# **Human-in-the-loop** clustering **reduces** Passive Acoustic Monitoring (PAM) annotation effort by **7 times** while retaining **80-90%** expert agreement in biodiversity assessments.

## *Creating Generalizable Data-Driven Approaches for Biodiversity Monitoring via Acoustics*

**Background:** 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 PAM a fast-growing paradigm, yet robust, generalizable sound distinction in complex soundscapes remain a central obstacle.

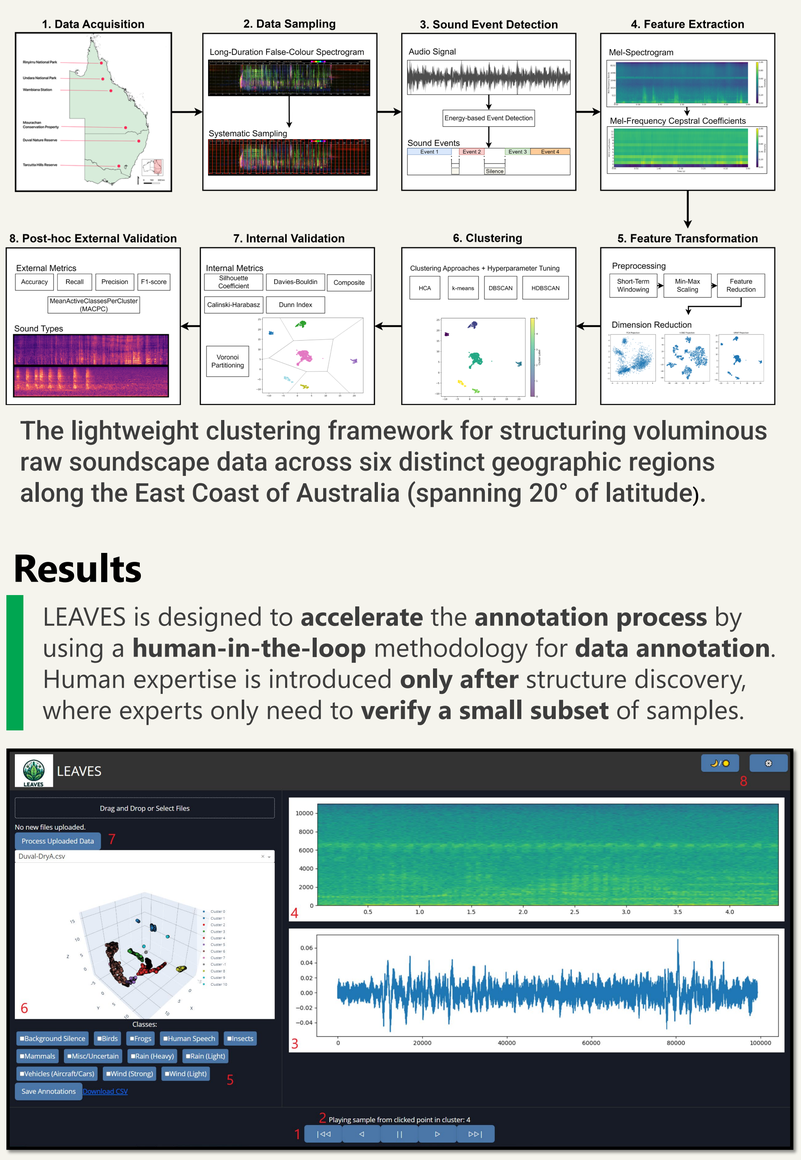
## Motivation

× **Manual slog:** Painfully slow to label and review endless recordings (hours of audio = weeks of work).
× **Scalability gap:** Too much sound, not enough experts
× **Standard techniques fail:** Models struggle with noisy real-world soundscapes.

## Methodology

1. Unsupervised clustering first: Let the data **speak for itself** *before* we involve a human.
2. Design an open-source **human-in-the-loop** tool that clusters similar sounds designed for **high-throughput annotation**.
3. Infuse ecological knowledge to ensure clusters are **domain-tuned** while still **generalizable**.

**Raw audio** is segmented into **sound events**, represented using lightweight **MFCC features**, and organised using **non-linear embedding** and **density-based clustering**. This exposes the inherent acoustic structure of complex soundscapes, **grouping similar sounds** together.




The lightweight clustering framework for structuring voluminous raw soundscape data across six distinct geographic regions along the East Coast of Australia (spanning 20° of latitude).

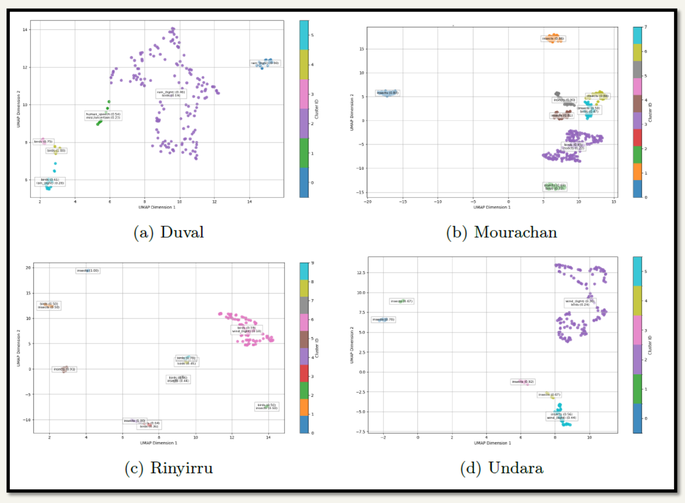
## Results

LEAVES is designed to **accelerate** the **annotation process** by using a **human-in-the-loop** methodology for **data annotation**. Human expertise is introduced **only after** structure discovery, where experts only need to **verify a small subset** of samples.



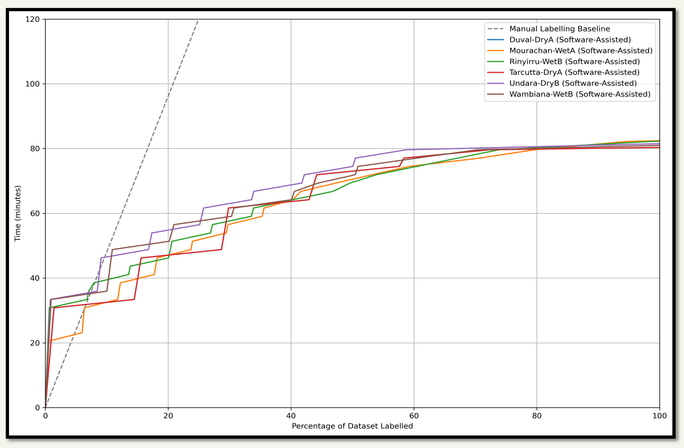

The Large-Scale Ecoacoustics Annotation and Visualisation with Efficient Segmentation (LEAVES) Open-Source Platform.

Across **four ecologically distinct sites**, LEAVES **consistently** reveals coherent acoustic groupings that **align with dominant sound sources**. This structure emerges **prior to labeling**, demonstrating that similar sounds cluster together **reliably** despite differences in habitat and soundscape composition.




UMAP embeddings of sound events from four Australian ecosystems, coloured by expert labels, showing strong correspondence between unsupervised clusters and ground truth.

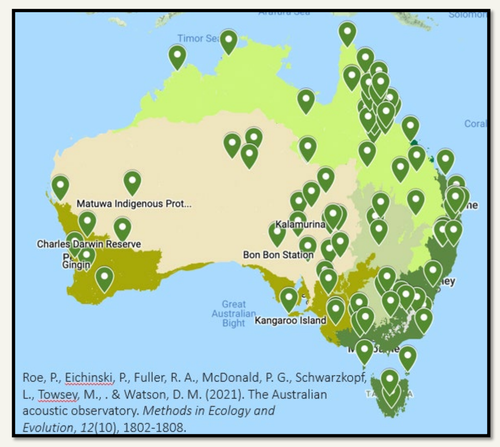
LEAVES uses the discovered acoustic clusters to enable **targeted expert verification**, where a label from a small sample is propagated to an entire cluster. This method proved to be up to **7.12 times faster** than traditional manual labeling.




A comparison of annotation efficiency. The traditional baseline represents an average of 11.3s per 4.5s sample. In contrast, the LEAVES-assisted lines demonstrate significant efficiency gains.

## Discussion

1. Unsupervised clustering reveals stable acoustic structure prior to labelling, and it generalises across sites without retraining.
2. Expert effort shifts from manual annotation to targeted verification..
3.. LEAVES functions as a scalable pre-filter rather than a species classifier.
4.. Together, these properties support large-scale soundscape analysis and long-term monitoring by reducing annotation effort while preserving interpretability.



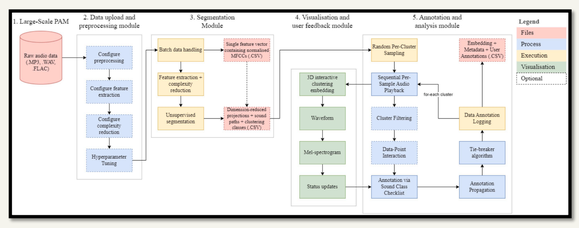
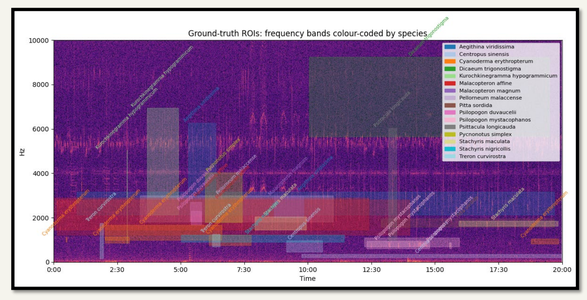




Table 2: Post-hoc external validation metrics for unsupervised clustering across six ecoacoustic datasets. Each site underwent cluster-based annotation via majority-vote label propagation from a manually labelled subset. Accuracy, precision, recall, and F1-score assess the alignment between cluster assignments and the propagated classes. MACPC quantifies the average number of active sound types per sample within each cluster (based only on sampled points), serving as a proxy for within-cluster ecological complexity. Also shown are the number of unique sound types and clusters identified.

| Dataset | Accuracy | Precision | Recall | F1-score | MACPC | # Sound Types | # Clusters | Sound Types Present |
|---|---|---|---|---|---|---|---|---|
| Undara | 0.848 | 0.839 | 0.848 | 0.837 | 1.525 | 8 | 18 | birds, frogs, human_speech, insects, misc/uncertain, rain_light, vehicles, wind_light |
| Duval | 0.898 | 0.849 | 0.898 | 0.870 | 1.430 | 9 | 15 | birds, human_speech, insects, mammals, misc/uncertain, rain_heavy, rain_light, vehicles, wind_light |
| Rinyirru | 0.940 | 0.959 | 0.940 | 0.945 | 1.227 | 6 | 12 | birds, insects, mammals, misc/uncertain, vehicles, wind_light |
| Mourachan | 0.954 | 0.914 | 0.954 | 0.933 | 1.723 | 10 | 19 | birds, frogs, insects, mammals, misc/uncertain, rain_heavy, rain_light, vehicles, wind_strong, wind_light |
| Wambiana | 0.985 | 0.982 | 0.985 | 0.983 | 1.100 | 4 | 6 | insects, birds, wind_light, wind_strong |
| Tarcutta | 0.930 | 0.920 | 0.930 | 0.925 | 1.350 | 5 | 10 | insects, birds, rain_heavy, wind_light, wind_strong |

**Table 3**
External validation results for the selected clusters in each dataset.

| Dataset | Ground truth labels | Software-assisted propagated label | FMI (Non-hierarchical) | ARI (Hierarchical) | NMI (Hierarchical) | FMI (Hierarchical) | Total labels | Incorrect labels (%) |
|---|---|---|---|---|---|---|---|---|
| Tarcutta-DryA | {birds, birds + vehicles} | birds | 0.8723 | 1.0000 | 1.0000 | 1.0000 | 30 | 4 (13.33%) |
| Undara-DryB | {insects + human speech, insects} | insects | 0.8450 | 1.0000 | 1.0000 | 1.0000 | 198 | 34 (17.17%) |
| Wambiana-WetB | {insects, insects + misc, insects + birds} | insects | 0.8301 | 1.0000 | 1.0000 | 1.0000 | 223 | 40 (17.94%) |
| Duval-DryA | {light rain, light rain + birds} | light rain | 0.9027 | 1.0000 | 1.0000 | 1.0000 | 98 | 10 (10.20%) |
| Mourachan-WetA | {insects, insects + birds} | insects | 0.7987 | 1.0000 | 1.0000 | 1.0000 | 212 | 50 (23.58%) |
| Rinyirru-WetB | {birds, birds + insects} | birds + insects | 0.8123 | 1.0000 | 1.0000 | 1.0000 | 121 | 26 (21.48%) |




**Limitations:** It's not magic: clusters can be coarse and require interpretation. Misclustering of sounds that are acoustically similar can occur. Can be computationally heavy. Overlapping sounds (in both frequency and time) are a non-trivial issue.

**Thomas Napier**

thomas.napier@jcu.edu.au
www.thomasnapier.com

**ACKNOWLEDGEMENTS:** Special thanks to my advisors Prof. Ickjai Lee, Prof. Lin Schwarzkopf, Dr. Euijoon Ahn and Dr. Slade Allen-Ankins. Big thanks also to Cornell University for inviting me to speak for their spring BioacousTalks series.

Cornell University
K. Lisa Yang Center for Conservation Bioacoustics

JAMES COOK UNIVERSITY AUSTRALIA