Received: 17 April 2022 | Accepted: 10 July 2022

DOI: 10.1111/2041-210X.13960

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

APPLICATION

soundscape_IR: A source separation toolbox for exploring acoustic diversity in soundscapes

Yi-Jen Sun[1] | Shih-Ching Yen[2] | Tzu-Hao Lin[1]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan (R.O.C)

[2]Center for General Education, National Tsing Hua University, Hsinchu, Taiwan (R.O.C)

Correspondence
Tzu-Hao Lin
Email: lintzuhao@gate.sinica.edu.tw

Funding information
Observer Ecological Consultant; Industrial Technology Research Institute; Ministry of Science and Technology, Taiwan, Grant/Award Number: MOST 109-2621-B-001-007-MY3

Handling Editor: Thomas White

Abstract

1. Soundscapes contain rich acoustic information associated with animal behaviours, environmental characteristics and human activities, providing opportunities for predicting biodiversity changes and associated drivers. However, assessing the diversity of animal vocalizations remains challenging due to the interference of environmental and anthropogenic noise. A tool for separating sound sources and delineating changes in acoustic signals is crucial for an effective assessment of acoustic diversity.
2. We present `soundscape_IR`, an open-source Python toolbox dedicated to soundscape information retrieval in which nonnegative matrix factorization is applied. This toolbox provides algorithms for supervised and unsupervised source separation (SS). It also enables the use of a snapshot recording for model training and subsequently applying adaptive and semi-supervised SS when target species produce sounds with varying features and when unseen sound sources are encountered.
3. Our results demonstrated that SS could enhance the vocalizations of target species, characterize the complexity of vocal repertoires and investigate the spatio-temporal divergence of soundscapes. In tropical forest soundscapes, the application of SS effectively detected the rutting vocalizations of sika deer and revealed a graded structure in their acoustic characteristics. In subtropical estuarine soundscapes, SS automated the process of identifying distinct biotic and abiotic sounds, and the result uncovered divergent sound compositions between inshore and offshore waters.
4. Implementation of SS in soundscape analysis offers a promising method for streamlining the assessment of acoustic diversity in diverse environments. Future application of SS will open new directions to acoustically quantify ecological interactions across individual, species and ecosystem levels.

KEYWORDS
acoustic diversity, denoising, information retrieval, nonnegative matrix factorization, soundscape dynamics, vocal repertoire

1 | INTRODUCTION

Soundscape monitoring is an emerging sensing technique for biodiversity, it enables researchers to remotely evaluate the conditions of ecosystems and species by listening to biophony, geophony and anthropophony (Pijanowski et al., 2011; Sethi et al., 2020). Recent development of autonomous sound recorders facilitates the acquisition of soundscape recordings (Ross et al., 2018). The substantial acoustic data generated open new directions for computing biodiversity changes across environmental gradients, as demonstrated in subtropical forests across an urban–rural gradient (Ross et al., 2018), coral reefs present at shallow and mesophotic depths (Lin, Akamatsu, Sinniger, & Harii, 2021), and maerl beds under different management conditions (Coquereau et al., 2017).

The diversity of animal vocalizations has been used as a proxy for biodiversity (Mooney et al., 2020). However, its measurement may be influenced by environmental and anthropogenic sounds. In areas such as rainforests and coastal waters, sounds generated from weather events and human activities can change soundscape characteristics and introduce bias in acoustic analysis (Bedoya et al., 2017; Guan et al., 2015). Moreover, animals can alter their vocal repertoires among behaviours, sex, life stages and noise conditions (Laiolo, 2010). Therefore, a practical assessment of acoustic diversity requires computing techniques that can separate sound sources and delineate changes in acoustic signals.

Recent advances in machine learning have enabled several breakthroughs in audio signal processing. Supervised learning methods based on deep neural networks have been used to detect animal vocalizations in noisy environments and classify species with near-human-level accuracy (Shiu et al., 2020; Stowell et al., 2019). These models are designed to learn task-specific rules without explicit programming; however, successful applications still require a substantial amount of annotated audio for model training (Zhong et al., 2020). To reduce the need for a comprehensive audio library, unsupervised learning has been employed for grouping sounds with similar spectro-temporal representations (Keen et al., 2021; Sainburg et al., 2020; Sethi et al., 2020; Ulloa et al., 2018). Despite that, the result of unsupervised learning may still be deteriorated by noise.

This study introduced `soundscape_IR` and associated algorithms for source separation (SS). SS is a signal processing technique that reconstructs independent sound sources from a mixture (Denton et al., 2022; Lin & Tsao, 2020). On the basis of `soundscape_IR` and two real-world datasets, we present our novel approaches of (a) separating sounds of target species from noisy recordings; (b) characterizing repertoire diversity; (c) assessing the composition of biotic and abiotic sounds; and (d) evaluating spatio-temporal changes in acoustic diversity.

2 | SOURCE SEPARATION

2.1 | Package overview

The package `soundscape_IR` is an open-source Python toolbox that utilizes nonnegative matrix factorization (NMF) in SS (Figure 1). The SS algorithms built into `soundscape_IR` depend on spectro-temporal representations generated using the class `audio_visualization`. The class `audio_visualization` transforms audio into a log-scaled spectrogram and enables the use of Welch's averaging method and spectrogram prewhitening (Lin, Akamatsu, & Tsao, 2021) in noise reduction. The procedure of SS, which consists of a training phase and a prediction phase, is operated by using the class `source_separation`. In the training phase, `source_separation` supports NMF and periodicity-coded NMF (PC-NMF) for supervised and unsupervised feature learning respectively. In the prediction phase, it integrates adaptive and semi-supervised SS to challenge highly variable environments. After completion of SS, regions of interest (ROIs) can be identified using an energy thresholding method provided in the class `spectrogram_detection`.

2.2 | Model-based SS through NMF

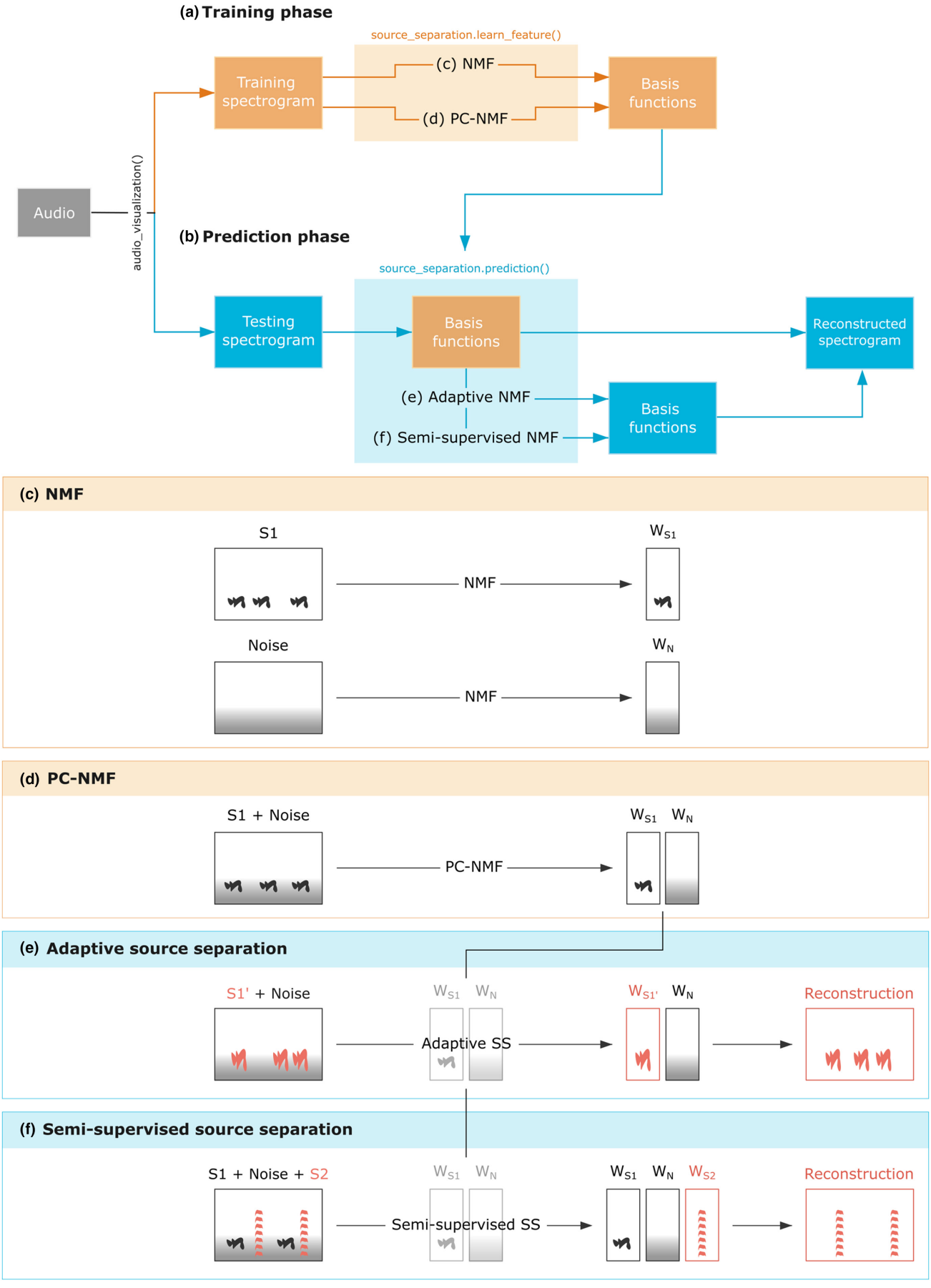
NMF is a machine learning algorithm that iteratively learns to reconstruct a nonnegative input matrix V by finding a set of basis functions W and encoding vectors H (Lin et al., 2017).

$$V_{ft} \approx (WH)_{ft} = \sum_{a=1}^{r} W_{fa} H_{at} \quad (1)$$

In SS, V is a spectrogram produced by computing discrete Fourier transforms over short overlapping/nonoverlapping windows of audio signals, f and t are the frequency and time of the spectrogram respectively and r is the number of basis functions. In `soundscape_IR`, V can also be constructed by cascading the spectrogram across consecutive frames, enabling basis functions to learn time-varying spectral features within a specified duration (Fu et al., 2017). Because NMF requires all entries to be nonnegative, a constraint that reflects the additive nature of sounds, the basis functions and encoding vectors learned through NMF often correspond to the spectral features of sound sources and their temporal activations respectively (Lin et al., 2017).

Model-based SS consists of a supervised training phase (Figure 1a) and a prediction phase (Figure 1b). In the training phase, NMF is applied to learn a set of basis functions from one target source, and this procedure is repeated for all target sources (Figure 1c). In the prediction phase, the sets of basis functions are used for predicting the temporal activations of target sources in a

FIGURE 1 Overview of using `soundscape_IR` for SS. The workflow starts with the transformation of audio recordings into spectrograms. The procedure of SS consists of (a) a model training phase and (b) a prediction phase. In the training phase, an SS model can be trained using (c) supervised NMF or (d) unsupervised PC-NMF depending on the quality of training data. In the prediction phase, (e) adaptive SS is applied if target sources alter their acoustic characteristics and (f) semi-supervised SS is used when unseen sources are encountered. After the prediction phase, the SS model generates reconstructed spectrograms of trained sources for further analysis.



mixture. This method involves fixing the basis functions but iteratively updating their encoding vectors. A time–frequency ratio mask can be subsequently estimated to reconstruct the spectrogram P_s of target source.

$$P_s = \frac{W_s H_s}{WH} \times V. \quad (2)$$

2.3 | Unsupervised and weakly supervised SS through PC-NMF

The use of noise-corrupted audio may affect the quality of feature learning in the supervised training phase (Lin & Tsao, 2020). To reduce the need for preparing clean audio, Lin et al. (2017) developed PC-NMF, an unsupervised SS model assumes that sound sources possess unique periodicities. PC-NMF consists of two NMF layers. The first layer decomposes a spectrogram into basis functions and encoding vectors, whereas the second layer separates the basis functions into groups by identifying group-specific periodicities from encoding vectors. For a mixture spectrogram V containing two sources, PC-NMF can generate two groups of basis functions, W_1 and W_2, in accordance with the periodicity hidden in H_1 and H_2 respectively (Figure 1d).

PC-NMF may not perform well when target sources do not display evident periodicity patterns (Lin, Akamatsu, & Tsao, 2021). To improve separation performance, `soundscape_IR` enables the use of PC-NMF in a weakly supervised manner. This method involves preparing annotations using Raven software (https://ravensoundsoftware.com) and generating a concatenated spectrogram of annotated fragments through `audio_visualization`. In accordance with the periodicity created, PC-NMF can identify two sets of basis functions for the target source and background noise respectively.

2.4 | Adaptive SS

Characterizing spectral variation is a key step in studying the repertoire of animal vocalizations. However, spectral variation (e.g. S_1' in Figure 1e) may reduce SS performance because the trained basis functions cannot represent the spectral features of target sources. To learn changes in spectral features, the constraint of having fixed basis functions in the prediction phase must be relaxed (Kwon et al., 2015). In `soundscape_IR`, trained basis functions can be adaptive (Figure 1e) by applying an alpha value α_s to control the ratio of updates in each iteration:

$$W_s^i = \alpha_s \widehat{W}_s^i + (1 - \alpha_s) W_s^{i-1}, \quad (3)$$

where s represents the target source, i indicates the ith iteration and $\widehat{W}_s^i$ is the estimation obtained using the standard NMF update rule.

The choice of α_s depends on the prior knowledge regarding whether the trained basis functions are representative of the target sources. If α_s equals 0, the prediction phase is based on conventional model-based SS, which assumes that the spectral features of target sources are invariant. If α_s equals 1, the basis functions are set to be freely updated (Figure S1).

2.5 | Semi-supervised SS

Inadequate coverage of all sound sources by a training dataset is a challenge commonly encountered in soundscape analysis. Such an issue may be overcome by employing semi-supervised SS, a technique that predicts the spectral features and temporal activations of sounds when only a few sound sources are known (Smaragdis et al., 2007). Semi-supervised SS assumes that new sound sources possess distinct spectral features from those already covered in the training data (e.g. S_2 in Figure 1f). This method involves adding a set of new basis functions initiated by random values into an SS model.

In `soundscape_IR`, semi-supervised SS is implemented in the prediction phase. During the iterative updating procedure, the trained basis functions are fixed (if adaptive SS is inactivated), but the newly added basis functions can update themselves through the standard NMF update rule (Figure 1f). The number of newly added basis functions is a key parameter in semi-supervised SS. For mixtures containing many new sources, a higher number of newly added basis functions can give more building blocks to perform spectrogram reconstruction (Figure S2).

3 | EXAMPLE APPLICATIONS

This study used two real-world datasets to illustrate the application of `soundscape_IR` in acoustic diversity assessment. The first dataset was a collection of tropical forest sounds that contained the rutting vocalizations of sika deer *Cervus nippon*. The second dataset was a collection of underwater sounds from a subtropical estuary.

3.1 | Rutting vocalizations of sika deer

The first case study investigated the application of SS for detecting rutting vocalizations of sika deer and assessing their repertoire diversity (see Figure S3 for an overview of analysis workflow). Rutting vocalizations play a crucial role in the breeding behaviour of sika deer, and information on their acoustic diversity can shed light on population status (Yen et al., 2013). In 2017, one Song Meter SM4 recorder (Wildlife Acoustics Inc.) was deployed at Sheding Nature Park, Pingtung, Taiwan (21°58′02.7″N, 120°49′02.4″E). The recorder was scheduled to run one 5-min recording every 15 min, with a sampling frequency of 44.1 kHz. Here, we only analysed the data recorded on 11 November 2017.

At first, we used Raven Lite 2 to annotate 28 rutting vocalizations from one 5-min recording and used `audio_visualization` to generate a concatenated spectrogram (Figure 2a). The concatenated spectrogram was subsequently used as the input of

`source_separation.learn_feature` to train an SS model in a weakly supervised manner, which was conducted by applying PC-NMF to learn two sets of basis functions for rutting vocalizations and noise respectively (Figure 2b,c). Two parameters, namely the number and duration of basis functions, may affect the outcome of feature learning. Here, we used 10 basis functions with a duration of 2.4 s, a setting that allows for capturing the most invariant region of rutting vocalizations, particularly the downsweep structure in frequencies <4 kHz. In our experience, using a larger number of basis functions is expected to learn more diverse features but may generate a set of time-shifting functions sharing the same spectral structure and reduce the abstraction of invariant features. For the duration of basis functions, we suggest choosing a minimum length that can cover the basic unit of animal vocalizations (such as a note or syllable of bird songs). Choosing a shorter duration may result in learning fragmented signals, but choosing a longer duration will slow down the computation speed.

Audio recordings collected in this case study were noisy because of various avian and insect species in the tropical forest (Figure 2d). To improve the separation performance of long-duration recordings, we enabled adaptive and semi-supervised SS in the operation of `source_separation.prediction`. The model effectively separated rutting vocalizations from other signals (Figure 2e) and yielded a true-positive rate of 83% at a 10% false-positive rate. Furthermore, our analysis revealed that a similar performance could be obtained using only five annotations for model training (Figure S4).

Due to the linear decomposition nature of NMF, feature learning performs best for sources with time-invariant structure or stereotyped frequency modulation. For highly variable sources like rutting vocalizations, we suggest applying adaptive SS for each call to deliver a good representation (Figure S3). This study applied `spectrogram_detection` to automatically identify ROIs from the reconstructed spectrograms of rutting vocalizations (Figure 2e). One adaptive basis function was selected for each detection to capture its spectral feature. The selected basis functions were analysed using uniform manifold approximation and projection (UMAP, see McInnes et al., 2020) to project the acoustic variation into a one-dimensional coordinate system. As shown in Figure 3, the distribution of UMAP coordinates summarized the graded structure of rutting vocalizations, with signals comprising a tonal component and a burst-pulse component that fell to the left and signals comprising tonal components with rich harmonics that fell to the right.

3.2 | Subtropical estuarine soundscapes

The second case study investigated the application of SS in assessing the spatio-temporal dynamics of estuarine soundscapes along an inshore–offshore gradient (analysis workflow in Figure S5). Estuaries

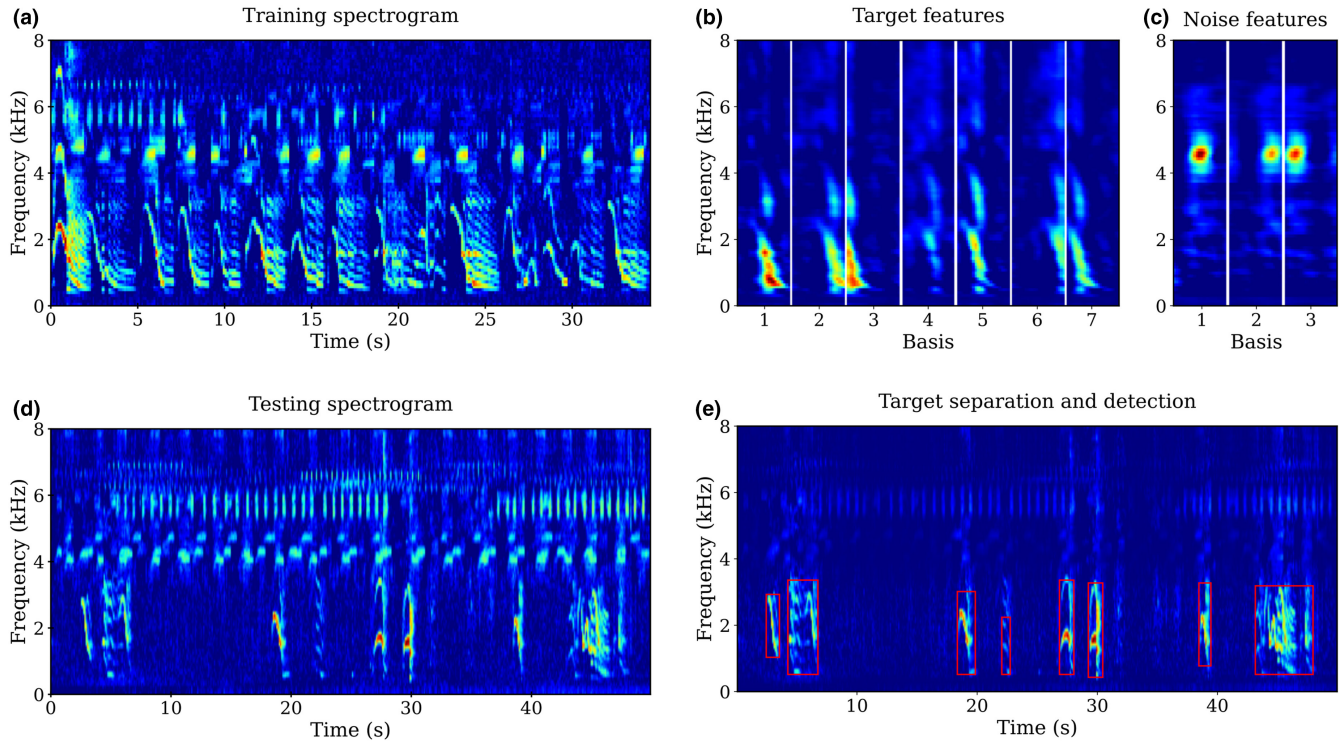


FIGURE 2 An example of applying SS in the automatic detection of sika deer calls. (a) The concatenated spectrogram of sika deer calls (time resolution: 0.1 s, frequency resolution: 86.1 Hz, frequency range: 0–8 kHz) used for weakly supervised model training. (b) Basis functions learned from sika deer calls and (c) noise using PC-NMF. (d) A testing spectrogram contains sounds produced from sika deer and insects. (e) The reconstructed spectrogram of sika deer calls. Red rectangles represent the duration and frequency range automatically detected using `spectrogram_detection`.

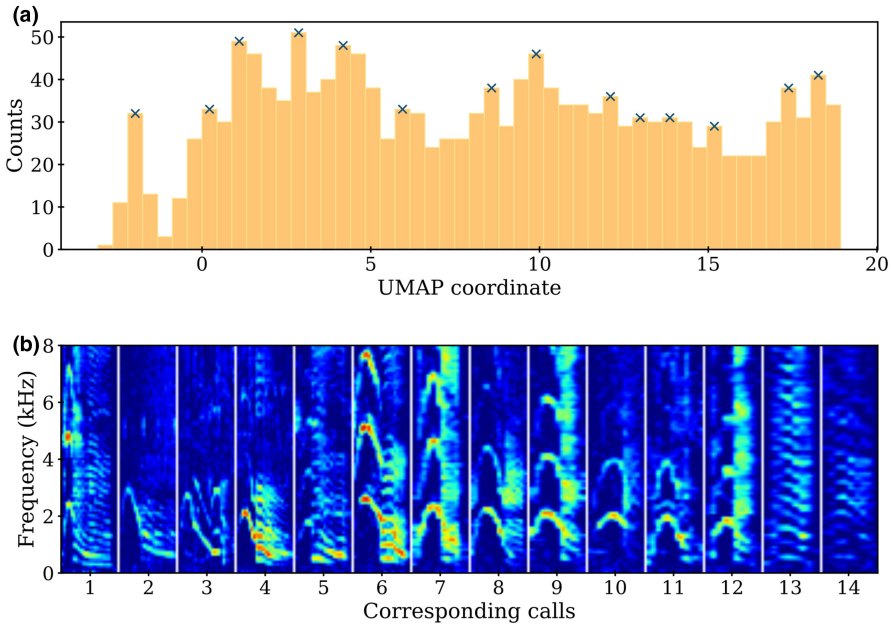


FIGURE 3 Diversity in sika deer calls. (a) Distribution of deer calls according to the 1D UMAP of adaptive basis functions. The *x*-axis represents the coordinates generated by performing 1D UMAP. The *y*-axis shows the number of calls after grouping the UMAP coordinates into 50 bins. (b) Examples of deer calls selected from coordinates marked by ×.

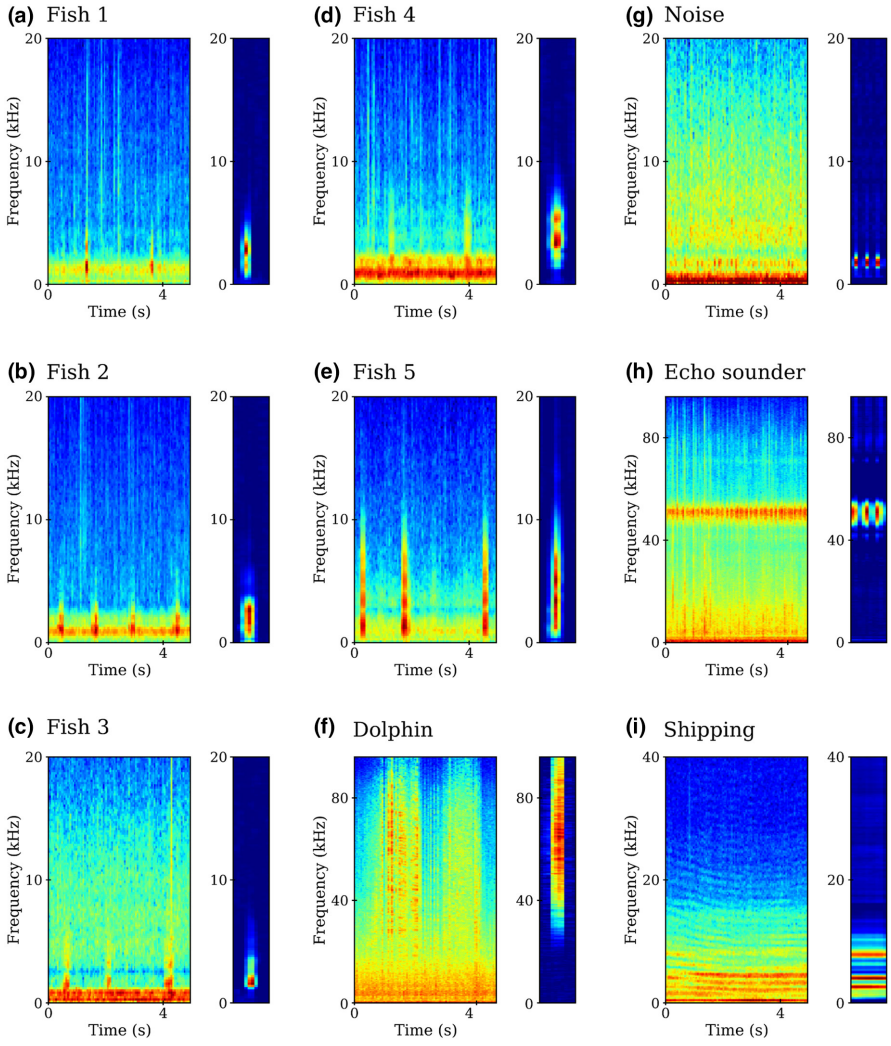


FIGURE 4 Diversity of underwater sounds recorded in Chunggang River estuary. (a)–(e) Five types of croaker calls with unique peak frequencies and frequency bandwidths. (f) Echolocation clicks of *Sousa chinensis*. Anthropogenic noise associated with (g) machine operations, (h) echo sounders and (i) shipping. For each sound type, one example was selected to show its spectrogram (left) and the associated basis function learned using semi-supervised SS (right).

support diverse aquatic species, and soundscape monitoring allows for a better understanding of human-biodiversity interactions (Guan et al., 2015). In 2018, two SoundTrap ST300 HF recorders (Ocean Instruments) were deployed in the inshore (24°40′48.4″N, 120°49′07.1″E, water depth: 9 m) and offshore waters (24°41′32.3″N, 120°48′48.8″E, water depth: 17.5 m) of Chunggang River Estuary,

Miaoli, Taiwan. Each recorder was bottom-mounted and scheduled to record sounds every 5 min using a sampling frequency of 192 kHz. This study only analysed the recordings collected on 26 August and 27 August.

To develop a generalized detector of underwater sounds, we selected one 5-min recording containing only ambient sounds in the training phase for learning 35 basis functions (Figure S6). The model was subsequently applied in the prediction phase by using semi-supervised SS to learn three new basis functions from each 5-min recording. By doing this, we assumed that sources distinct from the trained ambient sounds were captured in the new basis functions. However, this approach generated many basis functions with similar spectral structures. To challenge this issue, all basis functions learned from semi-supervised SS were organized into clusters through a density-based clustering algorithm (Sainburg et al., 2020). Based on spectral structure retained in each cluster, we identified five types of croaker calls, echolocation clicks produced by dolphins, sounds produced by machine operations, echo sounders and shipping (Figure 4).

The temporal variation of basis functions revealed a clear diurnal shift in sound types at both sites, with more fish calls observed during the night-time and more anthropogenic sounds noted during the daytime (Figure 5). However, the two recording sites exhibited distinct sound compositions. Odontocete clicks were only detected at the inshore site, which is located in the primary distribution range of Indo-Pacific humpback dolphins *Sousa chinensis*. In addition, more croaker calls (9.7%) and anthropogenic sounds (15%) were noted at the inshore site than at the offshore site (croaker calls: 4.8%, anthropogenic sounds: 1.6%).

4 | PERSPECTIVES

`soundscape_IR` is the first open-source Python toolbox that utilizes SS in soundscape information retrieval. As demonstrated in the first case study, SS can be used as a denoising filter to enhance sounds of target species. The outcome not only improves the detection of animal vocalizations but also facilitates the investigation of repertoire structure. The second case study showed that SS could uncover biotic and abiotic sounds hidden in long-duration recordings. Therefore, SS enables a comprehensive and thorough assessment of acoustic diversity.

Assessing acoustic diversity in high-biodiversity ecosystems can be challenging because of the interference from various biotic and abiotic sounds. In addition, many sound types may not be described in existing libraries (Potamitis, 2014). This study demonstrated the use of a snapshot recording for training an SS model and a procedure to generate a concatenated spectrogram of target signals for improving feature learning. Preparing a concatenated spectrogram may not be necessary if one can find a recording dominated by target signals, such as the ambient sound recording used in the second case study. Furthermore, the adaptive and semi-supervised learning

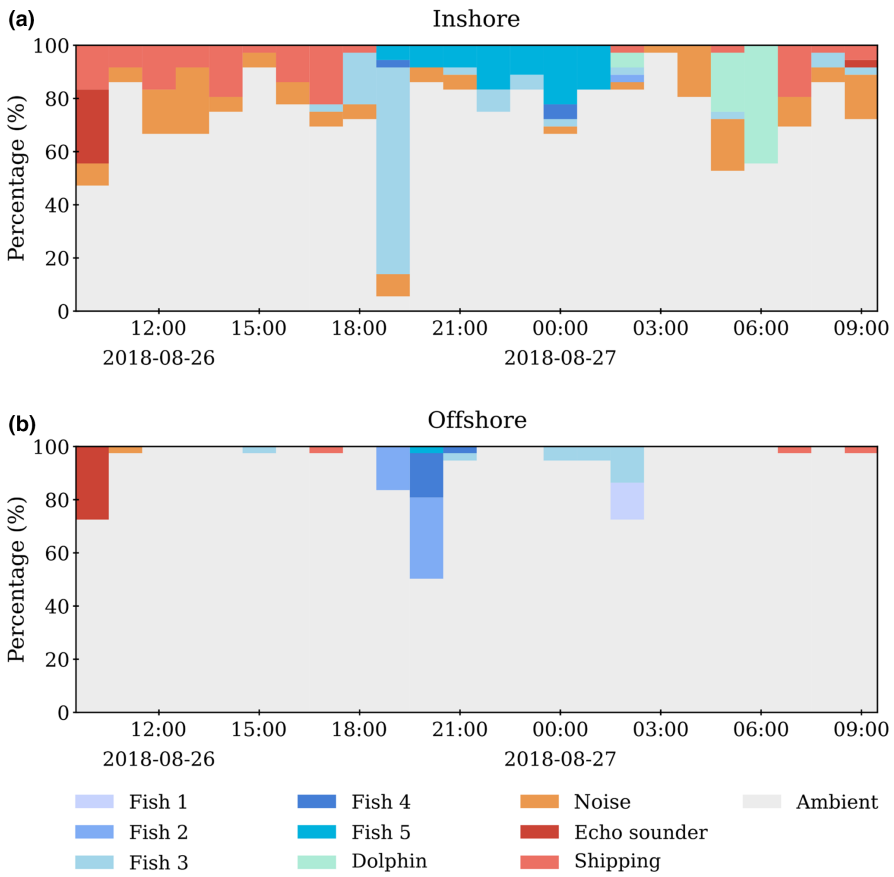


FIGURE 5 Spatio-temporal variation in underwater sounds detected in Chunggang River estuary. Data were separated into (a) the inshore site and (b) the offshore site. Colours indicate different sound types. Percentages were calculated by dividing the number of basis functions associated with a specific sound type by the total number of basis functions in each hour.

methods improve the robustness of SS, particularly when target signals deviate from training data or when data contain unseen sounds. Such capability greatly streamlines the work required in analysing long-duration recordings and facilitates acoustic diversity assessment when SS is applied in conjunction with other machine learning techniques (Denton et al., 2022).

Soundscapes contain rich information for studying ecological complexity across a full range of scales, from individuals and species to landscapes (Farina, 2019). In the first case study, we successfully used an SS model to characterize the diversity of rutting vocalizations. Spectral modulation of rutting vocalizations is associated with individual traits and social communication behaviours in sika deer, offering great opportunities for studying male competition, territory advertisement and intersexual behaviour (Yen et al., 2013). The same computing approach may be applied to species depending on sounds for communication and breeding. In the second case study, the SS result effectively revealed spatio-temporal changes in estuarine soundscapes. Diversity of biological sounds has been known to reflect taxonomic diversity and can provide insights into community structure (Mooney et al., 2020). Information on anthropogenic noise can also help identify human activities that may threaten wildlife (Sethi et al., 2020). Future application of SS will facilitate studying interactions among animal communities and their responses to anthropogenic stressors.

Implementation of SS offers a promising technique for assessing acoustic diversity in diverse environments. With an increasing number of projects now including soundscapes as an integral component of biodiversity monitoring, `soundscape_IR` also integrates Soundscape Viewer (Lin, Akamatsu, & Tsao, 2021) for visualizing soundscape dynamics and provides high-level functions in the module `batch_processing` for supporting the analysis of large-scale datasets. SS is an active research field, but challenges such as separating individual calls from social aggregation groups remain understudied and require collaborative input from multiple disciplines (Bermant, 2021). Future research efforts are critical for expanding the application of soundscape SS in biodiversity monitoring and conservation management.

AUTHOR CONTRIBUTIONS

Yi-Jen Sun and Tzu-Hao Lin conceptualized the present study. Shih-Ching Yen and Tzu-Hao Lin contributed to data acquisition. Yi-Jen Sun and Tzu-Hao Lin carried out data analysis and toolbox development. All authors contributed to the draft and approved the final submitted manuscript.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Technology, Taiwan. Field works at Sheding Nature Park were supported by the Kenting National Park Headquarters. We also thank Industrial Technology Research Institute and Observer Ecological Consultant for providing underwater audio recordings of Miaoli waters.

CONFLICT OF INTEREST

We declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All the acoustic data used in this study have been archived in the Depositar repository: https://pid.depositar.io/ark:37281/k516n1b94 (Sun et al., 2022). Source codes of `soundscape_IR` and tutorials for running the SS procedures of the two example applications are hosted on GitHub (https://github.com/meil-brcas-org/soundscape_IR) and at the open-access repository Zenodo: https://zenodo.org/record/6859143 (Lin & Sun, 2022).

ORCID

Yi-Jen Sun https://orcid.org/0000-0002-0200-8212
Shih-Ching Yen https://orcid.org/0000-0003-2219-1285
Tzu-Hao Lin https://orcid.org/0000-0002-6973-3953

REFERENCES

Bedoya, C., Isaza, C., Daza, J. M., & López, J. D. (2017). Automatic identification of rainfall in acoustic recordings. *Ecological Indicators*, *75*, 95–100. https://doi.org/10.1016/j.ecolind.2016.12.018

Bermant, P. C. (2021). BioCPPNet: Automatic bioacoustic source separation with deep neural networks. *Scientific Reports*, *11*(1), 23502. https://doi.org/10.1038/s41598-021-02790-2

Coquereau, L., Lossent, J., Grall, J., & Chauvaud, L. (2017). Marine soundscape shaped by fishing activity. *Royal Society Open Science*, *4*(1), 160606. https://doi.org/10.1098/rsos.160606

Denton, T., Wisdom, S., & Hershey, J. R. (2022). Improving bird classification with unsupervised sound separation. *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 636–640. https://doi.org/10.1109/ICASSP43922.2022.9747202

Farina, A. (2019). Ecoacoustics: A quantitative approach to investigate the ecological role of environmental sounds. *Mathematics*, *7*(1), 21. https://doi.org/10.3390/math7010021

Fu, S.-W., Li, P.-C., Lai, Y.-H., Yang, C.-C., Hsieh, L.-C., & Tsao, Y. (2017). Joint dictionary learning-based non-negative matrix factorization for voice conversion to improve speech intelligibility after oral surgery. *IEEE Transactions on Biomedical Engineering*, *64*(11), 2584–2594. https://doi.org/10.1109/TBME.2016.2644258

Guan, S., Lin, T.-H., Chou, L.-S., Vignola, J., Judge, J., & Turo, D. (2015). Dynamics of soundscape in a shallow water marine environment: A study of the habitat of the Indo-Pacific humpback dolphin. *The Journal of the Acoustical Society of America*, *137*(5), 2939–2949.

Keen, S. C., Odom, K. J., Webster, M. S., Kohn, G. M., Wright, T. F., & Araya-Salas, M. (2021). A machine learning approach for classifying and quantifying acoustic diversity. *Methods in Ecology and Evolution*, *12*(7), 1213–1225. https://doi.org/10.1111/2041-210X.13599

Kwon, K., Shin, J. W., & Kim, N. S. (2015). NMF-based speech enhancement using bases update. *IEEE Signal Processing Letters*, *22*(4), 450–454. https://doi.org/10.1109/LSP.2014.2362556

Laiolo, P. (2010). The emerging significance of bioacoustics in animal species conservation. *Biological Conservation*, *143*(7), 1635–1645. https://doi.org/10.1016/j.biocon.2010.03.025

Lin, T.-H., Akamatsu, T., Sinniger, F., & Harii, S. (2021). Exploring coral reef biodiversity via underwater soundscapes. *Biological Conservation*, *253*, 108901. https://doi.org/10.1016/j.biocon.2020.108901

Lin, T.-H., Akamatsu, T., & Tsao, Y. (2021). Sensing ecosystem dynamics via audio source separation: A case study of marine soundscapes off northeastern Taiwan. *PLoS Computational Biology*, *17*(2), e1008698. https://doi.org/10.1371/journal.pcbi.1008698

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), 4547. https://doi.org/10.1038/s41598-017-04790-7

Lin, T.-H., & Sun, Y.-J. (2022). Meil-brcas-org/soundscape_IR: Release v1.0 (v1.0). *Zenodo*, https://doi.org/10.5281/zenodo.6859143

Lin, T.-H., & Tsao, Y. (2020). Source separation in ecoacoustics: A roadmap towards versatile soundscape information retrieval. *Remote Sensing in Ecology and Conservation*, *6*(3), 236–247. https://doi.org/10.1002/rse2.141

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform manifold approximation and projection for dimension reduction. *ArXiv:1802.03426*. Retrieved from http://arxiv.org/abs/1802.03426

Mooney, T. A., Di Iorio, L., Lammers, M., Lin, T.-H., Nedelec, S. L., Parsons, M., Radford, C., Urban, E., & Stanley, J. (2020). Listening forward: Approaching marine biodiversity assessments using acoustic methods. *Royal Society Open Science*, *7*(8), 201287. https://doi.org/10.1098/rsos.201287

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. *Bioscience*, *61*(3), 203–216. https://doi.org/10.1525/bio.2011.61.3.6

Potamitis, I. (2014). Automatic classification of a taxon-rich community recorded in the wild. *PLoS ONE*, *9*(5), e96936. https://doi.org/10.1371/journal.pone.0096936

Ross, S. R. P.-J., Friedman, N. R., Dudley, K. L., Yoshimura, M., Yoshida, T., & Economo, E. P. (2018). Listening to ecosystems: Data-rich acoustic monitoring through landscape-scale sensor networks. *Ecological Research*, *33*(1), 135–147. https://doi.org/10.1007/s11284-017-1509-5

Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Computational Biology*, *16*(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

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

Shiu, Y., Palmer, K. J., Roch, M. A., Fleishman, E., Liu, X., Nosal, E.-M., Helble, T., Cholewiak, D., Gillespie, D., & Klinck, H. (2020). Deep neural networks for automated detection of marine mammal species. *Scientific Reports*, *10*(1), 607. https://doi.org/10.1038/s41598-020-57549-y

Smaragdis, P., Raj, B., & Shashanka, M. (2007). Supervised and semi-supervised separation of sounds from single-channel mixtures. In *Independent component analysis and signal separation* (pp. 414–421). Springer. https://doi.org/10.1007/978-3-540-74494-8_52

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

Sun, Y.-J., Yen, S.-C., & Lin, T.-H. (2022). *Soundscape recordings for source separation research (version 2022-07-11T10:01:25.320431)* [data set]. Retrieved from https://pid.depositar.io/ark:37281/k516n1b94

Ulloa, J. S., Aubin, T., Llusia, D., Bouveyron, C., & Sueur, J. (2018). Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis. *Ecological Indicators*, *90*, 346–355. https://doi.org/10.1016/j.ecolind.2018.03.026

Yen, S.-C., Shieh, B.-S., Wang, Y.-T., & Wang, Y. (2013). Rutting vocalizations of Formosan sika deer *Cervus nippon taiouanus*-acoustic structure, seasonal and diurnal variations, and individuality. *Zoological Science*, *30*(12), 1025–1031. https://doi.org/10.2108/zsj.30.1025

Zhong, M., LeBien, J., Campos-Cerqueira, M., Dodhia, R., Lavista Ferres, J., Velev, J. P., & Aide, T. M. (2020). Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling. *Applied Acoustics*, *166*, 107375. https://doi.org/10.1016/j.apacoust.2020.107375

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sun, Y-J, Yen, S-C, & Lin, T-H (2022). `soundscape_IR`: A source separation toolbox for exploring acoustic diversity in soundscapes. *Methods in Ecology and Evolution*, *13*, 2347–2355. https://doi.org/10.1111/2041-210X.13960