Creating Generalizable Data-Driven Approaches for Biodiversity Monitoring via Acoustics

Thomas Napier
James Cook University, Townsville, Queensland 4811, AUSTRALIA
thomas.napier@jcu.edu.au

Abstract

Global biodiversity is declining at unprecedented rates, yet traditional monitoring at the necessary scales remains costly and biased toward what can be seen. Sound offers a complementary lens: many species are detected more reliably by their vocalizations, microphones are inexpensive and unobtrusive, and they can cover greater spatial and temporal scales. These advantages have made passive acoustic monitoring a fast-growing paradigm, yet robust, generalizable sound distinction in complex soundscapes remain a central obstacle. My thesis addresses this by combining data-driven human-inspired representation learning with knowledge-guided unsupervised learning to prioritize hierarchical organization and structure discovery prior to labelling. Human-in-the-loop oversight is incorporated as targeted verification under uncertainty, drawing on active learning and weak supervision to direct effort where it has the highest value.

Introduction

Global biodiversity is undergoing decline at unprecedented rates, necessitating scalable and continuous monitoring solutions beyond traditional, often limited and bias, manual survey methods (Cardinale et al. 2012). Passive Acoustic Monitoring (PAM) has emerged, and now the deployment of inexpensive, non-invasive microphones can continuously capture entire soundscapes (Sueur and Farina 2015). Continental-scale initiatives, such as the Australian Acoustic Observatory (A2O), and the U.S. Northeast Passive Acoustic Sensing Network (NEPAN) exemplify this success, providing wide spatial and temporal coverage crucial for comprehensive ecological assessment (Roe et al. 2021).

This technological success, however, has created a severe technological bottleneck. The sheer volume of data renders manual analysis impossible. Existing machine learning and deep learning solutions, particularly supervised models, perform sub-optimally in this complex domain because they are typically trained on small, sanitized, and species-specific datasets (Stowell, 2022). Thus, these models are useful for narrow-scope research questions but cannot handle the complications of real-world heterogeneous soundscapes nor generalize beyond the circumscription of the dataset upon which they are trained. Consequently, existing approaches may suffer from severe domain shift when applied across diverse ecological contexts, rendering the scarcity of robust, labelled data a primary obstacle (Gibb et al. 2018).

The core thrust of my doctoral research addresses this impasse by proposing a perspective shift: instead of pursuing premature classification with inadequate labels, I propose prioritizing unsupervised structure discovery prior to labelling. My work investigates scalable frameworks that first reveal the hierarchical organization of soundscapes by treating them as compositions of events. Two key issues facing my current research can thus be summarized.

RQ1: How can I develop a scalable unsupervised framework to automatically discover the compositional structure of complex, multi-source soundscapes, thereby breaking the manual annotation bottleneck?

Traditional supervised models fail to scale without large, labeled datasets that are difficult to obtain. An unsupervised framework is essential to transform the intractable task of manual labeling into efficient, targeted verification, allowing experts to confirm structure rather than create it from scratch (Lin et al. 2017). For this, combining non-linear and adaptive density-based clustering is ideally positioned to assist in organizing the complexity and scale of continuous soundscapes.

RQ2: How can an unsupervised framework be designed to create acoustically pure and ecologically coherent groupings of sound events, particularly in the presence of polyphony, by implementing domain-adaptive criteria for event windowing and clustering?

Existing unsupervised approaches underperform across a variety of real-world situations due to polyphony and environmental noise and thus often produce "ecologically blind" clusters because standard metrics prioritize mathematical

compactness over biological relevance (e.g., mixing unrelated sounds or fragmenting a species vocalization over multiple clusters) (Michaud et al. 2023). The framework must therefore integrate knowledge from auditory scene analysis and ecological reasoning to ensure the groupings are not just mathematically optimal but represent distinct, biologically relevant sound sources.

Current Progress

Initial progress involved establishing a lightweight and robust core unsupervised architecture capable of structuring voluminous raw, heterogeneous soundscape data across diverse geographic regions. This addresses the challenge of building a robust scalable framework (RQ1) despite inevitable environmental complexity and domain shift.

I first validated the Mel-Frequency Cepstral Coefficient (MFCC) features for their ability to distinguish between major ecological sound categories (biophony, geophony, anthrophony, silence), achieving exceptional accuracy (~98%) across two ecologically distinct sites with a simple classifier (Napier et al. 2023). This proved that computationally inexpensive, human-inspired features were sufficient for the task and the complete UMAP-HDBSCAN pipeline successfully organized a large dataset from six biodiverse soundscapes into well-formed clusters.

To resolve the primary annotation bottleneck, I operationalized this framework and thus created LEAVES (Large-Scale Ecoacoustics Analysis and Visualization with Efficient Segmentation), an open-source, human-in-the-loop decision support tool (Napier et al. 2025). LEAVES uses the discovered acoustic clusters to enable targeted expert verification, where a label from a small sample is propagated to an entire cluster. This method proved to be up to 7.12 times faster than traditional manual labeling while maintaining around 79-90% similarity to ground truth.

Future Work and Research Timeline

Objective	Timeline
RQ1: Scalable Framework	Feb 2022-Mar 2024
RQ2: Ecologically Coherent Groupings	Mar 2024-Mar 2026
2.1: Adaptive Sound Event Detection	Mar 2024-Jan 2025
2.2: Source Separation	Jan 2025-Mar 2026
Thesis Writing & Defense	Mar 2026-July 2026

Table 1: Research Timeline

My future work, as shown in Table 1, will address another critical limitation in soundscape analysis: the reliance on fixed-length window sizes. This approach is suboptimal as it causes call fragmentation, and even state-of-the-art deep learning models continue to struggle with variable-length, rapidly shifting vocalizations (Kahl et al. 2021). To overcome this, I will explore a strategy involving embedding drift to create an adaptive windowing system that dynamically aligns segmentation with the start and end of actual acoustic events, ensuring their temporal integrity.

This refined segmentation is the first step toward solving the more complex problem of separating overlapping sounds in the frequency domain. Current source separation techniques are often inadequate for ecological data as they are computationally expensive, assume a fixed number of sound sources, or fail to generalize across the noisy, non-stationary conditions of unseen environments. Thus, my next steps, if successful, would be addressing this.

References

Cardinale, J. B.; Duffy, E. J.; Gonzalez, A.; Hooper, U. D.; Perrings, C.; Venail, P.; Narwani, A.; et al. 2012. Biodiversity Loss and Its Impact on Humanity. *Nature* 486 (7401): 59–67.

Gibb, R.; Browning, E.; Glover-Kapfer, P.; Jones, E. K. 2018. Emerging Opportunities and Challenges for Passive Acoustics in Ecological Assessment and Monitoring. *Methods in Ecology and Evolution* 10 (2): 169–85.

Kahl, S.; Wood, M. C.; Eibl, M.; Klinck, H. 2021. BirdNET: A Deep Learning Solution for Avian Diversity Monitoring. *Ecologica Informatics* 61: 101236.

Lin, T.-H.; Fang, S.-H.; Tsao, Y. 2017. Improving Biodiversity Assessment via Unsupervised Separation of Biological Sounds from Long-Duration Recordings. *Scientific Reports* 7 (1).

Michaud, F.; Sueur, J.; Le Cesne, M.; Haupert, S. 2023. Unsupervised Classification to Improve the Quality of a Bird Song Recording Dataset. *Ecological Informatics* 74: 101952–52.

Napier, T.; Ahn, E.; Allen-Ankins, S.; and Lee, I. 2023. An Optimised Grid Search Based Framework for Robust Large-Scale Natural Soundscape Classification. In *Proceedings of the Australasian Joint Conference of Artificial Intelligence*, 468–479.

Napier, T.; Ahn, E.; Allen-Ankins, S.; Schwarzkopf, L.; Lee, I. 2025. LEAVES: An Open-Source Web-Based Tool for the Scalable Annotation and Visualisation of Large-Scale Ecoacoustic Datasets Using Cluster Analysis. *Ecological Informatics* 87: 103026.

Roe, P.; Eichinski, P.; Fuller, A. R.; McDonald, G. P.; Schwarzkopf, L.; Towsey, M.; Truskinger, A.; Tucker, D.; Watson, M. D. 2021. The Australian Acoustic Observatory. *Methods in Ecology and Evolution* 12 (10): 1802–8.

Stowell, D. 2022. Computational Bioacoustics with Deep Learning: A Review and Roadmap. *PeerJ* 10: e13152.

Sueur, J.; Farina, A. 2015. Ecoacoustics: The Ecological Investigation and Interpretation of Environmental Sound. *Biosemiotics* 8 (3): 493–502.