
Decoding Insect Song: A Multitask Semisupervised Orthoptera Bioacoustic Classifier

Olga Isupova ^{*1} Danil Kuzin ^{*1} Ella Browning ¹ Tom Mills ¹ Steven Reece ¹

Abstract

Passive acoustic monitoring holds great promise for ecological inference, yet existing automated tools are typically narrowly trained and non-transferable. We address these limitations with PULSE, a semi-supervised, multi-task framework for Orthoptera bioacoustics, combining weakly-supervised species classification, self-supervised learning on unlabelled field audio, and knowledge distillation from a general-purpose bioacoustic model. Our domain-adapted specialist model outperforms a state-of-the-art general model across all metrics (macro F1: 0.21 vs. 0.07; AUC: 0.74 vs. 0.45; AP: 0.32 vs. 0.19), with active learning further raising F1 to 0.34 and AUC to 0.84. Beyond classification, the learned embeddings encode ecologically meaningful structure, exposed through an interactive visualisation tool for ecological discovery.

1. Introduction

Passive acoustic monitoring (PAM) is increasingly used to collect ecoacoustic data to assess habitat health and biodiversity responses to environmental change (Gibb et al., 2019; Ross et al., 2023). By extracting taxa-specific signals, spatial and temporal patterns in species presence, activity, and density can be identified, informing land management and conservation strategies. Automating this using machine learning (ML) has enabled large-scale PAM for birds (Kahl et al., 2021; Hamer et al., 2023) and bats (Aodha et al., 2022). However, despite the ecological importance of insects, there are no open-source ML tools for acoustically identifiable insects such as Orthoptera (crickets and grasshoppers), which use songs and calls for various ecological functions (Riede & Balakrishnan, 2025). A key barrier is the lack of large open-access sound libraries to train ML algorithms, though

new datasets on platforms such as Xeno-canto (Vellinga, 2024) and the Global Biodiversity Information Facility are beginning to address this (Funosas et al., 2026). Combining these data with ML approaches that reduce the need for large training sets, such as agile modelling and self-supervised learning, presents meaningful opportunities to address PAM taxonomic biases.

Extraction of Orthoptera signals from ecoacoustic data is more challenging than for birds or bats: many species call near-continuously whilst active, and overlapping calls are common (Funosas et al., 2026). Furthermore, reference calls on bioacoustic platforms are often unrepresentative of PAM recordings, with higher audio quality and limited noise, causing ML tools to perform poorly on field recordings. Combining “clean” sound library data with “messy” field recordings can improve model performance. To address these challenges, we present PULSE, which leverages a large collection of unlabelled UK field recordings — released alongside this work — to bridge the gap between curated sound libraries and real-world deployment.

2. Related Work

Recent approaches for Orthoptera PAM are based on deep learning. OrthopterOSS (Bennett et al., 2025) uses the AM-ResNet architecture (Xiao et al., 2022) but, like earlier models (Faiß & Stowell, 2023; Newson et al., 2017), relies solely on supervised classification. We argue that supplementing training with unsupervised tasks is essential to overcome the scarcity of labelled Orthoptera data.

Large models that were trained on soundscape datasets, such as BirdNET (Kahl et al., 2021) and Perch (Hamer et al., 2023), have demonstrated good performance on birds. Embeddings extracted from these models capture general-purpose audio representations from soundscapes and are useful for training downstream tasks. We distilled generalised knowledge from BirdNET by matching its embeddings.

Models trained on large global datasets often have lower performance on local datasets due to domain shift. Using multi-task training (Caruana, 1997), and combining supervision for known taxonomic classification, distillation for general ecoacoustic knowledge, and self-supervised learn-

¹Leverhulme Centre for Nature Recovery, School of Geography and the Environment, University of Oxford, Oxford, UK. Correspondence to: Olga Isupova <olga.isupova@ouce.ox.ac.uk>.

ing (SSL) for local acoustic profile, we address the data scarcity and domain shift problems.

3. Methodology

We propose Passive acoUstic Latent-Space Encoder (PULSE): a multi-task framework that includes unsupervised and supervised training.

3.1. Data

We used publicly available labelled data for Orthoptera and supplemented it with a large, unlabelled dataset of UK field recordings. A small subset of field recordings included labels provided by experts through active learning.

For our labelled data we used the recently compiled ECOSoundSet, which collates Orthopteran and other ecoacoustic data (Funosas et al., 2026), and online datasets such as Xeno-canto (Vellinga, 2024) and iNaturalist, collected by the script provided by the authors of Funosas et al. (2026). Refer to Appendix A.3 for details. We selected a subset of 19 Orthoptera species from these sources that are endemic to, or present in the UK (Table 3).

We supplemented the labelled data with over 130 GB collection of unlabeled field recordings collected across 10 sites in Oxfordshire, UK. We have released this data together with labels collected with active learning (Section 3.5). It is available at <https://doi.org/10.5281/zenodo.20625749>.

3.2. Backbone

First, an audio is transformed into a mel spectrogram. To process input spectrograms into embeddings e we use a VGGish (Hershey et al., 2017) backbone. The architecture was modified to match our spectrogram shape for input and BirdNET embedding size for output since we used BirdNET as ecological prior.

3.3. Multi-task Framework

Backbone embeddings serve as an input to our multi-task framework, that combines three optimisation objectives. Figure 1 shows the framework architecture.

Supervised classification. Using available labelled data, we optimised the classification head h_{cls} . Since recordings can contain several species vocalising at the same time, we have a multilabel problem, therefore, h_{cls} has the binary cross-entropy loss \mathcal{L}_{cls} .

Ecological prior. The pretrained BirdNET model provides useful representations for ecoacoustics. We encouraged our embeddings e to match BirdNET embeddings on unlabelled data. This was achieved by optimising the mean squared L_2 loss \mathcal{L}_{mse} between our embeddings and the frozen BirdNET

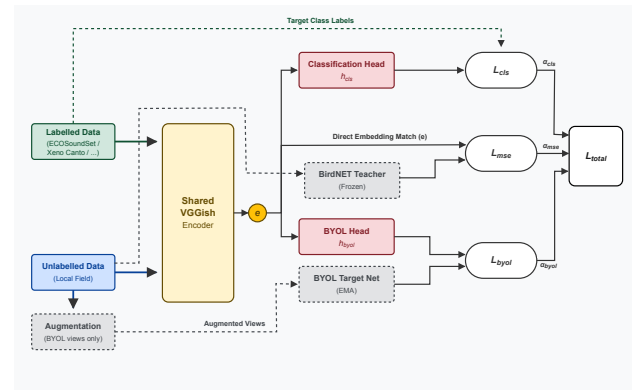


Figure 1. PULSE framework architecture with three tasks.

output. Since we matched BirdNET, we also work with 3-sec input clips.

Self-supervision. To adopt to our domain soundscapes, h_{byol} head of the framework used the unlabelled field data and applied the Bootstrap Your Own Latent (Grill et al., 2020) method. We augmented spectrograms with time and frequency masking to create two views of the same data. Using the online and target networks, updated with exponential moving average, we optimised the L_2 loss \mathcal{L}_{byol} between the normalised predictions of the online network and the projections of the target network.

3.4. Joint Optimisation

During training, we created batches sampled for each of three tasks simultaneously. We then optimised the weighted sum of the three loss functions: $\mathcal{L}_{total} = \alpha_{cls}\mathcal{L}_{cls} + \alpha_{mse}\mathcal{L}_{mse} + \alpha_{byol}\mathcal{L}_{byol}$, where we set $\alpha_{cls} = 0.9$, $\alpha_{mse} = 0.1$, $\alpha_{byol} = 0.1$. These weights accommodate significant imbalance between labelled and unlabelled data.

3.5. Active Learning

We used PULSE as a base model for labelling a part of field data with active learning using the DIRECT method (Nuggehalli et al., 2025) designed for imbalanced data. We followed the standard cycle: train PULSE, predict on unlabelled data, select the most informative batch with DIRECT, manually label it, and repeat. Labelling was conducted using Whombat (Balvanera et al., 2025)¹.

4. Experiments

In this section, we evaluate PULSE on both classification performance and embedding quality; implementation details are in Appendix A. Test data was from 3 held-out sites entirely excluded from training – including unlabelled data

¹Can be found at <https://github.com/mbsantiago/whombat/>

Table 1. Benchmark models and data they use for training.

Model	PULSE species library	Perch species library	Unlabelled field	Labelled field
PULSE	✓	—	✓	—
active PULSE	✓	—	✓	✓
pretrain Perch	—	✓	—	—
probe Perch	—	✓	—	✓

Table 2. Benchmark results on held out test data.

Model	Macro F1	Macro AUC	Macro AP
PULSE	0.212	0.742	0.322
active PULSE	0.338	0.838	0.556
pretrain Perch	0.071	0.454	0.189
probe Perch	0.329	0.832	0.439

– providing a direct assessment of model transferability to unseen environments.

4.1. Benchmark Against Baselines

As a baseline, we used Perch 2.0 (van Merriënboer et al., 2025), a recent extension of Perch (Hamer et al., 2023). Perch 2.0 has been shown to be very adaptable to new scenarios and species, even outperforming the specialist model for particular species (van Merriënboer et al., 2025).

We compared two versions of PULSE with two versions of Perch depending on whether they used the labelled field data or not (Table 1): *PULSE* and *pretrain Perch* only used labels from species library data, while *active PULSE* and *probe Perch* also used labels from the field data. Table 2 shows the performance results. We found that PULSE significantly outperformed pretrain Perch, and active PULSE outperformed probe Perch. In both cases — with and without local labels — the specialist PULSE was more adaptive

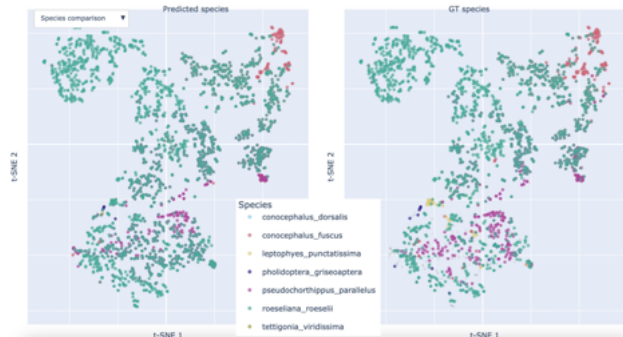


Figure 2. Embedding visualisation tool for ecological analysis.

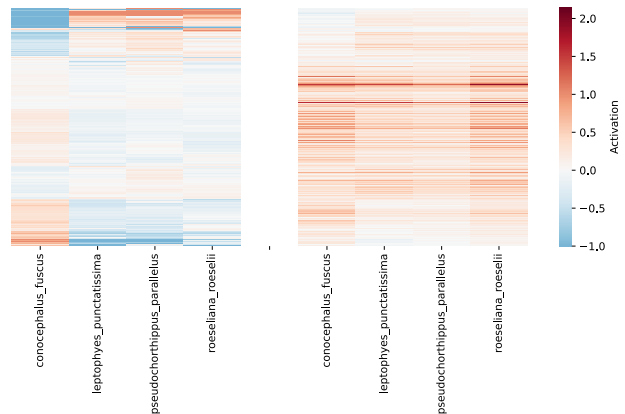


Figure 3. Selectivity matrix of 4 dominant species (left 4 columns) and examples of embeddings with individual species present (right 4 columns).

to the new environment than the general-purpose Perch.

PULSE generalises beyond Orthoptera: Appendix C presents results of PULSE extended to background classes such as cars and birds.

4.2. Embedding Analysis

As well as classification, the aim of our work was to build a model producing informative embeddings that will help ecologists with scientific discovery. We created an interactive visualisation of the embedding space (Figure 2), which can colour data points by predicted species; and, if ground truth data is available, by species labels; the sites where the recording were made and time of day. This tool will help ecologists to explore and understand their data. Qualitative analysis of the embedding space with this visualisation tool is in Appendix D.

To further characterise the embedding structure, we investigated species-specific information encoded in the embedding dimensions. For each of the four dominant species, we computed a selectivity index per embedding dimension — measuring how strongly each dimension responds to that species relative to all others (Morcos et al., 2018) — and visualised the resulting selectivity matrix as a clustered heatmap (Figure 3). Individual sample embeddings (on the right of Figure 3) overlay naturally onto this structure, confirming that per-recording activations reflect their ground-truth species labels. One-way ANOVA across species confirmed that the learned embeddings were broadly informative: 891 of 1024 dimensions showed significant between-species variation ($F > 10$), with a median F-statistic of 39, indicating that discriminative structure was distributed across the full embedding space rather than concentrated in

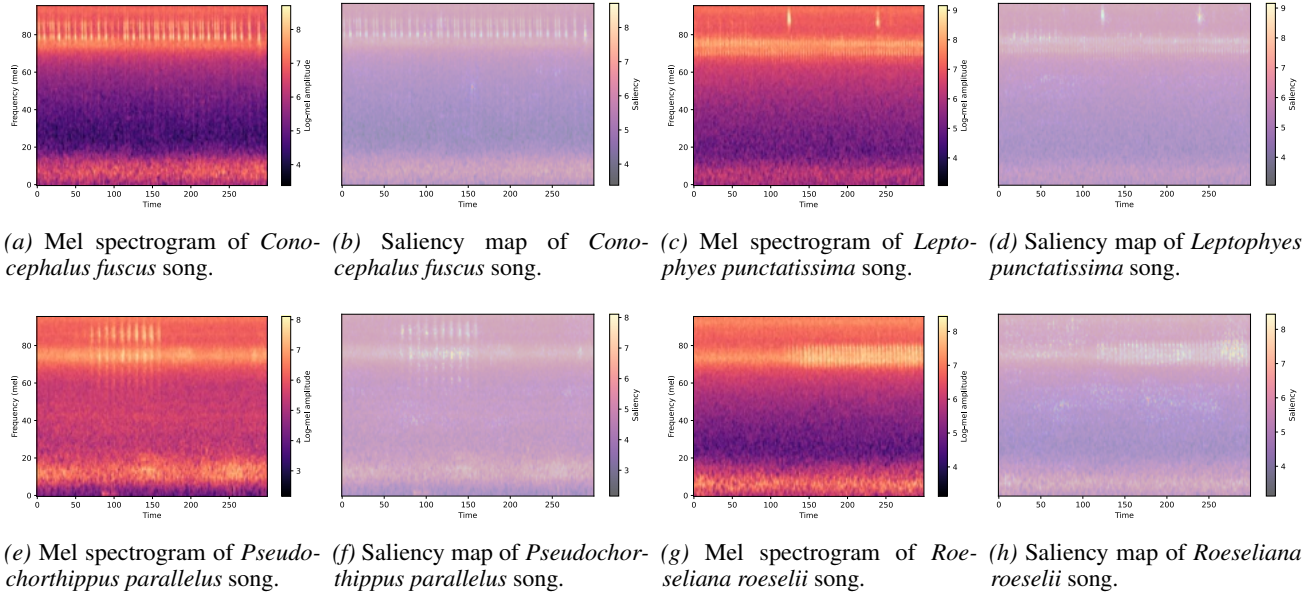


Figure 4. Saliency maps of single species recordings picturing all 4 dominant classes.

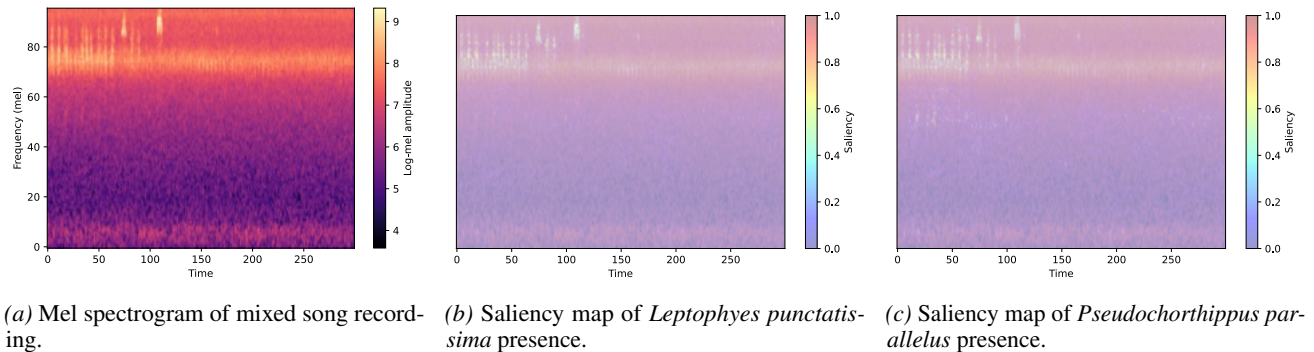


Figure 5. Saliency maps of detected components of a mixed song recording. NNLS method on mean profiles determines correct combination of species for this recording: *Leptophyes punctatissima* and *Pseudochorthippus parallelus*.

a small subset of dimensions.

We validated the selectivity profiles by computing Integrated Gradients saliency maps (Sundararajan et al., 2017) for single-species recordings, weighted by each species’ selectivity profile (Figure 4). The high-saliency regions corresponded closely to the key features of the call, namely the core frequency band and the areas of greatest amplitude, as seen in Figure 4b, suggesting the model was well attuned to the distinguishing parts of the song.

We further applied this analysis to unmix recordings with multiple species (a proxy to source separation). For a recording with multiple species, we first decomposed its embedding via non-negative least squares (NNLS) onto per-species mean prototype vectors to identify which species were present. For each detected species, we then computed a selectivity-weighted saliency map over the input spectrogram. Figure 5 shows an example of two-species recording

where NNLS correctly identifies the species. Figure 5b shows elevated saliency for both species, whereas Figure 5c correctly assigns low saliency to the other species’ song. This asymmetry may reflect differences in how distinctly each species’ call is represented, or patterns in the training data, such as common combinations with other species.

5. Conclusion

Using PULSE and active learning, we combined labelled and unlabelled data from newly available insect sound libraries to build a bioacoustic classifier tailored to UK Orthoptera. The learned embeddings revealed intraspecific behavioural differences and their environmental drivers, and enabled unmixing of overlapping calls using the embedding space alone. Ultimately, PULSE will facilitate large-scale insect PAM, providing vital information for conservation and habitat management.

Acknowledgements

The authors are members of the Leverhulme Centre for Nature Recovery and are grateful for the support of the Leverhulme Trust (grant number RC-2021-076). We would also like to thank the anonymous reviewers for their constructive comments and suggestions, which helped improve this paper.

Impact Statement

This paper presents work for which the goal is to advance the field of Machine Learning and the automatization of bioacoustic signals for monitoring ecological health. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

- Aodha, O. M., Balvanera, S. M., Damstra, E., Cooke, M., Eichinski, P., Browning, E., Barataud, M., Boughey, K., Coles, R., Giacomini, G., et al. Towards a general approach for bat echolocation detection and classification. *bioRxiv*, pp. 2022–12, 2022.
- Balvanera, S. M., Mac Aodha, O., Weldy, M. J., Pringle, H., Browning, E., and Jones, K. E. Whombat: An open-source audio annotation tool for machine learning assisted bioacoustics. *Methods in Ecology and Evolution*, 16(1): 19–28, 2025.
- Bennett, D., Nissen, H., Maschke, M. A., Reck, H., and Diekötter, T. Recent technological developments allow for passive acoustic monitoring of orthoptera (grasshoppers and crickets) in research and conservation across a broad range of temporal and spatial scales. *Basic and Applied Ecology*, 84:147–157, 2025.
- Caruana, R. Multitask learning. *Machine learning*, 28(1): 41–75, 1997.
- Faiß, M. and Stowell, D. Adaptive representations of sound for automatic insect recognition. *PLOS Computational Biology*, 19(10):e1011541, 2023.
- Fonseca, E., Favory, X., Pons, J., Font, F., and Serra, X. FSD50K: an open dataset of human-labeled sound events. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:829–852, 2021.
- Frommolt, K.-H. The archive of animal sounds at the Humboldt-University of Berlin. *Bioacoustics*, 6(4):293–296, 1996.
- Funosas, D., Massol, E., Bas, Y., Schmidt, S., Bennett, D., Arend, D., Gebhard, A., Barbaro, L., König, S., Font, R. C., et al. A finely annotated dataset for the automated acoustic identification of european orthoptera and cicadidae. *Scientific Data*, 2026.
- Gibb, R., Browning, E., Glover-Kapfer, P., and Jones, K. E. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2):169–185, 2019.
- Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al. Bootstrap your own latent—a new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33:21271–21284, 2020.
- Hamer, J., Triantafillou, E., Van Merriënboer, B., Kahl, S., Klinck, H., Denton, T., and Dumoulin, V. Birb: A generalization benchmark for information retrieval in bioacoustics. *arXiv preprint arXiv:2312.07439*, 2023.
- Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., et al. CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 131–135. IEEE, 2017.
- Kahl, S., Wood, C. M., Eibl, M., and Klinck, H. Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 2021.
- Morcos, A. S., Barrett, D. G., Rabinowitz, N. C., and Botvinick, M. On the importance of single directions for generalization. In *International Conference on Learning Representations*, 2018.
- Newson, S. E., Bas, Y., Murray, A., and Gillings, S. Potential for coupling the monitoring of bush-crickets with established large-scale acoustic monitoring of bats. *Methods in Ecology and Evolution*, 8(9):1051–1062, 2017.
- Nuggehalli, S., Zhang, J., Jain, L. K., and Nowak, R. D. Improved algorithm for deep active learning under imbalance via optimal separation. In *International Conference on Machine Learning*, pp. 46815–46836. PMLR, 2025.
- Riede, K. and Balakrishnan, R. Acoustic monitoring for tropical insect conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 380(1928), 2025.
- Ross, S. R.-J., O’Connell, D. P., Deichmann, J. L., Desjonquères, C., Gasc, A., Phillips, J. N., Sethi, S. S., Wood, C. M., and Burivalova, Z. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology*, 37(4):959–975, 2023.

- Sundararajan, M., Taly, A., and Yan, Q. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, pp. 3319–3328. PMLR, 2017.
- van Merriënboer, B., Dumoulin, V., Hamer, J., Harrell, L., Burns, A., and Denton, T. Perch 2.0: The bittern lesson for bioacoustics. *arXiv preprint arXiv:2508.04665*, 2025.
- Vellinga, W. Xeno-canto – orthoptera sounds from around the world. *Xeno-canto Foundation for Nature Sounds. Occurrence dataset <https://doi.org/10.15468/76yp6j>*, pp. 06–06, 2024.
- Walker, S. E. and Cade, W. H. The effects of temperature and age on calling song in a field cricket with a complex calling song, *Teleogryllus oceanicus* (orthoptera: Gryllidae). *Canadian Journal of Zoology*, 81(8):1414–1420, 2003.
- Xiao, H., Liu, D., Chen, K., and Zhu, M. AMResNet: An automatic recognition model of bird sounds in real environment. *Applied Acoustics*, 201:109121, 2022.

Table 3. List of target Orthoptera species

Latin name	Common name
<i>Acheta domesticus</i>	house cricket
<i>Chorthippus albomarginatus</i>	lesser marsh grasshopper
<i>Chorthippus brunneus</i>	common field grasshopper
<i>Conocephalus dorsalis</i>	short-winged conehead
<i>Conocephalus fuscus</i>	long-winged conehead
<i>Gryllus bimaculatus</i>	two-spotted cricket
<i>Leptophyes punctatissima</i>	speckled bush-cricket
<i>Meconema meridionale</i>	southern oak bush cricket
<i>Meconema thalassinum</i>	oak bush-cricket
<i>Metrioptera brachyptera</i>	bog bush cricket
<i>Myrmeleotettix maculatus</i>	mottled grasshopper
<i>Omocestus rufipes</i>	woodland grasshopper
<i>Omocestus viridulus</i>	common green grasshopper
<i>Pholidoptera griseoptera</i>	dark bush-cricket
<i>Pseudochorthippus parallelus</i>	meadow grasshopper
<i>Roeseliana roeselii</i>	European bush-cricket
<i>Stenobothrus lineatus</i>	stripe-winged grasshopper
<i>Stethophyma grossum</i>	large marsh grasshopper
<i>Tettigonia viridissima</i>	great green bush-cricket

Table 4. List of target Orthoptera species

Background category
Anthropophony
Aves
Bat
Geophony
Passeriformes
Strigiformes

A. Implementation details

In this section we present all technical details of data and approaches we used.

A.1. Field data collection

Our field data was collected in summer 2024 across 10 sites in Oxfordshire, UK. On each site we installed 2 Audiomoths, a low-cost acoustic logger, that recorded 15 sec segments every minute at specified times: 4:00-6:00, 10:00-12:00, 17:00-19:00, 21:00-22:00 — for 4-6 days at 96kHz sampling rate. Note that there were some failures and not all segments were recorded with one Audiomoth failing completely.

A.2. List of species of interest

Table 3 gives the full list of 19 Orthoptera species used in our study. Table 4 presents the list of additional 6 background classes we used for extended PULSE in Appendix C.

A.3. Public data details

For labelled species library data we used a subset of ECOSoundSet data (Funosas et al., 2026), subsampling only data containing our selected list of species (Table 3). ECOSoundSet has a mix of strong (a bounding box for each sound event) and weak (indicator that a recording contains a species) labels. For this work, we only worked with weak labels, therefore, all strong labels were converted to weak labels. Recordings in ECOSoundSet were of varying length, we split them into 3-sec clips.

Table 5. Species with significant presence in test data

Species	Number of test 15-sec recordings where it is present
<i>Conocephalus fuscus</i>	41
<i>Leptophyes punctatissima</i>	14
<i>Pseudochorthippus parallelus</i>	46
<i>Roeseliana roeselii</i>	105

In addition to ECOSoundSet we use other online sources with the download script provided by (Funosas et al., 2026). Online sources include Xeno-canto (Vellinga, 2024), iNaturalist, observation.org, SNSB (Bavarian State Collections of Natural History) and MinIO. Here we again used a subset of the data with the selected species (Table 3). This data include only weak labels. Recordings in these datasets are of varying length with a label available for the whole recording only. For each recording we cropped up to 5 clips of 3 sec around the peaks² (similar to Perch (van Merriënboer et al., 2025) and BirdNET (Kahl et al., 2021)).

Perch 2.0 (van Merriënboer et al., 2025), that we used as a baseline, was originally trained also on Xeno-Canto and iNaturalist, as well as additionally on the Tierstimmenarchiv (Frommolt, 1996) and FSD50K (Fonseca et al., 2021) datasets.

A.4. Training details

We held out three of the 10 sites for the test dataset to evaluate transferability, while data from remaining seven sites was randomly split between train and validation datasets in the ratio 8/1, respectively. A part of the train data was labelled with the active learning approach (Section 3.5), obtaining 552 labelled 15-sec recordings. We used the label taxonomy from ECOSoundSet (Funosas et al., 2026). Parts of the validation and test datasets were labelled at random, resulting in 67 and 197 labelled 15-sec recordings, respectively. These 15-sec recordings were split into 3-sec clips with overlap of 1 sec to be fed into the classification head of PULSE.

All recordings (including unlabeled) were split into 3-sec clips to be used in the ecological prior head for comparison with the BirdNET embeddings. Additionally, clips of three seconds around one peak of the full 15-sec recording were extracted for the BYOL head.

ECOSoundSet and open online data were split at random into train, validation and test datasets in the ratio 8/1/1, respectively. Data from ECOSoundSet and online sources was used by the classifier head.

We trained PULSE for 20 epochs, Adam optimiser with learning rate $1e - 5$, exponential decay scheduled every 100,000 steps with rate 0.96, early stopping with patience 5 evaluated on validation data loss.

We compared PULSE with pretrained Perch (van Merriënboer et al., 2025). We used two versions of Perch. For the first version (*pretrain Perch*), we used the model’s direct predictions for our target classes. Since Perch was trained on the same Xeno-canto and iNaturalist datasets, it has all our selected species (Table 3). Secondly, for *probe Perch* we used embeddings from pretrained Perch and trained a linear probe connecting embeddings to logits for 19 selected species with binary cross-entropy loss. Linear probe Perch was trained for 50 epochs, Adam optimiser with learning rate $1e - 3$, early stopping with patience 5 evaluated on validation data loss.

For both methods we also used the validation dataset to tune classification threshold for positive predictions per class.

A.5. Evaluation metrics

For main results in Table 2, we used macro metrics — AUC, AP — over all classes that have non-zero support in labelled test data. For macro F1 we also included classes with only false positives.

For embedding analysis (Section 4.2), we focussed on 4 classes that had significant presence in the test data (Table 5). The selectivity matrix, F-statistics and mean prototypes were computed over a balanced train data, whereas all examples in Section 4.2 are from the test data. For balancing of the train data we subsampled it to the size of the smallest class using a rarest-first, purity-prioritised procedure. Classes were processed in ascending order of sample count, so minority

²We used `scipy.signal.findpeaks_cwt()` function in Python to extract peaks.

Table 6. PULSE with background classes results on held out test data.

Model	Macro F1-O	Macro AUC-O	MacroAP-O	Macro F1	Macro AUC	Macro AP
PULSE with back-ground classes	0.383	0.827	0.465	0.363	0.749	0.458

classes had first pick of examples. Within each class, candidates were prioritised by the number of active labels they carry — single-label recordings are selected first, followed by two-label recordings, and so on — with random selection within each purity tier to avoid index bias. Samples already selected for a previously processed class counted toward the quota of subsequent classes, preventing double-counting of overlapping multilabel recordings.

We computed a selectivity matrix as $S_{d,k} = \frac{\mu_{d,k} - \mu_{d,-k}}{|\mu_{d,k}| + |\mu_{d,-k}| + \epsilon}$, where d is a dimensionality of the embedding space, k is a class, $\mu_{d,k}$ is the mean for dimensionality d and class k among all recordings, $\mu_{d,-k}$ is the mean for dimensionality d for all classes other than k . sch

It is worth mentioning, whilst [Morcos et al. \(2018\)](#) caution that class selectivity in individual units does not causally explain network performance, our use of the selectivity index was purely diagnostic — we employed it as an interpretive tool to characterise the structure of learned representations rather than to make claims about which dimensions are necessary or sufficient for classification.

For Integrated Gradients ([Sundararajan et al., 2017](#)) we used zero baseline and 50 training steps. Note that the target signal is the selectivity-weighted sum of embedding dimensions rather than a class logit.

Mean prototype vectors used for unmixing multi-species recordings are just the mean of embeddings containing a species in the balanced train data.

Species presence in mixed recordings was determined via NNLS: species with coefficients below 0.01 were deemed absent.

B. Data source comparison

The recordings from our field study sites provided extensive audio data for analysis. However, the passive nature of this method and use of relatively low-cost devices (AudioMoths) resulted in some noticeable differences compared with publicly available audio databases such as Xeno-canto.

The annotated 15-sec clips contained a single species in 51% of cases, two-species in 38% and more than two species in the remaining 11%. Xeno-canto recordings in comparison tended to be more focused on a single species, and whilst multi-species recordings did occur, they appeared to be much less frequent.

Recordings on Xeno-canto were generally collected using higher-specification microphones and were selectively collected at locations and timings to maximise the clarity and isolation of individual species’ songs. In contrast, our field study recordings captured numerous additional sound sources alongside the target Orthoptera, including birds, wind, rain, and traffic.

C. Additional Results

Here we present the results of PULSE when we extended classes to include both selected Orthoptera species (Table 3) and background classes (Table 4). Both ECOSoundSet data and our labelled field data contain training examples for background classes. Results are given in Table 6. We present both metrics computed for Orthoptera classes only (metrics with “-O” postfix) and for all classes. The results show that performance on Orthoptera classes remained competitive in comparison to PULSE trained on Orthoptera classes only (see *active PULSE* in Table 2) and that performance on all classes remained at a comparable level. This suggests that PULSE can generalise to broader soundscape analysis, not only Orthoptera classification.

D. Qualitative Embedding Analysis

This section provides analysis based on interactive embedding visualisation of the labelled train data.

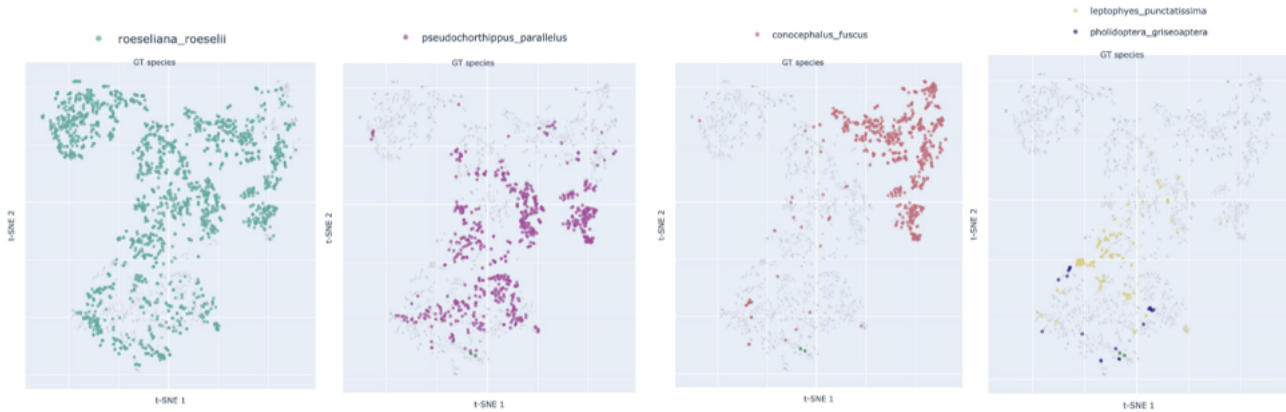


Figure 6. Embedding visualisation of labelled train data. The three left-hand plots are coloured by presence of the three most prevalent species; the right-most plot shows the less common species.

D.1. Embedding Space Reveals Clear Separation for Two of the Three Most Common Species but Wider Dispersion for the Third

Three species were prevalent across the annotated dataset: *Roeseliana roesellii* (Rr), *Conocephalus fuscus* (Cf), and *Pseudochorthippus parallelus* (Pp). A further two species were less frequently observed: *Pholidoptera griseoaptera* (Pg) and *Leptophyes punctatissima* (Lp)³ (Figure 6). Among the three dominant species, Pp and Cf occupied relatively distinct regions of the embedding space, with some overlap where both species co-occurred within a single embedding (see Appendix D.2). Embeddings containing Rr, the most frequent of the three, were distributed more widely. Some of the contributing factors for this are discussed in subsequent sections.

D.2. Embeddings with Multi-Species Composition Showed a Tendency to Cluster

Clustering of embeddings appeared to reflect both the number of species present and their particular combinations. Furthermore, multi-species embeddings tended to be surrounded by single-species embeddings of one constituent species, suggesting that one species acted as the dominant feature in the embedding (Figure 7). For instance, the cluster of Cf in the top-right of Figure 7 (solid purple ellipse) is next to a cluster of predominantly Cf and Rr immediately below (yellow dotted ellipse).

D.3. Multi-Species Embeddings with Varying Composition Can Cluster by Recording Site

Clustering of different species in certain areas of the embedding space aligned in some instances with the recording site. For example, the top-right corner of the plot contains many single-species Cf embeddings, almost all from a single site. Analysis of the spectrograms indicated that these were predominantly loud Cf songs that persist throughout the clips (Figure 8). Further examples of site-specific clusters, linked to particular species combinations or song characteristics (e.g. constant or brief), are illustrated in Figure 8.

D.4. Single-Species Embeddings, Such as *Roeseliana Roesellii*, Also Cluster due to Variations in Song

Roeseliana roesellii clustered by multi-species combination (see Appendix D.2), but the spread of their embeddings across most of the plot appeared to be influenced by multiple environmental and ecological factors that vary by site and individual. The most noticeable of these were the loudness and the speed (or pulse rate) of the song, which both vary widely as reflected in the spread of embeddings towards the edges of the plot. The duration of the song within the clip and number of individuals singing appear to also have an effect. This was also observed for *Pseudochorthippus parallelus*, however, this was less clear for other species (Figure 9).

³*Pholidoptera griseoaptera* has a very little support, only 1 recording in the test data, whereas *Leptophyes punctatissima* has enough support to be included in analysis in Section 4.2

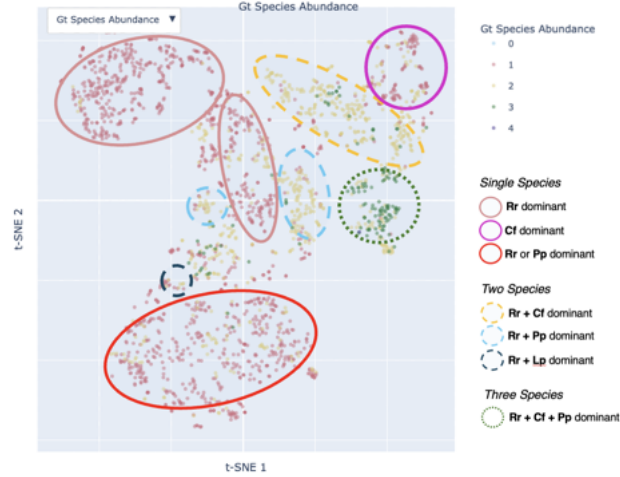


Figure 7. Embeddings coloured by the number of species present in each embedding. Manually drawn ellipses indicate visually identified clusters, with line style and colour denoting the predominant species combination within each cluster, as determined from the ground truth annotations.

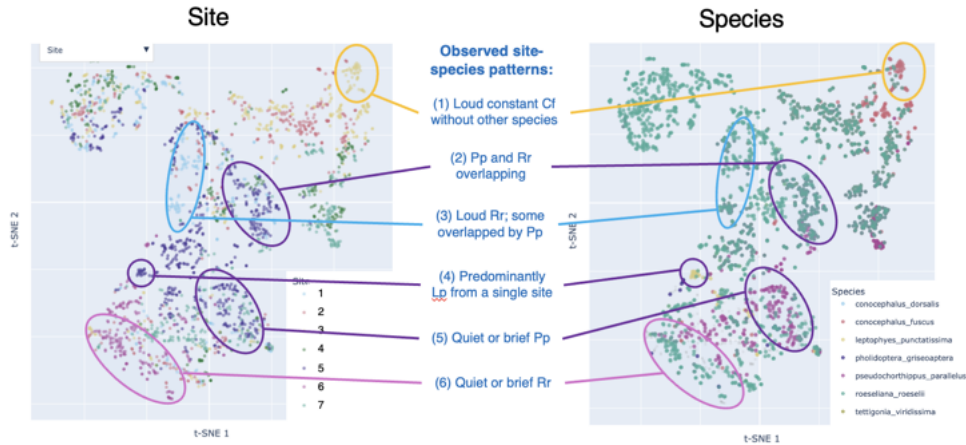


Figure 8. Embeddings coloured by recording site (left), with ellipses highlighting visible site clusters, linked to a plot coloured by species (right), illustrating how some site clusters correspond to particular species combinations or song characteristics.

D.5. Time of Day Produces Some Clear Clusters That Appear to Reflect Song Variation, Particularly in Chirp Rate, for Example *Pseudochorthippus Parallelus*

Temperature has been identified as a key factor influencing the chirp rate of various Orthoptera species (Walker & Cade, 2003). This is largely attributable to temperature variation and was observed in our data for *Pseudochorthippus parallelus* and *Roeseliana roeselii* in particular. The dawn and dusk chirp rates of these species were noticeably slower than mid-morning and late-afternoon rates (Figure 10).

D.6. AudioMoth Tonal Band (Around 15-25kHz) Can Result in Clustering Together of Brief or Quiet Songs From Different Species

A consistent tonal band, visible as a brighter region on spectrograms between 15–25kHz, appeared throughout the field dataset. This is a known characteristic of AudioMoth recordings and was present regardless of environment or recording settings. As this frequency range overlaps with the core calling frequencies of some Orthoptera species, it could reduce the distinctiveness of their songs on the spectrogram. For distant or brief calls, the tonal band became the more prominent feature, and this appeared to be the common factor linking otherwise dissimilar songs in the same cluster. This is reflected

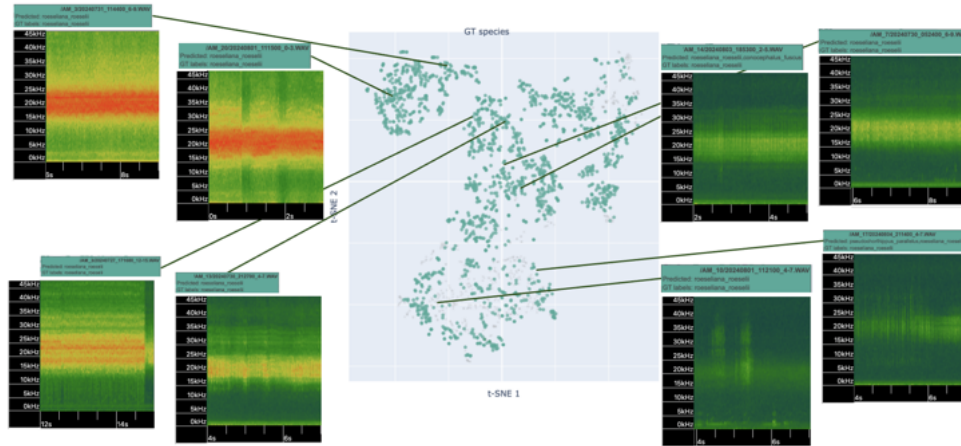


Figure 9. *Roeseliana roeselii* embeddings with a random selection of points linked to their spectrograms, illustrating the variation in song features across the embedding space.

in certain areas of the plot, where embeddings containing a single species, either Rr or Pp, sat in close proximity despite the two songs being acoustically distinct, suggesting that quieter calls were harder to differentiate at the embedding level (Figure 11).

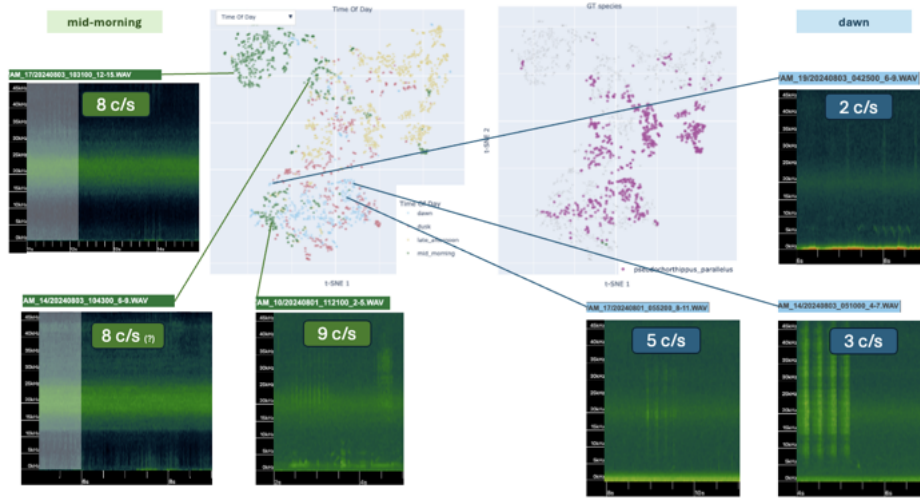


Figure 10. Embeddings with Time of Day for all annotations (left) and the spread of embeddings with *Pseudochorthippus parallelus* (right). The selected spectrograms show three from the mid-morning and three from around dawn with the approximate chirp rate added to each plot. Note the left hand two plots show the preceding 1-second as the selected clip only catches the end of the song.

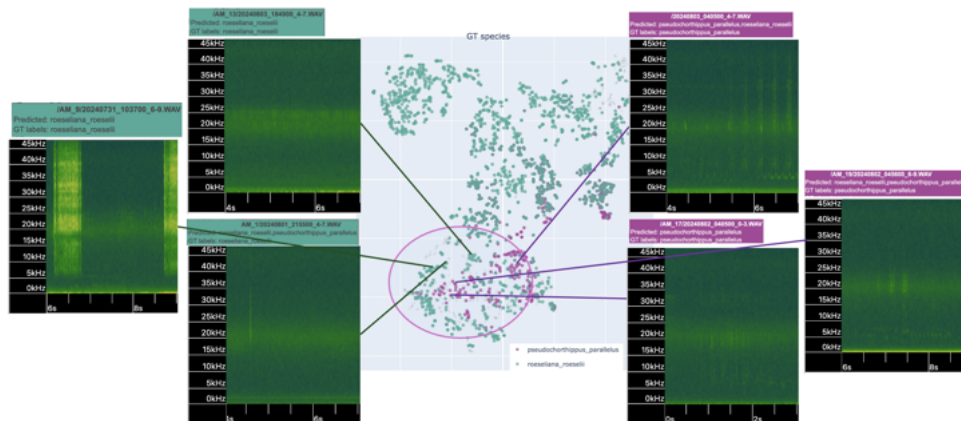


Figure 11. Single species embeddings with a focus on spectrograms where the embeddings with a single species of either *Roeseliana roeselii* (green) or *Pseudochorthippus parallelus* (purple) are placed in a similar area. The AudioMoth tonal band can be seen between 15-25 kHz in all the spectrograms.