers want to extend this analysis to other taxa, frequency ranges can be altered, and other taxa-specific classification algorithms can be included. Finally, DOA estimation works well at a distance beyond a couple of metres, but as individuals approach the MAARU device in the mid-canopy, any call made in the mid-canopy but directly above/ below the MAARU device would register as a call in the upper canopy/ understory. Incorporating species-specific attenuation information into the pipeline would not only resolve this but also approximate a distance and direction. This attenuation information may also allow researchers to perform a crude positional localisation without requiring cross-device synchronisation. To do so would be an exciting and incredibly useful application of this work but is beyond the scope of this chapter.

Further, the analysis used in this pipeline are dependent on the accuracy of the associated software. While HARKBird worked well in lab tests, it is unknown how it would fare in a noisier environment such as a woodland. Similarly, bird detections are dependent on BirdNET. No ground-truth data was available in this study, so a degree of approximation was required. As such, I opted to use a somewhat conservative confidence level with BirdNET (0.5). However, some surprising species were still detected (Marsh Harrier and Short-Toed Treecreeper). Confirming or rejecting these detections requires the contribution of an expert listener, which does somewhat divert from the autonomy of the process. That being said, it would also be possible to avoid this kind of immediate inaccuracy with the use of a custom bird list built upon citizen science observations from eBird or similar.

I used MAARU and the suggested pipeline to also explore the use of vertical space between species and locations, determining the different calling heights of species at either location. The examples I present in this analysis are crude but evidence of the usefulness of DOA estimation. The positional information from a single DOA recorder could also be applied to many novel investigations, including investigating species movement patterns, species interactions, communication networks, predictors of vertical space usage (such as diet/ size etc.) and perhaps a temporal variation of vertical space usage.

4.6. CONCLUSION

In this chapter, I aimed to investigate whether data collected by a single DOA recorder (MAARU, Chapter 3) could be used to approximate species abundance and appraise the use of vertical space. Over the course of this chapter, I: (1) Document a pipeline for call detection, localisation, and classification using a novel combination of emerging technologies. (2) Approximate species abundance through novel calling azimuth clustering. And (3) Evidence using a single DOA recorder as a stand-alone means of enhanced acoustic monitoring, exploring the behavioural dynamics of 19 bird species – all without the need for any manual data annotation.

4.7. DATA ACCESSIBILITY

Data Available Upon Request

Code Available at: https://github.com/BeckyHeath/DOA_HeightTracking

5. THESIS CONCLUSION

5.1. AIMS, OUTCOMES AND LIMITATIONS

This thesis aims to assess and develop methods in (spatial) sound ecology. I began the thesis by exploring the founding principles of sound ecology, exploring the necessity of research in this field and identifying key literature gaps. Through this current state-of-the-art review, I highlighted three important gaps in the literature and potential challenges in the field. These gaps were data storage limitations, low-cost spatial acoustics, and underexplored applications of single-device localisation. As the thesis concludes, I will explore how these gaps were addressed over the course of my PhD and some of the critical limitations.

(1) Aim 1: Investigate Whether Data Saving Practice in Ecoacoustics Affects the Quantification of Soundscapes

Here I explored two data-saving practices: lossy MP3 compression and temporal down-sampling, and their effects on Analytical Indices and a CNN-derived AudioSet Fingerprint. In this chapter, I provided detailed patterns in index alteration as a result of compression and showed that the AudioSet Fingerprint was less susceptible to variations in experimental protocol than Analytical Indices. Considering just compression, the conclusions I drew largely agreed with other works that looked specifically at the impact of compression on targeted calls (Araya-Salas et al., 2019; Stowell & Plumbley, 2014), except on a whole soundscape scale.

While I had initially wanted to assess this practice as I suspected it might be doing more harm than people expected, I was surprised and somewhat relieved to see that was not the case. While this provides evidence that mp3 compression (particularly low rates) can reduce data storage and transmission costs, these results should be treated with some critical caveats.

Firstly, these results are only relevant for audible range surveys, which immediately exclude many environmentally essential species from the analysis. Secondly, all acoustic data in the study was from one field site, and it is unclear whether these results are transferrable to other systems. Perhaps these results evidence that mp3 compression does not automatically render acoustic data redundant but rather that if you test for the use case, you may find it is possible to draw the same conclusions with compressed and uncompressed data.

This work has also added to the increasing body of literature advocating for the use of DL-based soundscape appraisals, owing to the evidence we found of its resilience to experimental changes. The DL methodology we used in this study was the AudioSet Fingerprint from the VGGish CNN (Gemmeke et al., 2017; Hershey et al., 2017). However, there have since been several key DL developments that may pose an advantage even over AudioSet. Primarily, BirdNET, an rNN species classification algorithm (Kahl et al., 2021), was recently released and has been met with increasing uptake and may be one of the main routes of monitoring species in countries within its range. Exploring whether compression impacts BirdNET would be crucial as the algorithm's uptake becomes more popular. Similarly, it would be essential to consider the effects of data-saving practice on other DL algorithms such as wav2vec, NLP and others (Baevski et al., 2020; Stowell, 2022; Vaswani et al., 2017) as their use case in bioacoustics increases.

Ultimately the results from this study pose the idea that perhaps the use of compression is not as immediately damaging as previously thought. But this should not be taken as a ubiquitous go-ahead, especially as methods continue to develop.

(2) Aim 2: Develop an Autonomous Multi-Channel Acoustic Recorder Capable of Sound-Source Localisation

The second aim of this thesis was to develop a low-cost, easily accessible alternative to the available multichannel spatial recorders, with a vital aim of this device being fully autonomous.

The device itself, nicknamed MAARU, costs £205, excluding powering at the time of writing, under £20 more than an equivalent omnidirectional recorder (Sethi et al., 2018). Compared to broader literature, including non-autonomous systems, this costs more than an AudioMoth recorder (£75 (Hill et al., 2018)) but less than recorders offered by Wildlife Acoustics (£700 (Wildlife Acoustics , 2022)). Both of these recorders have the advantage of working "out of the box" but lack the advantages of autonomy and single-device localisation. So, while MAARU is not the cheapest acoustic recorder currently available, I feel the novel analysis and deployment strategies it offers far outweigh its cost.

Similarly, MAARU is somewhat intermediate in terms of ease of access. Specifically, it is not as easy to get running as a prebuilt recorder (AudioMoth/ Wildlife Acoustics) – but offers an advantage over specialist recorders developed just for science which relies on custom

MAARU provides a truly exciting next step in autonomous acoustic recording and spatial acoustics as a version that is more widely available than many of the previous models. MAARU has many advantages, which I describe in detail in the chapter, but it also has some important drawbacks. Not least, MAARU has some onboard issues with microphone power, mitigated but at the cost of an increase in false positives. Further, MAARU is currently limited to audible range acoustics (sample rate: 16kHz) and requires external effort to set up, which may be off-putting for less technically minded individuals or researchers looking to include ultrasound in their survey. Despite this, MAARU offers and relatively cheap and straightforward solution to an otherwise expensive and complicated problem.

Since completing MAARU, AudioMoth (one of the largest equipment providers in sound ecology) has launched a GPS-synced adaptation of their base device. GPS-synced AudioMoths are an exciting development in large array-based spatial acoustics. MAARU answers a slightly different question that GPS-Synced AudioMoths in that the exact locations of individual vocalisations are not as important as their relative directions. MAARU also has the advantage of true autonomy over AudioMoths but lacks the “out-of-the-box” setup of AudioMoths. It will nevertheless be interesting to see how extensively this tech is taken up in the field in the coming years as the true breadth of application of spatial acoustics is realised.

As may be expected from a somewhat low-cost microphone array, I also had some problems with the audio quality from MAARU. Not least is the sample rate and resultant dynamic range. MAARU is limited to 16kHz (8kHz dynamic range), which fails to cover even the full audible range. By comparison, most, if not all, other recorders sample at least CD quality (44kHz, 22kHz dynamic range) which covers the entire audible range. This should be considered carefully by those hoping to use MAARU recorders as immediately many taxa of animals will go undetected. As I previously discussed, MAARU has been my first step in low-cost, easy-to-use multichannel acoustics. As the technology continues to develop, more low-cost arrays may be released, potentially those which can record well into ultrasound.

As an open and customisable piece of equipment, MAARU is open to switched and updated components, which may open the device to exciting future developments. For example, ReSpeaker has an integrated audio interface and processing platform (the ReSpeaker Core and Core V2.0 (Seeed Studio , 2022)). This platform is a dedicated wi-fi-enabled audio interface which can be used with other dedicated microphone arrays which boast slightly higher sensitivity (-26dbFS compared to -22dBFS (Seeed Studio , 2022)). The Core has DOA algorithms, which could be used to run edge signal localisation.

The current setup is also limited to areas of wireless connection via internet routers or access to mobile networks (3/4G). Even my field deployment, which was at a field site near a residential area, had minimal areas of sufficient connection, which largely influenced where I could place the recorders. Deployment ranges would increase dramatically if devices could connect to the internet through other means, such as through widespread and low-latency satellites like Starlink (Starlink, 2022), low-power wide-area radio connectivity like LoRa (LoRa Alliance, 2022), or together via a MESH network. Demonstrating usability through these means would be a compelling use case – especially as a considerable amount of ecological fieldwork takes place in some of the world's most remote environments.

The applications of a device like MAARU are many and may be even more in future. In this thesis and associated resources, I provide instructions to create simple, cheap and field-tested means of doing sonic localisation in sound ecology. While this tech may benefit from development in a few areas, MAARU is a great starting point.

(3) Aim 3: Explore Whether a Single Recorder Capable of DOA Estimation Can Be Used To Approximate Abundance and Monitor Behaviour

Here I applied the findings and base analysis pipelines developed when lab testing MAARU to the field and began to explore the kind of additional information that MAARU could determine. Most importantly, here, I was able to provide a pipeline for spatially clustering detections for better abundance measuring alongside looking at what this meant for vertical space occupancy. This work was largely exploratory and did not compare directly to the wider literature. But offers some unique solutions to the questions of abundance and vertical acoustics.

To fully understand the dynamics of an ecosystem, it is important to determine populations alongside presence/ absence (Farr et al., 2019). Overestimating the number of individuals is a unique challenge to sound ecology. In in-person field surveys, counts are often determined by sight through the number of individuals of the same species present simultaneously. It is easy for a person with binaural hearing to distinguish between 3˚ and 10˚ depending on where exactly the sound has come from (Risoud et al., 2018). In that sense, researchers can easily approximate how many calling individuals there are for close-by detections even without seeing an individual. Similarly, camera traps require a line of sight, so the minimum number of individuals can easily be determined.

In sound ecology, particularly without directionality, the problem is more complex and requires researchers to factor in detection radius, and it is impossible to tell whether two calls from the same species are one individual calling twice or two individuals. While a coarse investigation into this, I was able to cluster a species' calls every 10 minutes and determine the number of 10˚ clusters in the recording as an approximator of abundance. This also allowed me to approximate the calling rate of species, which may be useful to the wider field. A fundamental limitation of this spatially clustered abundance is the lack of temporal dimension. The measures are taken every 10 minutes, but an exciting improvement would be to consider time more realistically as a continuous variable rather than 10-minute chunks.

Another key limitation of this work is the absence of ground-truth species richness and abundance data. While most detections were logical, the pipeline identified that two surprising detections (Marsh Harriers and Short-Toed Treecreepers) were present despite a somewhat conservative confidence level. Validating the results of this survey would require external input from expert listeners but would be vital to confirm the results of this exploratory analysis, so these results should be treated with caution. In the pipeline I present, I just use BirdNET as a classifying algorithm which does limit the scope of the data analysis. The MAARU recorder already has a limited dynamic range, but using BirdNET as the sole classifier also removes other audible-range vocalising species, such as mammals, frogs, and insects, from the analysis, which all play vital roles in forest ecosystems.

Altogether the pipeline represents a novel investigation into single-recorder spatial localisation and begins to unpick some interesting characteristics. We start to see some differences in vertical space use alongside differences in calling rates between species.

5.2. CONCLUDING REMARKS

Over the course of this thesis, I have quantified the previously unknown impact of commonly used experimental variation on soundscape quantification, vouching for compression as a realistic and valuable solution to many of the data management problems sound ecology researchers face. I provide permanent access to the data and scripts used in this analysis and have presented this work in multiple formats. In the latter chapters of this work, I present an exciting, low-cost, and completely open-source platform for users who wish to get set up with spatial acoustic recording quickly and easily. I provide comprehensive testing procedures, scripts and data alongside novel applications of DOA acoustics and defined analysis pipelines.

By enhancing research in sound ecology with these advances in data management and spatial acoustics, we can build an increasing wealth of nuanced understanding of some of the world's most sensitive ecosystems and ultimately aid in their protection.

6. REFERENCES

Alcocer, I., Lima, H., Sugai, L. S. M., & Llusia, D. (2022). Acoustic indices as proxies for biodiversity: a meta-analysis. *Biological Reviews*. https://doi.org/10.1111/brv.12890

Allen, M., Girod, L., Newton, R., Madden, S., Blumstein, D. T., & Estrin, D. (2008). VoxNet: An interactive, rapidly-deployable acoustic monitoring platform. *Proceedings - 2008 International Conference on Information Processing in Sensor Networks, IPSN 2008*, 371–382. https://doi.org/10.1109/IPSN.2008.45

Andrei, V. (2015). Considerations on Developing a Chainsaw Intrusion Detection and Localization System for Preventing Unauthorized Logging. *Journal of Electrical and Electronic Engineering*, *3*(6), 202. https://doi.org/10.11648/j.jeee.20150306.15

Araya-Salas, M., Smith-Vidaurre, G., & Webster, M. (2019). Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. *Bioacoustics*, *28*(1), 57–73. https://doi.org/10.1080/09524622.2017.1396498

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, *2020-Decem*(Figure 1), 1–12.

Baker, E., & Vincent, S. (2019). A deafening silence: A lack of data and reproducibility in published bioacoustics research? *Biodiversity Data Journal*, *7*. https://doi.org/10.3897/BDJ.7.E36783

Baudry, J., Bunce, R. G. H., & Burel, F. (2000). Hedgerows: An international perspective on their origin, function and management. *Journal of Environmental Management*, *60*(1), 7–22. https://doi.org/10.1006/jema.2000.0358

Bernard, K. S., & Steinberg, D. K. (2013). Krill biomass and aggregation structure in relation to tidal cycle in a penguin foraging region off the Western Antarctic Peninsula. *ICES Journal of Marine Science*, *70*(4), 834–849. https://doi.org/10.1093/icesjms/fst088

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., Mccowan, B., Ali, A. M., & Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, *48*(3), 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Boelman, N. T., Asner, G. P., Hart, P. J., & Martin, R. E. (2007). Multi-trophic invasion resistance in Hawaii: Bioacoustics, field surveys, and airborne remote sensing. *Ecological Applications*, *17*(8), 2137–2144. https://doi.org/10.1890/07-0004.1

Boers, N. (2021). Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation. *Nature Climate Change*, *11*(8), 680–688. https://doi.org/10.1038/s41558-021-01097-4

Boers, N., & Rypdal, M. (2021). Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21), 1–7. https://doi.org/10.1073/pnas.2024192118

Bohnenstiehl, D. R., Lyon, R. P., Caretti, O. N., Ricci, S. W., & Eggleston, D. B. (2018). Investigating the utility of ecoacoustic metrics in marine soundscapes. *Journal of Ecoacoustics*. https://doi.org/10.22261/jea.r1156l

Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., Teboul, O., Grangier, D., Tagliasacchi, M., & Zeghidour, N. (2022). *AudioLM: a Language Modeling Approach to Audio Generation*. 1–11. http://arxiv.org/abs/2209.03143

Boulton, C. A., Lenton, T. M., & Boers, N. (2022). Pronounced loss of Amazon rainforest resilience since the early 2000s. *Nature Climate Change*, *12*(3), 271–278. https://doi.org/10.1038/s41558-022-01287-8

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J. B., Kiszka, J. J., Kulbicki, M., Manel, S., Robbins, W. D., Vigliola, L., & Mouillot, D. (2018). Environmental DNA illuminates the dark diversity of sharks. *Science Advances*, *4*(5). https://doi.org/10.1126/sciadv.aap9661

Bowler, D. E., Buyung-Ali, L., Knight, T. M., & Pullin, A. S. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning*, *97*(3), 147–155. https://doi.org/10.1016/j.landurbplan.2010.05.006

Bradfer-Lawrence, T., Gardner, N., & Dent, D. H. (2018). Canopy bird assemblages are less influenced by habitat age and isolation than understory bird assemblages in Neotropical secondary forest. *Ecology and Evolution*, *8*(11), 5586–5597. https://doi.org/10.1002/ece3.4086

Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S. G., & Dent, D. H. (2019). Guidelines for the use of acoustic indices in environmental research. *Methods in Ecology and Evolution*, *10*(10), 1796–1807. https://doi.org/10.1111/2041-210x.13254

Brooke, M. D. L., Butchart, S. H. M., Garnett, S. T., Crowley, G. M., Mantilla-Beniers, N. B., & Stattersfield, A. J. (2008). Rates of movement of threatened bird species between IUCN Red List categories and toward extinction. *Conservation Biology*, *22*(2), 417–427. https://doi.org/10.1111/j.1523-1739.2008.00905.x

Browning, E., Gibb, R., Glover-Kapfer, P., & Jones, K. E. (2017). Passive acoustic monitoring in ecology and conservation. *WWF Conservation Technology Series 1*, *2*(October), 1–75. https://doi.org/10.13140/RG.2.2.18158.46409

Bruck, J. N., Walmsley, S. F., & Janik, V. M. (2022). Cross-modal perception of identity by sound and taste in bottlenose dolphins. *Science Advances*, *8*(20), 40–42. https://doi.org/10.1126/sciadv.abm7684

Bruggemann, L., Schutz, B., & Aschenbruck, N. (2021). Ornithology meets the IoT: Automatic Bird Identification, Census, and Localization. *7th IEEE World Forum on Internet of Things, WF-IoT 2021*, 765–770. https://doi.org/10.1109/WF-IoT51360.2021.9595401

Butchart, S. H. M., Stattersfield, A. J., & Collar, N. J. (2006). How many bird extinctions have we prevented? *Oryx*, *40*(3), 266–278. https://doi.org/10.1017/S0030605306000950

Buxton, R. T., Lendrum, P. E., Crooks, K. R., & Wittemyer, G. (2018). Pairing camera traps and acoustic recorders to monitor the ecological impact of human disturbance. *Global Ecology and Conservation*, *16*, e00493. https://doi.org/10.1016/j.gecco.2018.e00493

Buxton, R. T., McKenna, M. F., Clapp, M., Meyer, E., Stabenau, E., Angeloni, L. M., Crooks, K., & Wittemyer, G. (2018). Efficacy of extracting indices from large-scale acoustic recordings to monitor biodiversity. *Conservation Biology*, *32*(5), 1174–1184. https://doi.org/10.1111/cobi.13119

Buxton, R. T., McKenna, M. F., Mennitt, D., Fistrup, K., Crooks, K., Angeloni, L. M., &

Wittemyer, G. (2017). Noise pollution is pervasive in U.S.protected areas. *Science*, *356*, 531–533. https://doi.org/10.1126/science.aah4783

Carruthers-Jones, J., Eldridge, A., Guyot, P., Hassall, C., & Holmes, G. (2019). The call of the wild: Investigating the potential for ecoacoustic methods in mapping wilderness areas. *Science of the Total Environment*, *695*, 133797. https://doi.org/10.1016/j.scitotenv.2019.133797

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), 9–13. https://doi.org/10.1126/sciadv.1400253

Chandler, C. J., Van Helden, B. E., Close, P. G., & Speldewinde, P. C. (2020). 2D or not 2D? Three-dimensional home range analysis better represents space use by an arboreal mammal. *Acta Oecologica*, *105*(March), 103576. https://doi.org/10.1016/j.actao.2020.103576

Chereskin, E., Connor, R. C., Friedman, W. R., Jensen, F. H., Allen, S. J., Sørensen, P. M., Krützen, M., & King, S. L. (2022). Allied male dolphins use vocal exchanges to “bond at a distance.” *Current Biology*, *32*(7), 1657-1663.e4. https://doi.org/10.1016/j.cub.2022.02.019

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*(10), 1632–1644. https://doi.org/10.1111/2041-210X.13256

Clare, E. L., Economou, C. K., Bennett, F. J., Dyer, C. E., Adams, K., McRobie, B., Drinkwater, R., & Littlefair, J. E. (2022). Measuring biodiversity from DNA in the air. *Current Biology*, *32*(3), 693-700.e5. https://doi.org/10.1016/j.cub.2021.11.064

Costello, M. J., Beard, K. H., Corlett, R. T., Cumming, G. S., Devictor, V., Loyola, R., Maas, B., Miller-Rushing, A. J., Pakeman, R., & Primack, R. B. (2016). Field work ethics in biological research. *Biological Conservation*, *203*, 268–271. https://doi.org/10.1016/j.biocon.2016.10.008

Cousins, I. T., Johansson, J. H., Salter, M. E., Sha, B., & Scheringer, M. (2022). Outside the Safe Operating Space of a New Planetary Boundary for Per- and Polyfluoroalkyl Substances (PFAS). *Environmental Science and Technology*, *56*(16), 11172–11179.

https://doi.org/10.1021/acs.est.2c02765

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*(2), 1–24. https://doi.org/10.18637/jss.v034.i02

Crunchant, A.-S., Isaacs, J. T., & Piel, A. K. (2022). *Localizing wild chimpanzees with passive acoustics*. *January*, 1–12. https://doi.org/10.1002/ece3.8902

Darras, K., Batáry, P., Furnas, B. J., Grass, I., Mulyani, Y. A., & Tscharntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. *Ecological Applications*, *29*(6). https://doi.org/10.1002/eap.1954

Darras, K., Pütz, P., Fahrurrozi, Rembold, K., & Tscharntke, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biological Conservation*, *201*, 29–37. https://doi.org/10.1016/j.biocon.2016.06.021

Davies, A. B., & Asner, G. P. (2014). Advances in animal ecology from 3D-LiDAR ecosystem mapping. *Trends in Ecology and Evolution*, *29*(12), 681–691. https://doi.org/10.1016/j.tree.2014.10.005

Dent, M. L., Screven, L. A., & Kobrina, A. (2018). *Hearing in Rodents*. https://doi.org/10.1007/978-3-319-92495-3_4

Derryberry, E. P., Phillips, J. N., Derryberry, G. E., Blum, M. J., & Luther, D. (2020). Singing in a silent spring: Birds respond to a half-century soundscape reversion during the COVID-19 shutdown. *Science*, *370*(6516), 575–579. https://doi.org/10.1126/SCIENCE.ABD5777

Desjonquères, C., Gifford, T., & Linke, S. (2020). Passive acoustic monitoring as a potential tool to survey animal and ecosystem processes in freshwater environments. *Freshwater Biology*, *65*(1), 7–19. https://doi.org/10.1111/fwb.13356

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., Keitt, T. H., Kenney, M. A., Laney, C. M., Larsen, L. G., Loescher, H. W., Lunch, C. K., Pijanowski, B. C., Randerson, J. T., Read, E. K., Tredennick, A. T., Vargas, R., Weathers, K. C., & White, E. P. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1424–1432. https://doi.org/10.1073/pnas.1710231115

Doser, J. W., Finley, A. O., Weed, A. S., & Zipkin, E. F. (2021). Integrating automated acoustic vocalization data and point count surveys for estimation of bird abundance. *Methods in Ecology and Evolution*, *12*(6), 1040–1049. https://doi.org/10.1111/2041-210X.13578

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, *10*(9), 1412–1430. https://doi.org/10.1111/2041-210X.13234

Dunlop, K. M., Jarvis, T., Benoit-Bird, K. J., Waluk, C. M., Caress, D. W., Thomas, H., & Smith, K. L. (2018). Detection and characterisation of deep-sea benthopelagic animals from an autonomous underwater vehicle with a multibeam echosounder: A proof of concept and description of data-processing methods. *Deep-Sea Research Part I: Oceanographic Research Papers*, *134*(January), 64–79. https://doi.org/10.1016/j.dsr.2018.01.006

Eldridge, A., Guyot, P., Moscoso, P., Johnston, A., Eyre-Walker, Y., & Peck, M. (2018). Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats. *Ecological Indicators*, *95*, 939–952. https://doi.org/10.1016/j.ecolind.2018.06.012

Elise, S., Urbina-Barreto, I., Pinel, R., Mahamadaly, V., Bureau, S., Penin, L., Adjeroud, M., Kulbicki, M., & Bruggemann, J. H. (2019). Assessing key ecosystem functions through soundscapes: A new perspective from coral reefs. *Ecological Indicators*, *107*(April), 105623. https://doi.org/10.1016/j.ecolind.2019.105623

Elliott, D., Otero, C. E., Wyatt, S., & Martino, E. (2021). *Tiny Transformers for Environmental Sound Classification at the Edge*. 1–12. http://arxiv.org/abs/2103.12157

Ewers, R. M., Didham, R. K., Fahrig, L., Ferraz, G., Hector, A., Holt, R. D., Kapos, V., Reynolds, G., Sinun, W., Snaddon, J. L., & Turner, E. C. (2011). A large-scale forest fragmentation experiment: The stability of altered forest ecosystems project. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1582), 3292–3302. https://doi.org/10.1098/rstb.2011.0049

Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating ecology as a big-data science: Current advances, challenges, and solutions. *BioScience*, *68*(8), 563–576. https://doi.org/10.1093/biosci/biy068

Farr, M. T., Green, D. S., Holekamp, K. E., Roloff, G. J., & Zipkin, E. F. (2019). Multispecies hierarchical modeling reveals variable responses of African carnivores to management alternatives. *Ecological Applications*, *29*(2), 1–11. https://doi.org/10.1002/eap.1845

Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, *19*(7), 1673–1679.

Frommolt, K. H., & Tauchert, K. H. (2014). Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. *Ecological Informatics*, *21*, 4–12. https://doi.org/10.1016/j.ecoinf.2013.12.009

Fuller, S., Axel, A. C., Tucker, D., & Gage, S. H. (2015). Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? *Ecological Indicators*, *58*, 207–215. https://doi.org/10.1016/j.ecolind.2015.05.057

Furnas, B. J., & Callas, R. L. (2015). Using automated recorders and occupancy models to monitor common forest birds across a large geographic region. *Journal of Wildlife Management*, *79*(2), 325–337. https://doi.org/10.1002/jwmg.821

Furnas, B. J., Newton, D. S., Capehart, G. D., & Barrows, C. W. (2019). Hierarchical distance sampling to estimate population sizes of common lizards across a desert ecoregion. *Ecology and Evolution*, *9*(6), 3046–3058. https://doi.org/10.1002/ece3.4780

Gallacher, S., Wilson, D., Fairbrass, A., Turmukhambetov, D., Firman, M., Kreitmayer, S., Mac Aodha, O., Brostow, G., & Jones, K. (2021). Shazam for bats: Internet of Things for continuous real-time biodiversity monitoring. *IET Smart Cities*, *3*(3), 171–183. https://doi.org/10.1049/smc2.12016

Gámez, S., & Harris, N. C. (2022). Conceptualizing the 3D niche and vertical space use. *Trends in Ecology & Evolution*, 1–10. https://doi.org/10.1016/j.tree.2022.06.012

Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., Depraetere, M., & Pavoine, S. (2013). Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecological Indicators*, *25*, 279–287. https://doi.org/10.1016/j.ecolind.2012.10.009

Gatti, L. V., Basso, L. S., Miller, J. B., Gloor, M., Gatti Domingues, L., Cassol, H. L. G., Tejada, G., Aragão, L. E. O. C., Nobre, C., Peters, W., Marani, L., Arai, E., Sanches, A. H., Corrêa, S. M., Anderson, L., Von Randow, C., Correia, C. S. C., Crispim, S. P., & Neves,

R. A. L. (2021). Amazonia as a carbon source linked to deforestation and climate change. *Nature*, *595*(7867), 388–393. https://doi.org/10.1038/s41586-021-03629-6

Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., & Ritter, M. (2017). Audio Set: An ontology and human-labeled dataset for audio events. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 776–780. https://doi.org/10.1109/ICASSP.2017.7952261

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, *10*(2), 169–185. https://doi.org/10.1111/2041-210X.13101

Girod, L., Lukac, M., Trifa, V., & Estrin, D. (2006). The design and implementation of a self-calibrating distributed acoustic sensing platform. *SenSys'06: Proceedings of the Fourth International Conference on Embedded Networked Sensor Systems*, 71–84. https://doi.org/10.1145/1182807.1182815

Gómez, W. E., Isaza, C. V., & Daza, J. M. (2018). Identifying disturbed habitats: A new method from acoustic indices. *Ecological Informatics*, *45*(May 2017), 16–25. https://doi.org/10.1016/j.ecoinf.2018.03.001

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., … Fargione, J. (2017). Natural climate solutions. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(44), 11645–11650. https://doi.org/10.1073/pnas.1710465114

Grondin, F., Létourneau, D., Godin, C., Lauzon, J.-S., Vincent, J., Michaud, S., Faucher, S., & Michaud, F. (2022). ODAS: Open embeddeD Audition System. *Frontiers in Robotics and AI*, *9*(May), 1–9. https://doi.org/10.3389/frobt.2022.854444

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*(3), 156–162. https://doi.org/10.1890/120103

Harvey, C. A., & González Villalobos, J. A. (2007). Agroforestry systems conserve species-rich but modified assemblages of tropical birds and bats. *Biodiversity and Conservation*, *16*(8), 2257–2292. https://doi.org/10.1007/s10531-007-9194-2

Heald, O. J. N., Fraticelli, C., Cox, S. E., Stevens, M. C. A., Faulkner, S. C., Blackburn, T. M., & Le Comber, S. C. (2020). Understanding the origins of the ring-necked parakeet in the UK. *Journal of Zoology*, *312*(1), 1–11. https://doi.org/10.1111/jzo.12753

Heath, B. E., Sethi, S. S., Orme, C. D. L., Ewers, R. M., & Picinali, L. (2020). Autonomous Rainforest Soundscape Identification : Quantifying the Impacts of Data Lossy Compression , Recording Length , and Index Selection. *E-Forum Acousticum 2020*, 3207–3208. https://doi.org/10.48465/fa.2020.0365f

Heath, B. E., Sethi, S. S., Orme, C. D. L., Ewers, R. M., & Picinali, L. (2021). How index selection, compression, and recording schedule impact the description of ecological soundscapes. *Ecology and Evolution*, *11*(19), 13206–13217. https://doi.org/10.1002/ece3.8042

Hedley, R., & Bayne, E. (2020). *Testing the Accuracy of a BirdNET , Automatic bird song Classifier*.

Hedley, R. W., Huang, Y., & Yao, K. (2017). Direction-of-arrival estimation of animal vocalizations for monitoring animal behavior and improving estimates of abundance. *Avian Conservation and Ecology*, *12*(1). https://doi.org/10.5751/ACE-00963-120106

Hermosilla, T., Ruiz, L. A., Kazakova, A. N., Coops, N. C., & Moskal, L. M. (2014). Estimation of forest structure and canopy fuel parameters from small-footprint full-waveform LiDAR data. *International Journal of Wildland Fire*, *23*(2), 224–233. https://doi.org/10.1071/WF13086

Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., & Wilson, K. (2017). CNN architectures for large-scale audio classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 131–135. https://doi.org/10.1109/ICASSP.2017.7952132

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, *9*(5), 1199–1211. https://doi.org/10.1111/2041-210X.12955

Hödl, W. (1977). Call differences and calling site segregation in anuran species from central Amazonian floating meadows. *Oecologia*, *28*(4), 351–363. https://doi.org/10.1007/BF00345990

Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H. M., Carpenter, K. E., Chanson, J., Collen, B., Cox, N. A., Darwall, W. R. T., Dulvy, N. K., Harrison, L. R., Katariya, V., Pollock, C. M., Quader, S., Richman, N. I., Rodrigues, A. S. L., Tognelli, M. F., … Stuart, S. N. (2010). The impact of conservation on the status of the world's vertebrates. *Science*, *330*(6010), 1503–1509. https://doi.org/10.1126/science.1194442

Honrado, J. P., Pereira, H. M., & Guisan, A. (2016). Fostering integration between biodiversity monitoring and modelling. *Journal of Applied Ecology*, *53*(5), 1299–1304. https://doi.org/10.1111/1365-2664.12777

Huff, M. H., Bettinger, K. A., Ferguson, H. L., Brown, M. J., & Altman, B. (2000). A habitat-based point-count protocol for terrestrial birds, emphasizing Washington and Oregon. *General Technical Reports of the US Department of Agriculture, Forest Service*, *PNW-GTR-501*, 2–30.

Huo, W., Ling, W., Wang, Z., Li, Y., Zhou, M., Ren, M., Li, X., Li, J., Xia, Z., Liu, X., & Huang, X. (2021). Miniaturized DNA Sequencers for Personal Use: Unreachable Dreams or Achievable Goals. *Frontiers in Nanotechnology*, *3*(February), 1–17. https://doi.org/10.3389/fnano.2021.628861

Jain, M., & Balakrishnan, R. (2012). Does acoustic adaptation drive vertical stratification? A test in a tropical cricket assemblage. *Behavioral Ecology*, *23*(2), 343–354. https://doi.org/10.1093/beheco/arr191

Johnston, A., Newson, S. E., Risely, K., Musgrove, A. J., Massimino, D., Baillie, S. R., & Pearce-Higgins, J. W. (2014). Species traits explain variation in detectability of UK birds. *Bird Study*, *61*(3), 340–350. https://doi.org/10.1080/00063657.2014.941787

Jones, G., & Teeling, E. C. (2006). The evolution of echolocation in bats. *Trends in Ecology and Evolution*, *21*(3), 149–156. https://doi.org/10.1016/j.tree.2006.01.001

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*(January), 101236.

https://doi.org/10.1016/j.ecoinf.2021.101236

Kasten, E. P., Gage, S. H., Fox, J., & Joo, W. (2012). The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. *Ecological Informatics*, *12*, 50–67. https://doi.org/10.1016/j.ecoinf.2012.08.001

King, S. L., & Janik, V. M. (2013). Bottlenose dolphins can use learned vocal labels to address each other. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 13216–13221. https://doi.org/10.1073/pnas.1304459110

Krause, B. (2008). Anatomy of the soundscape: Evolving perspectives. *AES: Journal of the Audio Engineering Society*, *56*(1–2), 73–80.

Laiolo, P., & Tella, J. L. (2006). Landscape bioacoustics allow detection of the effects of habitat patchiness on population structure. *Ecology*, *87*(5), 1203–1214. https://doi.org/10.1890/0012-9658(2006)87[1203:LBADOT]2.0.CO;2

Leahy, L., Scheffers, B. R., Williams, S. E., & Andersen, A. N. (2022). Arboreality drives heat tolerance while elevation drives cold tolerance in tropical rainforest ants. *Ecology*, *103*(1), 1–11. https://doi.org/10.1002/ecy.3549

Levin, S. A. (1970). Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle. *The American Naturalist*, *104*(939), 413–423.

Li, W., Hou, X., Xu, C., Qin, M., Wang, S., Wei, L., Wang, Y., Liu, X., & Li, Y. (2021). Validating eDNA measurements of the richness and abundance of anurans at a large scale. *Journal of Animal Ecology*, *90*(6), 1466–1479. https://doi.org/10.1111/1365-2656.13468

Linke, S., & Deretic, J. A. (2020). Ecoacoustics can detect ecosystem responses to environmental water allocations. *Freshwater Biology*, *65*(1), 133–141. https://doi.org/10.1111/fwb.13249

Lohchab, V., Kumar, M., Suryan, G., Gautam, V., & Das, R. K. (2018). A Review of IoT based Smart Farm Monitoring. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, *Icicct*, 1620–1625. https://doi.org/10.1109/ICICCT.2018.8473337

Lostanlen, V., Bernabeu, A., Béchennec, J. L., Briday, M., Faucou, S., & Lagrange, M. (2021). Energy efficiency is not enough:Towards a batteryless internet of sounds. *ACM*

International Conference Proceeding Series, 147–155. https://doi.org/10.1145/3478384.3478408

Lovejoy, T. E., & Nobre, C. (2018). Amazon tipping point. *Science Advances*, *4*(2), 1–2. https://doi.org/10.1126/sciadv.aat2340

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, A. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Mammides, C., Goodale, E., Dayananda, S. K., Kang, L., & Chen, J. (2017). Do acoustic indices correlate with bird diversity? Insights from two biodiverse regions in Yunnan Province, south China. *Ecological Indicators*, *82*(March), 470–477. https://doi.org/10.1016/j.ecolind.2017.07.017

McCormick, M. I., Allan, B. J. M., Harding, H., & Simpson, S. D. (2018). Boat noise impacts risk assessment in a coral reef fish but effects depend on engine type. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-22104-3

Mckay, D. I. A., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., & Lenton, T. M. (2022). *Exceeding 1.5C Global Warming Could Trigger Multiple Tipping Points*. *1171*(September). https://doi.org/10.1126/science.abn7950

Mcloughlin, M. P., Stewart, R., & McElligott, A. G. (2019). Automated bioacoustics: Methods in ecology and conservation and their potential for animal welfare monitoring. *Journal of the Royal Society Interface*, *16*(155). https://doi.org/10.1098/rsif.2019.0225

Medina-García, A., Araya-Salas, M., & Wright, T. F. (2015). Does vocal learning accelerate acoustic diversification? Evolution of contact calls in Neotropical parrots. *Journal of Evolutionary Biology*, *28*(10), 1782–1792. https://doi.org/10.1111/jeb.12694

Mennill, D. J., Battiston, M., Wilson, D. R., Foote, J. R., & Doucet, S. M. (2012). Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. *Methods in Ecology and Evolution*, *3*(4), 704–712. https://doi.org/10.1111/j.2041-210X.2012.00209.x

Metcalf, O. C., Barlow, J., Devenish, C., Marsden, S., Berenguer, E., & Lees, A. C. (2021). Acoustic indices perform better when applied at ecologically meaningful time and frequency scales. *Methods in Ecology and Evolution*, *12*(3), 421–431. https://doi.org/10.1111/2041-210X.13521

Metcalf, O. C., Lees, A. C., Barlow, J., Marsden, S. J., & Devenish, C. (2020). hardRain: An R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach. *Ecological Indicators*, *109*(October 2019), 105793. https://doi.org/10.1016/j.ecolind.2019.105793

Mitchell, S. L., Edwards, D. P., Bernard, H., Coomes, D., Jucker, T., Davies, Z. G., & Struebig, M. J. (2018). Riparian reserves help protect forest bird communities in oil palm dominated landscapes. *Journal of Applied Ecology*, *55*(6), 2744–2755. https://doi.org/10.1111/1365-2664.13233

Moore, J. F., Pine, W. E., Mulindahabi, F., Niyigaba, P., Gatorano, G., Masozera, M. K., & Beaudrot, L. (2020). Comparison of species richness and detection between line transects, ground camera traps, and arboreal camera traps. *Animal Conservation*, *23*(5), 561–572. https://doi.org/10.1111/acv.12569

Morfi, V., Nolasco, I., Lostanlen, V., Singh, S., Strandburg-Peshkin, A., Gill, L., Pamuła, H., Benvent, D., & Stowell, D. (2021). Few-Shot Bioacoustic Event Detection: A New Task at the DCASE 2021 Challenge. *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, *November*, 145–149.

Mori, E., Onorati, G., & Giuntini, S. (2020). Loud callings limit human tolerance towards invasive parakeets in urban areas. *Urban Ecosystems*, *23*(4), 755–760. https://doi.org/10.1007/s11252-020-00954-y

Morton, E. S. (1975). The American Society of Naturalists Ecological Sources of Selection on Avian Sounds. *The American Naturalist*, *109*(965), 17–34. http://www.jstor.org/stable/2459634

Mouterde, S. C., Elie, J. E., Mathevon, N., & Theunissen, F. E. (2017). Single neurons in the avian auditory cortex encode individual identity and propagation distance in naturally degraded communication calls. *Journal of Neuroscience*, *37*(13), 3491–3510. https://doi.org/10.1523/JNEUROSCI.2220-16.2017

N. Mitianoudis, M. E. D. (2003). USING BEAMFORMING IN THE AUDIO SOURCE SEPARATION PROBLEM Nikolaos Mitianoudis and Michael E . Davies DSP Group , Queen Mary. *Seventh International Symposium on Signal Processing and Its Applications*, *2*, 2–5. https://doi.org/10.1109/ISSPA.2003.1224822

Nakadai, K., Okuno, H. G., & Mizumoto, T. (2017). Development, deployment and applications of robot audition open source software HARK. *Journal of Robotics and Mechatronics*, *29*(1), 16–25. https://doi.org/10.20965/jrm.2017.p0016

Nakadai, K., Takahashi, T., Okuno, H. G., Nakajima, H., Hasegawa, Y., & Tsujino, H. (2010). Design and implementation of robot audition system "HARK" - Open source software for listening to three simultaneous speakers. *Advanced Robotics*, *24*(5–6), 739–761. https://doi.org/10.1163/016918610X493561

Nobre, C. A., Sampaio, G., Borma, L. S., Castilla-Rubio, J. C., Silva, J. S., & Cardoso, M. (2016). Land-use and climate change risks in the amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(39), 10759–10768. https://doi.org/10.1073/pnas.1605516113

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*, 1–6. https://doi.org/10.1038/ncomms4379

Oliveira, B. F., & Scheffers, B. R. (2019). Vertical stratification influences global patterns of biodiversity. *Ecography*, *42*(2), 249. https://doi.org/10.1111/ecog.03636

Papin, M., Aznar, M., Germain, E., Guérold, F., & Pichenot, J. (2019). Using acoustic indices to estimate wolf pack size. *Ecological Indicators*, *103*(October 2018), 202–211. https://doi.org/10.1016/j.ecolind.2019.03.010

Paul, C., Weber, M., & Knoke, T. (2017). Agroforestry versus farm mosaic systems – Comparing land-use efficiency, economic returns and risks under climate change effects. *Science of the Total Environment*, *587–588*, 22–35. https://doi.org/10.1016/j.scitotenv.2017.02.037

Peck, H. L. (2013). *Investigating ecological impacts of the non-native population of rose-ringed parakeets (Psittacula krameri) in the UK*. *December*, 212.

Pérez-Granados, C., & Traba, J. (2021). Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. *Ibis*, *163*(3), 765–783. https://doi.org/10.1111/ibi.12944

Pettorelli, N., Wegmann, M., Skidmore, A., Mücher, S., Dawson, T. P., Fernandez, M., Lucas, R., Schaepman, M. E., Wang, T., O'Connor, B., Jongman, R. H. G., Kempeneers, P., Sonnenschein, R., Leidner, A. K., Böhm, M., He, K. S., Nagendra, H., Dubois, G., Fatoyinbo, T., … Geller, G. N. (2016). Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sensing in Ecology and Conservation*, *2*(3), 122–131. https://doi.org/10.1002/rse2.15

Pfeifer, M., Kor, L., Nilus, R., Turner, E., Cusack, J., Lysenko, I., Khoo, M., Chey, V., Chung, A., & Ewers, R. M. (2016). Mapping the structure of Borneo's tropical forests across a degradation gradient. *Remote Sensing of Environment*, *176*, 84–97.

Pieretti, N., Farina, A., & Morri, D. (2011). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, *11*(3), 868–873. https://doi.org/10.1016/j.ecolind.2010.11.005

Pijanowski, B. C., Villanueva-rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., Gage, S. H., & Pieretti, N. (2011). *Soundscape Ecology : The Science of Sound in the Landscape*. *61*(3), 203–216. https://doi.org/10.1525/bio.2011.61.3.6

Porto Marques, T., Cote, M., Rezvanifar, A., Branzan Albu, A., Ersahin, K., Mudge, T., & Gauthier, S. (2021). Instance segmentation-based identification of pelagic species in acoustic backscatter data. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *DI*, 4373–4382. https://doi.org/10.1109/CVPRW53098.2021.00494

Rapport, D. J. (1989). What constitutes ecosystem health? *Perspectives in Biology and Medicine*, *33*(1), 120–132. https://doi.org/10.1353/pbm.1990.0004

Rapport, D. J., Costanza, R., & McMichael, A. J. (1998). Assessing ecosystem health. *Trends in Ecology and Evolution*, *13*(10), 397–402. https://doi.org/10.1016/S0169-5347(98)01449-9

Rempel, R. S., Hobson, K. A., Holborn, G., Van Wilgenburg, S. L., & Elliott, J. (2005). Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory. *Journal of Field Ornithology*, *76*(1), 1–11. https://doi.org/10.1648/0273-8570-76.1.1

Rhinehart, T. A., Chronister, L. M., Devlin, T., & Kitzes, J. (2020). Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution*, *10*(13), 6794–6818. https://doi.org/10.1002/ece3.6216

Risoud, M., Hanson, J. N., Gauvrit, F., Renard, C., Lemesre, P. E., Bonne, N. X., & Vincent, C. (2018). Sound source localization. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *135*(4), 259–264. https://doi.org/10.1016/j.anorl.2018.04.009

Roca, I. T., & Proulx, R. (2016). Acoustic assessment of species richness and assembly rules in ensiferan communities from temperate ecosystems. *Ecology*, *97*(1), 116–123. https://doi.org/10.1890/15-0290.1

Roe, P., Eichinski, P., Fuller, R. A., McDonald, P. G., Schwarzkopf, L., Towsey, M., Truskinger, A., Tucker, D., & Watson, D. M. (2021). The Australian Acoustic Observatory. *Methods in Ecology and Evolution*, *12*(10), 1802–1808. https://doi.org/10.1111/2041-210X.13660

Round, P. D., Gale, G. A., & Brockelman, W. Y. (2006). A comparison of bird communities in mixed fruit orchards and natural forest at Khao Luang, southern Thailand. *Biodiversity and Conservation*, *15*(9), 2873–2891. https://doi.org/10.1007/s10531-005-2006-7

Royle, J. A. (2004). N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. *Biometrics*, *60*(1), 108–115. https://doi.org/10.1111/j.0006-341X.2004.00142.x

Saito, K., Nakamura, K., Ueta, M., Kurosawa, R., Fujiwara, A., Kobayashi, H. H., Nakayama, M., Toko, A., & Nagahama, K. (2015). Utilizing the Cyberforest live sound system with social media to remotely conduct woodland bird censuses in Central Japan. *Ambio*, *44*(4), 572–583. https://doi.org/10.1007/s13280-015-0708-y

Scheffers, B. R., Phillips, B. L., Laurance, W. F., Sodhi, N. S., Diesmos, A., & Williams, S. E. (2013). Increasing arboreality with altitude: A novel biogeographic dimension. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1770). https://doi.org/10.1098/rspb.2013.1581

Scheffers, B. R., Shoo, L., Phillips, B., Macdonald, S. L., Anderson, A., VanDerWal, J., Storlie, C., Gourret, A., & Williams, S. E. (2017). Vertical (arboreality) and horizontal (dispersal) movement increase the resilience of vertebrates to climatic instability. *Global Ecology and Biogeography*, *26*(7), 787–798. https://doi.org/10.1111/geb.12585

Schmidt, R. O. (1979). Multiple Emitter Location and Parameter Estimation. *Ieee*, *March*, 243–258.

Scyphers, S. B., Powers, S. P., Heck, K. L., & Byron, D. (2011). Oyster reefs as natural breakwaters mitigate shoreline loss and facilitate fisheries. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0022396

Seddon, N., Chausson, A., Berry, P., Girardin, C. A. J., Smith, A., & Turner, B. (2020). Understanding the value and limits of nature-based solutions to climate change and other global challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1794). https://doi.org/10.1098/rstb.2019.0120

Sethi, S. S., Ewers, R. M., Jones, N., Picinali, L., Orme, C. D. L., Sleutel, J., Shabrani, A., Zulkifli, N., & Bernard, H. (2020). *Avifaunal and Herpetofaunal point counts with recorded acoustic data*. Zenodo. https://doi.org/10.5281/zenodo.3997172

Sethi, S. S., Ewers, R. M., Jones, N. S., Orme, C. D. L., & Picinali, L. (2018). Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. *Methods in Ecology and Evolution*, *2018*(December 2017), 1–5. https://doi.org/10.1111/2041-210X.13089

Sethi, S. S., Ewers, R. M., Jones, N. S., Signorelli, A., Picinali, L., & Orme, C. D. L. (2020). SAFE Acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. *Methods in Ecology and Evolution*, *11*(10), 1182–1185. https://doi.org/10.1111/2041-210X.13438

Sethi, S. S., Fossoy, F., Cretois, B., & Rosten, C. (2021). *Management-relevant applications of acoustic monitoring for Norwegian nature: The Sound of Norway*.

Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., Orme, C. D. L., Wrege, P. H., & Ewers, R. M. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 17049–17055. https://doi.org/10.1073/pnas.2004702117

Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Jane, D., & Klinck, H. (2020). *Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set*. *24*, 1–7. https://doi.org/10.1073/pnas.2004702117

Sethi, S. S., Kovac, M., Wiesemüller, F., Miriyev, A., & Boutry, C. M. (2022). Biodegradable sensors are ready to transform autonomous ecological monitoring. *Nature Ecology and Evolution*, *6*(September), 1245–1247. https://doi.org/10.1038/s41559-022-01824-w

Shonfield, J., & Bayne, E. M. (2017). Autonomous recording units in avian ecological research: current use and future applications. *Avian Conservation and Ecology*, *12*(1). https://doi.org/10.5751/ace-00974-120114

Smith, B. R., Root-Gutteridge, H., Butkiewicz, H., Dassow, A., Fontaine, A. C., Markham, A., Owens, J., Schindler, L., Wijers, M., & Kershenbaum, A. (2021). Acoustic localisation of wildlife with low-cost equipment: Lower sensitivity, but no loss of precision. *Wildlife Research*. https://doi.org/10.1071/WR21089

Sousa-Lima, R. S., Norris, T. F., Oswald, J. N., & Fernandes, D. P. (2013). A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals. *Aquatic Mammals*, *39*(1), 23–53. https://doi.org/10.1578/AM.39.1.2013.23

Spillmann, B., van Noordwijk, M. A., Willems, E. P., Mitra Setia, T., Wipfli, U., & van Schaik, C. P. (2015). Validation of an acoustic location system to monitor Bornean orangutan (Pongo pygmaeus wurmbii) long calls. *American Journal of Primatology*, *77*(7), 767–776. https://doi.org/10.1002/ajp.22398

Stevenson, B. C., Borchers, D. L., Altwegg, R., Swift, R. J., Gillespie, D. M., & Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, *6*(1), 38–48. https://doi.org/10.1111/2041-210X.12291

Stowell, D. (2022). Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, *10*(2017), e13152. https://doi.org/10.7717/peerj.13152

Stowell, D., & Plumbley, M. D. (2014). Large-scale analysis of frequency modulation in birdsong data bases. *Methods in Ecology and Evolution*, *5*(9), 901–912. https://doi.org/10.1111/2041-210X.12223

Stowell, D., Wood, M. D., Pamuła, H., Stylianou, Y., & Glotin, H. (2019). Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *Methods in Ecology and Evolution*, *10*(3), 368–380. https://doi.org/10.1111/2041-210X.13103

Struebig, M. J., Turner, A., Giles, E., Lasmana, F., Tollington, S., Bernard, H., & Bell, D. (2013). Quantifying the Biodiversity Value of Repeatedly Logged Rainforests. Gradient and Comparative Approaches from Borneo. In *Advances in Ecological Research* (1st ed., Vol. 48, Issue April 2020). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-417199-2.00003-3

Sueur, Jerome, Aubin, T., & Simonis, C. (2008). Equipment review: Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*(2), 213–226. https://doi.org/10.1080/09524622.2008.9753600

Sueur, Jérôme, Farina, A., Gasc, A., Pieretti, N., & Pavoine, S. (2014). Acoustic indices for biodiversity assessment and landscape investigation. *Acta Acustica United with Acustica*, *100*(4), 772–781. https://doi.org/10.3813/AAA.918757

Sueur, Jérôme, Krause, B., & Farina, A. (2019a). Climate Change Is Breaking Earth’s Beat. *Trends in Ecology and Evolution*, *34*(11), 971–973. https://doi.org/10.1016/j.tree.2019.07.014

Sueur, Jérôme, Krause, B., & Farina, A. (2019b). Climate Change Is Breaking Earth’s Beat. *Trends in Ecology & Evolution*, *34*(11), 971–973. https://doi.org/10.1016/j.tree.2019.07.014

Sueur, Jérôme, Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS ONE*, *3*(12). https://doi.org/10.1371/journal.pone.0004065

Sugai, L. S. M., Desjonquères, C., Silva, T. S. F., & Llusia, D. (2019). A roadmap for survey designs in terrestrial acoustic monitoring. *Remote Sensing in Ecology and Conservation*, 1–16. https://doi.org/10.1002/rse2.131

Sumitani, S., Suzuki, R., Chiba, N., Matsubayashi, S., Arita, T., Nakadai, K., & Okuno, H. G. (2019). An Integrated Framework for Field Recording, Localization, Classification and Annotation of Birdsongs Using Robot Audition Techniques - Harkbird 2.0. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2019-May*(1), 8246–8250. https://doi.org/10.1109/ICASSP.2019.8683743

Suzuki, R., Matsubayashi, S., Hedley, R. W., Nakadai, K., & Okuno, H. G. (2017). Harkbird: Exploring acoustic interactions in bird communities using a microphone array. *Journal of Robotics and Mechatronics*, *29*(1), 213–223. https://doi.org/10.20965/jrm.2017.p0213

Sweger, A. L., & Uetz, G. W. (2016). Characterizing the vibratory and acoustic signals of the "purring" wolf spider, Gladicosa gulosa (Araneae: Lycosidae). *Bioacoustics*, *25*(3), 293–303. https://doi.org/10.1080/09524622.2016.1160328

Thiel, S., Tschapka, M., Heymann, E. W., & Heer, K. (2021). Vertical stratification of seed-dispersing vertebrate communities and their interactions with plants in tropical forests. *Biological Reviews*, *96*(2), 454–469. https://doi.org/10.1111/brv.12664

Thode, A. M., Gerstoft, P., Burgess, W. C., Sabra, K. G., Guerra, M., Stokes, M. D., Noad, M., & Cato, D. H. (2006). A portable matched-field processing system using passive acoustic time synchronization. *IEEE Journal of Oceanic Engineering*, *31*(3), 696–710. https://doi.org/10.1109/JOE.2006.880431

Tilman, D. (2004). Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(30), 10854–10861. https://doi.org/10.1073/pnas.0403458101

Towsey, M. (2018). *The calculation of acoustic indices derived from long-duration recordings of the natural environment*. *0*(August 2017), 0–12.

Towsey, M. W., Truskinger, A. M., & Roe, P. (2016). The Navigation and Visualisation of Environmental Audio Using Zooming Spectrograms. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015*, 788–797. https://doi.org/10.1109/ICDMW.2015.118

Towsey, M., Wimmer, J., Williamson, I., & Roe, P. (2014). The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics*, *21*(100), 110–119. https://doi.org/10.1016/j.ecoinf.2013.11.007

Towsey, M., Znidersic, E., Broken-Brow, J., Indraswari, K., Watson, D. M., Phillips, Y., Truskinger, A., & Roe, P. (2018). Long-duration, false-colour spectrograms for detecting species in large audio data-sets. *Journal of Ecoacoustics*. https://doi.org/10.22261/jea.iuswui

Trifa, V., Girod, L., Collier, T., Blumstein, D. T., & Taylor, C. E. (2007). Automated Wildlife Monitoring Using Self-Configuring Sensor Networks Deployed in Natural Habitats. *UCLA: Center for Embedded Network Sensing. Retrieved From*. https://escholarship.org/uc/item/2059b99k

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-27980-y

Vanden Hole, C., Van Daele, P. A. A. G., Desmet, N., Devos, P., & Adriaens, D. (2014). Does sociality imply a complex vocal communication system? A case study for Fukomys micklemi (Bathyergidae, Rodentia). In *Bioacoustics* (Vol. 23, Issue 2, pp. 143–160). Taylor & Francis. https://doi.org/10.1080/09524622.2013.841085

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS)*.

Venkatesh, S., Moffat, D., & Miranda, E. R. (2022). You Only Hear Once: A YOLO-like Algorithm for Audio Segmentation and Sound Event Detection. *Applied Sciences (Switzerland)*, *12*(7), 1–20. https://doi.org/10.3390/app12073293

Vesna, I. (2009). Understanding Bland Altman Analysis. *Biochemia Medica*, *19*(1), 10–16. https://doi.org/10.11613/BM.2013.003

Villanueva-Rivera, L. J., & Pijanowski, B. C. (2016). Package "soundecology." *Http://Ljvillanueva.Github.Io/Soundecology/, CRAN*, *Package version 1.3.3*, 14. http://ljvillanueva.github.io/soundecology/

Villanueva-Rivera, L. J., Pijanowski, B. C., Doucette, J., & Pekin, B. (2011). A primer of acoustic analysis for landscape ecologists. *Landscape Ecology*, *26*(9), 1233–1246. https://doi.org/10.1007/s10980-011-9636-9

Villén-, C. D. S., Marco, P. P. De, Márquez, R., Beltrán, J. F., & Llusia, D. (2022). *Acoustic species distribution models (aSDMs): A framework to forecast shifts in calling behaviour under climate change*. *2022*(May), 1–14. https://doi.org/10.1111/2041-210X.13923

Warren, P. S., Katti, M., Ermann, M., & Brazel, A. (2006). Urban bioacoustics: It's not just noise. *Animal Behaviour*, *71*(3), 491–502. https://doi.org/10.1016/j.anbehav.2005.07.014

Wijers, M., Loveridge, A., Macdonald, D. W., & Markham, A. (2019). CARACAL: a versatile passive acoustic monitoring tool for wildlife research and conservation. *Bioacoustics*, *00*(00), 1–17. https://doi.org/10.1080/09524622.2019.1685408

Wijers, M., Loveridge, A., Macdonald, D. W., & Markham, A. (2021). CARACAL: a versatile passive acoustic monitoring tool for wildlife research and conservation. *Bioacoustics*, *30*(1), 41–57. https://doi.org/10.1080/09524622.2019.1685408

Wilson, S. J., & Bayne, E. M. (2018). Use of an acoustic location system to understand how presence of conspecifics and canopy cover influence ovenbird (Seiurus aurocapilla) space use near reclaimed wellsites in the boreal forest of alberta. *Avian Conservation and Ecology*, *13*(2). https://doi.org/10.5751/ACE-01248-130204

Wolters, P., Daw, C., Hutchinson, B., & Phillips, L. (2021). *Proposal-based Few-shot Sound Event Detection for Speech and Environmental Sounds with Perceivers*. 1–7. http://arxiv.org/abs/2107.13616

Wood, C. M., Kahl, S., Rahaman, A., & Klinck, H. (2022). The machine learning-powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. *PLoS Biology*, *20*(6), 1–6. https://doi.org/10.1371/journal.pbio.3001670

Yamahara, K. M., Preston, C. M., Birch, J., Walz, K., Marin, R., Jensen, S., Pargett, D., Roman, B., Ussler, W., Zhang, Y., Ryan, J., Hobson, B., Kieft, B., Raanan, B., Goodwin, K. D., Chavez, F. P., & Scholin, C. (2019). In situ autonomous acquisition and preservation of marine environmental dna using an autonomous underwater vehicle. *Frontiers in Marine Science*, *6*(JUL), 1–14. https://doi.org/10.3389/fmars.2019.00373

Zhang, L., Towsey, M., Zhang, J., & Roe, P. (2016). Classifying and ranking audio clips to support bird species richness surveys. *Ecological Informatics*, *34*, 108–116. https://doi.org/10.1016/j.ecoinf.2016.05.005

Zipkin, E. F., Thorson, J. T., See, K., Lynch, H. J., Grant, E. H. C., Kanno, Y., Chandler, R. B., Letcher, B. H., & Royle, J. A. (2014). Modeling structured population dynamics using data from unmarked individuals. *Ecology*, *95*(1), 22–29. https://doi.org/10.1890/13-1131.1

7. APPENDIX A: Chapter 2 Supplementary Material

Contents

1 - Above Ground Biomass

Table 1 shows the AGB data for points surrounding the recording locations. GPS locations are given to +/- 1.11km (0.01°). The average for each site was 66.16 t.ha^{-1} in primary forest, 30.74 t.ha^{-1} in secondary forest at 17.27 t.ha^{-1} in logged forest.

Location	Co-Ordinates	AGB (t.ha^{-1})
Primary	**117.53513, 4.66443**	**77.8151982458024**
Primary	117.53513, 4.65443	54.9446402261185
Primary	117.52513, 4.66443	87.2290056404617
Primary	117.52513, 4.65443	3.18829485141835
Primary	117.53513, 4.67443	49.5375612006372
Primary	117.54513, 4.66443	116.326089963978
Primary	117.54513, 4.67443	74.0737999980781
Logged	**117.58118, 4.69372**	**13.3752792205534**
Logged	117.58118, 4.68372	4.77470660981513
Logged	117.57118, 4.69372	61.6814645965346
Logged	117.57118, 4.68372	67.7614569739226
Logged	117.58118, 4.70372	42.4952102365264
Logged	117.59118, 4.69372	18.5855492747789
Logged	117.59118, 4.70372	6.5248825548366
Cleared	**117.59141, 4.70272**	**9.85390659254801**
Cleared	117.59141, 4.69272	7.9067100668437
Cleared	117.58141, 4.70272	48.796127019185
Cleared	117.58141, 4.69272	3.96428134965722
Cleared	117.59141, 4.71272	5.65620070374987
Cleared	117.60141, 4.70272	38.4582510289304
Cleared	117.60141, 4.71272	6.92104474036953

2 – Recording Period Information

Table 2.1 gives the total recordings, and total length of recordings at each of the sites. In total 648 20 minute recordings were collected equating to 216 hours of raw acoustic data.

Site	Recording Period (Feb-March 2019)	Hours and Minutes in rec period	Recording number (20min)
Cleared	26th 8:40- 1st 11:07	74h 40min	224
Logged	27th 10 – 2nd 9:00	71h 00min	213
Primary	26th 11 – 1st 9:07	70h 20min	211

2.2 gives the number of recordings from which indices were successfully calculated from. This is out of 648, 1296, 2592, and 5184 for the 20-, 10-, 5- and 2.5- minute recordings respectively. The total number of recordings used in the study was 87,480. Of which, 0.17% of files were lost in AudioSet Calculation, 0.31% of files were lost in Analytical Index Calculation – Mostly in the 2.5-minute set.

Compression	Recording Length	AudioSet Fingerprint Readings	Analytical Index Readings
Raw	20min	647	645
CBR320	20min	648	646
CBR256	20min	648	647
CBR128	20min	648	648
CBR64	20min	647	648
CBR32	20min	648	648
CBR16	20min	648	646
CBR8	20min	648	648
VBR0	20min	648	645
Raw	10min	1292	1292
CBR320	10min	1292	1292
CBR256	10min	1292	1291
CBR128	10min	1292	1292
CBR64	10min	1292	1291
CBR32	10min	1278	1292
CBR16	10min	1275	1291
CBR8	10min	1292	1292
VBR0	10min	1283	1291
Raw	5min	2584	2584
CBR320	5min	2584	2583
CBR256	5min	2584	2578
CBR128	5min	2584	2579
CBR64	5min	2584	2584
CBR32	5min	2584	2584
CBR16	5min	2584	2584
CBR8	5min	2584	2584
VBR0	5min	2584	2583
Raw	2.5min	5184	5176
CBR320	2.5min	5179	5160
CBR256	2.5min	5184	5180
CBR128	2.5min	5183	5168
CBR64	2.5min	5184	5164
CBR32	2.5min	5189	5168
CBR16	2.5min	5184	5183
CBR8	2.5min	5184	5168
VBR0	2.5min	5184	5156
TOTAL		87,329	87,211
Files Lost:		151	269

3 – Analytical Index Meanings

Table 3 was adapted from (Bradfer-Lawrence et al., 2019), provided here with permission.

Index and Acronym	Measures	High Value Indicates	Low Value Indicates	Reference
Acoustic Complexity Index (ACI)	Frequency band dependant changes in amplitude over time. A measure to designed to quantify acoustic irregularity typical of birdsong.	Storms and intermittent raindrops. High levels of avian vocalisations and insect stridulation	Constant noise filling the whole spectrogram e.g. cicadas or wind	(Pieretti et al., 2011)
Acoustic Diversity Index (ADI)	A frequency band dependant measure of the proportion of signals above -50dBFS, ADI is calculated from the Shannon index of each band	High levels of geophonic or anthrophonic noise Very quiet recordings with a minimal signal in any frequency band	The dominance of a particular frequency band such as insect noise	(Villanueva-Rivera et al., 2011)
Acoustic Evenness (AEve)	A frequency band dependant measure of the proportion of signals above -50dBFS, AEve calculated from the Gini index of each band	The reverse of ADI, the dominance of a narrow frequency band relating to insect noise	Large amounts of geophony or near silence Sometimes acoustically saturated soundscapes	(Villanueva-Rivera et al., 2011)
Bioacoustic Index (Bio)	A measure of both amplitude and number of occupied frequency bands between 2 and 11kHz, relative to the quietest 1kHz frequency band	The greater dissimilarity between quiet and loud bands Loud tonal insect noise such as cicadas	No sound in the given range, although some biophony does occur outside of this range	(Boelman et al., 2007)
Acoustic Entropy (H)	Measures spread of amplitude across frequency bands and/or time steps. Between 1 and 0, (diffuse and concentrated sound respectively)	Near silent recordings and faint bird calls Rain and wind	When insect noise dominates a single frequency band	(Jérôme Sueur et al., 2008)
Median of the Amplitude Envelope (M)	Measure of recording amplitude	High levels of geophony, particularly storms	Quiet recordings	(Jérôme Sueur et al., 2014)
Normalised Difference Soundscape Index (NDSI)	Ratio of anthrophony (1-2kHz) and biophony(2-11kHz)	High levels of biophony, particularly insects	When insect biophony dominates 1-2kHz range instead	(Kasten et al., 2012)

4 – Index Autocorrelation (Figure 2)

Figure 4.1 shows a correlation matrix of Analytical Indices, this is the same as the figure in the main test, provided here for comparisson.

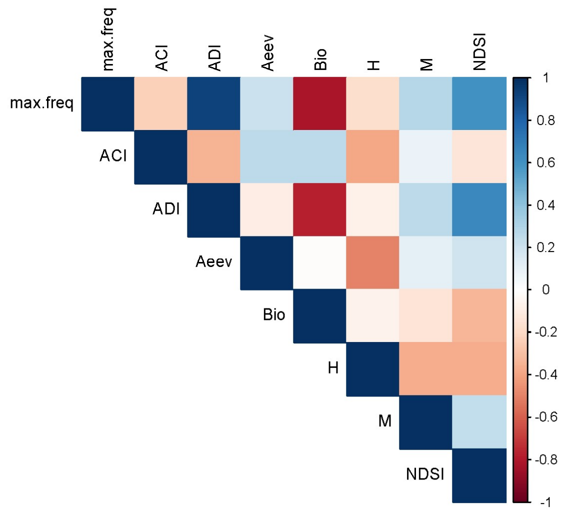
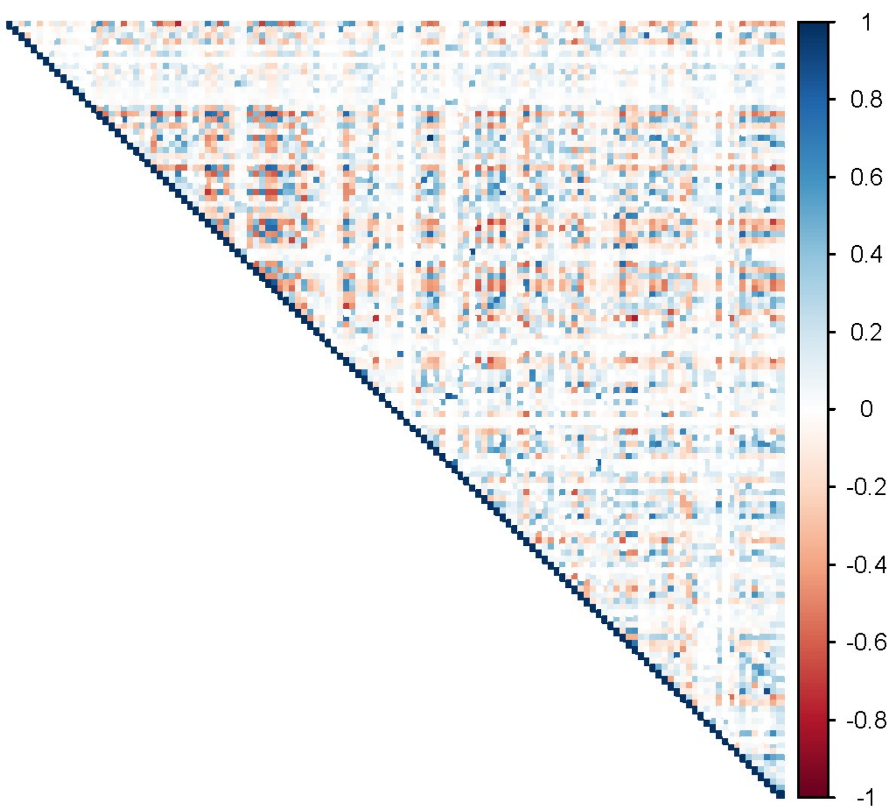


Figure 4.2 shows a correlation matrix of all the features given in the AudioSet Fingerprint (feat1, feat2, feat3…) etc.



5 – Impact of Compression: Like-for-Like Differences

AudioSet Fingerprint

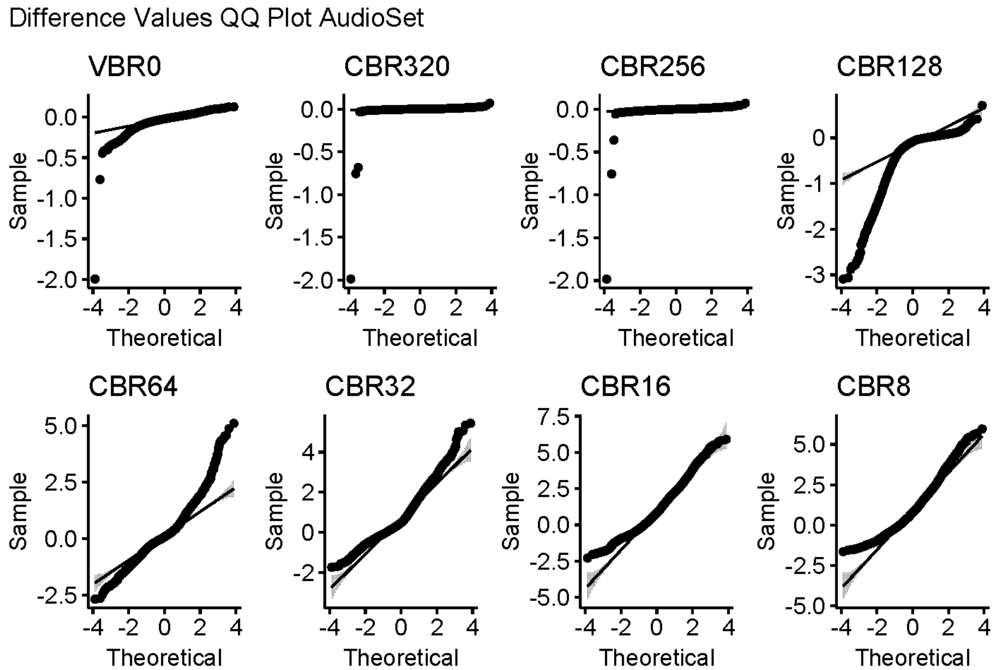


Figure 5.1.1 Q-Q plots of the like-for-like differences between AudioSet Fingerprint values from compressed vs. raw audio files.

Analytical Indices

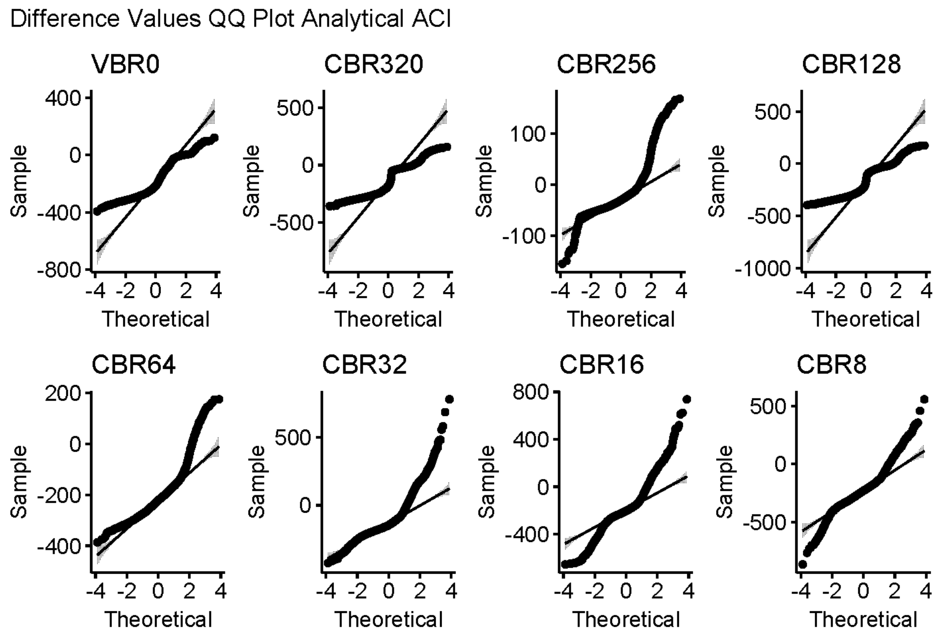


Figure 5.1.2 Q-Q plots of the like-for-like differences between ACI values from compressed vs. raw audio files.

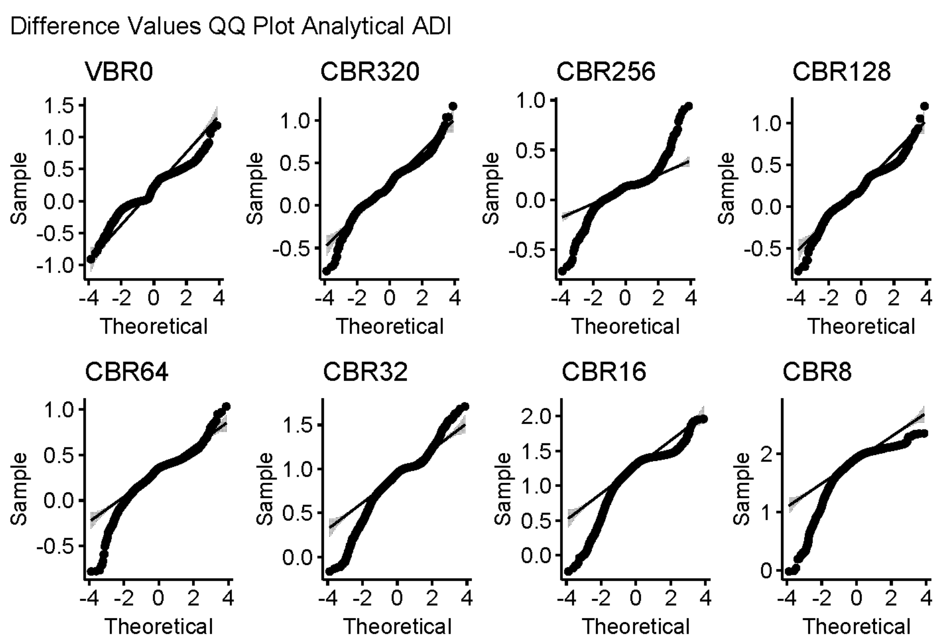


Figure 5.1.3 Q-Q plots of the like-for-like differences between ADI values from compressed vs. raw audio files.

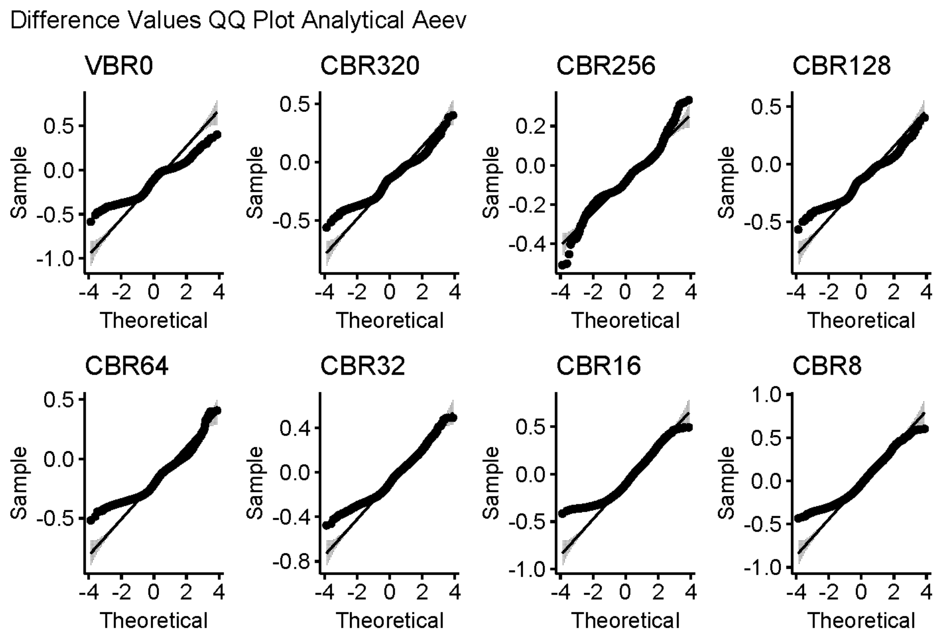


Figure 5.1.4 Q-Q plots of the like-for-like differences between AEve values from compressed vs. raw audio files.

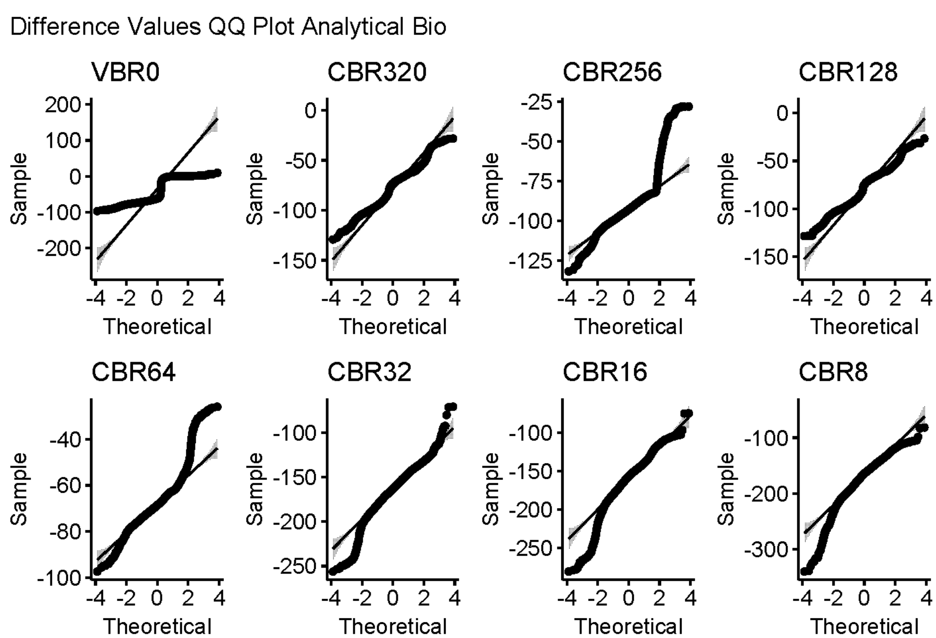


Figure 5.1.5 Q-Q plots of the like-for-like differences between Bio values from compressed vs. raw audio files.

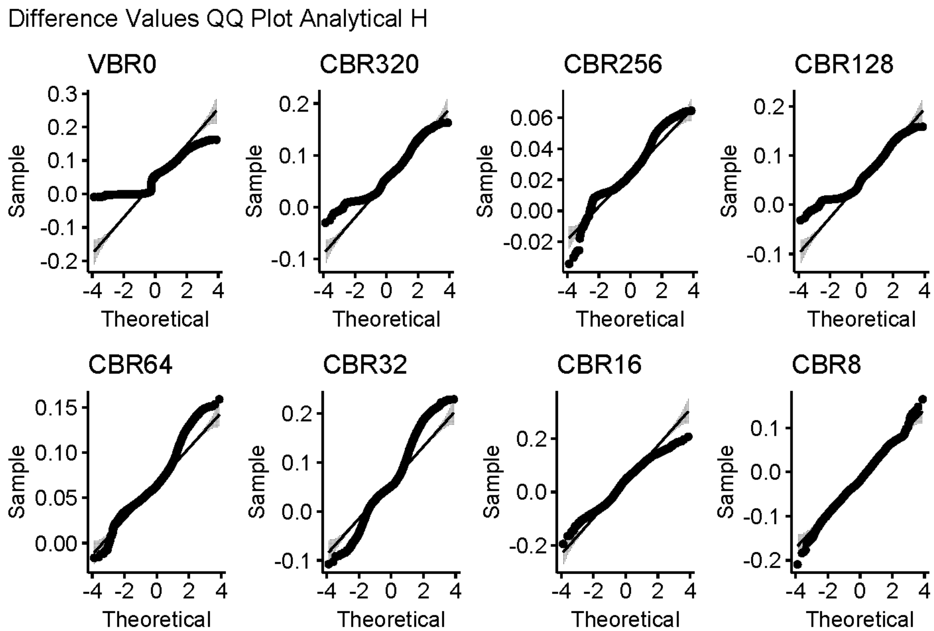


Figure 5.1.6 Q-Q plots of the like-for-like differences between H values from compressed vs. raw audio files

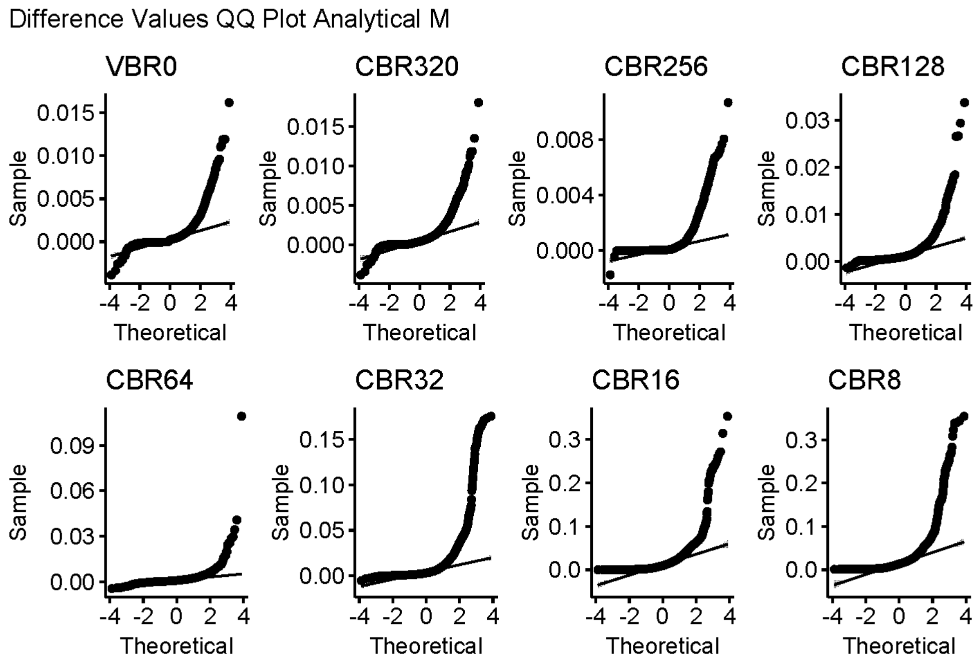


Figure 5.1.7 Q-Q plots of the like-for-like differences between M values from compressed vs. raw audio files.

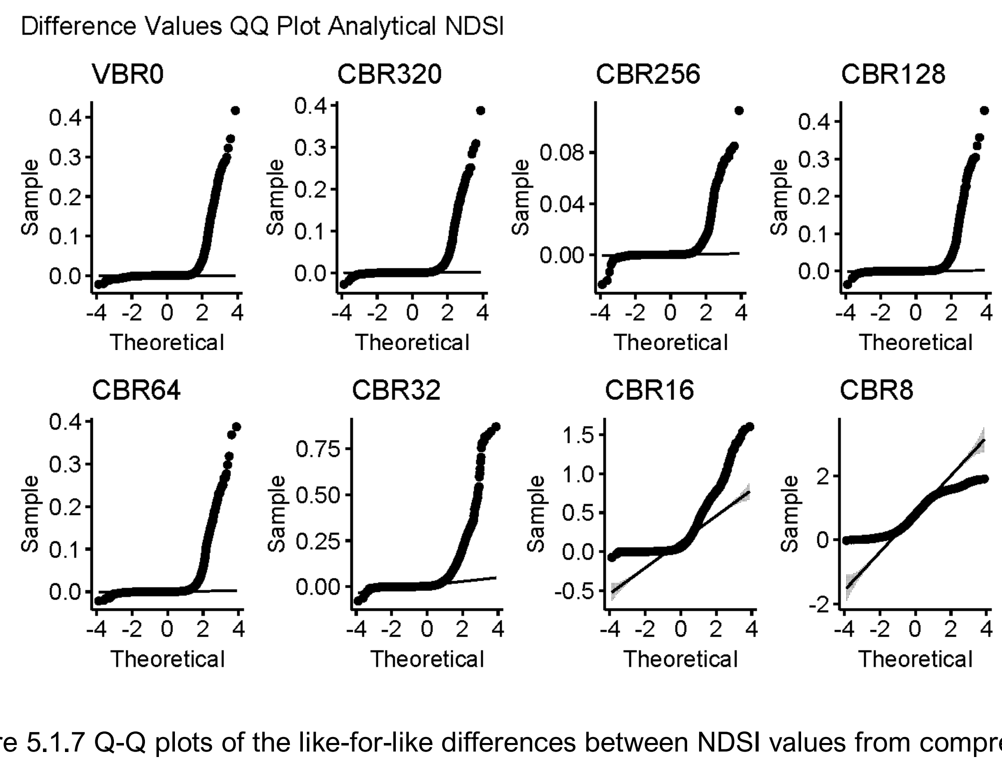


Figure 5.1.7 Q-Q plots of the like-for-like differences between NDSI values from compressed vs. raw audio files.

5b – Bland-Altmann plots from all recording lengths

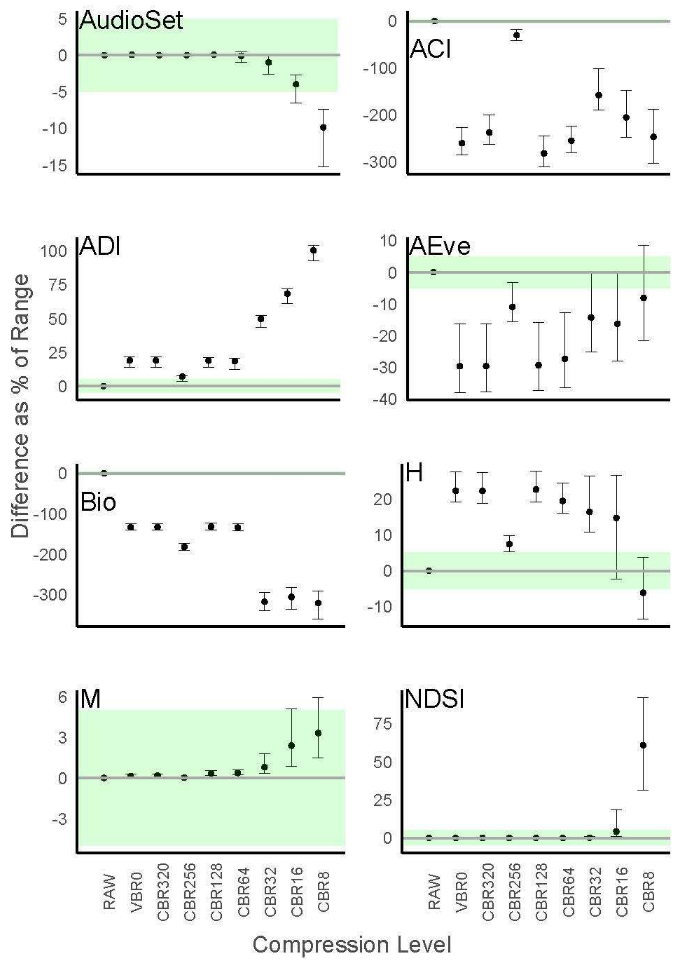


Figure 5.2.1 Scaled difference in acoustic indices from raw audio with increasing compression in **2.5-minute** audio samples. The horizontal green region shows the ±5% Difference. Dots and whiskers show the median and interquartile range of D (difference) from different indices under increasing levels of compression.

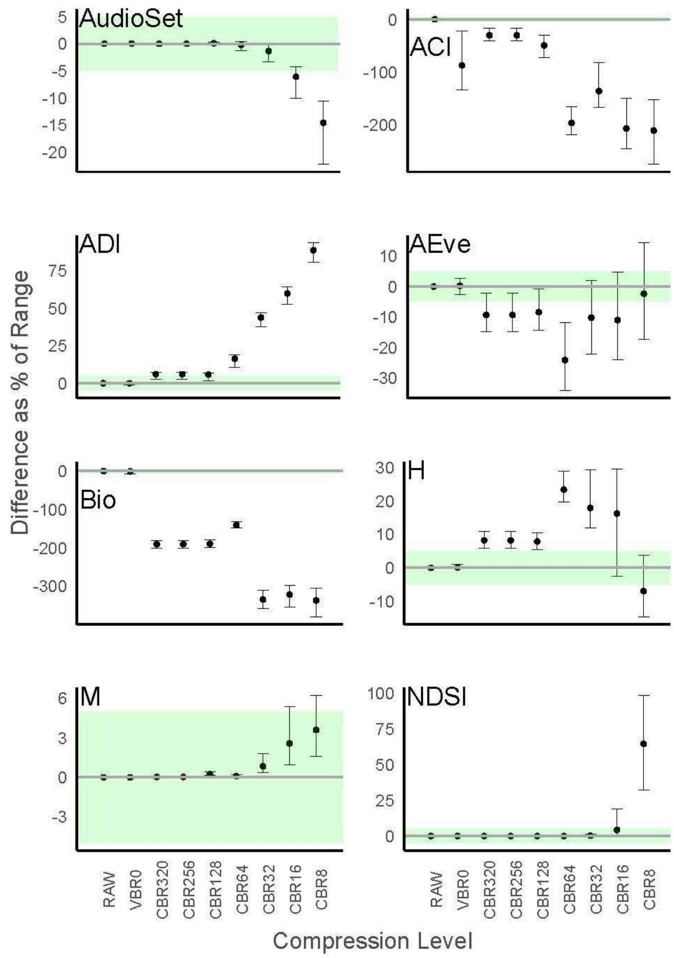


Figure 5.2.2 Scaled difference in acoustic indices from raw audio with increasing compression in **5-minute** audio samples (same as in main thesis). The horizontal green region shows the ±5% Difference. Dots and whiskers show the median and interquartile range of D (difference) from different indices under increasing levels of compression.

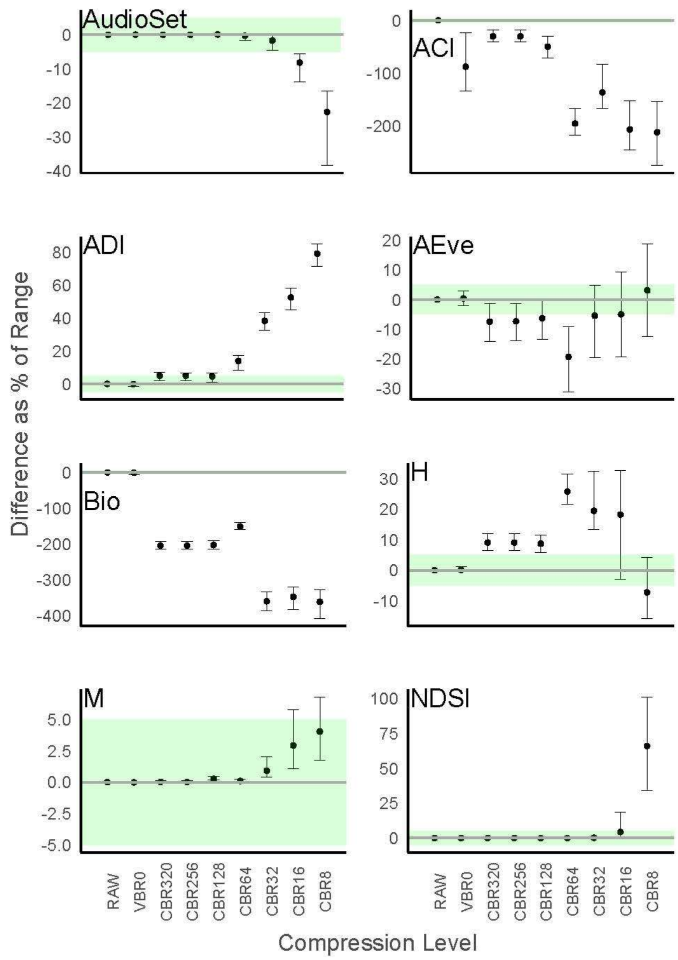


Figure 5.2.3 Scaled difference in acoustic indices from raw audio with increasing compression in **10-minute** audio samples. The horizontal green region shows the ±5% Difference. Dots and whiskers show the median and interquartile range of D (difference) from different indices under increasing levels of compression.

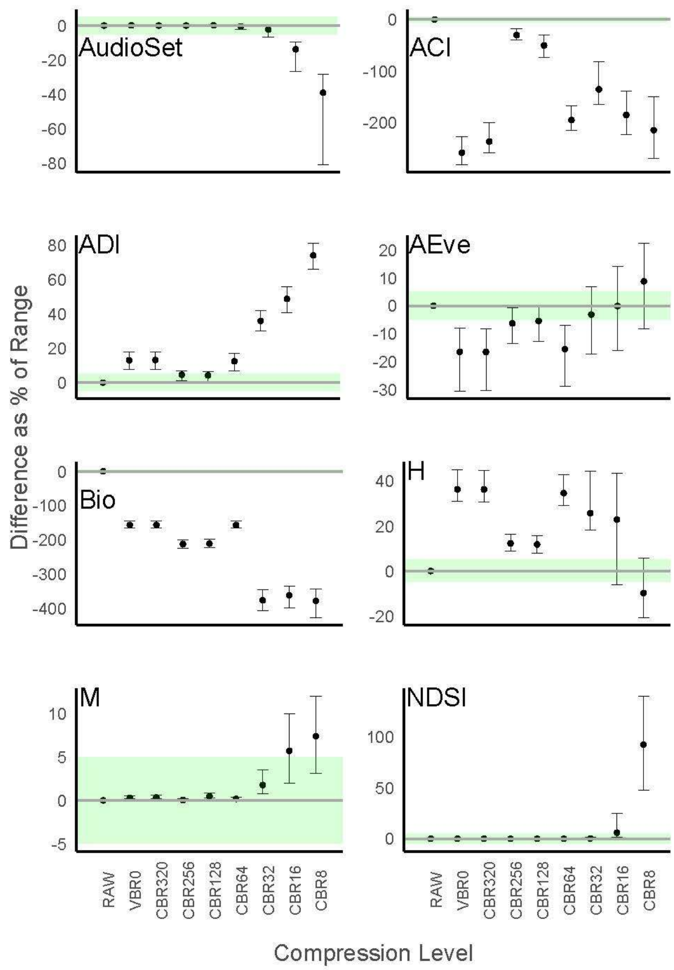


Figure 5.2.4 Scaled difference in acoustic indices from raw audio with increasing compression in **20-minute** audio samples. The horizontal green region shows the ±5% Difference. Dots and whiskers show the median and interquartile range of D (difference) from different indices under increasing levels of compression.

5c – Model Outputs (Spearman's Rank Correlation Rho)

Table 5.3 gives the model outputs from a Spearman's Rank Correlation.

Index	Recording Length	p-value	S	rho	Alternative hypothesis: Rho is not equal to 0
ACI	All	< 2.2e-16	5.9908e+13	0.4528838	True
ADI	All	< 2.2e-16	2.0076e+1	-0.8334471	True
AEve	All	< 2.2e-16	1.0431e+14	0.04732955	True
Bio	All	< 2.2e-16	2.7865e+13	0.7455166	True
H	All	< 2.2e-16	1.1409e+14	-0.04198094	True
M	All	< 2.2e-16	1.9495e+14	-0.7803825	True
NDSI	All	< 2.2e-16	1.8306e+14	-0.6718236	True
AudioSet (+/-)	All	< 2.2e-16	1.4664e+14	-0.3216693	True
AudioSet (absolute)	All	< 2.2e-16	2.1968e+14	-0.9800273	True
ACI	20min	< 2.2e-16	2.3062e+10	0.2897066	True
ADI	20min	< 2.2e-16	5.7586e+10	-0.7736128	True
AEve	20min	< 2.2e-16	3.7679e+10	-0.1604924	True
Bio	20min	< 2.2e-16	6581545650	0.7972928	True
H	20min	0.0001731	3.0867e+10	0.04930388	True
M	20min	< 2.2e-16	5.678e+10	-0.7487885	True
NDSI	20min	< 2.2e-16	5.4404e+10	-0.6756143	True
AudioSet (+/-)	20min	< 2.2e-16	4.3576e+10	-0.324907	True
AudioSet (absolute)	20min	< 2.2e-16	6.5184e+10	-0.9818801	True
ACI	10min	< 2.2e-16	7.7613e+10	0.6993322	True
ADI	10min	< 2.2e-16	4.6993e+11	-0.820476	True
AEve	10min	7.569e-05	2.4864e+11	0.03678488	True
Bio	10min	< 2.2e-16	7.1987e+10	0.7211259	True
H	10min	< 2.2e-16	3.011e+11	-0.1664376	True
M	10min	< 2.2e-16	4.5137e+11	-0.7485807	True
NDSI	10min	< 2.2e-16	4.3252e+11	-0.6755327	True
AudioSet (+/-)	10min	< 2.2e-16	3.4289e+11	-0.3224749	True
AudioSet (absolute)	10min	< 2.2e-16	5.1372e+11	-0.9813791	True
ACI	5min	< 2.2e-16	6.2663e+11	0.6977385	True
ADI	5min	< 2.2e-16	3.8019e+12	-0.8339001	True
AEve	5min	< 2.2e-16	1.7817e+12	0.1405795	True
Bio	5min	< 2.2e-16	5.7816e+11	0.7211196	True
H	5min	< 2.2e-16	2.4202e+12	-0.1674052	True
M	5min	< 2.2e-16	3.6248e+12	-0.7484442	True
NDSI	5min	< 2.2e-16	3.4694e+12	-0.6734986	True
AudioSet (+/-)	5min	< 2.2e-16	2.7754e+12	-0.3239534	True
AudioSet (absolute)	5min	< 2.2e-16	4.1512e+12	-0.9802393	True
ACI	2.5min	< 2.2e-16	1.1169e+13	0.3289954	True
ADI	2.5min	< 2.2e-16	3.0666e+13	-0.8423228	True

AEve	2.5min	0.00296	1.6416e+13	0.01379665	True
Bio	2.5min	< 2.2e-16	4.0057e+12	0.7593467	True
H	2.5min	< 2.2e-16	1.5846e+13	0.04798813	True
M	2.5min	< 2.2e-16	3.0169e+13	-0.8124801	True
NDSI	2.5min	< 2.2e-16	2.7774e+13	-0.6686069	True
AudioSet (+/-)	2.5min	< 2.2e-16	2.2333e+13	-0.3199297	True
AudioSet (absolute)	2.5min	< 2.2e-16	3.3499e+13	-0.979805	True

6 – Impact of Recording Schedule: Recording Length on Variance

Table 6 shows the outputs of a Levene's test for homogeneity of variance. Whereby the variance each index/feature is compared across all recording lengths. For Analytical Indices 3/7 has unequal variances across recording lengths. For the AudioSet Fingerprint, 46/128 features have unequal variance across recording lengths.

index	P value
ACI	0.841637
ADI	1.78E-39
AEve	0.028428
Bio	0.794537
H	0.004092
M	0.067758
NDSI	0.714233
feat1	0.837382
feat2	0.959618
feat3	1.58E-06
feat4	0.988474
feat5	0.981484
feat6	0.912722
feat7	0.980983
feat8	0.99822
feat9	0.800522
feat10	0.119387
feat11	0.999692
feat12	0.901465
feat13	0.989909
feat14	0.000972
feat15	8.75E-06
feat16	0.781888
feat17	2.41E-14
feat18	0.407886
feat19	0.955162
feat20	0.406722
feat21	4.56E-11
feat22	0.092009
feat23	0.99988
feat24	0.01768
feat25	2.69E-12
feat26	0.996596
feat27	0.001689
feat28	0.185621
feat29	0.50037
feat30	0.410133

Index	P value
feat42	0.706519
feat43	2.68E-05
feat44	3.54E-07
feat45	0.779526
feat46	0.892285
feat47	0.000662
feat48	1.05E-09
feat49	0.000274
feat50	9.91E-06
feat51	0.999517
feat52	0.002494
feat53	0.723206
feat54	0.820143
feat55	0.462347
feat56	2.27E-06
feat57	4.58E-13
feat58	0.99986
feat59	0.987711
feat60	0.005519
feat61	0.522512
feat62	0.968115
feat63	0.970316
feat64	0.996904
feat65	0.001259
feat66	0.002933
feat67	0.73536
feat68	0.000104
feat69	0.999244
feat70	0.993015
feat71	0.934927
feat72	0.03893
feat73	0.930866
feat74	0.972913
feat75	0.563142
feat76	0.000475
feat77	7.73E-05
feat78	1.63E-08

Index	P value
feat90	0.15218
feat91	0.932792
feat92	1.96E-15
feat93	0.906691
feat94	1.01E-08
feat95	0.86466
feat96	2.87E-09
feat97	0.824457
feat98	0.979221
feat99	0.797975
feat100	0.985947
feat101	0.998016
feat102	1.46E-07
feat103	2.92E-05
feat104	0.093257
feat105	0.997949
feat106	0.999947
feat107	0.925252
feat108	0.997711
feat109	0.334764
feat110	6.68E-05
feat111	0.998584
feat112	1.28E-08
feat113	7.91E-14
feat114	0.284966
feat115	0.974201
feat116	0.99686
feat117	0.999671
feat118	0.519933
feat119	0.304281
feat120	5.41E-11
feat121	0.050387
feat122	0.983873
feat123	0.000172
feat124	8.04E-10
feat125	0.693315
feat126	0.994528

feat31	0.255612
feat32	0.974623
feat33	0.01168
feat34	2.89E-10
feat35	0.867947
feat36	3.25E-09
feat37	0.995959
feat38	6.07E-06
feat39	0.092141
feat40	3.58E-05
feat41	5.58E-11

feat79	3.75E-14
feat80	0.526298
feat81	0.924814
feat82	0.800698
feat83	0.939481
feat84	0.320144
feat85	0.289219
feat86	4.56E-08
feat87	0.998176
feat88	3.07E-05
feat89	0.981046

feat127	0.004084
feat128	0.553217

7 – Impact of Index Type: Confusion Matrices (Figure 2)

7a – Confusion Matrices for all Recording Lengths

Table 7.1.1. Confusion matrices from random forest classifiers trained on AudioSet Fingerprint (left) and Analytical Indices (right) using uncompressed raw audio. Confusion matrices are organised in rows by recording length.

Analytical	**20min**			AudioSet			
Grassland	124	22	6	Grassland	147	4	1
Secondary	29	99	11	Secondary	2	126	13
Primary	5	11	123	Primary	2	4	133

Analytical	**10min**			AudioSet			
Grassland	235	49	13	Grassland	292	6	2
Secondary	46	213	20	Secondary	3	253	26
Primary	7	23	248	Primary	6	8	264

Analytical	**5min**			AudioSet			
Grassland	484	67	49	Grassland	585	9	11
Secondary	97	421	46	Secondary	11	508	44
Primary	9	61	486	Primary	17	14	521

Analytical	**2.5min**			AudioSet			
Grassland	1011	126	75	Grassland	1160	24	32
Secondary	193	837	97	Secondary	28	996	104
Primary	13	138	957	Primary	33	39	1040

7b – Accuracy, Precision and Recall Statistics

Table 7.1.2. Accuracy, precision and recall of random forest classifiers trained on each test group using uncompressed raw audio.

Index Type	Recording Length	Accuracy	Precision	Recall
Analytical	20Min	80.47	80.45	80.43
Analytical	10Min	81.50	81.53	81.56
Analytical	5Min	80.87	80.79	80.91
Analytical	2.5Min	81.38	81.28	81.35
AudioSet	20Min	93.98	93.95	93.92
AudioSet	10Min	94.06	94.06	94.00
AudioSet	5Min	93.84	93.85	93.77
AudioSet	2.5min	92.48	93.85	93.92

8 – Impact of Temporal Splitting (Figure 4)

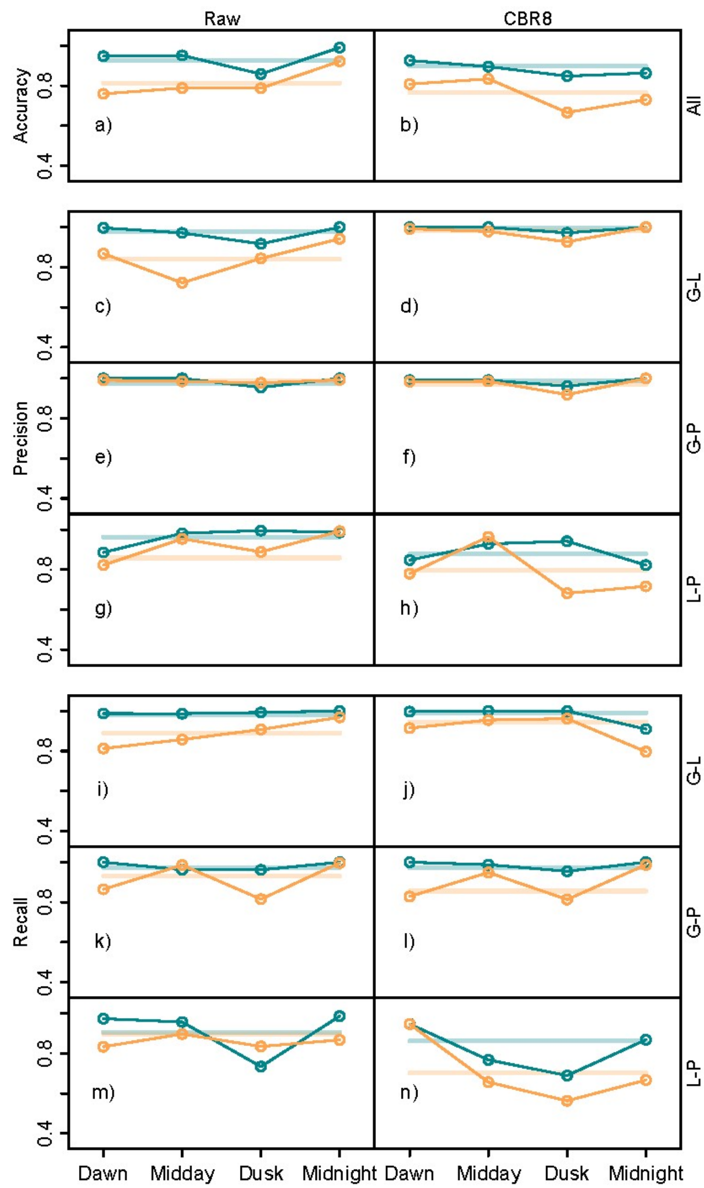


*Figure 8.1.1 Classification model performance as a function of temporal sectioning (x-axis), compression (raw audio, left column; CBR8 compression, right column) and index choice (AudioSet Fingerprint: blue; Analytical Indices: orange). This includes just models train on **2.5-minute audio recordings.** Pale horizontal lines show performance without temporal*

sectioning. Precision and recall are partitioned into pairwise performance by site (C, cleared forest; L, logged forest; P, primary forest).

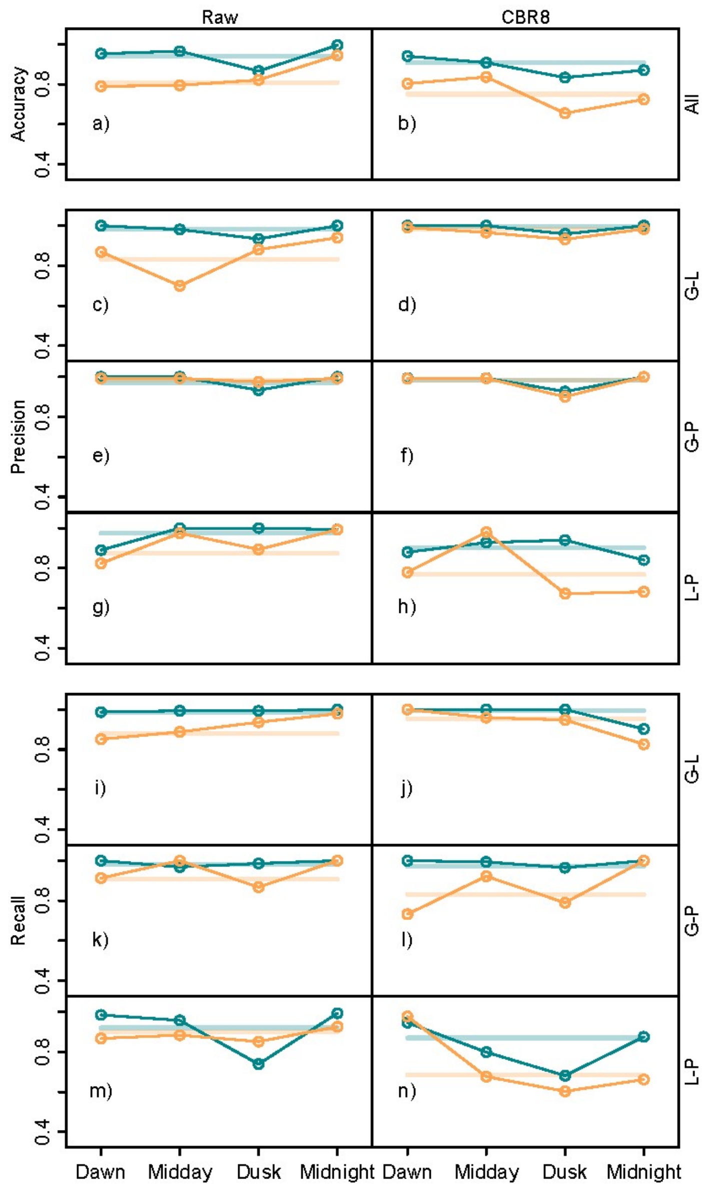


*Figure 8.1.2 Classification model performance as a function of temporal sectioning (x-axis), compression (raw audio, left column; CBR8 compression, right column) and index choice (AudioSet Fingerprint: blue; Analytical Indices: orange). This includes just models train on **5-***

minute audio recordings. *Pale horizontal lines show performance without temporal sectioning. Precision and recall are partitioned into pairwise performance by site (C, cleared forest; L, logged forest; P, primary forest).*

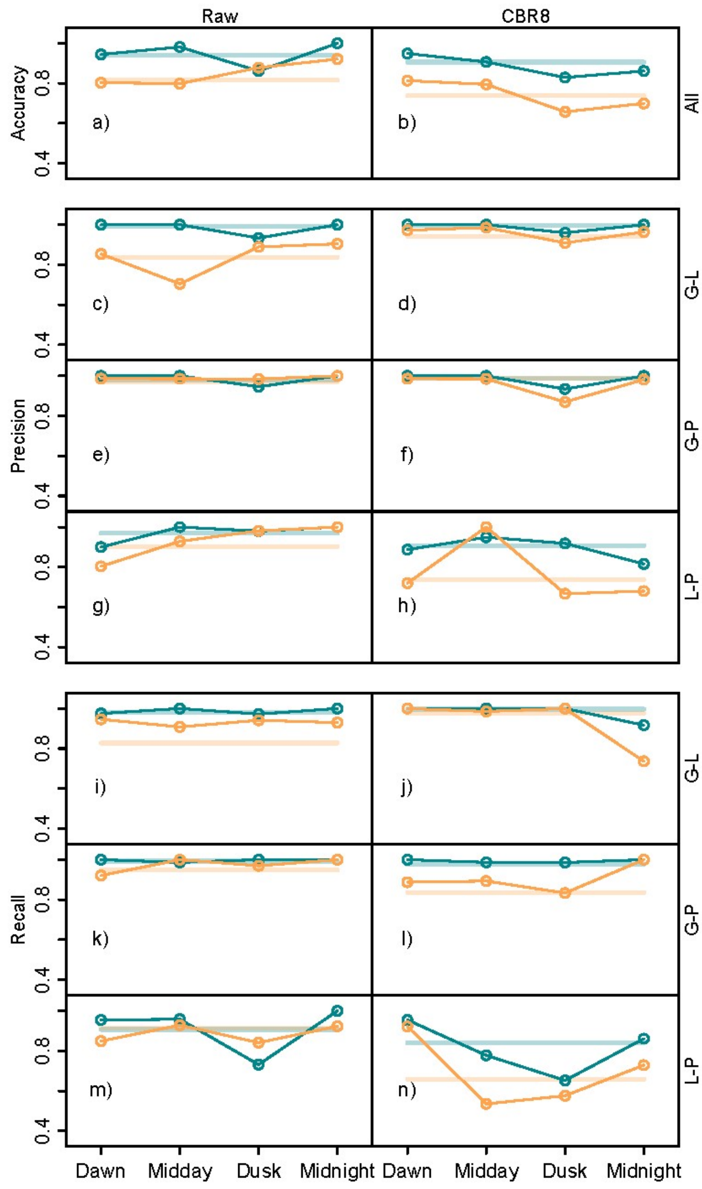


*Figure 8.1.3 Classification model performance as a function of temporal sectioning (x-axis), compression (raw audio, left column; CBR8 compression, right column) and index choice (AudioSet Fingerprint: blue; Analytical Indices: orange). This includes just models train on **10-minute audio recordings.*** *Pale horizontal lines show performance without temporal*

sectioning. Precision and recall are partitioned into pairwise performance by site (C, cleared forest; L, logged forest; P, primary forest).

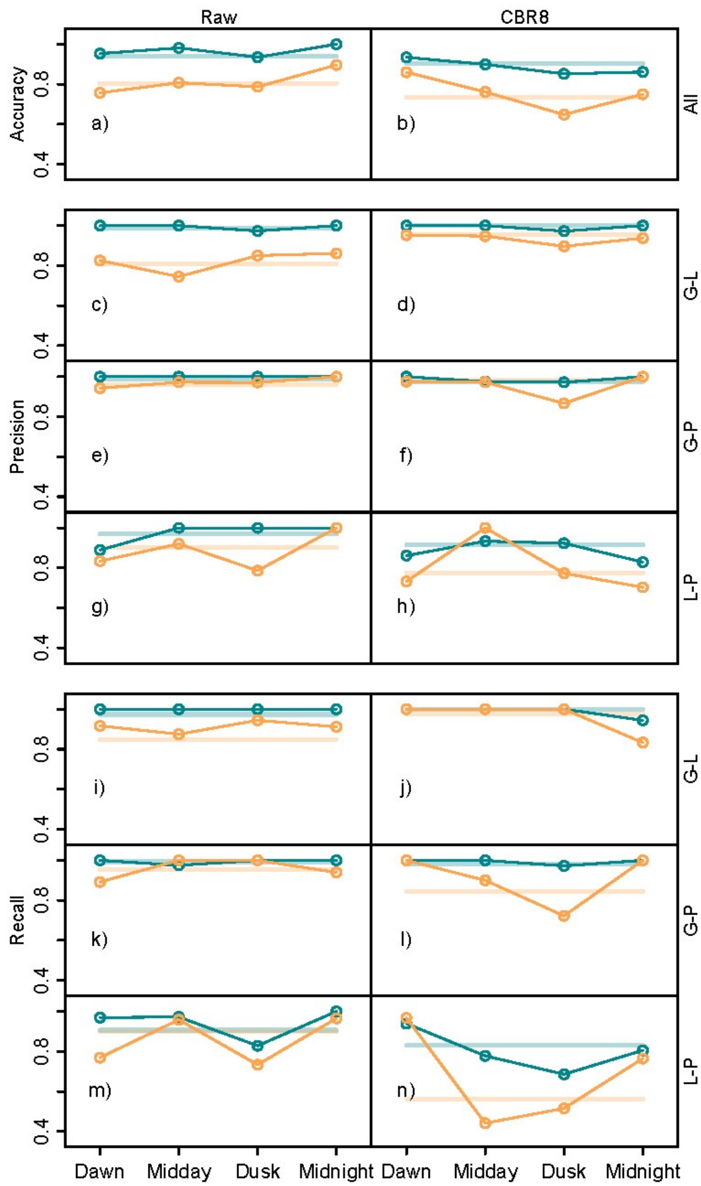


*Figure 8.1.4 Classification model performance as a function of temporal sectioning (x-axis), compression (raw audio, left column; CBR8 compression, right column) and index choice (AudioSet Fingerprint: blue; Analytical Indices: orange). This includes just models train on **20-***

minute audio recordings. *Pale horizontal lines show performance without temporal sectioning. Precision and recall are partitioned into pairwise performance by site (C, cleared forest; L, logged forest; P, primary forest).*

9 – Synthesis: Beta Regression Modelling the Contribution of all Parameters (Figure 5)

9a – Data Variance (ϕ component)

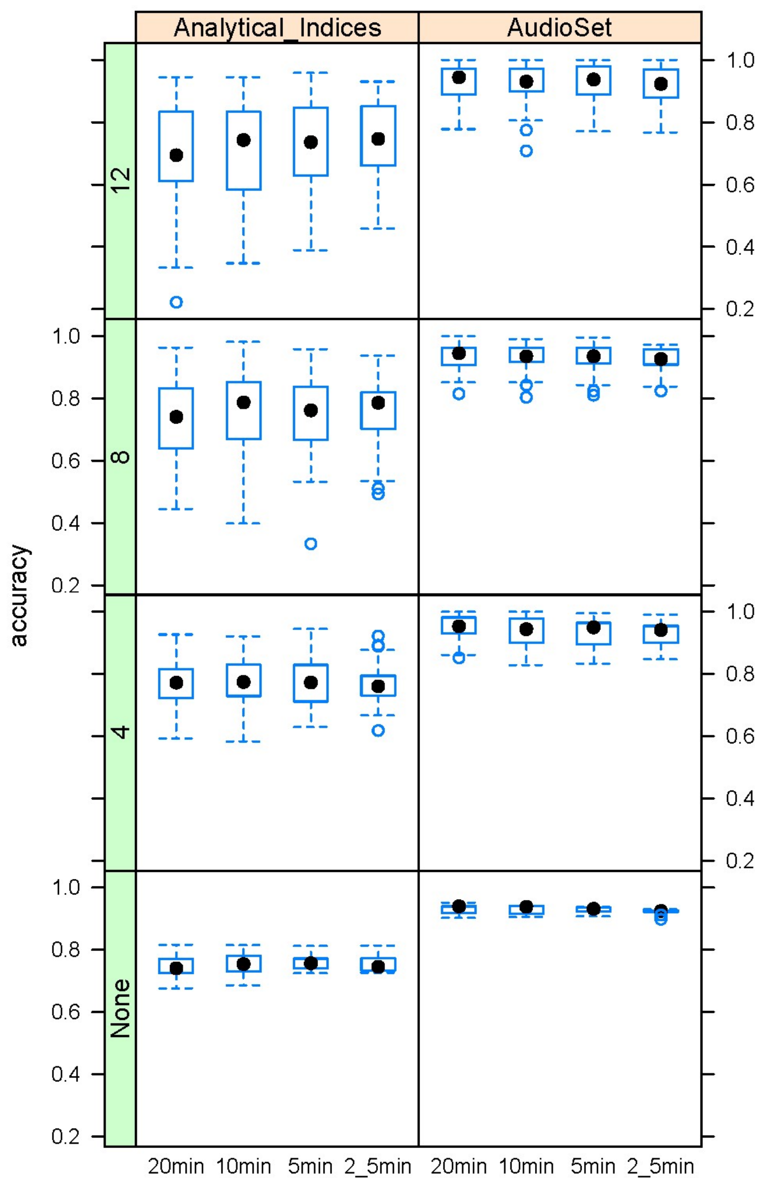


Figure 9.1 shows the range of accuracies of models trained on either Analytical Indices or the AudioSet Fingerprint, for recordings across all recording lengths. Data in this figure is all from raw uncompressed recordings.

9b – AIC values of models

Table 9.2 shows the Akaike Information Criterion (AIC) for different models used to explain variation in accuracy as a result of experimental variation. The best model is indicated by the highest AIC, in this case Beta Regression + φ.

Model	Df	AIC
Linear Model	47	-3428.738
Beta Regression	47	-4368.424
Beta Regression + φ	**54**	**-4682.399**
Beta Regression + φ (continuous temporal splits + recording length)	18	-4516.453

Maximal Model: accuracy ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type

9c Precision and Recall Figure Outputs

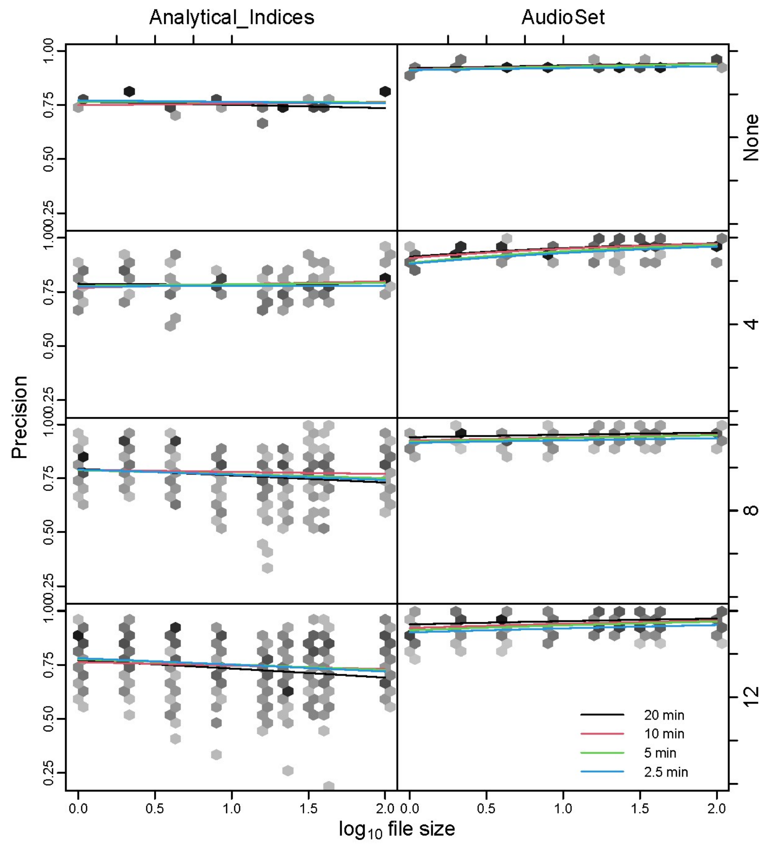


Figure 9.3.1 Classifier **precision** for model predictions as a function of file size (x-axis), index type (columns), temporal subsetting (rows), and Recording Length (colours, see legend) Hexagon binning is used to show the distribution and density of the underlying data.

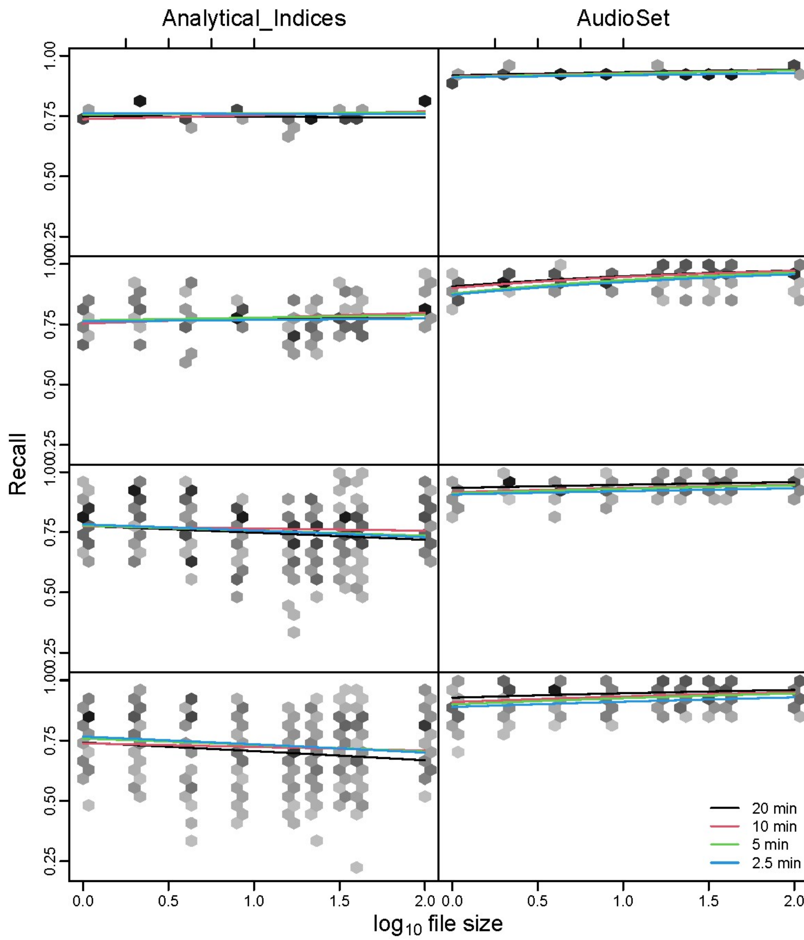


Figure 9.3.1 Classifier **recall** for model predictions as a function of file size (x-axis), index type (columns), temporal subsetting (rows), and Recording Length (colours, see legend) Hexagon binning is used to show the distribution and density of the underlying data.

9d – Accuracy Output Statistics (using Discrete Variables)

Table 9.4.1 AOV outputs for the multivariate linear model (**accuracy**)

	df	Sum Sq	Mean Sq	F Value	Pr(>F)
log10(file.size)	1	0.000	0.000	0.000	0.99740
chunks	3	0.243	0.081	9.644	2.59e-06 ***
frame.size	3	0.007	0.002	0.271	0.84646
index.type	1	15.980	15.980	1899.119	< 2e-16 ***
log10(file.size):chunks	3	0.049	0.016	1.947	0.12002
log10(file.size):frame.size	3	0.008	0.003	0.301	0.82446
chunks:frame.size	9	0.016	0.002	0.215	0.99238
log10(file.size):index.type	1	0.288	0.288	34.231	5.83e-09 ***
chunks:index.type	3	0.119	0.040	4.726	0.00274 **
frame.size:index.type	3	0.076	0.025	3.009	0.02920 *
log10(file.size):chunks: index.type	3	0.040	0.013	1.599	0.18771
log10(file.size):frame.size: index.type	3	0.016	0.005	0.634	0.59286
chunks:frame.size:index.type	9	0.009	0.001	0.121	0.99919
Residuals	1753	14.751	0.008		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Betareg Anova of the Maximal Model:

(accuracy ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type | index.type * chunks)

Table 9.4.2 Analysis of Deviance Table (Type II tests) for the beta regression

	Df	Chisq	Pr(>Chisq)
log10(file.size)	1	26.2128	3.058e-07 ***
chunks	3	31.6818	6.107e-07 ***
frame.size	3	15.7820	0.0012568 **
index.type	1	2985.9825	< 2.2e-16 ***
log10(file.size):chunks	3	18.0278	0.0004341 ***
log10(file.size):frame.size	3	2.9280	0.4028558
chunks:frame.size	9	6.3156	0.7079609
log10(file.size):index.type	1	59.0065	1.572e-14 ***
chunks:index.type	3	7.1061	0.0685927 .
frame.size:index.type	3	36.2699	6.566e-08 ***
log10(file.size):chunks: index.type	3	13.0715	0.0044844 **

log10(file.size):frame.size:index.type	3	0.8071	0.8477715
chunks:frame.size:index.type	9	7.1524	0.6212537

Maxima Model Summary:

betareg(formula = accuracy_t ~ log10(file.size) + chunks + frame.size + index.type +
log10(file.size):chunks + log10(file.size):frame.size + chunks:frame.size +
log10(file.size):index.type + chunks:index.type + frame.size:index.type +
log10(file.size):chunks:index.type + log10(file.size):frame.size:index.type+
chunks:frame.size:index.type | index.type * chunks, data = dat)

Table 9.4.3. Standardized weighted residuals 2:

Min	1Q	Median	3Q	Max
-2.8771	-0.7693	-0.2106	0.5468	6.6999

Table 9.4.4 ChiSq test for significance of relationship between accuracy, variables, and interactions.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.104521	0.100877	10.949	< 2e-16 ***
log10(file.size)	-0.015526	0.072085	-0.215	0.82946
chunks4	0.083034	0.130390	0.637	0.52425
chunks8	0.138091	0.131426	1.051	0.29339
chunks12	-0.058062	0.121445	-0.478	0.63259
frame.size10min	-0.070346	0.124861	-0.563	0.57317
frame.size5min	0.031137	0.125578	0.248	0.80417
frame.size2_5min	0.055264	0.125603	0.440	0.65995
index.typeAudioSet	1.334952	0.128606	10.380	< 2e-16 ***
log10(file.size):chunks4	0.046597	0.080474	0.579	0.56257
log10(file.size):chunks8	-0.138090	0.080784	-1.709	0.08738 .
log10(file.size):chunks12	-0.160987	0.074907	-2.149	0.03162 *
log10(file.size):frame.size10min	0.101558	0.079187	1.283	0.19967
log10(file.size):frame.size5min	0.038708	0.079340	0.488	0.62564
log10(file.size):frame.size2_5min	0.007661	0.079262	0.097	0.92300
chunks4:frame.size10min	-0.006499	0.139389	-0.047	0.96281
chunks8:frame.size10min	0.056064	0.139326	0.402	0.68740
chunks12:frame.size10min	0.061023	0.128661	0.474	0.63529
chunks4:frame.size5min	-0.043339	0.139628	-0.310	0.75627
chunks8:frame.size5min	-0.032370	0.139174	-0.233	0.81608
chunks12:frame.size5min	0.061505	0.129137	0.476	0.63388
chunks4:frame.size2_5min	-0.091839	0.139014	-0.661	0.50884
chunks8:frame.size2_5min	-0.025140	0.139089	-0.181	0.85656
chunks12:frame.size2_5min	0.082253	0.129125	0.637	0.52412
log10(file.size):index.typeAudioSet	0.202437	0.095711	2.115	0.03442 *

chunks4:index.typeAudioSet	-0.234469	0.210918	-1.112	0.26628
chunks8:index.typeAudioSet	0.060060	0.176105	0.341	0.73307
chunks12:index.typeAudioSet	0.166156	0.178544	0.931	0.35205
frame.size10min: index.typeAudioSet	0.002228	0.163241	0.014	0.98911
frame.size5min:index.typeAudioSet	-0.122698	0.163322	-0.751	0.45249
frame.size2_5min:index.typeAudioSet	-0.165346	0.162560	-1.017	0.30909
log10(file.size):chunks4: index.typeAudioSet	0.420832	0.130630	3.222	0.00127 **
log10(file.size):chunks8: index.typeAudioSet	0.177120	0.106759	1.659	0.09710 .
log10(file.size):chunks12: index.typeAudioSet	0.273251	0.107516	2.541	0.01104 *
log10(file.size):frame.size10min: index.typeAudioSet	-0.085818	0.112621	-0.762	0.44605
log10(file.size):frame.size5min: index.typeAudioSet	-0.027098	0.112236	-0.241	0.80921
log10(file.size):frame.size2_5min: index.typeAudioSet	-0.077750	0.111174	-0.699	0.48433
chunks4:frame.size10min: index.typeAudioSet	-0.004506	0.233281	-0.019	0.98459
chunks8:frame.size10min: index.typeAudioSet	-0.240160	0.188273	-1.276	0.20210
chunks12:frame.size10min: index.typeAudioSet	-0.234850	0.188868	-1.243	0.21370
chunks4:frame.size5min: index.typeAudioSet	-0.171315	0.229595	-0.746	0.45557
chunks8:frame.size5min: index.typeAudioSet	-0.176300	0.187565	-0.940	0.34725
chunks12:frame.size5min: index.typeAudioSet	-0.315394	0.188434	-1.674	0.09418 .
chunks4:frame.size2_5min: index.typeAudioSet	-0.148262	0.227108	-0.653	0.51387
chunks8:frame.size2_5min: index.typeAudioSet	-0.227651	0.185826	-1.225	0.22055
chunks12:frame.size2_5min: index.typeAudioSet	-0.433908	0.186840	-2.322	0.02021 *

Table 9.4.5 Phi Coefficients (model with log link)

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	4.9232	0.2351	20.944	< 2e-16 ***
index.typeAudioSet	1.6943	0.3330	5.088	3.61e-07 ***
chunks4	-1.6147	0.2623	-6.155	7.50e-10 ***
chunks8	-2.4408	0.2487	-9.815	< 2e-16 ***
chunks12	-2.6482	0.2441	-10.851	< 2e-16 ***
index.typeAudioSet:chunks4	-1.7679	0.3741	-4.725	2.30e-06 ***

index.typeAudioSet:chunks8	-0.6451	0.3532	-1.826	0.0678 .
index.typeAudioSet:chunks12	-1.5081	0.3476	-4.339	1.43e-05 ***

Accuracy Model Outputs:

Type of estimator:	ML (maximum likelihood)
Log-likelihood:	2395 on 54 Df
Pseudo R-squared:	0.427
Number of iterations:	66 (BFGS) + 2 (Fisher scoring)

9e – Precision Output Statistics (using Discrete Variables)

Table 9.5.1 AOV outputs for the multivariate linear model (**precision**)

	df	**Sum Sq**	**Mean Sq**	**F Value**	**Pr(>F)**
log10(file.size)	1	0.004	0.004	0.586	0.44408
chunks	3	0.122	0.041	5.355	0.00114 **
frame.size	3	0.004	0.001	0.197	0.89841
index.type	1	14.264	14.264	1883.934	< 2e-16 ***
log10(file.size):chunks	3	0.040	0.013	1.741	0.15658
log10(file.size):frame.size	3	0.005	0.002	0.238	0.86978
chunks:frame.size	9	0.009	0.001	0.130	0.99892
log10(file.size):index.type	1	0.276	0.276	36.392	1.96e-09 ***
chunks:index.type	3	0.075	0.025	3.323	0.01906 *
frame.size:index.type	3	0.066	0.022	2.893	0.03416 *
log10(file.size):chunks:index.type	3	0.022	0.007	0.987	0.39801
log10(file.size):frame.size:index.type	3	0.017	0.006	0.748	0. 52322
chunks:frame.size:index.type	9	0.007	0.001	0.101	0.99962
Residuals	1753	13.273	0.008		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Betareg Anova of the Maximal Model:

(precision ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type | index.type * chunks)

Table 9.5.2 Analysis of Deviance Table (Type II tests) for the Beta regression

	Df	**Chisq**	**Pr(>Chisq)**
log10(file.size)	1	21.9304	2.827e-06 ***
chunks	3	29.6890	1.604e-06 ***
frame.size	3	19.7551	0.0001908 ***
index.type	1	3047.6774	< 2.2e-16 ***
log10(file.size):chunks	3	16.5601	0.0008703 ***
log10(file.size):frame.size	3	2.8885	0.4091298
chunks:frame.size	9	8.2666	0.5075198
log10(file.size):index.type	1	70.1343	< 2.2e-16 ***
chunks:index.type	3	10.0553	0.0181017 *
frame.size:index.type	3	37.7504	3.192e-08 ***
log10(file.size):chunks:index.type	3	12.2136	0.0066863 **
log10(file.size):frame.size:index.type	3	1.2606	0.7385054
chunks:frame.size:index.type	9	7.0018	0.6369337

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Maxima Model Summary:

Call:
betareg(formula = precision_t ~ log10(file.size) + chunks + frame.size + index.type + log10(file.size):chunks +
log10(file.size):frame.size + chunks:frame.size + log10(file.size):index.type + chunks:index.type + frame.size:index.type +
log10(file.size):chunks:index.type + log10(file.size):frame.size:index.type + chunks:frame.size:index.type |
index.type * chunks, data = dat)

Table 9.5.3 Standardized weighted residuals 2:

Min	1Q	Median	3Q	Max
-3.4204	-0.7584	-0.2320	0.5653	6.9159

Table 9.5.4 ChiSq test for significance of relationship between precision, variables, and interactions.

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	1.181140	0.102133	11.565	< 2e-16 ***
log10(file.size)	-0.078810	0.072539	-1.086	0.27728
chunks4	0.128346	0.132032	0.972	0.33101
chunks8	0.162282	0.132533	1.224	0.22078
chunks12	0.033704	0.122106	0.276	0.78253
frame.size10min	-0.088089	0.125963	-0.699	0.48435
frame.size5min	0.003497	0.126647	0.028	0.97797
frame.size2_5min	0.035712	0.126740	0.282	0.77812
index.typeAudioSet	1.258609	0.129528	9.717	< 2e-16 ***
log10(file.size):chunks4	0.048827	0.081153	0.602	0.54740
log10(file.size):chunks8	-0.095730	0.081253	-1.178	0.23873
log10(file.size):chunks12	-0.126488	0.074918	-1.688	0.09134 .
log10(file.size):frame.size10min	0.118991	0.079432	1.498	0.13413
log10(file.size):frame.size5min	0.069972	0.079556	0.880	0.37912
log10(file.size):frame.size2_5min	0.039836	0.079465	0.501	0.61616
chunks4:frame.size10min	-0.019950	0.140254	-0.142	0.88689
chunks8:frame.size10min	0.061675	0.140020	0.440	0.65959
chunks12:frame.size10min	0.042094	0.128671	0.327	0.74356
chunks4:frame.size5min	-0.058843	0.140566	-0.419	0.67550
chunks8:frame.size5min	-0.039343	0.139823	-0.281	0.77842
chunks12:frame.size5min	0.026060	0.129143	0.202	0.84008
chunks4:frame.size2_5min	-0.106103	0.140104	-0.757	0.44886
chunks8:frame.size2_5min	-0.055055	0.139700	-0.394	0.69351
chunks12:frame.size2_5min	0.020382	0.129112	0.158	0.87456
log10(file.size):index.typeAudioSet	0.262357	0.095836	2.738	0.00619 **
chunks4:index.typeAudioSet	-0.225830	0.209074	-1.080	0.28008
chunks8:index.typeAudioSet	0.145267	0.175036	0.830	0.40658
chunks12:index.typeAudioSet	0.225946	0.177678	1.272	0.20349

frame.size10min: index.typeAudioSet	0.027811	0.163868	0.170	0.86523
frame.size5min:index.typeAudioSet	-0.100711	0.163807	-0.615	0.53868
frame.size2_5min: index.typeAudioSet	-0.145433	0.163105	-0.892	0.37258
log10(file.size):chunks4: index.typeAudioSet	0.421647	0.129360	3.259	0.00112 **
log10(file.size):chunks8: index.typeAudioSet	0.134391	0.105921	1.269	0.20452
log10(file.size):chunks12: index.typeAudioSet	0.236834	0.106627	2.221	0.02634 *
log10(file.size):frame.size10min: index.typeAudioSet	-0.105328	0.112208	-0.939	0.34789
log10(file.size):frame.size5min: index.typeAudioSet	-0.049108	0.111744	-0.439	0.66032
log10(file.size):frame.size2_5min: index.typeAudioSet	-0.106684	0.110651	-0.964	0.33497
chunks4:frame.size10min: index.typeAudioSet	0.005458	0.230916	0.024	0.98114
chunks8:frame.size10min: index.typeAudioSet	-0.262740	0.186811	-1.406	0.15959
chunks12:frame.size10min: index.typeAudioSet	-0.233415	0.187695	-1.244	0.21365
chunks4:frame.size5min: index.typeAudioSet	-0.154663	0.227259	-0.681	0.49615
chunks8:frame.size5min: index.typeAudioSet	-0.189694	0.186022	-1.020	0.30785
chunks12:frame.size5min: index.typeAudioSet	-0.310507	0.187116	-1.659	0.09703 .
chunks4:frame.size2_5min: index.typeAudioSet	-0.135128	0.224834	-0.601	0.54783
chunks8:frame.size2_5min: index.typeAudioSet	-0.224587	0.184237	-1.219	0.22284
chunks12:frame.size2_5min: index.typeAudioSet	-0.409874	0.185403	-2.211	0.02706 *

Table 9.5.5 Phi Coefficients (model with log link)

	Estimate	**Std.Error**	**Z Value**	**Pr(>|z|)**
(Intercept)	4.9166	0.2351	20.916	< 2e-16 ***
index.typeAudioSet	1.6912	0.3330	5.079	3.79e-07 ***
chunks4	-1.5953	0.2624	-6.080	1.20e-09 ***
chunks8	-2.4061	0.2488	-9.672	< 2e-16 ***
chunks12	-2.5716	0.2442	-10.532	< 2e-16 ***
index.typeAudioSet:chunks4 -	1.6599	0.3740	-4.438	9.10e-06 ***
index.typeAudioSet:chunks8	-0.4915	0.3533	-1.391	0.164
index.typeAudioSet:chunks12	-1.3908	0.3477	-4.001	6.32e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator:	ML (maximum likelihood)
Log-likelihood:	2529 on 54 Df
Pseudo R-squared:	0.4346
Number of iterations:	66 (BFGS) + 2 (Fisher scoring)

9f – Recall Output Statistics (using Discrete Variables)

Table 9.6.1 AOV outputs for the multivariate linear model (**recall**)

	Df	Sum. Sq	Mean. Sq	F Value	Pr(>F)
log10(file.size)	1	0.000	0.000	0.027	0.86948
chunks	3	0.241	0.080	9.581	2.83e-06 ***
frame.size	3	0.010	0.003	0.394	0.75750
index.type	1	16.000	16.000	1904.901	< 2e-16 ***
log10(file.size):chunks	3	0.045	0.015	1.799	0.14531
log10(file.size):frame.size	3	0.007	0.002	0.288	0.83382
chunks:frame.size	9	0.016	0.002	0.214	0.99248
log10(file.size):index.type	1	0.300	0.300	35.664	2.83e-09 ***
chunks:index.type	3	0.131	0.044	5.198	0.00142 **
frame.size:index.type	3	0.084	0.028	3.340	0.01862 *
log10(file.size):chunks: index.type	3	0.041	0.014	1.625	0.18156
log10(file.size):frame.size: index.type	3	0.017	0.006	0.662	0.57556
chunks:frame.size:index.ty pe	9	0.009	0.001	0.116	0.99931
Residuals	1753	14.724	0.008		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Betareg Anova of the Maximal Model:

(recall ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type **|** index.type * chunks)

Table 9.6.2 Analysis of Deviance Table (Type II tests) for the Beta regression

	Df	Chisq	Pr(>Chisq)
log10(file.size)	1	21.9304	2.827e-06 ***
chunks	3	29.6890	1.604e-06 ***
frame.size	3	19.7551	0.0001908 ***
index.type	1	3047.6774	< 2.2e-16 ***
log10(file.size):chunks	3	16.5601	0.0008703 ***
log10(file.size):frame.size	3	2.8885	0.4091298
chunks:frame.size	9	8.2666	0.5075198
log10(file.size):index.type	1	70.1343	< 2.2e-16 ***
chunks:index.type	3	10.0553	0.0181017 *
frame.size:index.type	3	37.7504	3.192e-08 ***
log10(file.size):chunks:index.type	3	12.2136	0.0066863 **
log10(file.size):frame.size:index.type	3	1.2606	0.7385054
chunks:frame.size:index.type	9	7.0018	0.6369337

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Maxima Model Summary

Call:

betareg(formula = recall_t ~ log10(file.size) + chunks + frame.size + index.type + log10(file.size):chunks + log10(file.size):frame.size + chunks:frame.size + log10(file.size):index.type + chunks:index.type + frame.size:index.type + log10(file.size):chunks:index.type + log10(file.size):frame.size:index.type + chunks:frame.size:index.type | index.type * chunks, data = dat)

Table 9.5.3 Standardized weighted residuals 2

Min	1Q	Median	3Q	Max
-2.8994	-0.7530	-0.2073	0.5620	6.7930

Table 9.5.4 ChiSq test for significance of relationship between recall, variables, and interactions.

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	1.098143	0.101391	10.831	< 2e-16 ***
log10(file.size)	-0.015687	0.072412	-0.217	0.82849
chunks4	0.090868	0.131462	0.691	0.48943
chunks8	0.141348	0.131676	1.073	0.28307
chunks12	-0.043664	0.121815	-0.358	0.72001
frame.size10min	-0.067906	0.125422	-0.541	0.58822
frame.size5min	0.036177	0.126179	0.287	0.77433
frame.size2_5min	0.061919	0.126215	0.491	0.62372
index.typeAudioSet	1.330295	0.130164	10.220	< 2e-16 ***
log10(file.size):chunks4	0.041886	0.081174	0.516	0.60586
log10(file.size):chunks8	-0.136666	0.080990	-1.687	0.09152 .
log10(file.size):chunks12	-0.161706	0.075030	-2.155	0.03114 *
log10(file.size):frame.size10min	0.102739	0.079372	1.294	0.19553
log10(file.size):frame.size5min	0.038478	0.079556	0.484	0.62863
log10(file.size):frame.size2_5min	0.006807	0.079479	0.086	0.93175
chunks4:frame.size10min	-0.006198	0.140491	-0.044	0.96481
chunks8:frame.size10min	0.058158	0.139619	0.417	0.67701
chunks12:frame.size10min	0.051066	0.128978	0.396	0.69216
chunks4:frame.size5min	-0.038412	0.140795	-0.273	0.78499
chunks8:frame.size5min	-0.028125	0.139501	-0.202	0.84022
chunks12:frame.size5min	0.054384	0.129486	0.420	0.67449
chunks4:frame.size2_5min	-0.087356	0.140184	-0.623	0.53319
chunks8:frame.size2_5min	-0.022520	0.139418	-0.162	0.87168
chunks12:frame.size2_5min	0.071848	0.129468	0.555	0.57893
log10(file.size):	0.201887	0.096762	2.086	0.03694 *

index.typeAudioSet				
chunks4:index.typeAudioSet	-0.245409	0.212017	-1.158	0.24707
chunks8:index.typeAudioSet	0.058168	0.176588	0.329	0.74186
chunks12:index.typeAudioSet	0.163712	0.179160	0.914	0.36083
frame.size10min: index.typeAudioSet	0.003025	0.165081	0.018	0.98538
frame.size5min: index.typeAudioSet	-0.130519	0.165120	-0.790	0.42926
frame.size2_5min: index.typeAudioSet	-0.173535	0.164352	-1.056	0.29103
log10(file.size):chunks4: index.typeAudioSet	0.431582	0.131410	3.284	0.00102 **
log10(file.size):chunks8: index.typeAudioSet	0.188465	0.107162	1.759	0.07863 .
log10(file.size):chunks12: index.typeAudioSet	0.288612	0.107881	2.675	0.00747 **
log10(file.size):frame.size10min: index.typeAudioSet	-0.088741	0.113542	-0.782	0.43447
log10(file.size):frame.size5min: index.typeAudioSet	-0.024313	0.113140	-0.215	0.82985
log10(file.size):frame.size2_5min: index.typeAudioSet	-0.075012	0.112071	-0.669	0.50329
chunks4:frame.size10min: index.typeAudioSet	-0.003125	0.234520	-0.013	0.98937
chunks8:frame.size10min: index.typeAudioSet	-0.242249	0.188956	-1.282	0.19983
chunks12:frame.size10min: index.typeAudioSet	-0.230048	0.189657	-1.213	0.22514
chunks4:frame.size5min: index.typeAudioSet	-0.173004	0.230870	-0.749	0.45364
chunks8:frame.size5min: index.typeAudioSet	-0.181134	0.188242	-0.962	0.33593
chunks12:frame.size5min: index.typeAudioSet	-0.316069	0.189201	-1.671	0.09481 .
chunks4:frame.size2_5min: index.typeAudioSet	-0.151230	0.228365	-0.662	0.50783
chunks8:frame.size2_5min: index.typeAudioSet	-0.229031	0.186516	-1.228	0.21947
chunks12:frame.size2_5min: index.typeAudioSet	-0.428817	0.187602	-2.286	0.02227 *

Table 9.5.5 Phi Coefficients (model with log link)

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	4.9081	0.2350	20.881	< 2e-16 ***
index.typeAudioSet	1.6413	0.3330	4.929	8.26e-07 ***
chunks4	-1.6202	0.2623	-6.176	6.56e-10 ***
chunks8	-2.4222	0.2487	-9.740	< 2e-16 ***
chunks12	-2.6224	0.2441	-10.744	< 2e-16 ***

index.typeAudioSet:chunks4	-1.7007	0.3741	-4.546	5.47e-06 ***
index.typeAudioSet:chunks8 -	0.5817	0.3532	-1.647	0.0996 .
index.typeAudioSet:chunks12 -	1.4477	0.3476	-4.165	3.12e-05 ***

Type of estimator: ML (maximum likelihood)

Log-likelihood: 2401 on 54 Df

Pseudo R-squared: 0.4293

Number of iterations: 66 (BFGS) + 2 (Fisher scoring)

9g – Accuracy Output Statistics (using Continuous Variables)

Linear Model:

max_model <- accuracy_t ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type
Linear Model (as in early analysis): mod_lm <- lm(max_model, data= dat)

Table 9.7.1 AOV outputs for the multivariate linear model (**accuracy**)

	Df	Sum Sq	Mean Sq	F Value	Pr(>|z|)
log10(file size)	1	0.000	0.000	0.000	0.99738
Chunks	1	0.194	0.194	23.308	1.50e-06 ***
Frame.size	1	0.005	0.005	0.657	0.41767
Index.type	1	15.980	15.980	1917.358	< 2e-16 ***
Log10(file.size):chunks	1	0.026	0.026	3.091	0.07892 .
Log10(file.size):frame.size	1	0.000	0.000	0.036	0.84963
Chunks:frame.size	1	0.012	0.012	1.429	0.23213
log10(file.size):index.type	1	0.288	0.288	34.563	4.91e-09 ***
chunks:index.type	1	0.103	0.103	12.356	0.00045 ***
frame.size:index.type	1	0.074	0.074	8.826	0.00301 **
log10(file.size):chunks: index.type	1	0.035	0.035	4.236	0.03972 *
log10(file.size):frame.size: index.type	1	0.003	0.003	0.339	0.56030
chunks:frame.size:index.type	1	0.006	0.006	0.763	0.38257
Residuals	1785	14.876	0.008		

Betareg Model:

mod_br <- betareg(max_model, data=dat)
mod_br_phi <- update(mod_br, . ~ . | index.type * chunks)

betareg(formula = accuracy_t ~ log10(file.size) + chunks + frame.size + index.type + log10(file.size):chunks + log10(file.size):frame.size + chunks:frame.size + log10(file.size):index.type + chunks:index.type + frame.size:index.type + log10(file.size):chunks:index.type + log10(file.size):frame.size:index.type + chunks:frame.size:index.type |
index.type * chunks, data = dat)

Table 9.7.2 Standardized weighted residuals 2

Min	1Q	Median	3Q	Max
-2.9508	-0.7654	-0.2208	0.4747	6.2269

Table 9.7.3 Coefficients of Mean Model with Log Link.

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	**1.147e+00**	**1.121e-01**	**10.233**	**< 2e-16 *****
Log10(file.size)	9.695e-02	8.097e-02	1.197	0.23121
Chunks	5.030e-03	1.210e-02	0.416	0.67755
Frame.size	4.086e-05	1.323e-04	0.309	0.75745
Index.type	**9.345e-01**	**1.762e-01**	**5.302**	**1.14e-07 *****
Log10(file.size):chunks	**-1.999e-02**	**8.032e-03**	**-2.489**	**0.01280 ***
Log10(file.size):frame.size	-2.056e-05	7.761e-05	-0.265	0.79114
Chunks:frame.size	-1.010e-05	1.222e-05	-0.827	0.40822
Log10(file.size):index.type	**3.420e-01**	**1.315e-01**	**2.601**	**0.00929 ****
Chunks:index.type	-1.252e-03	1.908e-02	-0.066	0.94765
Frame.size:index.type	1.300e-04	2.144e-04	0.607	0.54417
Log10(file.size):chunks: index.type	4.399e-04	1.299e-02	0.034	0.97298
Log10(file.size):frame.size: index.type	1.370e-04	1.269e-04	1.079	0.28045
Chunks:frame.size:index.type	2.517e-05	2.027e-05	1.242	0.21435

Table 9.7.3 Phi Coefficients (model with log link)

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	**3.686964**	**0.129049**	**28.570**	**<2e-16 *****
Index.typeAudioSet	**0.384924**	**0.185972**	2.070	**0.0385 ***
Chunks	**-0.125049**	**0.013364**	-9.357	**<2e-16 *****
Index.type:AudioSet	-0.001346	0.019434	-0.069	0.9448

Type of estimator:	ML (maximum likelihood)
Log-likelihood:	2301 on 18 Df
Pseudo R-squared:	0.4209
Number of iterations:	30 (BFGS) + 3 (Fisher scoring)

Table 9.7.4 shows the Akaike Information Criterion (AIC) for different models used to explain variation in accuracy as a result of experimental variation.

	df	Estimate
Mod_lm	15	-3491.238
Mod_br	15	4401.508
Mod_br_phi	18	-4565.338

Table 9.7.5. Analysis of Deviance Table (Type II tests)

	Df	Chisq	Pr(>Chisq)
Log10(file.size)	1	14.6258	0.0001311 ***
Chunks	1	17.2050	3.355e-05 ***
Frame.size	1	8.6459	0.0032780 **
Index.type	**1**	1912.4822	< 2.2e-16 ***
Log10(file.size):chunks	**1**	9.8664	0.0016832 **
Log10(file.size):frame.size	1	0.2492	0.6176605
Chunks:frame.size	1	0.0097	0.9213576
Log10(file.size):index.type	**1**	69.2583	< 2.2e-16 ***
Chunks:index.type	1	2.0376	0.1534518
Frame.size:index.type	1	34.8975	3.475e-09 ***
Log10(file.size):chunks: index.type	1	0.0011	0.9729774
Log10(file.size):frame.size: index.type	1	1.1649	0.2804523
Chunks:frame.size:index.type	1	1.5418	0.2143490

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

9h – Precision Output Statistics (using Continuous Variables)

Linear Model:

max_model <- precision_t ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type
Linear Model (as in early analysis): mod_lm <- lm(max_model, data= dat)

Table 9.8.1 AOV outputs for the multivariate linear model (**precision**)

	Df	Sum Sq	Mean Sq	F Value	Pr(>|z|)
log10(file size)	1	0.004	0.004	0.592	0.44176
Chunks	1	0.074	0.074	9.931	0.00165 **
Frame.size	1	0.002	0.002	0.207	0.64929
Index.type	1	14.264	14.264	1903.190	< 2e-16 ***
Log10(file.size):chunks	1	0.023	0.023	3.072	0.07980 .
Log10(file.size):frame.size	1	0.000	0.000	0.042	0.83792
Chunks:frame.size	1	0.006	0.006	0.750	0.38654
log10(file.size):index.type	1	0.276	0.276	36.768	1.62e-09 ***
chunks:index.type	1	0.061	0.061	8.203	0.00423 **
frame.size:index.type	1	0.064	0.064	8.533	0.00353 **
log10(file.size):chunks:index.type	1	0.021	0.021	2.741	0.09800.
log10(file.size):frame.size: index.type	1	0.007	0.007	0.939	0.33259
chunks:frame.size:index.type	1	0.005	0.005	0.649	0.42042
Residuals	1785	13.378	0.007		

Beta regression model:

mod_br <- betareg(max_model, data=dat)
mod_br_phi <- update(mod_br, . ~ . | index.type * chunks)

Call:
betareg(formula = precision_t ~ log10(file.size) + chunks + frame.size + index.type + log10(file.size):chunks + log10(file.size):frame.size + chunks:frame.size + log10(file.size):index.type + chunks:index.type + frame.size:index.type + log10(file.size):chunks:index.type + log10(file.size):frame.size:index.type + chunks:frame.size:index.type | index.type * chunks, data = dat)

Table 9.8.2 Standardized weighted residuals 2

Min	1Q	Median	3Q	Max
-3.2818	-0.7450	-0.2431	0.4942	6.4455

Table 9.8.3 Coefficients of Mean Model with Log Link.

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	**1.199e+00**	**1.134e-01**	**10.576**	**< 2e-16 *****
Log10(file.size)	7.168e-02	8.170e-02	0.877	0.38030
Chunks	6.611e-03	1.217e-02	0.543	0.58714
Frame.size	6.663e-05	1.337e-04	0.498	0.61830
Index.type	**8.978e-01**	**1.742e-01**	**5.155**	**2.54e-07 *****
Log10(file.size):chunks	**-1.622e-02**	**8.077e-03**	**-2.008**	**0.04468 ***
Log10(file.size):frame.size	-5.560e-05	7.791e-05	-0.714	0.47547
Chunks:frame.size	-6.978e-06	1.227e-05	-0.569	0.56964
Log10(file.size):index.type	**3.686e-01**	**1.298e-01**	**2.839**	**0.00452 ****
Chunks:index.type	5.268e-03	1.880e-02	0.280	0.77928
Frame.size:index.type	9.215e-05	2.118e-04	0.435	0.66345
Log10(file.size):chunks: index.type	-3.549e-03	1.280e-02	-0.277	0.78164
Log10(file.size):frame.size: index.type	1.709e-04	1.253e-04	1.364	0.17246
Chunks:frame.size:index.type	2.604e-05	1.999e-05	1.302	0.19276

Table 9.8.4 Phi Coefficients (model with log link)

	Estimate	Std.Error	Z Value	Pr(>|z|)
(Intercept)	3.667224	0.129159	28.393	<2e-16 ***
Index.type	0.477444	0.185956	2.568	0.0102 *
Chunks	-0.118561	0.013389	-8.855	<2e-16 ***
Index.type:chunks	0.002017	0.019442	0.104`	0.9174

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood: 2431 on 18 Df
Pseudo R-squared: 0.4288

Number of iterations: 30 (BFGS) + 2 (Fisher scoring)

Table 9.8.5 shows the Akaike Information Criterion (AIC) for different models used to explain variation in precision as a result of experimental variation.

	df	Estimate
Linear Model	15	-3682.240
Betareg Model	15	-4654.656
Betareg Phi Model	18	-4825.003

Table 9.8.6. Analysis of Deviance Table (Type II tests)

	Df	Chisq	Pr(>Chisq)
Log10(file.size)	1	13.3449	0.0002591 ***
Chunks	1	4.4200	0.0355203 *
Frame.size	1	12.0242	0.0005251 ***
Index.type	**1**	2014.0679	< 2.2e-16 ***
Log10(file.size):chunks	**1**	7.9126	0.0049092 **
Log10(file.size):frame.size	1	0.0297	0.8631529
Chunks:frame.size	1	0.0855	0.7699322
Log10(file.size):index.type	**1**	76.1929	< 2.2e-16 ***
Chunks:index.type	1	3.2384	0.0719289 .
Frame.size:index.type	1	36.4003	1.607e-09 ***
Log10(file.size):chunks: index.type	1	0.0768	0.7816436
Log10(file.size):frame.size: index.type	1	1.8614	0.1724631
Chunks:frame.size:index.type	1	1.6964	0.1927592

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

9i – Recall Output Statistics (using Continuous Variables)

Linear Model:

max_model <- recall_t ~ (log10(file.size) + chunks + frame.size) ^ 2 * index.type
Linear Model (as in early analysis): mod_lm <- lm(max_model, data= dat)

Table 9.9.1 AOV outputs for the multivariate linear model (**recall**)

	Df	**Sum Sq**	**Mean Sq**	**F Value**	**Pr(>|z|)**
log10(file size)	1	0.000	0.000	0.027	0.868874
Chunks	1	0.189	0.189	22.319	2.03e-06 ***
Frame.size	1	0.008	0.008	1.008	0.315415
Index.type	1	15.999	15.999	1922.764	< 2e-16 ***
Log10(file.size):chunks	1	0.023	0.023	2.819	0.093339
Log10(file.size):frame.size	1	0.000	0.000	0.024	0.876837
Chunks:frame.size	1	0.012	0.012	1.445	0.229471
log10(file.size):index.type	1	0.300	0.300	36.002	2.38e-09 ***
chunks:index.type	1	0.112	0.112	13.489	0.000247 ***
frame.size:index.type	1	0.082	0.082	9.880	0.001699 **
log10(file.size):chunks: index.type	1	0.037	0.037	4.410	0.035864 *
log10(file.size):frame.size: index.type	1	0.003	0.003	0.312	0.576262
chunks:frame.size: index.type	1	0.006	0.006	0.765	0.381986
Residuals	1785	14.853	0.008		

Betareg Model:

mod_br <- betareg(max_model, data=dat)
mod_br_phi <- update(mod_br, . ~ . | index.type * chunks)

betareg(formula = recall_t ~ log10(file.size) + chunks + frame.size + index.type + log10(file.size):chunks + log10(file.size):frame.size + chunks:frame.size + log10(file.size):index.type + chunks:index.type + frame.size:index.type + log10(file.size):chunks:index.type + log10(file.size):frame.size:index.type + chunks:frame.size:index.type | index.type * chunks, data = dat)

Table 9.9.2 Standardized weighted residuals 2

Min	**1Q**	**Median**	**3Q**	**Max**
-2.9469	-0.7504	-0.2228	0.4802	6.2298

Table 9.9.3 Coefficients of Mean Model with Log Link.

	Estimate	**Std.Error**	**Z Value**	**Pr(>|z|)**
(Intercept)	**1.158e+00**	**1.128e-01**	**10.262**	**< 2e-16 *****
Log10(file.size)	9.237e-02	8.146e-02	1.134	0.25681
Chunks	4.649e-03	1.216e-02	0.382	0.70212
Frame.size	3.086e-05	1.330e-04	0.232	0.81649
Index.type	**9.083e-01**	**1.766e-01**	**5.142**	**2.71e-07 *****
Log10(file.size):chunks	**-1.964e-02**	**8.068e-03**	**-2.435**	**0.01491 ***
Log10(file.size):frame.size	-1.806e-05	7.780e-05	-0.232	0.81648
Chunks:frame.size	-1.033e-05	1.226e-05	-0.842	0.39970
Log10(file.size):index.type	**3.512e-01**	**1.318e-01**	**2.665**	**0.00769 ****
Chunks:index.type	-5.484e-04	1.908e-02	-0.029	0.97708
Frame.size:index.type	1.406e-04	2.147e-04	0.655	0.51246
Log10(file.size):chunks: index.type	1.253e-03	1.300e-02	0.096	0.92325
Log10(file.size):frame.size: index.type	1.291e-04	1.269e-04	1.018	0.30884
Chunks:frame.size:index.type	2.624e-05	2.029e-05	1.293	0.19587

Table 9.9.4 Phi Coefficients (model with log link)

	Estimate	**Std.Error**	**Z Value**	**Pr(>|z|)**
(Intercept)	3.6677091	0.1290430	**28.422**	<2e-16 ***
Index.typeAudioSet	0.3899579	0.1859748	2.097	0.036 *
Chunks	-0.1232328	-9.221	-9.221	<2e-16 ***
Index.type:AudioSet:chunks	0.0003653	0.0194351	-0.019	0.985

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)

Log-likelihood: 2305 on 18 Df

Pseudo R-squared: 0.4235

Number of iterations: 30 (BFGS) + 2 (Fisher scoring)

Table 9.9.5 shows the Akaike Information Criterion (AIC) for different models used to explain variation in accuracy as a result of experimental variation.

	df	Estimate
Linear Model	15	-3494.087
Betareg Model	15	-4412.078
Betareg Phi Model	18	-4573.955

Table 9.9.6. Analysis of Deviance Table (Type II tests)

	Df	Chisq	Pr(>Chisq)
Log10(file.size)	1	16.0094	6.303e-05 ***
Chunks	1	15.9911	6.364e-05 ***
Frame.size	1	7.9690	0.004758 **
Index.type	**1**	**1913.3500**	< 2.2e-16 ***
Log10(file.size):chunks	**1**	**9.1696**	0.002461 **
Log10(file.size):frame.size	1	0.2459	0.619967
Chunks:frame.size	1	0.0057	0.939644
Log10(file.size):index.type	**1**	**73.0189**	< 2.2e-16 ***
Chunks:index.type	1	2.7647	0.096362
Frame.size:index.type	1	**36.5939**	1.455e-09 ***
Log10(file.size):chunks: index.type	1	0.0093	0.923248
Log10(file.size):frame.size: index.type	1	1.0356	0.308839
Chunks:frame.size:index.type	1	1.6729	0.195866

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

APPENDIX B: Chapter 3 Supplementary Material

Contents:

1 Popular/ Emerging Spatial ARUs

Table 1 charts some of the emerging methods in spatial acoustics. These range from commercial “out of the box” solutions, to DIY and custom recorders. The MAARU recorder developed in this thesis is in the bottom row.

Name	**Hyper -bolic/ DOA**	**Mics** Num **ARU**s	**Base unit**	**min cost for study**	**Data** transfer	**Mic Sync**	**Power**	**Ref**
Wildlife Acoustics	Hyper-bolic	2 mic (1 as failsafe) 3+ ARUs	SM2+	£700 (£2,100)	Manual	GPS	Battery	(Mennill et al., 2012)
Audio Moth Dev	Hyper-bolic	1 mic 3+ ARUs	Audio Moth	£75 (£225)	Manual	GPS	Battery	(Hill et al., 2018)
SAFE Acoustics	Hyper-bolic	1 mic 3+ ARUs	Rasp-berry Pi	£200	Cloud	NA	Self-charging	(Sethi et al., 2018)
VoxNet	DOA+ Hyper-bolic	4 mics 3+ ARUs	Custom	?	Cloud	Network	Battery	(Allen et al., 2008).
CARACAL	DOA+ Hyper-bolic	4 mics 3+ ARUs	Custom	£150 (£450)	Manual	?	Battery	(Wijers et al., 2019).
TAMAGO/ Dev-Audio	DOA+ Hyper bolic	7/8 mics 3+ ARUs	Laptop GUI	?	Laptop required	Cable	Laptop required	(Suzuki et al., 2017)
Crunchant et al. 2022	Hyper-bolic	1 mic 4 ARUs	Rasp-berry Pi	?	Manual	GPS	Self-charging	Crunchant et al., 2022
WASN	Hyper-bolic	W mic 5 ARUs	Rasp-berry Pi	?	Cloud	Network	Battery	Bruggemann et al., 2021
This one	**DOA**	**6 mics 1ARU**	**Rasp-berry Pi**	**£226 (£226)**	**Cloud**	**Short Cable**	**Self-charing**	**This**

2 Device Costs (Accurate as of July 2021)

Table 2.1. per component costs of the MAARU recorder. Solar panel and battery require ments are location (sunlight) dependant so costs will vary depending on latitude of field site (London, GB used here). SD card size is also variable, the price given is for a 128GB micro S D.

Item	Price GBP
Powering	
Battery *	119
12v to 5v step-down converter	9.33
10m of AWG10 cable (10m)	21.06
Hooks and Slings (x4)	14.00
10L Dri-Bag	5.99
Electronics	
Raspberry Pi 4 (4GB)	54.99
Respeaker 6 Mic Array (incl. soundcard)	36.20
Huawei Mobile Dongle	38.98
Memory Card**	14.99
USB-USB extension cable	8.99
Sistema Tupperware	2.50
IPRO Vents (x6)	18.00
Sniper Tape	14.99
General	
Sniper Tape	4.99
Sealant	7.95
Polystyrene	0.00
Silica Gel (each)	0.29
Kwik-Lok Tie (each)	3.15
TOTALS	
Powering Total	255.37
Electronics Total	189.64
Whole Device Total	**461**
Whole Device (Excluding Powering)	**205.63**

Table 2.2. per component costs of an autonomous omnidirectional recorder (Sethi et al., 2018). Powering costs are excluded here as powering requirements across both devices are the same.

Item	Cost GBP
Electronics	
Rasberry Pi B	35.27
Rode SmartLav+ Microphone	47.00
TRRS to TRS audio jack splitter	6.59
UASNW1 USB audio card	8.99
Memory Card 64GB***	11.99
Huawei Dongle	38.98
1m USB Extension Cable	3.30
USB to Micro USB Cable	4.69
Dri-Box	10.99
Dry Bag	5.99
General	
Cable Ties	5.39
Kwik-Lok (x2)	6.30
TOTALS	
Device Total (Excluding Powering)	**185.49**

3 Device Technical Information and Weatherproofing

Table 3.1.1. MAARU device technical specifications.

Microphone Part Dimensions	15x15x13cm
Microphone Part Weight	<1kg
Solar Panel Part Dimensions	84x58x15cm (+ suspension equipment)
Solar Panel Part Weight	~11kg
Microphone Part Attachment Height	2/3m
Solar Panel Part Attachment Height	Canopy where possible, dependant on Sun exposure
Total Weight	~ 12kg
Total Number Recorders	4

Table 3.1.2. IPRO vent details (B0.2). Data taken and adapted from

www.ipromembrane.com

IP Rating	IP67, IP68 (5m water immersion for 1h)
WEP (60s)	300kPA
Maximum Trasmission Loss (Max Value 100 – 10,000 Hz)	<1.5dB
Material Type and Colour	White ePTFE
Characteristic	Hydrophobic
Reference Thickness (mm)	0.4
Temperature Range (^{O}C)	-40 to 85
Adhesive Type	Acrylic
Reccommended Orientation	Internal Mount
ROHS Compliance	Yes

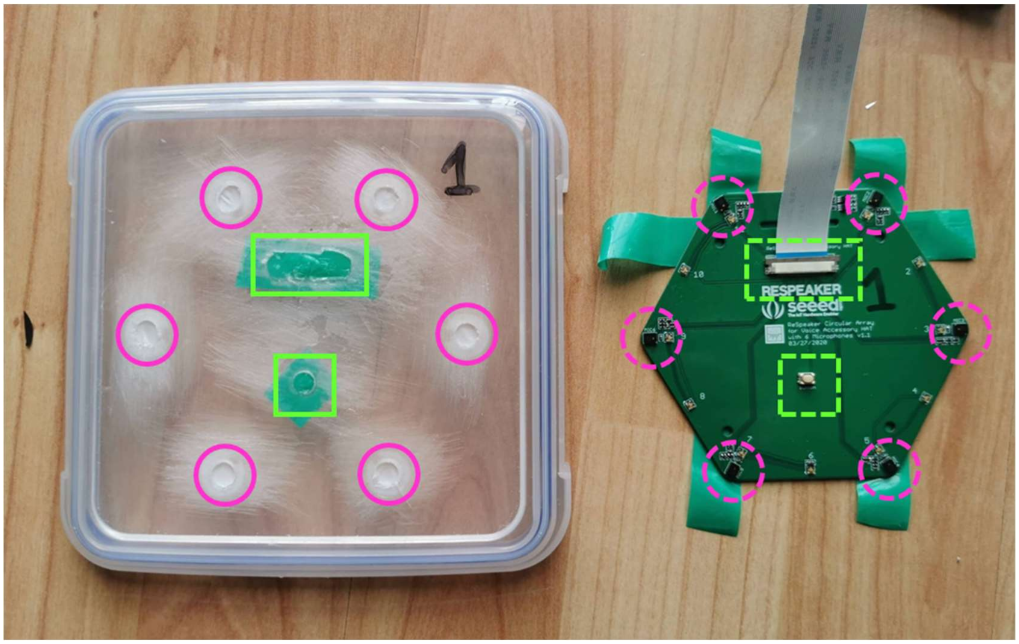


Figure 3 The sound-permitting device weatherproofing. The full pink circles correspond to holes in the container lid covered by ePTFE waterproof acoustic vents, and these holes line up with MEMS microphones (dashed pink circles) on the array. The array is held flush to the container lid by additional holes made for other raised components on the PCB (green rectangles), which are instead sealed flexibly with an air-and-water tight sealant.

4 Powering Assay

To determine the power usage of the multi-channel and omnidirectional devices, regular watt readings were taken from both devices (via a USB power reader) during continuous and non-continuous recording. In one test, a screen (showing status information) was attached to the recorder so the exact status of the device could be determined; in another, the device was left as it would be *in situ,* and an average reading was found. In both cases, readings were taken every 15 seconds for 5 minutes. We found that the addition of multiple microphones had no significant impact on the power usage compared to the omnidirectional device, in all phases: start-up (PO), recording (REC), and simultaneous recording and upload (REC+UP) (Respectively: $t = 0.87_{10.51}$, $p = 0.400$, $t = 1.88_{20.5}$, $p = 0.075$ and $t = 0.253_{9.39}$, $p = 0.8054$, Welch's T-Test.). Data transmission is almost instantaneous in the omnidirectional recorder, so no measured time points fell during a period where recording and upload were happening simultaneously in the non-continuous test.

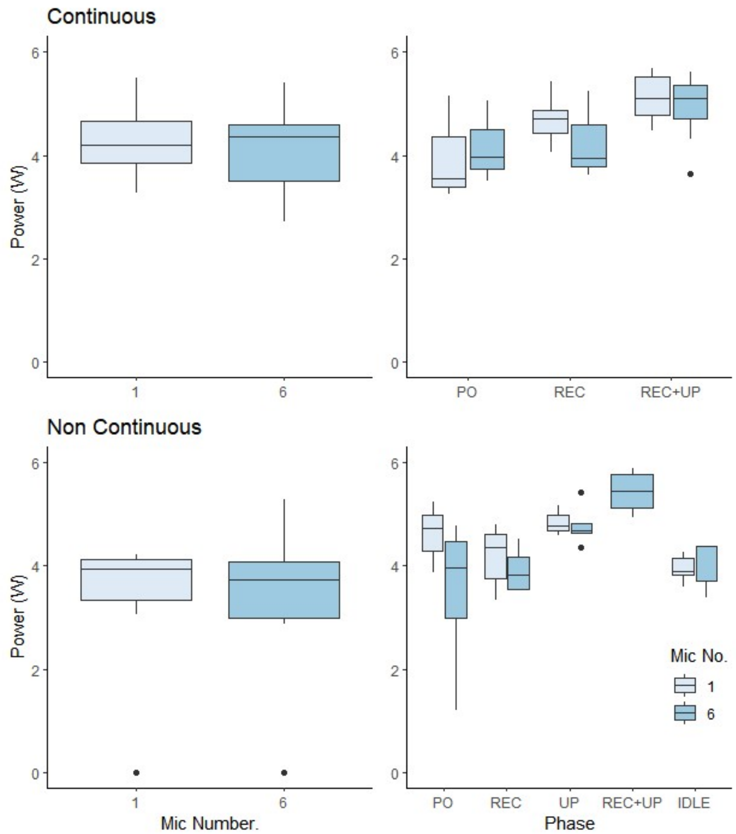


Figure 4 shows the comparative power usage between the 1- and 6- Mic recorders under different running conditions. The left panels show values averaged over a 5-minute period. The right panels show the individual running periods separately PO = Powering on, REC = Recording audio data, UP= Uploading data, REC+UP = Simultaneous recording and upload. Power consumption appears slightly higher for both in the continuous recording. Black dots show outliers. Overall, there is no significant difference in power consumption between the devices.

5 Power Spectra Tests

1- Device Test Details:

Table 5.1.1. Lab tests codes and conditions

Test Code	Tone Type	Weatherproofing?	Device Number
1a	Pink Noise	N	2
1b	Eurasian Wren	N	2
3a	Pink Noise	Y	3
5a	Pink Noise	Y	2
5b	Eurasian Wren	Y	2
7a	Pink Noise	N	3

Table 5.1.2. Compared test groups and outcomes.

Test Type	Devices Used
Device Number	3a vs. 5a
Waterproofing	1b vs. 5b AND 1a vs. 5a
Signal Type	1a vs. 1b AND 5a vs. 5b
Test Number	5a vs. 5b

2- Spectral Information from Lab Tests:

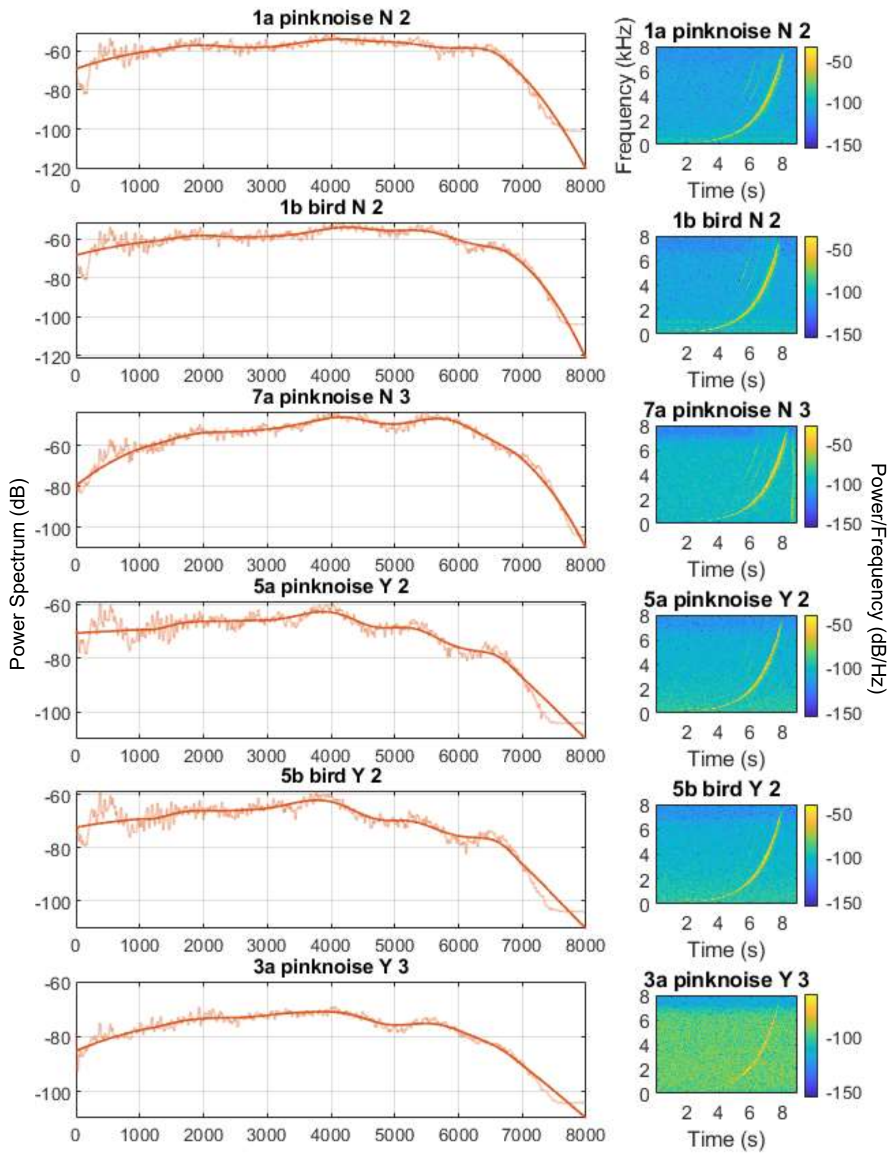


Figure 5.2.1. shows the recorded power spectra of a sweep recorded in lab tests. Test codes shown above power spectra, and conditions can be determined through Table 5.1.1

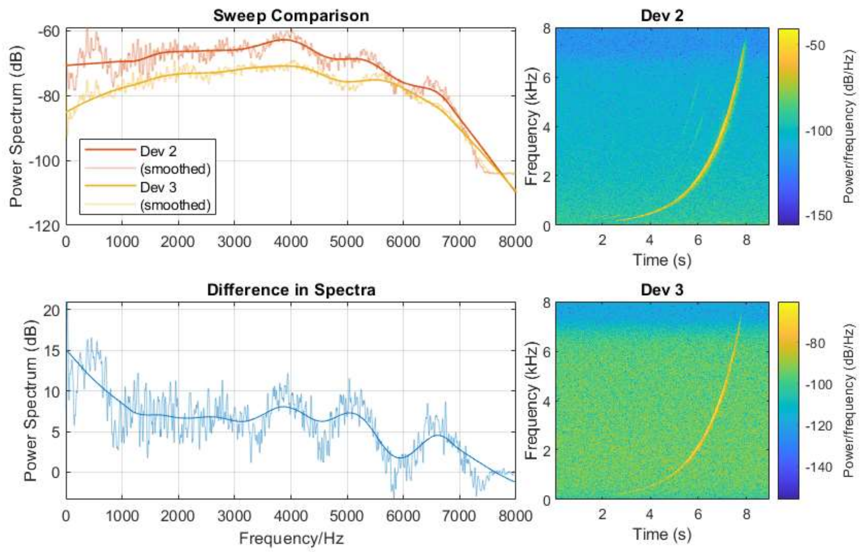


Figure 5.2.2. compares the power spectra recorded in the same test conditions from two different devices to determine the impact of the device on the quality of recorded signal. A >10dB difference is observed below 1kHz, but higher than that all differences are below 10dB.

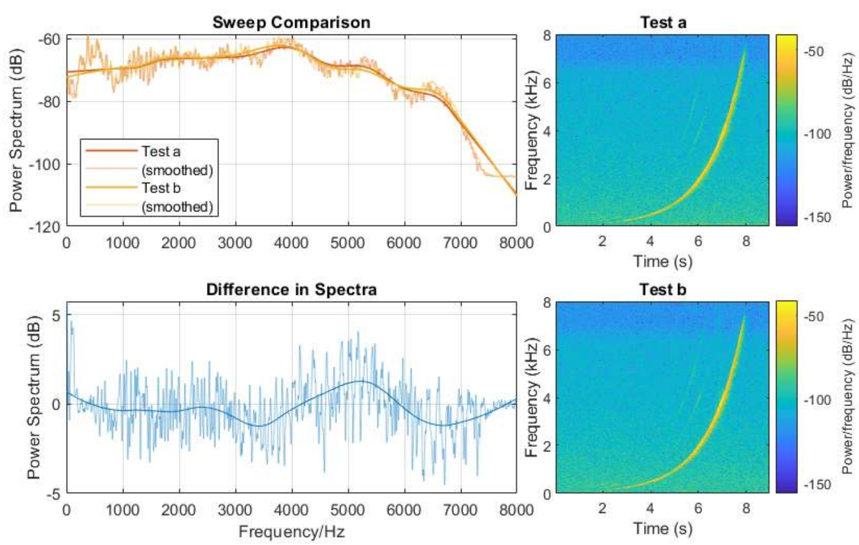


Figure 5.2.3. compares the power spectra recorded in the same test conditions from the same device at two different times to determine the impact of the test on the quality of recorded signal. The difference in spectra stays below 5dB across all frequencies.

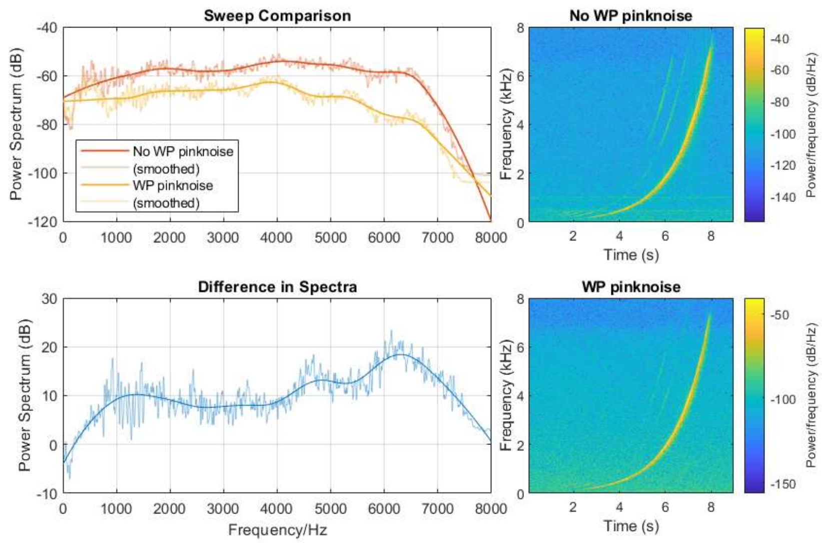


Figure 5.2.4. compares the power spectra recorded from weatherproofed and un-weatherproofed devices to determine the impact of weatherproofing on the quality of recorded signal. A <10dB difference is observed below 4kHz, and <20dB under 6.5kHz, beyond which the difference tends to zero. (Test 1)

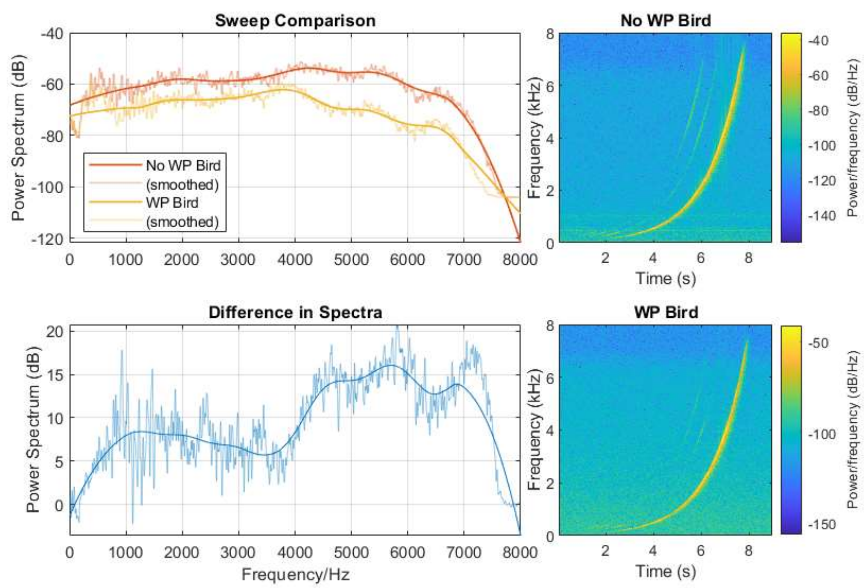


Figure 5.2.5. compares the power spectra recorded from weatherproofed and un-weatherproofed devices to determine the impact of weatherproofing on the quality of recorded signal. A <10dB difference is observed below 4kHz, and <20dB under 6.5kHz, beyond which the difference tends to zero. (Test 2)

6 Gain Adjustments

Table 6.1 Maximum absolute dB FS values in each channel and in each recording – scaled b y colour. Used to determine an appropriate threshold for an appropriate gain shift.

fileName	ch1	ch2	ch3	ch4	ch5	ch6
1a_pinknoise2_N_2.wav	0.31079	0.24704	0.25882	0.22009	0.2562	0.22284
1a_pinknoise_N_2.wav	0.30157	0.31351	0.23456	0.24158	0.25079	0.25195
1b_bird_N_2.wav	0.30795	0.24701	0.24536	0.24503	0.27225	0.26205
3a_pinknoise_Y_3.wav	0.015961	0.018219	0.069366	0.015808	0.018341	0.26749
5a_pinknoise_Y_2.wav	0.11502	0.14166	0.062836	0.065399	0.068756	0.14896
5b_bird_Y_2.wav	0.11044	0.13443	0.052612	0.059875	0.066376	0.15988
7a_pinknoise_N_3.wav	0.53217	0.65256	0.60669	0.72601	0.69006	0.6456

fileName	ch1	ch2	ch3	ch4	ch5	ch6
ASNW2_PostMortemLoc_bird01.wav	0.0336	0.03717	0.000519	0.000488	0.024689	0.028503
ASNW2_PostMortemLoc_bird02.wav	0.035217	0.030243	0.000549	0.000488	0.020172	0.027374
ASNW2_PostMortemLoc_pink01.wav	0.029572	0.027222	0.000488	0.000488	0.016724	0.028473
ASNW2_Repowered_bird01.wav	0.000488	0.000488	0.023163	0.049896	0.045471	0.038849
ASNW2_Repowered_bird02.wav	0.000549	0.000488	0.020905	0.042633	0.034271	0.02948
ASNW2_Repowered_pink01.wav	0.000519	0.000549	0.019897	0.045868	0.042145	0.041565
ASNW2_Repowered_pink02.wav	0.000488	0.000488	0.021027	0.047913	0.052032	0.033508
PAWS2_PostMortem_bird01.wav	1	1	0.019897	0.029358	0.032013	0.021545
PAWS2_PostMortem_bird02.wav	1	1	0.022156	0.028992	0.030029	0.018524
PAWS2_PostMortem_bird03.wav	0.98846	0.89392	0.015106	0.020599	0.019653	0.017761
PAWS2_PostMortem_pink01.wav	1	0.92575	0.019165	0.027161	0.026794	0.018677
PAWS2_PostMortem_pink02.wav	1	0.84131	0.017151	0.023529	0.024628	0.018433
PAWS2_flipped_bird01.wav	1	1	0.020325	0.027344	0.027557	0.022339
PAWS2_flipped_pink01.wav	1	0.87335	0.020477	0.026062	0.027344	0.019501
PAWS1_PostMortemLoc_bird01.wav	0.06958	0.11835	0.13934	0.11206	0.084686	0.070709
PAWS1_PostMortemLoc_bird02.wav	0.067444	0.099731	0.16461	0.14154	0.099518	0.079712
PAWS1_PostMortemLoc_pink02.wav	0.055023	0.096924	0.14304	0.13043	0.091949	0.06662
PAWS1_PostMortemLoc_pink03.wav	0.049805	0.0802	0.093475	0.091461	0.073853	0.0495

Based on the Above the "AddGain.m" script does the following:

Channel Status	Range max(abs(x))	Outcome
overpower	0.5=< **max(abs(x))**	If any channels have this run gain adjustment method
ok	0.01< **max(abs(x))** <0.5	No action
dead	**max(abs(x))** <= 0.01	Remove recording from sample

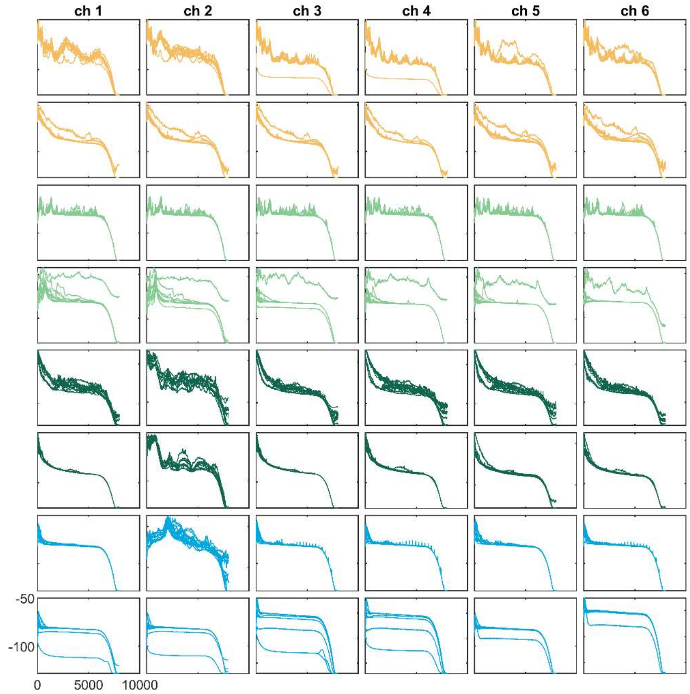


Figure 6.2 cross microphone power spectra for each recorder (colours). The top row of each recorder is a week after deployment while the bottom row is a month later. Several dead channels present and variation in level between microphones.

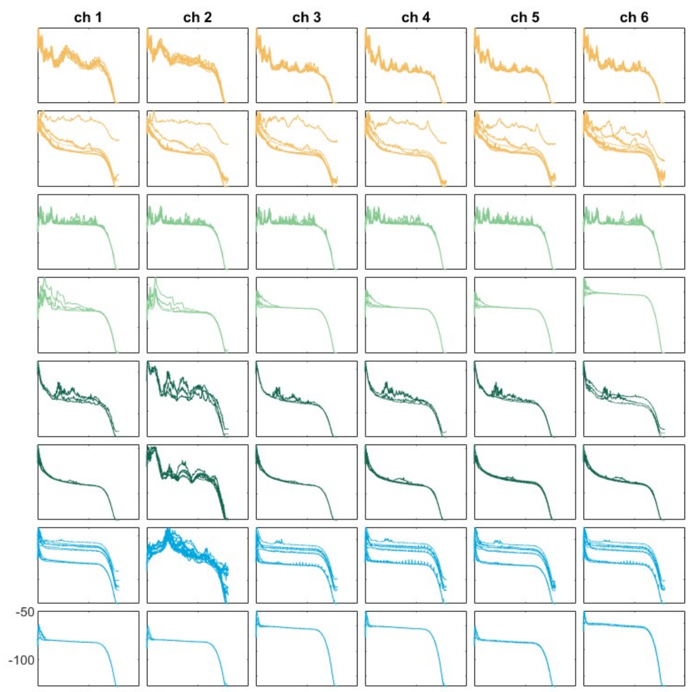


Figure 6.3 cross microphone power spectra for each recorder (colours) after automated gain shifts. The top row of each recorder is a week after deployment while the bottom row is a month later. Several dead channels have been removed and the variation in signal between channels is reduced.

7 Localisation Tests

1 – HARKBird Thresholds

8 Table 7.1.1. Pre-deployment test conditions and HARKBird thresholds

Test Code	Tone Type	Weatherproofing?	Device Number	Threshold
1a	Pink Noise	N	2	29
1b	Eurasian Wren	N	2	29
3a	Pink Noise	Y	3	25
5a	Pink Noise	Y	2	25
5b	Eurasian Wren	Y	2	25
7a	Pink Noise	N	3	29

Table 7.1.2. Post-deployment test conditions and HARKBird threshold.

Recorder	Tone Type	Test Number*	Threshold
ASNW2	Eurasian Wren	01	24
ASNW2	Eurasian Wren	02	24
ASNW2	Pink Noise	01	24
ASNW2Repowered	Eurasian Wren	01	24
ASNW2Repowered	Eurasian Wren	02	24
ASNW2Repowered	Pink Noise	01	24
ASNW2Repowered	Pink Noise	02	24
PAWS2	Eurasian Wren	01	24
PAWS2	Eurasian Wren	02	24
PAWS2	Pink Noise	01	24
PAWS2	Pink Noise	02	24
PAWS2Flipped	Eurasian Wren	01	24
PAWS2Flipped	Pink Noise	01	24
PAWS1	Eurasian Wren	01	24
PAWS1	Eurasian Wren	02	24
PAWS1	Pink Noise	02	24
PAWS1	Pink Noise	03	24

2 –Pre and Post Localisation (No Gain Adjustment):

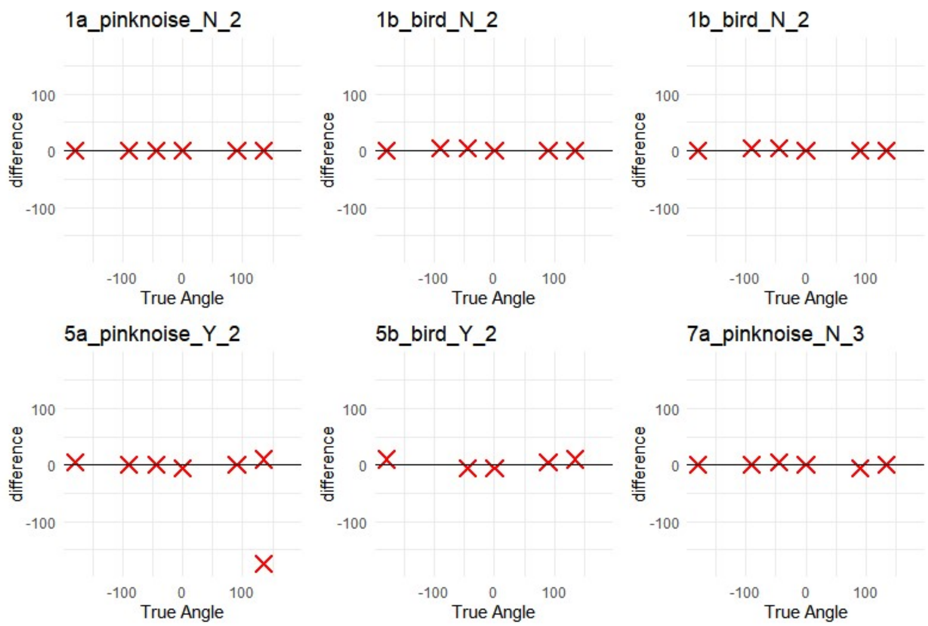


Figure 7.2.1 localisation error for tests on recorders pre-deployment. Test conditions are given above each of the true/ error plots. (No Gain Adjustment)

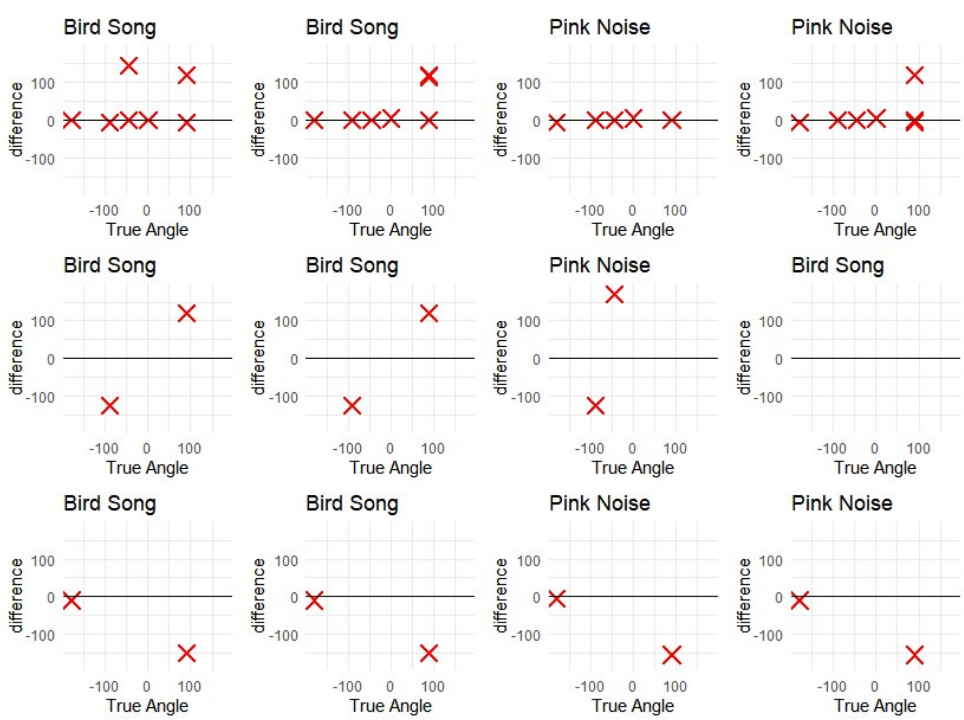


Figure 7.2.2 localisation error for tests on recorders post deployment. Each row denotes each of the recorders. The top row (PAWS1) is able to localise signals accurately, but the others are unable to detect or accurately localise any of the signals. (No Gain Adjustment).

3 –Post Localisation (with Gain Adjustment):

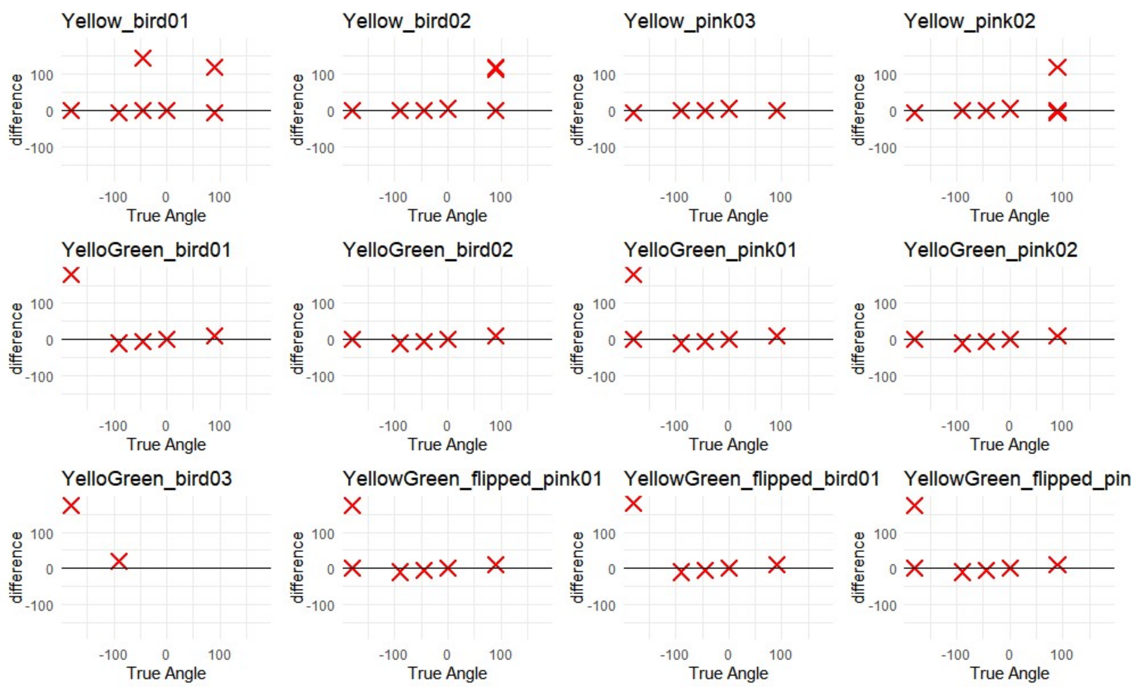


Figure 7.3 localisation error for tests on recorders post deployment. Each row denotes each of the recorders. After the autonomous gain adjustment is employed, all recorders were able to successfully detect and localise signals.

4 - Localisation Error Data:

Table 7.4 compares the error between true sound source locations and HARKBird predicted directions for different conditions. In several tests a number of false positives were present. In the below table we provide mean accuracy with and without false positives included.

test	group	mean	SD	maxEr	Mean (Echoes Removed)	SD (Echoes Re-moved)	Max Error (Echoes Re-moved)	Prop of Calls Detected	Tests	False Detections (Echoes)
pre	WP bird	4.17	7.03	10	3	7.58	10	1	1	0.17
pre	NoWP bird	1.43	2.31	5	NA	NA	NA	1	1	0
pre	WP pink noise	-23.5	61.7	-175	1.67	5.16	10	1	1	0.17
pre	NoWP pink noise	0.41	2.04	5	NA	NA	NA	1	3	0
post	PAWS1 Bird	35.35	59.2	145	-0.5	2.84	-5	1	2	0.29
post	PAWS1 Pink	9.58	34.93	-180	0	3.33	-5	1	2	0.17
post	PAWS2 Bird	27.62	84.48	180	-2	7.75	-10	1	3	0.29
post	PAWS2 Pink	33.33	72.51	180	-1	6.87	-10	1	3	0.29
post	ANSW2 Bird	-46.6	126.2	170	NA	NA	NA	0.4	3	0
post	ANSW2 Pink	-41.2	124.99	-150	NA	NA	NA	0.4	4	0
pre	**WP**	-10.7	49.7	-175	2.27	6.07	10	1	2	0.15
pre	**NoWP**	0.8	2.36	5	0.83	2.41	5	1	4	0.04
post	**PAWS1**	23.46	50.33	145	-0.25	3.02	5	1	4	0.23
post	**PAWS2**	17.5	75.345	180	-1.5	7.21	10	1	6	0.29
post	**ANSW2**	-43.5	114.8	155	NA	NA	NA	0.4	7	0

5 – Localisation Analysis Outputs

Table 7.5.1 mean and SD of localisation error of pre-deployment tests

Waterproofing?	Test Tone	False Positive Removed?	Mean	SD
No	bird	n	1.43	2.44
No	bird	y	1.67	2.58
No	pinknoise	n	0.56	2.36
No	pinknoise	y	4.16	2.36
Yes	bird	n	2.17	7.36
Yes	bird	y	3	7.58
Yes	pinknoise	n	-23	66.99
Yes	pinknoise	y	1.67	5.4

Table 7.5.2 mean and SD of localisation error of post-deployment tests of ASNW1

Tone	Raw mean dif (+SD)	Removed echoes raw mean (+SD)	Gain Test?	adjG mean dif (+SD)	Removed echoes mean (+SD)
bird	-41.24 (+107.7)	NA	remove	NA	NA
Pinknoise	-46.67 (+126.2)	NA	remove	NA	NA

Table 7.5.3 mean and SD of localisation error of post-deployment tests of PAWS1

Tone	Raw mean dif (+SD)	Removed echoes raw mean (+SD)	Gain Test?	adjG mean dif (+SD)	Removed echoes mean (+SD)
bird	35 (+58)	-0.5 (+2.8)	No action	Same as before	Same as before
Pinknoise	9.58 (+34.2)	0 (+3.24)	No action	Same as before	Same as before

Table 7.5.4 mean and SD of localisation error of post-deployment tests of PAWS2

Tone	Raw mean dif (+SD)	Removed echoes raw mean (+SD)	Gain Test?	adjG mean dif (+SD)	Removed echoes mean (+SD)
bird	NA	NA	Added Gain	27.61 (+87.47)	-2 (+7.75)
Pinknoise	NA	NA	Added Gain	33.33 (+72.51)	-1 (+6.87)

* PAWS2 flipped (both) and bird samples were actually generated at a THRESH of 23.5

8 – Pre- and Post- Deployment Sweep Comparisons

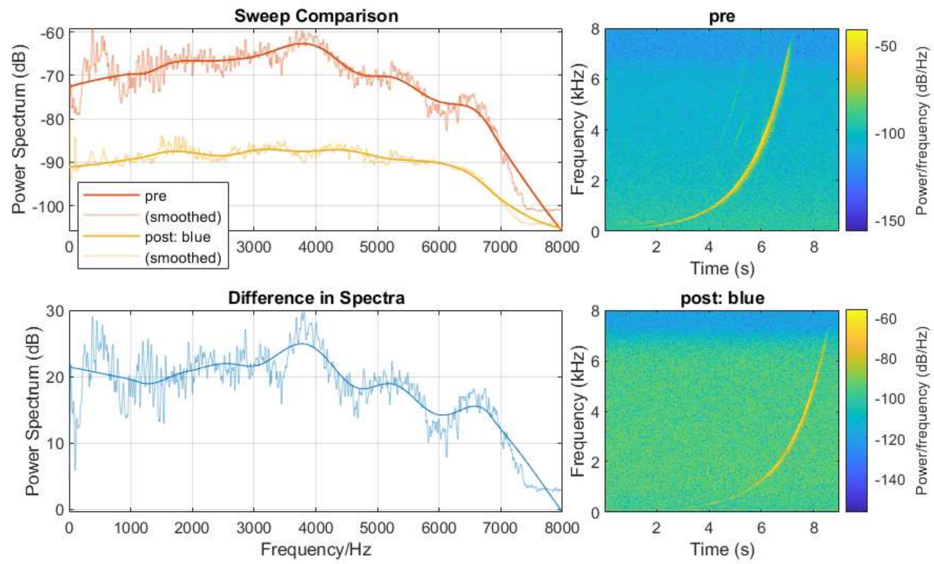


Figure 8.1 compare the spectra of a sweep recorded by the same device (PAWS2) pre- and post- 6-month deployment; the corresponding iii and iv show this as spectrograms.

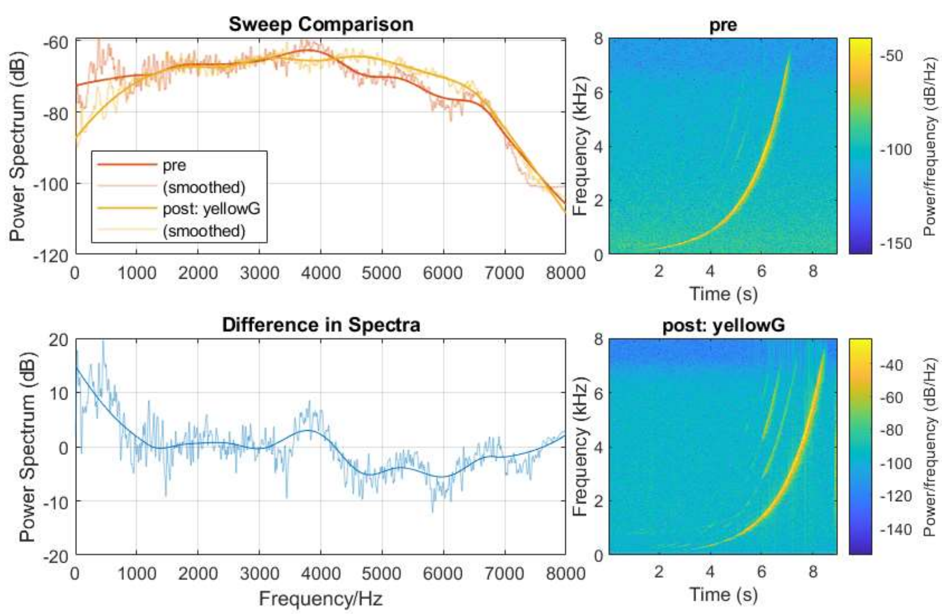


Figure 8.2 compare the spectra of a sweep recorded by the same device (PAWS2) pre- and post- 6-month deployment; the corresponding iii and iv show this as spectrograms.

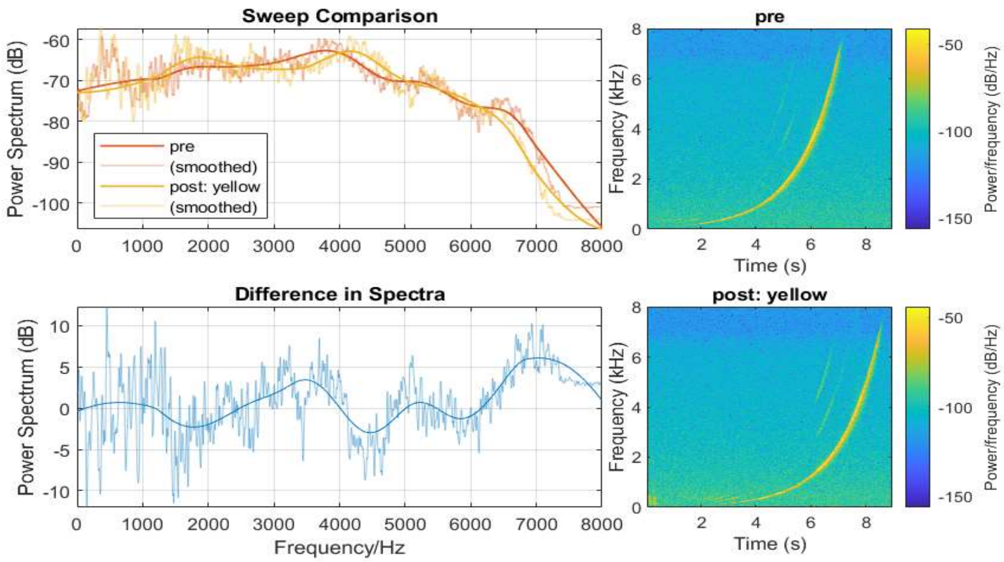


Figure 8.2 compare the spectra of a sweep recorded by the same device (PAWS1) pre- and post- 6-month deployment; the corresponding iii and iv show this as spectrograms

8. APPENDIX C: Chapter 4 Supplementary Material

Contents

1 – Field Site Aerial View

Figure 1.1 show clear forest differences between ASNW (light green/brown) and PAWS (darker green). Point colours represent signal strength as determined by mobile speed tests,

2 – HARKBird Parameters

Table 2.1 HARKBird config for ASNW Recorders

Group Sensitivity	THRESH (Detection threshold)	Frequency Range	Number of Sources	Gain?
High	24.5	2-8kHz	2	No
Medium	24.7	2-8kHz	2	No
Low	24.75	2-8kHz	2	No

Table 2.1 HARKBird config for PAWS Recorders

Group Sensitivity	THRESH (Detection threshold)	Frequency Range	Number of Sources	Gain?
High	25	2-8kHz	2	No
Medium	25.15	2-8kHz	2	No
Low	25.25	2-8kHz	2	No

3 – Bonferroni Significance Table

Species	Mean	DF	T-Value	Raw P	Bonferroni Significance (p<0.00263)
Blue Tit	3.92	26	4.45	0.000145	*
Bullfinch	2.33	5	1.87	0.121	
Buzzard	1	2	NA	NA	
Coal Tit	5.64	158	9.35	< 2.2e-16	*
Great Spotted Woodpecker	19.38	46	2.26	0.0284	
Great Tit	3.68	24	2.77	0.0108	
Green Woodpecker	4.18	26	2.44	0.0218	
Jay	2	1	NA	NA	
Long-Tailed Tit	3	16	5.50	4.869e-05	*
Magpie	6.36	35	4.94	1.907e-05	*
Marsh Harrier	1.5	1	1	0.5	
Mistle Thrush	2.75	3	2.33	0.11	
Nuthatch	2.8	4	1.36	0.24	
Robin	6.19	25	2.19	0.0381	
Rose-Ringed Parakeet	5.50	580	17.13	< 2.2e-16	*
Short-Toed Treecreeper	4.43	6	2.16	0.074	
Treecreeper	4.4	4	1.73	0.16	
Willow Tit	1	1	NA	NA	
Wren	3.57	20	4.25	0.000384	*

Table 2.1 Bonferroni significance table to determine whether per-individual call rate was non-zero for each species.