Received: 1 November 2022 | Accepted: 13 March 2023

DOI: 10.1111/2041-210X.14103

Methods in Ecology and Evolution

RESEARCH ARTICLE

Acoustic animal identification using unsupervised learning

Maria J. Guerrero[1] | Carol L. Bedoya[2] | José D. López[1] | Juan M. Daza[3] | Claudia Isaza[1]

[1]SISTEMIC, Facultad de Ingeniería, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia

[2]Atarau Sanctuary, PO BOX 2341, Christchurch, 8140, New Zealand

[3]Grupo Herpetológico de Antioquia, Instituto de Biología, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia

Correspondence
Maria J. Guerrero
Email: mariaj.guerrero@udea.edu.co

Funding information
Universidad de Antioquia, Grant/Award Number: 47/574 and 47/146; Alexander von Humboldt Institute for Research on Biological Resources; The Colombian National Fund; Francisco Jose de Caldas—MINCIENCIAS, Grant/Award Number: 111585269779

Handling Editor: Sarab Sethi

Abstract

1. Passive acoustic monitoring is usually presented as a complementary approach to monitoring wildlife communities and assessing ecosystem conditions. Automatic species detection methods support biodiversity monitoring and analysis by providing information on the presence–absence of species, which allows understanding the ecosystem structure. Therefore, different alternatives have been proposed to identify species. However, the algorithms are parameterized to identify specific species. Analysing multiple species would help to monitor and quantify biodiversity, as it includes the different taxonomic groups present in the soundscape.
2. We present an unsupervised methodology for multi-species call recognition from ecological soundscapes. The proposal is based on a clustering algorithm, specifically the learning algorithm for multivariate data analysis (LAMDA) 3pi algorithm, which automatically suggests the number of clusters associated with the sonotypes. Emphasis was made on improving the segmentation of the audio to analyse the whole soundscape without parameterizing the algorithm according to each taxonomic group.
3. To estimate the performance of our proposal, we used four datasets from different locations, years and habitats. These datasets contain sounds from the four major taxonomic groups that dominate terrestrial soundscapes (birds, amphibians, mammals and insects) in audible and ultrasonic spectra. The methodology presents performances between 75% and 96% in presence–absence species recognition.
4. Using the clusters proposed by our methodology, the whole soundscape biodiversity was measured and compared with the estimate of four acoustic indices (ACI, NP, SO and BI). Our approach performs biodiversity assessments similar to acoustic indices with the advantage of providing information about acoustic communities without the need for prior knowledge of the species present in the audio recordings.

KEYWORDS
automatic species identification, bioacoustics, clustering, passive acoustic monitoring, soundscape

1 | INTRODUCTION

The confluence of biotic, geophysical and anthropogenic sounds generates what is known as soundscape. These sounds are changing in time and space, reflecting important ecosystem processes (Pijanowski et al., 2011). Passive acoustic monitoring (PAM) is used to identify ecosystem conditions by collecting audio recordings of all the activities present in a soundscape, thus providing information about animal behaviour and the structure of the ecosystem (Dumyahn & Pijanowski, 2011; Stowell & Sueur, 2020; Sueur & Farina, 2015).

One way to carry out PAM is through species identification. Nowadays, different computational techniques allow automatic species detection of birds (Bedoya & Molles, 2021; Stowell & Sueur, 2020; Zhao et al., 2017), anurans (LeBien et al., 2020; Xie et al., 2017, 2018), mammals (Agranat, 2013; Dufourq et al., 2021; Ruff et al., 2021) and insects (Aide et al., 2017). Most automatic acoustic detection techniques perform a species-by-species analysis. However, a method that can simultaneously detect animal calls from multiple taxonomic groups is preferable to monitor and quantify biodiversity. Identification of multiple animal calls increases the difficulty of detection, especially in ecosystems with a large number of species, such as tropical ones.

Most automatic call detection techniques are based on probabilistic models (Ovaskainen et al., 2018) and machine learning techniques (Gan et al., 2020; Xie et al., 2020). In general, they follow a common four-step procedure for species identification: (i) pre-processing, (ii) segmentation, (iii) feature extraction and (iv) classification. In most frameworks based on deep learning, several of these steps are integrated into a single workflow, feature extraction and classification (Stowell, 2022).

Signal denoising and estimation of a time–frequency representation (e.g. spectrogram) are usually performed at the pre-processing stage (Stowell, 2022). In Xie et al. (2021), authors disclose different noise-reduction techniques based on the acoustic characteristics of the species of interest.

Audio segmentation is typically performed by (i) manually selecting the frequency range and time location in which an individual emitted sounds (Ducrettet et al., 2020; Premoli et al., 2021); (ii) creating templates with examples of the vocalizations (Araya-Salas & Smith-Vidaurre, 2017; Katz et al., 2016); (iii) using energy-based analysis (Bedoya, Isaza, et al., 2014; Ulloa et al., 2018); and (iv) using segmentation methods based on image analysis techniques (Potamitis, 2015; Xie et al., 2017). However, no specific segmentation technique has been studied for species belonging to different taxonomic groups that also allows segmentation for species at high frequencies.

Spectro-temporal features (e.g. call rate, dominant frequency, duration; Priyadarshani et al., 2018) are extensively used for both call description and clustering/classification tasks. Mel-frequency cepstral coefficients are commonly used as features in species identification (Bedoya, Isaza, et al., 2014; Potamitis, 2015). However, these coefficients are based on human auditory perception (30 Hz–3 kHz) and redistribute the frequency across the spectrum logarithmically, which is not useful for multi-species identification as most frequency bands are occupied.

Species identification usually relies upon supervised learning methods. Techniques based on convolutional neural networks (CNNs), support vector machines, and random forest algorithms dominate this area of research (LeBien et al., 2020; Ruff et al., 2021; Xie et al., 2018). However, working with supervised learning approaches requires labelled data, which implies prior knowledge of the species in a site of interest. This is particularly troublesome in tropical biomes and biodiversity hotspots, which are significantly understudied, have large Linnean and Wallacean shortfalls, and where most species are unknown to science (Giam et al., 2011; Joppa et al., 2011; Scheffers et al., 2012). In these regions, taxonomic and call descriptions are unavailable for most species (Brehm et al., 2008), and even for known taxa, the incompleteness of samples (Brehm et al., 2008) complicates the development of automatic recognition techniques (Acconcjaioco & Ntalampiras, 2021).

In the absence of training data, species identification and the estimation of bioacoustic inventories are foregone. Instead, acoustic indices (Sueur et al., 2008), which are proxies for biodiversity metrics, are used. These indices bypass the need for training species-specific models by generating a rapid all-encompassing biodiversity assessment based on the species acoustic energy distribution (Depraetere et al., 2012; Towsey et al., 2013). However, acoustic indices are highly dependent on the type of recorder, the recorder setup and the signal-processing parameters (Brodie et al., 2020), which makes their reliability to be continuously challenged (Bradfer-Lawrence et al., 2019; Mammides et al., 2017; Moreno-Gómez et al., 2019). Despite these flaws, there are no other viable approaches to acoustically assess biodiversity in the absence of training data. In 2018, Ulloa et al. proposed a methodology that allows decomposing acoustic communities to differentiate between sites. Their approach focuses on analysing the general soundscape composition and clustering similar sounds, but without associating clusters to species or taxonomic groups. Also, a critical stage in their approach is the selection of the number of clusters, which is challenging to estimate without prior knowledge of the area.

In this manuscript, we propose an unsupervised methodology for animal call identification and its potential applications as a biodiversity estimator. We show that by using unsupervised learning techniques, we can obtain biodiversity information similar to the one generated with acoustic indices, and recover highly accurate species-specific information. The core of our proposal is a clustering algorithm (learning algorithm for multivariate data analysis [LAMDA]; Aguilar-Martin & Mantaras, 1982), which does not require the number of classes as an input parameter. Here, LAMDA is combined with a species feature set and a segmentation stage that does not require species-specific parameter tuning. Our method is able to analyse multiple frequency bands of a recording, extract segments (possible vocalizations from species of different taxonomic groups) and cluster them using an unsupervised workflow. Thus, clusters represent sonotypes that respond to intra-species acoustic variability and allow differentiating between taxonomic groups. An expert associates the sonotypes with the calls of the species; then, this association is used to identify species-specific vocalizations in new recordings.

The number of proposed sonotypes reflects the biodiversity of the analysed site. To demonstrate the capability of our method in estimating biodiversity, we compare our results against four commonly used acoustic indices: Acoustic complexity index (ACI), bioacoustic index (BI), number of peaks (NP) and spectral occupancy (SO). Succinctly, using four case studies, we show that the species structure can be used as a biodiversity indicator. Specifically, we (1) evaluate our approach in a highly biodiverse location, where there are 39 sound-producing species; (2) compare our approach to other supervised and unsupervised methods; (3) test the generalizability of our method to independent samples from various species, sites and years; and (4) assess the ability of our approach to identify animal vocalizations in the ultrasonic spectrum without parameter settings.

2 | MATERIALS AND METHODS

2.1 | Study sites

We used four datasets seeking to measure the algorithm performance in different seasons, years, and habitats. These datasets are from tropical regions. Datasets A and B were collected in the protected area of Jaguas hydroelectric power plant (06°26 N, 075°05 W; 06°21 N, 074°59 W), on the eastern slope of the northern Cordillera Central in Antioquia, Colombia. The protected area comprises 50 km^2, including the San Lorenzo reservoir. It is dominated by different successional stages of secondary forest (70%), with the remaining areas of cropland/natural vegetation mosaic (23%), unnatural or degraded surfaces (5%) and grassland (2%). The area maintains rich communities of terrestrial vertebrates, including threatened and endemic species, and is considered paramount for biodiversity conservation at the regional scale (Sánchez-Giraldo et al., 2020).

Datasets C and D were collected from a rural area in Puerto Wilches, Santander, Colombia (7°21′52.5”N, 73°51′33.0”W). The characterized zone corresponds to a circular area delimited by a radius of 1500 m. Oil palm crops of different ages (75%), secondary vegetation (7.6%), some forest (6.13%), grassland (5.5%) and aquatic vegetation (3.2%) dominate the site. Dataset C focuses on species in the audible spectrum, whereas dataset D was acquired with an ultrasonic recorder to detect calls from bats and stridulations from orthopterans (see Section 2.2).

2.2 | Acoustics datasets

We evaluated our method in four different case studies (one dataset by each case) to demonstrate its capabilities for species identification and analyse the acoustic patterns which give an idea of the biodiversity of the site.

2.2.1 | Case 1—Dataset A

We tested the capabilities of our approach to recognize species of different taxonomical groups in a highly biodiverse site. This study case consists of 50 randomly selected recordings collected between November 2012 and February 2013 using a Song Meter SM2 device.[1] Each recording was 1-minute long, acquired every 10 minutes, with a sampling rate of 44.1 kHz using a single channel at 16-bit resolution. Species present in the soundscape were manually labelled by three experienced bioacoustic experts. They listened to the audio and reviewed the spectrograms generated in Raven.[2] The dataset contains calls from 39 species, including 21 species of birds (3 species with two different calls identified), seven species of anurans and 11 species of insects. There was disagreement between the bioacoustic experts regarding the data labels, which shows that multi-species identification is a complex issue, even when it is performed by humans (see Appendix A).

2.2.2 | Case 2—Dataset B

This dataset was used to compare our proposal with other available packages and software: Autodetec from WarbleR package (Araya-Salas & Smith-Vidaurre, 2017), MonitoR (Katz et al., 2016) and Kaleidoscope Pro[1]. The dataset includes 1000 one-minute audio recordings obtained every 15 min using a Song Meter SM4 device[1] with a 24 kHz sampling rate and 16-bit resolution. Data were manually labelled by experts using Raven[2] and Sonic Visualiser[3] . Four anuran species and one bird species were found.

To detect species using R packages, we divided the dataset into 200 recordings for each species. We then randomly split this dataset into clustering and test subsets for both Kaleidoscope Pro and our proposal, with 100 audios in each subset.

2.2.3 | Case 3—Dataset C

We use this dataset to test our method in different environments, sites and years. This is known as method-based validation (Ullmann et al., 2021). This type of external validation tests the stability of the clustering and focuses on the structural similarities of the clustering results generated by a method (Ullmann et al., 2021). It allows verifying that our results are not an artefact of our initial dataset. In addition, test the capabilities of our method to characterize the community composition and quantify each species acoustic contribution to the site biodiversity. We compare these results against acoustic indices. Dataset C was collected using a Song Meter Mini device[1], recording a minute every 10 min with a sampling rate of 48 kHz. It consists of 2638 audio recordings obtained between March and June 2021. This dataset was divided into two: a 207 recordings subset used for species identification and another 2431 recordings to compare our approach with acoustic indices. In the subset used to evaluate species identification, strict labelling work was done by experts where 11 species were labelled, including six species of birds, four species of anurans and a primate.

2.2.4 | Case 4—Dataset D

Test the capabilities to detect species in the ultrasonic spectrum. The dataset was acquired using a Song Meter Mini bat device[1] recording 15 s every 15 min with a sampling rate of 384 kHz. In all, 13 species of bats and 6 orthopterans were found. Table 2 in Appendix B presents the dataset information for each case study.

In the four case studies, we only used external validation (labels) to estimate the performance of our approach to identify the presence of the species (see Section 2.6). These labels did not partake in any stage of the process.

2.3 | Proposed approach

Our method (Figure 1) suggests sonotypes that can be associated with animal sounds present in a soundscape and group them based on acoustic similarities. The first stage (pre-processing) reduces the background noise and highlights the biotic acoustic activity. It facilitates the second stage (segmentation), where segments are extracted using Otsu thresholding (Otsu, 1979) and morphological operations. The third stage (feature extraction) estimates the dominant, minimum and maximum frequencies, and the linear-scale cepstral coefficients, which are used as input for the last stage (clustering). Then, a clustering algorithm analyses the extracted features and groups the segments based on their acoustic similarities. Finally, the clusters were associated with an animal call pattern.

This approach does not require either training or prior knowledge of the number of species and allows the estimation of acoustic structure of the site.

2.3.1 | Pre-processing

The pre-processing is performed on the magnitude spectrogram (spectrogram hereafter) $\mathbf{S} \in \mathbb{R}^{N_s \times N_t}$ of each recording, where N_s and N_t are the number of data in the axes that represent the frequency and time domains, respectively. We created a denoised spectrogram $\mathbf{S}' \in \mathbb{R}^{N_s \times N_t}$ by applying the noise-reduction method proposed by Xie et al. (2017). Spectrograms are convolved using a Gaussian kernel to remove small gaps and graininess in the audio signal. In our case, the size of the Gaussian Kernel was 3 × 3. Some noise sources cannot be reduced using Gaussian filtering; thus, the spectral subtraction technique proposed by Xie et al. (2017) was included. We only considered as background noise the sounds generated by geophonies such as wind and rain. This type of sound is broadband, with spectral components at all frequencies (Bedoya et al., 2017).

2.3.2 | Segmentation

To detect and isolate possible, we proposed modifying the (Xie et al., 2017) acoustic event detection (AED) algorithm to work on all

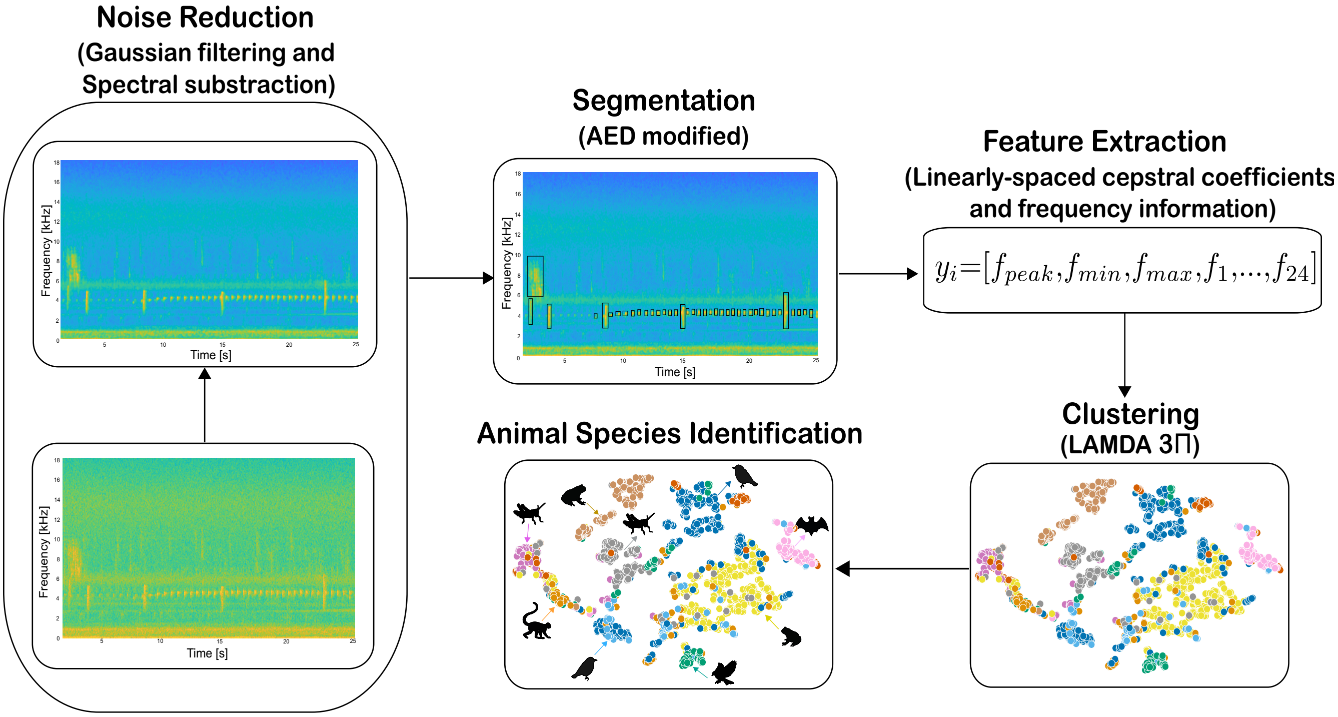

FIGURE 1 Proposed approach schema. Each recording is noise reduced and segmented. Then, the frequency information, and the linear-scale cepstral coefficients are extracted and clustered using LAMDA 3pi. The analysed segments are represented as a point for clustering, and each colour corresponds to one cluster. Each cluster has a pattern (sonotype) that will be associated with different species calls.

the frequency bands. To select the AED, different proposals were analysed (see Appendix C).

Our segmentation approach applies Otsu thresholding (Otsu, 1979) and morphological operations such as opening and closing in the $\mathbf{S}'$ spectrogram. The original method proposed by Xie et al. (2017) eliminates all segments that are not species of interest (e.g. segments above 5 kHz). Since we are interested in multiple taxonomic groups, we conserved the segments in all frequency bands. The resulting segments are stored in a matrix $\mathbf{G} \in \mathbb{R}^{N_b \times 4}$, where N_b is the number of segments, and used as spectral and temporal delimiters in the noise-reduced spectrogram during the feature extraction process.

As the segmentation we propose is based on the intensity of the vocalizations in the spectrogram and a noise reduction has been previously performed, most of the segments identified correspond to biophony. Appendix D contains more details about the segmentation process.

2.3.3 | Feature extraction

After obtaining $\mathbf{G}$, each segment is extracted directly from the noise-reduced spectrogram $\mathbf{S}'$ using the information from the bounding boxes. Then, a feature vector is computed for each segment. The first step of the feature extraction algorithm consists in estimating the logarithm of the energy for the segment $\mathbf{H} \in \mathbb{R}^{N_u \times N_h}$ (Equation (1)), where $N_u \in \mathbb{R}$ and $N_h \in \mathbb{R}$ are the length of the segmented call in the spectral and temporal domains, respectively. This operation extracts the relevant acoustic information from the temporal domain of the segment; and then, redistributes it across the spectral domain in a nonlinear way. The step is performed using windows (frames) of size w.

$$\mathbf{Q}_{i,m} = \log\left(\sum_{j=mw}^{m(w+1)} \left|\mathbf{H}_{i,j}\right|^2\right), \forall m = 1, \ldots, \frac{N_h}{w} \wedge \forall i = 1, \ldots, N_u, \qquad (1)$$

where $\mathbf{Q} \in \mathbb{R}^{N_u \times N_m}$ is a matrix with the logarithm of the energies calculated from $\mathbf{H}$, w is the size of the moving window (rectangular, no overlapping), m is the current frame and $N_m = N_h / w$ is the number of logarithms calculated in each window.

Afterward, the unitary discrete cosine transform (DCT) of $\mathbf{Q}$ is computed (Equation 2). The objective of this step is to reduce the dimensionality of $\mathbf{Q}$ and set a common length for the extracted feature vector in all animals sounds, regardless their duration or bandwidth.

$$\mathbf{Y}_{k,m} = N_p \sum_{i=1}^{N_u} \mathbf{Q}_{i,m} \cos\left(\frac{\pi}{N_u}(i-1)(k-0.5)\right), \qquad (2)$$
$$\forall k = 1, \ldots, N_k \wedge \forall m = 1, \ldots, N_m$$

where $\mathbf{Y}_{k,m}$ is a matrix that contains the DCT coefficients, k is the index of the frequency band, N_k is the number of coefficients and N_p is a normalization factor used to make the transformed matrix orthogonal. $N_p = \sqrt{1/N_u}$ for $k = 1$ and $N_p = \sqrt{2/N_u}$ for $2 \le k \le N_u$.

Usually, the first 12–13 coefficients are enough to detect a species. However, more coefficients are needed when a finer level of detail is required (Bedoya & Molles, 2021; Ntalampiras & Potamitis, 2021). In our work, 24 coefficients allowed us to have an adequate vocalization representation and, consequently, a high detection accuracy over all the studied taxonomic groups.

Finally, the extracted coefficients are concatenated with three spectral features (peak, minimum and maximum frequencies), creating a feature vector $\mathbf{y} \in \mathbb{R}^{N_f}$ for each segment $\mathbf{H}$ (sonotype). Each vector is represented as a hyper point of dimension 27 (N_f), which are then grouped in the clustering stage.

2.3.4 | Clustering

The hyper points represented by the extracted features are grouped according to their closeness in the $\mathbb{R}^{N_f}$ space. Thus, similar segments will be in a cluster. Each cluster groups the points corresponding to a call pattern, and intra-species call variability is preserved by having several clusters that can be associated with the same species. For this reason, we call each cluster a sonotype because it represents a call pattern associated with a species.

Since the number of species in each recording is unknown, it is necessary to use a clustering algorithm that does not require the number of clusters as an input parameter.

We used LAMDA (Aguilar-Martin & Mantaras, 1982) for the clustering of identified segments in soundscapes. LAMDA is a fuzzy-based method that does not require the number of classes (i.e. number of species) as an input parameter. This is in fact one of the biggest bottlenecks in the development of unsupervised approaches, as the initial number of classes is arguably the most important hyperparameter in clustering. We implemented the full-reinforcement version of LAMDA (Bedoya, Waissman, et al., 2014), which uses a fuzzy aggregation operator that naturally restricts the number of generated clusters and has been validated in bioacoustic data analyses (Bedoya & Molles, 2021; Bedoya, Waissman, et al., 2014). The features extracted from all segments were normalized before using them as input for the clustering algorithm.

The first step of LAMDA consists in calculating the marginal adequacy degrees (MADs) $\mathbf{M}$ (Equation 3), which are the contributions of the features extracted (cepstral coefficients) from each acoustic segment to each of the existent clusters (sonotype).

$$M_{c,f} = \rho_{c,f}^{\hat{y}_f}\left(1-\rho_{c,f}\right)^{1-\hat{y}_f}, \qquad (3)$$

where $\mathbf{M} \in \mathbb{R}^{N_c \times N_f}$ is a matrix with the values of the MADs extracted from the analysed element, $\rho \in \mathbb{R}^{N_c \times N_f}$ is a matrix with the mean values of the N_f features in each c-th cluster, $\hat{\mathbf{y}} \in \mathbb{R}^{N_f}$ is the vector with the normalized values of the features of the analysed element, $f = 1, \ldots, N_f$ is the current feature, $c = 1, \ldots, N_c$ is the current cluster, N_f is the number of features and N_c is the number of existent clusters.

Initially, the only predefined cluster is the non-information class (NIC), which accepts all elements equally ($\rho_{0,f} = 0.5 \, \forall f = 1, \ldots, N_f$). The first element is always assigned to the NIC; thus, it is considered unrecognized. Afterward, a new class is created with the NIC parameters modified by the values of the first element:

$$\rho_{1,f} = \frac{\left(\hat{y}_f + \rho_{0,f}\right)}{2}. \quad (4)$$

Every time a new element (a feature vector from sonotype) is analysed (using Equation 3), the obtained MADs are combined using a full reinforcement aggregation operator (Equation 5). The result of this operation is known as the global adequacy degree (GAD) $\mathbf{g}_c$ of an element to a cluster:

$$\mathbf{g}_c = \frac{\prod_{f=1}^{N_f} M_{c,f}}{\prod_{f=1}^{N_f} M_{c,f} + \prod_{f=1}^{N_f} \left(1 - M_{c,f}\right)}, \quad (5)$$

where $\mathbf{g} \in \mathbb{R}^{N_c}$ is calculated using the MADs of the new entry. Once the GADs of all clusters are obtained, the element is classified in the cluster with the maximum GAD. If such maximum GAD is the NIC class, a new cluster is created using the parameters of the NIC updated with the values of the element Equation (4). On the other hand, if an element is assigned to an existing class c, the parameters of the cluster are updated with the values of such element Equation (6)

$$\rho_{c,f}^{(k)} = \rho_{c,f}^{(k-1)} + \frac{\hat{y}_f - \rho_{c,f}^{(k-1)}}{n_c^{(k)}}, \quad (6)$$

where c is the current cluster, $n_c^{(k)}$ is the current number of elements classified in the cluster c and $\rho_{c,f}^{(k-1)}$ is the previous $k-1$ value of $\rho_{c,f}$ (before the update). A cluster is considered stable when the data do not change clusters from the previous to the current iteration. Stable clusters were obtained by iterating 10 times the LAMDA algorithm.

Each resulting cluster represents a specific sound pattern that we call sonotype. According to the sonotype represented, each group could be associated with an animal species.

The expert analyses the time–frequency information of the segments associated with the cluster, the call patterns and the most representative element obtained from the membership degrees GAD. Using this information, the expert associates sonotypes with species, allowing species identification in new audio recordings. Species with complex vocalizations, species with different call types or vocalizations with high variability will require multiple clusters to be represented.

Spectral information in the vocalization (peak, minimum and maximum frequencies), and the median and standard deviation of these three features, together with clustering information, are taken into account to identify species in new recordings.

2.4 | Bioacoustic tools parameter setting

2.4.1 | Warble R

Warble R is an R package used to analyse the structure of acoustic animal signals. It includes the Autodetect function for vocalization detection. Parameters were defined based on the acoustic characteristics of each species (frequency band, amplitude, and call length). Additionally, Autodetect has a parameter related to the amplitude threshold to differentiate the signal of interest from background noise; we performed different tests for this threshold: 10%, 15% and 20%.

2.4.2 | Monitor R

Monitor R is an R package for animal vocalization identification in large acoustic datasets. It works as an acoustic template detector. The parameters for creating the template include the time range and frequency bands (maximum and minimum) of the targeted vocalization. These were defined for each species.

2.4.3 | Kaleidoscope Pro (version 5.4.3) by Wildlife Acoustics Inc.

It is a licensed software for detecting animal vocalizations using a clustering algorithm. It requests parameters such as the frequency range, maximum and minimum detection duration, and the maximum time interval between vocalizations; we manually selected the parameters for each species. For clustering analysis, Kaleidoscope requires the maximum distance to the cluster centre, the FFT window, the number of maximum states and the maximum cluster number; these parameters were left as default, except for the FFT window for which we used 5.33 ms (128 @0–12 kHz, 256 @13–24 kHz, 512 @25–48 kHz and 1024 @49–96 kHz).

2.5 | Acoustics indices

For the third study case, we used four acoustic indices related to the species richness to compare the application of our method to estimate site biodiversity: ACI (Pieretti et al., 2010), BI (Boelman et al., 2008), NP (Gasc et al., 2013) and SO (Rojas et al., 2022; Xue et al., 2013). In general, these four indices aim to measure the contributions of biotic elements to the acoustic spectrum. We compare the trend of these indices with the application of our approach.

ACI quantifies spectral variations in the acoustic spectrum by penalizing similar energy values in adjacent frequency bins. The more heterogeneous the soundscape, the higher the value of the ACI. We use the ACIft, which is the ACI calculated along frequencies (Pieretti et al., 2010). BI quantifies the acoustic energy between 2 and 8 kHz, which is usually the frequency band with most biophonies (Boelman et al., 2008). NP measures the number of elements contributing to the acoustic spectrum (i.e. the number of peaks in the power spectral density; Gasc et al., 2013). Similarly, SO measures the percentage of the acoustic spectrum that is being used. It is done by aggregating the bandwidths of the occupied frequency bands; then, dividing by the total acoustic spectrum available (Xue et al., 2013).

A Hann-type window was used for the spectrogram calculation of the acoustic indices, with a window size of 512 and no overlapping. The frequency band was limited to 2–8 kHz for BI, and the signal was divided into 10 parts for NP. The computational tool in Python by Rendon et al. (2022) was used to estimate ACI, BI and NP, while the SO index was estimated using the implementation provided by Rojas et al. (2022).

2.6 | Evaluation metrics

The performance of each case study was evaluated according to the presence/absence detection of each species in a recording:

$$\text{Hit rate}\,[\,\%\,] = \frac{N_d}{N_a} \times 100, \tag{7}$$

where N_d is the number of audio files where the species was correctly detected, and N_a is the total number of audio recordings where the species is present according to the labels. In addition, we estimate the false-positive rate (FPR) = True Negatives/(True Negatives + False positives).

In the exploration stage, species from different taxonomic groups are associated with specific sonotypes proposed by our methodology (see Section 2.3.4). Based on this association, the species will be automatically recognized in new recordings. Then, presence (*Nd*) will be counted in the recordings where the algorithm identifies the vocalization of the species.

3 | RESULTS

3.1 | Case 1: Multi-species call recognition

Using dataset A, we tested our approach to characterize all the acoustic activities in a highly biodiverse site. First, the cepstral coefficients of each segmented sound were automatically extracted and clustered using LAMDA 3pi. Then, an experienced ecologist associated the resulting clusters with their respective species.

The median clustering hit rate performances were 85%, 96% and 89%, with a median FPR of 4%, 8% and 8% for Aves, Amphibians and Insects, respectively. Only five species were detected with hit rates below 60% on a call-by-call basis (Figure 2). Colombia is a biodiversity hotspot and many biophonies are still unknown; thus, there were several cases in which a species did not have a specific epithet. Nonetheless, their taxonomic group could be identified (Figure 2; e.g. Avian sp.1, Orthoptera sp. 1, etc.).

Our method was able to differentiate anuran species that call in similar frequency ranges (approximately 4.5 kHz) and had similar call patterns, such as *Leucostethus jota* and *Hyloxalus ramosi. Our approach identifies* them as different sonotypes, leaving the calls *of Leucostethus jota* in one cluster (red rectangles) and *Hyloxalus ramosi* in another one (blue rectangles), as is shown in Figure 3.

Species with high vocal complexity (i.e. calls of long duration with several notes in multiple frequency bands) were identified in most cases using the entire call or a significant part of their call, such as the case of Basileuterus sp. (Figure 4). This particular bird call has frequencies between 6 and 10 kHz and different notes. Our methodology focuses on the similarities among clusters, rather than searching for a specific pattern. That allows the detection of variations present in species calls and captures the diversity within species.

We hypothesize that the five species with low clustering accuracy (*Ramphocelus dimidiatus* call, *Pitangus sulphuratus*, Avian sp. 4, Otrhoptera sp. 2, Orthoptera sp. 10) were mostly due to the reduced number of call examples available to generate a highly confident cluster. Some species had either few call examples or appeared in a small number of recordings (e.g. *Pitangus sulphuratus*). Additionally, species detection in this dataset proved complex, even for experts. Table 1, Appendix A shows the difference between the partitions performed by the experts. Table 4 (Appendix E) summarizes the cases where the accuracy was below 60% and compares our approach with manual detection.

3.2 | Case 2: Comparison with other available software for species identification

We used dataset B to compare our proposal performance against Autodetect from WarbleR, MonitoR and Kaleidoscope Pro. Although the R packages (WarbleR and MonitoR) do not include a clustering step, we decided to include them as biologists widely use them for species identification. The dataset has five different species manually labelled and has some recordings with strong background noise.

Table 1 presents the comparison results of our proposed approach with three other species recognition proposals. Despite using a completely unsupervised method, our approach achieved the best species detection performance, with an average hit rate of 75%. To see the detailed performance of each species in each analysed methodology, see Table 5 in Appendix F.

Our approach does not require the tuning process necessary for R packages and Kaleidoscope Pro to obtain the best performance. Those tools require prior knowledge of signal characteristics such as each species frequency band, amplitude, call length, and in the case of Kaleidoscope, additional knowledge in the cluster analysis parameters.

Our approach outperformed all the other methodologies, achieving hit rates above 60%. Low species detection occurred due to high gain activity across multiple frequency bands, which could complicate signal detection. Analysis of audio recordings revealed the presence of rain, which likely masked the species signals. Nonetheless, our method was able to detect some of these masked signals (see Appendix G).

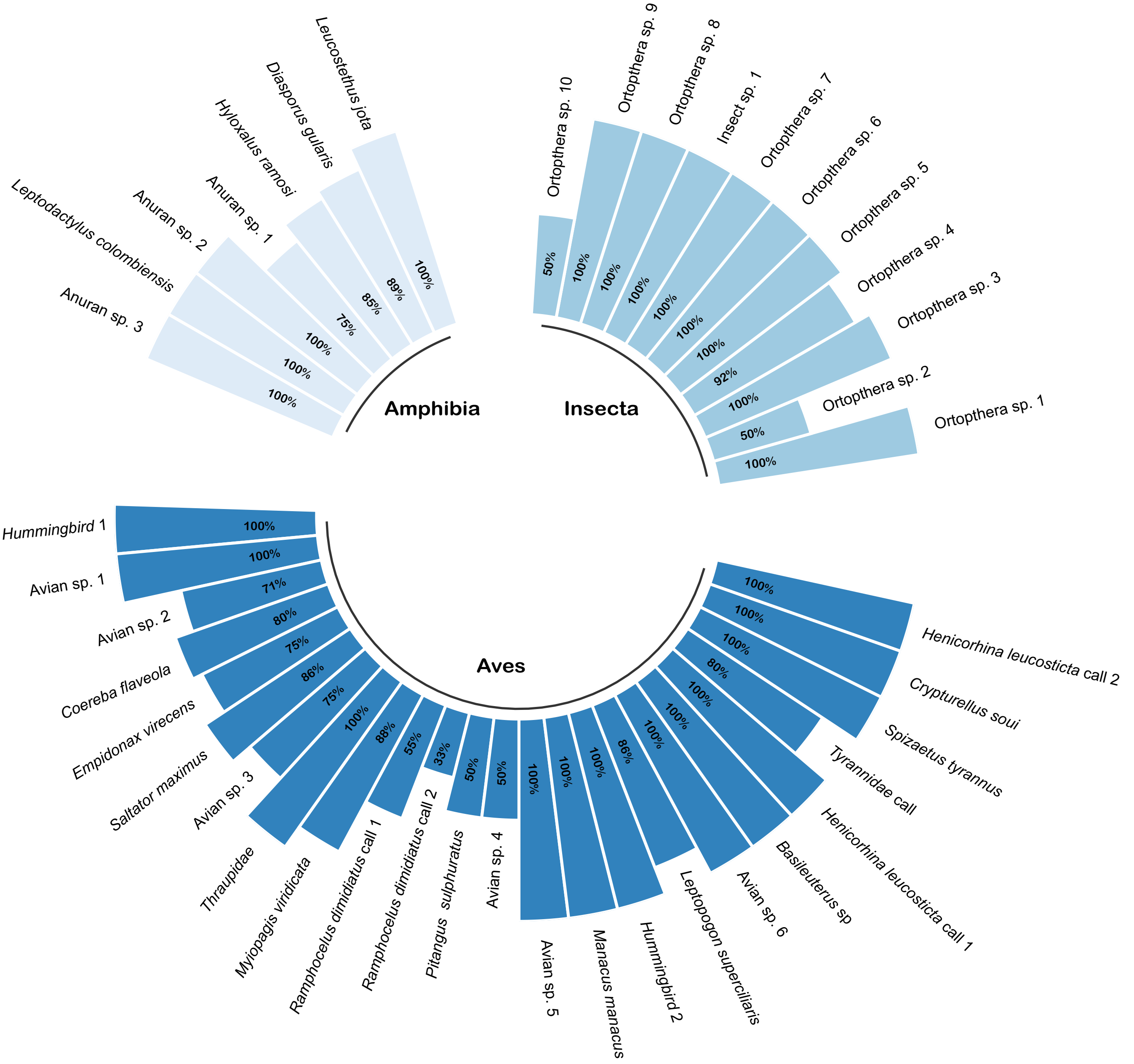

FIGURE 2 Clustering results of our method at a site with 39 species. Each bar represents the hit rate for each species. Results are coloured in accordance with their taxonomic group (Aves, Amphibia and Insecta).

3.3 | Case 3: Method validation on independent data and biodiversity assessments

To evaluate the performance of our unsupervised approach, a randomly selected subset of 207 audio recordings was manually labelled by a group of experts. Recordings include calls from birds, anurans and a primate species. Calls from all the recordings (from dataset C) were automatically segmented and clustered using LAMDA. The median hit rate of our method was 89%, with a median FPR of 17% for the 11 species (Figure 5).

In some cases, our methodology successfully detected vocalizations with low intensity in the spectrogram, such as those produced by the bird species Nyctidromus albicolis (see Appendix H). Furthermore, there are audio recordings where different species vocalized in the same frequency band, such as in the case of the avian species *Nyctidromus albicolis* and the anuran species *Leptodactylus fragilis*. Our approach correctly identified both and grouped them into different clusters (see Appendix I).

In this case study, we applied our approach as a biodiversity indicator, associating the number of sonotypes proposed by the method (clusters). For this purpose, we used 2431 audio recordings corresponding to two different locations at the study site (KLSA13 and KLSA14) registered during March 2021. The locations were chosen due to their biodiversity variety, where 10 species were selected by experts. We analysed the recordings of each site and our proposal

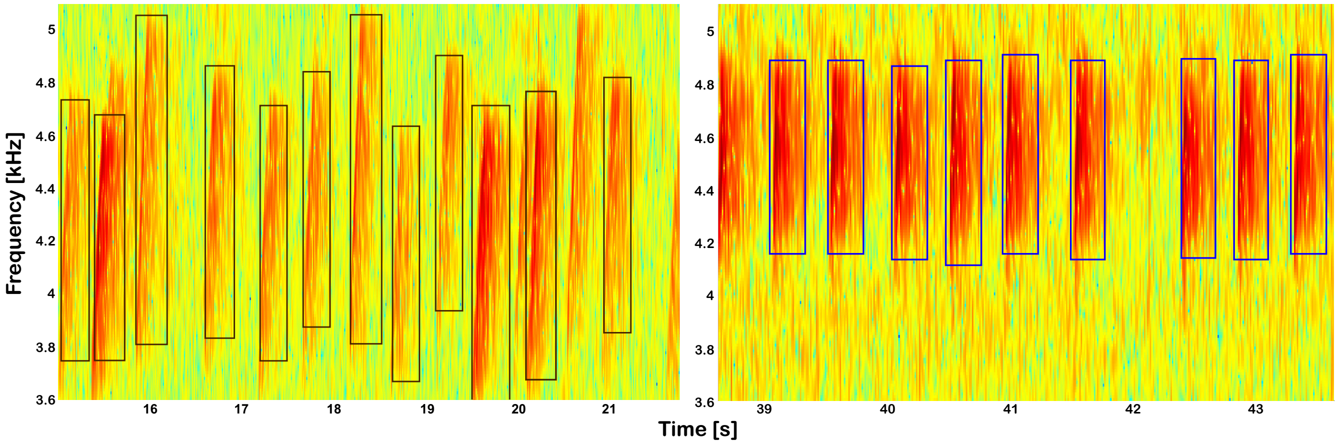

FIGURE 3 Anurans calls. Red segments are associated with *Leucostethus jota* and blue segments are associated with *Hyloxalus ramosi*. Our methodology was able to differentiate the calls and separate them into different clusters.

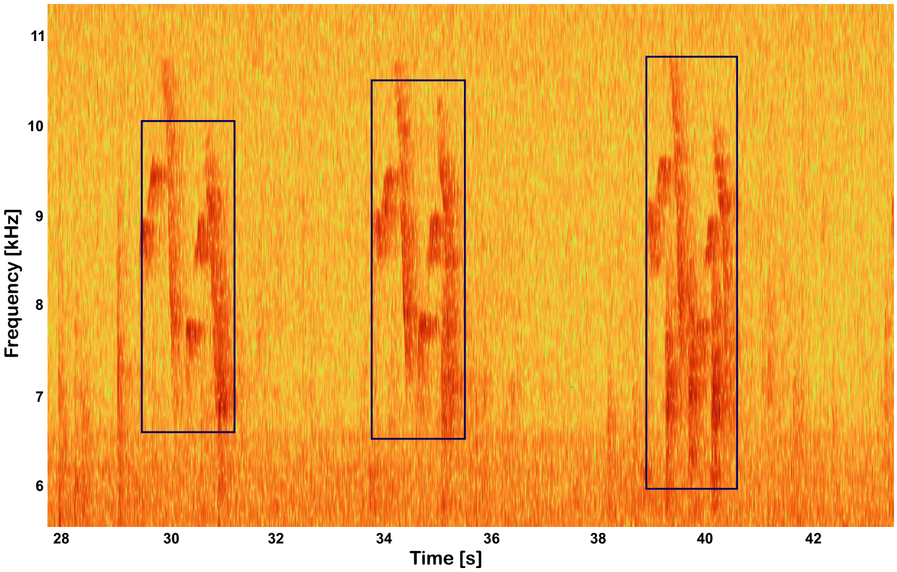

FIGURE 4 Bird complex call associated with *Basileuterus* sp. Each blue rectangle is a species call segment detected by our methodology.

suggests sonotypes. Then, we reviewed the number of different sonotypes in each recording and calculated the mean of the number of sonotypes per hour, generating the 24-hour acoustic pattern for biodiversity analysis. We compared our results with the acoustic indices used to measure biodiversity (ACI, NP, BI and SO). For each acoustic index and sonotypes found, the mean maximum value was taken to normalize each of the values.

Figure 6a,b shows the 24-hour pattern for each normalized acoustic index and our approach for sites KLSA13 and KLSA14, respectively. When the number of sonotypes found for each hour is counted (see black line in Figure 5), it is possible to see that our proposal characterizes biophony throughout the day in a similar way to acoustic indices. Nonetheless, our method allows going one step beyond, as we can perform automatic animal identification for each site and know the acoustic structure according to the selected species (Figure 6c,d) serving as an indicator of biodiversity.

In this case, it is evident that both sites present a similar biophony pattern (Figure 6a,b). Moreover, our method identifies the acoustic structure of each site showing that site KLSA13 and site KLSA14 (Figure 6c,d) are different in structure.

3.4 | Case 4: Detection of ultrasonic species

Our method identifies species in the ultrasonic spectrum (>20 kHz), with no parameter adjustment. This capability was tested using recordings collected with an ultrasonic recorder. Dataset included 13 species of bats and 6 species of orthopterans, all of them found at ultrasonic frequencies.

Figure 7 shows the detection results for each ultrasonic species. The median hit rate for all the species was 96%, with a median FPR of 15%. No parameter tuning was needed to achieve this result.

4 | DISCUSSION

We introduce an unsupervised methodology for animal identification in tropical soundscapes using a segmenter based on image analysis, cepstrals coefficients and frequency information as features, and clustering algorithm. As it is an unsupervised methodology, data for model training are not required. This proposal allows detecting species from multiple taxonomic groups in recordings with biophony in multiple frequency bands, high frequencies and low-intensity calls in spectrograms without significant background noise. Additionally, our approach analyses data recorded from different landscapes using different types of recorders. It is possible to use our methodology results to perform biodiversity assessments similar to acoustic indices, with the added benefit of identifying species from diverse taxonomic groups and determining the contribution of the acoustic richness structure associated with each species at different sites.

With our unsupervised approach, we detected the presence–absence of species achieving performances between 75% and 96%, with FPRs between 4% and 17%. This performance is comparable to the one achieved with supervised methodologies such as CNNs (LeBien et al., 2020; Ruff et al., 2020; Ruff et al., 2021), other machine learning methods (Bellisario et al., 2019; Brodie et al., 2020; Xie et al., 2018) that require labelled species data, and unsupervised approaches (Bedoya, Isaza, et al., 2014; Jancovic & Köküer, 2019; Potamitis, 2015). In comparison with other available packages and

TABLE 1 Presence–absence detection results in R libraries: WarbleR-Autodetec with a threshold of 10% (ATH10), WarbleR-Autodetec with a threshold of 15% (ATH15), WarbleR-Autodetec with a threshold of 20% (ATH20), MonitoR, Kaleidoscope Pro (KP) and our proposal.

Software	Average hit rate
ATH10	0.49
ATH15	0.52
ATH20	0.48
MonitoR	0.52
KP	0.51
Our proposal	**0.75**

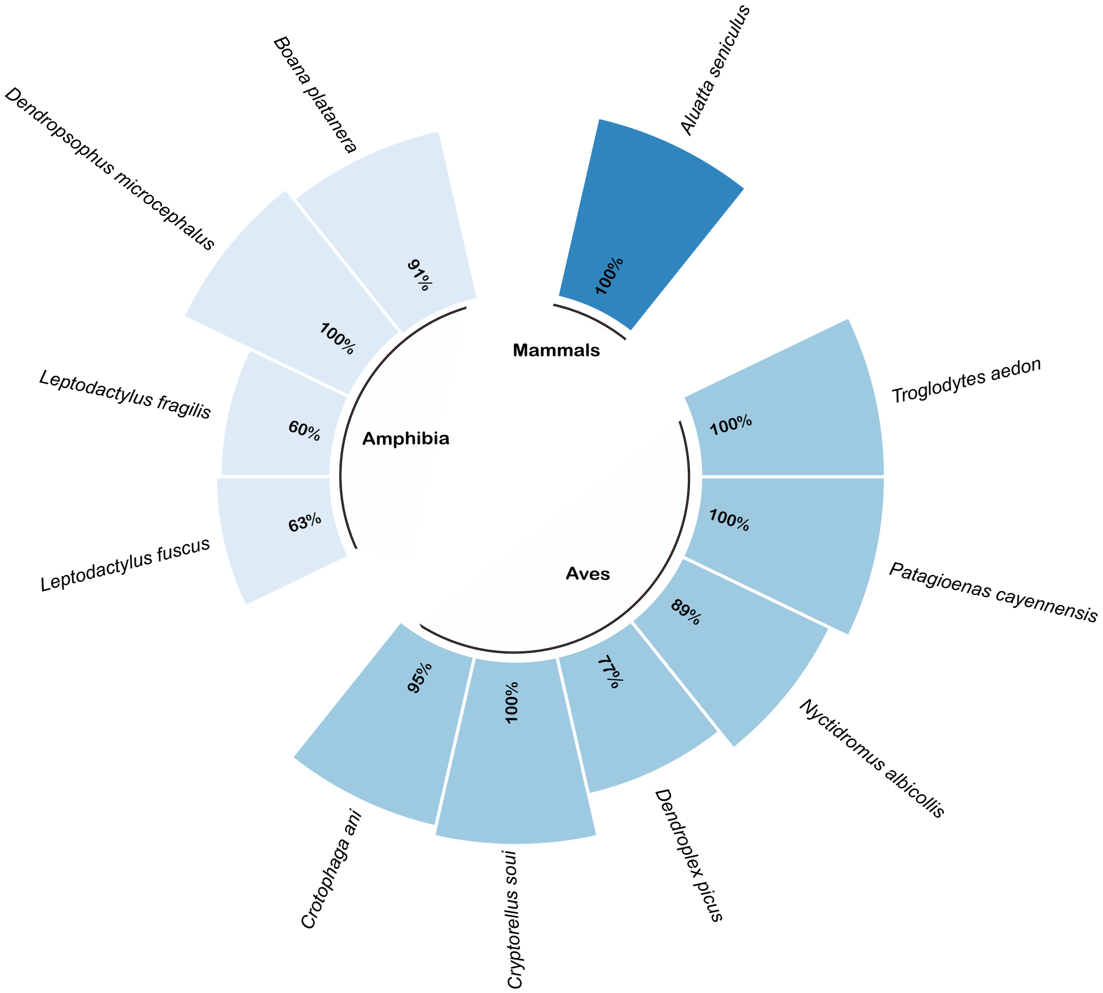

FIGURE 5 Clustering results of our method with 11 species. Bars indicate hit rates for each species in all the recordings for dataset C. Results are coloured in accordance with their taxonomic group (Aves, Amphibia and Mammals).

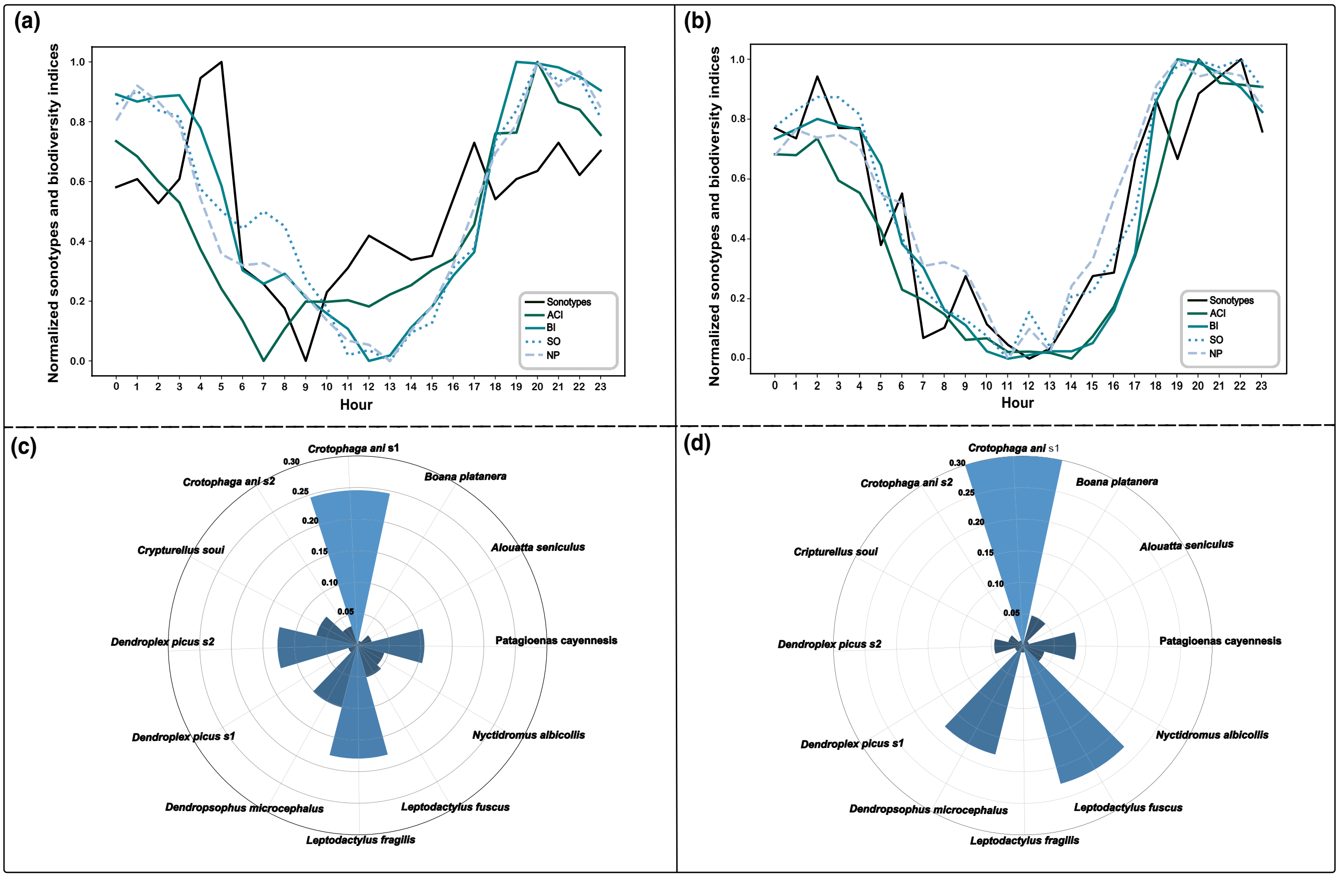

FIGURE 6 Generated acoustic pattern using our proposed approach with the data corresponding to sites KLSA13 (a) and KLSA14 (b) in March 2021 compared with four commonly used acoustic indices for the estimation of species richness: ACI, BI, NP and SO. The *y*-axis corresponds to the normalized value of each acoustic index. In the case of our proposal, it corresponds to the normalized value of the mean number of sonotypes found for each hour of the day. Figures (c and d) correspond to the acoustic structure of each site according to the selected species. The percentage presence of each species in each zone is shown here.

software, remarkable detections were obtained with our methodology despite the activity in multiple frequency bands that masked some species calls. Moreover, our proposal does not need to tune species-specific parameters, it only requires an expert to associate the clusters generated by the algorithm (sonotypes) to a species.

The LAMDA 3pi clustering method does not require the number of classes as an input parameter. Therefore, the number of generated clusters per hour within the analysed period informs about the biophony of the soundscape, and it is possible to identify the species present in the study area. In addition, we found cases where experts could identify the taxonomic group but not the specific species as some avians and orthopterans in the first and third case studies. This clustering algorithm can differentiate among species calls and allows finding sonotypes of species that were not expected a priori. It naturally creates clusters from the differences found among segments in the soundscape at audible and ultrasonic frequencies.

Among its limitations, as in the case of supervised classification algorithms, we found that clustering results may be affected by background noise in the recordings. In addition, FPRs increase when, in the exploration stage, the number of vocalizations of a species is low. Therefore, recordings with significant vocalizations should be used. There were cases in which several species were grouped in the same cluster. This can happen due to the lack of vocalizations needed to form confident clusters, calls being masked by background noise or having calls with low intensity (e.g. some animal vocalizing farther away from the microphone). This limitation could be solved using cluster validation indices or using the fuzzy logic properties from the LAMDA clustering algorithm. Cluster validity indices evaluate the quality of the partition generated by a clustering algorithm. A partition is considered of good quality when it produces compact and well-separated clusters. This evaluation is applied to the output of the clustering algorithm and does not require adjusting any parameters during the clustering process. A clustering validation stage could significantly help experts to associate clusters with species, as high-quality index values usually correlate with high-confidence clusters. Despite the availability of various cluster validity indices in the literature, these remain significantly underused in animal communication studies.

Recognizing multiple species is a complex task that depends on the adequate performance of previous stages. Signal pre-processing and segmentation stages are critical when several frequency bands are to be analysed simultaneously. This methodology can be improved with a more robust pre-processing stage using techniques that do not sacrifice frequency bands. This would help to segment

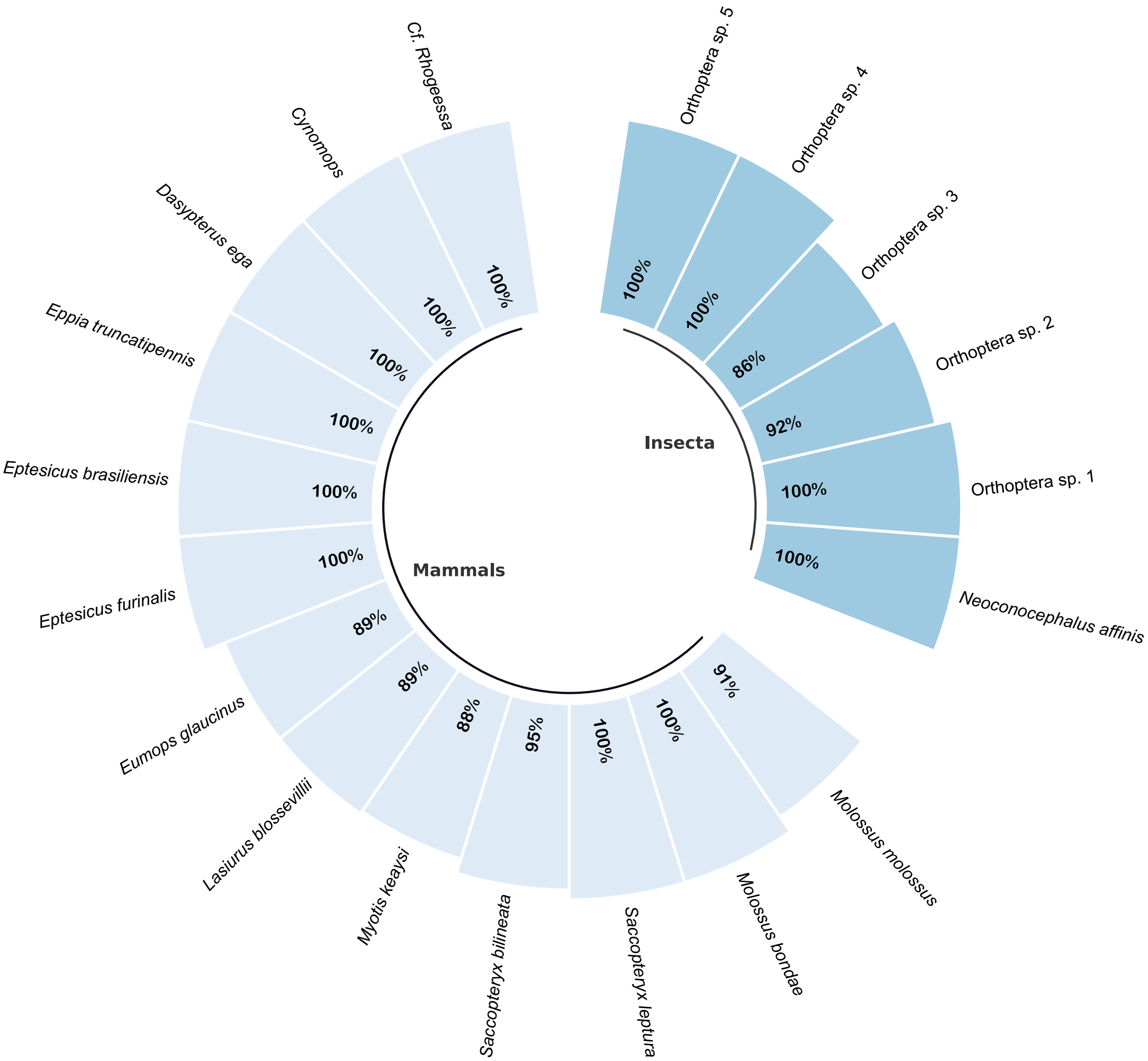

FIGURE 7 Unsupervised species recognition in the ultrasonic spectrum. Clustering results are organized in accordance with the taxonomic class of each species (Insecta, Mammalia). Bars indicate hit rates for each species in all the recordings.

more accurately and obtain more representative clusters. In addition, using cluster validation indices, as mentioned above, could help to improve the cluster–species relationship.

We present an unsupervised method for the simultaneous identification of animals from diverse taxonomic groups. Our approach is highly accurate, works in both the audible and ultrasonic spectra, and does not require the labelling of training data. These features significantly facilitate the analysis of massive acoustic datasets. In addition, our approach allows characterizing community composition in environments where species, or their vocalizations, are yet unknown to science, a common scenario in biodiversity hotspots, underdeveloped countries, freshwater and marine ecosystems, and vibroscapes.

AUTHOR CONTRIBUTIONS

Maria J. Guerrero conceptualization, methodology, algorithms, writing; Carol L. Bedoya and José D. López conceived ideas and algorithms; Juan M. Daza biology conceptualization; Claudia Isaza conceptualization, supervision, methodology, validation and project administration. All authors contributed to the drafts and gave the final approval for publication.

ACKNOWLEDGEMENTS

This work was supported by Universidad de Antioquia, Instituto Tecnológico Metropolitano de Medellín, Alexander von Humboldt Institute for Research on Biological Resources and The Colombian National Fund for Science, Technology and Innovation, Francisco Jose de Caldas—MINCIENCIAS (Program No. 111585269779). Data from Jaguas were funded by Universidad de Antioquia and Isagen under contracts 47/146 and 47/574. Data from Puerto Wilches were funded by Universidad de Antioquia, SGI and Ecopetrol under contract FOGR09. 2012 dataset was labelled by Karen Jaramillo (frogs), Dariel Martinez (birds) and Juan M. Daza (frogs and insects). 2017

dataset was labelled by Karen Jaramillo, Manuela Ospina and Julian Giron. 2021 dataset was labelled by Victor Martinez (bats), Andrea Lopera (birds and diurnal mammals), Andres Velez (insects) and Ana Sepulveda (frogs and diurnal mammals).

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14103.

DATA AVAILABILITY STATEMENT

The algorithms used in our methodology were implemented in Matlab R2022a. From this implementation, an executable was created (Matlab is not required to use it) for Windows OS users. The executable and datasets are available in: https://tinyurl.com/y2w2tkhw.

ORCID

Maria J. Guerrero https://orcid.org/0000-0003-0632-9176
Carol L. Bedoya https://orcid.org/0000-0002-7013-7083
José D. López https://orcid.org/0000-0003-2213-1186
Juan M. Daza https://orcid.org/0000-0002-3494-489X
Claudia Isaza https://orcid.org/0000-0003-1044-9429

ENDNOTES

[1] ®Wildlife Acoustics, Inc.

[2] ®The Cornell Lab of Ornitology

[3] Queen Mary University of London

REFERENCES

Acconcjaioco, M., & Ntalampiras, S. (2021). One-shot learning for acoustic identification of bird species in non-stationary environments. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 755–762). IEEE. https://doi.org/10.1109/ICPR48806.2021.9412005

Agranat, I. (2013). *Bat species identification from zero crossing and full spectrum echolocation calls using hidden Markov models, fisher scores, unsupervised clustering and balanced winnow pairwise classifiers*. 010016. https://doi.org/10.1121/1.4799403

Aguilar-Martin, J., & Mantaras, R. L. (1982). The process of classification and learning the meaning of linguistic descriptors or concepts. *Approximate Reasoning in Decision Analysis*, 165–175.

Aide, T. M., Hernández-Serna, A., Campos-Cerqueira, M., Acevedo-Charry, O., & Deichmann, J. L. (2017). Species richness (of insects) drives the use of acoustic space in the tropics. *Remote Sensing*, *9*, 1096. https://doi.org/10.3390/rs9111096

Araya-Salas, M., & Smith-Vidaurre, G. (2017). WarbleR an r package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution*, *8*, 184–191. https://doi.org/10.1111/2041-210X.12624

Bedoya, C., Isaza, C., Daza, J., & López, J. D. (2014). Automatic recognition of anuran species based on syllable identification. *Ecological Informatics*, *24*, 200–209. https://doi.org/10.1016/j.ecoinf.2014.08.009

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

Bedoya, C., Waissman, J., & Isaza, C. (2014). Yager–Rybalov triple pi operator as a means of reducing the number of generated clusters in unsupervised anuran vocalization recognition. In *Nature-inspired computation and machine learning* (Vol. *8857*, pp. 382–391). Springer International Publishing.

Bedoya, C. L., & Molles, L. E. (2021). Acoustic censusing and individual identification of birds in the wild. *bioRxiv*, 2021.10.29.466450. https://doi.org/10.1101/2021.10.29.466450

Bellisario, K. M., Broadhead, T., Savage, D., Zhao, Z., Omrani, H., Zhang, S., Springer, J., & Pijanowski, B. C. (2019). Contributions of MIR to soundscape ecology. Part 3: Tagging and classifying audio features using a multi-labeling k-nearest neighbor approach. *Ecological Informatics*, *51*, 103–111. https://doi.org/10.1016/j.ecoinf.2019.02.010

Boelman, N., Asner, G., Hart, P., & Martin, R. (2008). Multi-trophic invasion resistance in Hawaii: Bioacoustics, field surveys, and airborne remote sensing. *Ecological Applications: A Publication of the Ecological Society of America*, *17*, 2137–2144. https://doi.org/10.1890/07-0004.1

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

Brehm, G., Fiedler, K., Hauser, C., & Dalitz, H. (2008). Methodological challenges of a megadiverse ecosystem. *Gradients in a Tropical Mountain Ecosystem of Ecuador*, *198*, 41–47. https://doi.org/10.1007/978-3-540-73526-7_5

Brodie, S., Allen-Ankins, S., Towsey, M., Roe, P., & Schwarzkopf, L. (2020). Automated species identification of frog choruses in environmental recordings using acoustic indices. *Ecological Indicators*, *119*, 106852. https://doi.org/10.1016/j.ecolind.2020.106852

Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., & Sueur, J. (2012). Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. *Ecological Indicators*, *13*, 46–54. https://doi.org/10.1016/j.ecolind.2011.05.006

Ducrettet, M., Forget, P.-M., Ulloa, J., Yguel, B., Gaucher, P., Princé, K., Haupert, S., & Sueur, J. (2020). Acoustic monitoring of the white-throated toucan (*Ramphastos tucanus*) in disturbed tropical landscapes. *Biological Conservation*, *245*, 108574. https://doi.org/10.1016/j.biocon.2020.108574

Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., Stender, C. S., Li, W., Liu, Z., Chen, Q., Zhou, Z., & Turvey, S. T. (2021). Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sensing in Ecology and Conservation*, *7*, 475–487. https://doi.org/10.1002/rse2.201

Dumyahn, S. L., & Pijanowski, B. C. (2011). Soundscape conservation. *Landscape Ecology*, *26*, 1327–1344. https://doi.org/10.1007/s10980-011-9635-x

Gan, H., Zhang, J., Towsey, M., Truskinger, A., Stark, D., van Rensburg, B. J., Li, Y., & Roe, P. (2020). Data selection in frog chorusing recognition with acoustic indices. *Ecological Informatics*, *60*, 101160. https://doi.org/10.1016/j.ecoinf.2020.101160

Gasc, A., Sueur, J., Pavoine, S., Pellens, R., & Grandcolas, P. (2013). Biodiversity sampling using a global acoustic approach: Contrasting sites with microendemics in New Caledonia. *PLoS ONE*, *8*, e65311. https://doi.org/10.1371/journal.pone.0065311

Giam, X., Scheffers, B., Sodhi, N., Wilcove, D., Ceballos, G., & Ehrlich, P. (2011). Reservoirs of richness: Least disturbed tropical forests are centres of undescribed species diversity. *Proceedings Biological Sciences/The Royal Society*, *279*, 67–76. https://doi.org/10.1098/rspb.2011.0433

Jancovic, P., & Köküer, M. (2019). Bird species recognition using unsupervised modeling of individual vocalization elements. *IEEE/ACM*

Transactions on Audio, Speech, and Language Processing, *27*, 932–947. https://doi.org/10.1109/TASLP.2019.2904790

Joppa, L., Roberts, D., Myers, N., & Pimm, S. (2011). Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 13171–13176. https://doi.org/10.1073/pnas.1109389108

Katz, J., Hafner, S. D., & Donovan, T. (2016). Tools for automated acoustic monitoring within the R package monitoR. *Bioacoustics*, *25*, 197–210. https://doi.org/10.1080/09524622.2016.1138415

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J. P., Dodhia, R., Ferres, J. L., & Aide, T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. *Ecological Informatics*, *59*, 101113. https://doi.org/10.1016/j.ecoinf.2020.101113

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

Moreno-Gómez, F., Bartheld, J., Silva-Escobar, A., Briones, R., Márquez, R., & Penna, M. (2019). Evaluating acoustic indices in the Valdivian rainforest, a biodiversity hotspot in south America. *Ecological Indicators*, *103*, 1–8. https://doi.org/10.1016/j.ecolind.2019.03.024

Ntalampiras, S., & Potamitis, I. (2021). Acoustic detection of unknown bird species and individuals. *CAAI Transactions on Intelligence Technology*, *6*, 291–300. https://doi.org/10.1049/cit2.12007

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*, 62–66.

Ovaskainen, O., Moliterno de Camargo, U., & Somervuo, P. (2018). Animal sound identifier (ASI): Software for automated identification of vocal animals. *Ecology Letters*, *21*, 1244–1254. https://doi.org/10.1111/ele.13092

Pieretti, N., Farina, A., & Morri, D. (2010). A new methodology to infer the singing activity of an avian community: The acoustic complexity index (ACI). *Ecological Indicators*, *11*, 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. *Bioscience*, *61*, 203–216. https://doi.org/10.1525/bio.2011.61.3.6

Potamitis, I. (2015). Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity. *Ecological Informatics*, *26*, 6–17. https://doi.org/10.1016/j.ecoinf.2015.01.002

Premoli, M., Baggi, D., Bianchetti, M., Gnutti, A., Bondaschi, M., Mastinu, A., Migliorati, P., Signoroni, A., Leonardi, R., Memo, M., & Bonini, S. A. (2021). Automatic classification of mice vocalizations using machine learning techniques and convolutional neural networks. *PLoS ONE*, *16*, 1–16. https://doi.org/10.1371/journal.pone.0244636

Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: A review. *Journal of Avian Biology*, *49*, jav-01447. https://doi.org/10.1111/jav.01447

Rendon, N., Rodríguez-Buritica, S., Sanchez-Giraldo, C., Daza, J. M., & Isaza, C. (2022). Automatic acoustic heterogeneity identification in transformed landscapes from Colombian tropical dry forests. *Ecological Indicators*, *140*, 109017. https://doi.org/10.1016/j.ecolind.2022.109017

Rojas, E. C., Giraldo, C. S., Bedoya, C., & Rojas, J. M. D. (2022). Habitat and acoustic spectrum as determinant factors of the occupation of neotropical anurans. *Biota Colombiana*, *23*, e910. https://doi.org/10.21068/2539200X.910

Ruff, Z. J., Lesmeister, D. B., Appel, C. L., & Sullivan, C. M. (2021). Workflow and convolutional neural network for automated identification of animal sounds. *Ecological Indicators*, *124*, 107419. https://doi.org/10.1016/j.ecolind.2021.107419

Ruff, Z. J., Lesmeister, D. B., Duchac, L. S., Padmaraju, B. K., & Sullivan, C. M. (2020). Automated identification of avian vocalizations with deep convolutional neural networks. *Remote Sensing in Ecology and Conservation*, *6*, 79–92. https://doi.org/10.1002/rse2.125

Sánchez-Giraldo, C., Bedoya, C. L., Morán-Vásquez, R. A., Isaza, C. V., & Daza, J. M. (2020). Ecoacoustics in the rain: Understanding acoustic indices under the most common geophonic source in tropical rainforests. *Remote Sensing in Ecology and Conservation*, *6*, 248–261. https://doi.org/10.1002/rse2.162

Scheffers, B., Joppa, L., Pimm, S., & Laurance, W. (2012). What we know and don't know about earth's missing biodiversity. *Trends in Ecology & Evolution*, *27*, 501–510. https://doi.org/10.1016/j.tree.2012.05.008

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

Stowell, D., & Sueur, J. (2020). Ecoacoustics: Acoustic sensing for biodiversity monitoring at scale. *Remote Sensing in Ecology and Conservation*, *6*, 217–219. https://doi.org/10.1002/rse2.174

Sueur, J., & Farina, A. (2015). Ecoacoustics: The ecological investigation and interpretation of environmental sound. *Biosemiotics*, *8*(3), 493–502. https://doi.org/10.1007/s12304-015-9248-x

Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS ONE*, *3*(12), 4065. https://doi.org/10.1371/journal.pone.0004065

Towsey, M., Wimmer, J., Williamson, I., & Roe, P. (2013). The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics*, *21*, 110–119. https://doi.org/10.1016/j.ecoinf.2013.11.007

Ullmann, T., Hennig, C., & Boulesteix, A. L. (2021). Validation of cluster analysis results on validation data: A systematic framework. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(3), e1444. https://doi.org/10.1002/widm.1444

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

Xie, J., Colonna, J. G., & Zhang, J. (2021). Bioacoustic signal denoising: A review. *Artificial Intelligence Review*, *54*, 3575–3597. https://doi.org/10.1007/s10462-020-09932-4

Xie, J., Hu, K., Zhu, M., & Guo, Y. (2020). Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window. *Expert Systems with Applications*, *152*, 113390. https://doi.org/10.1016/j.eswa.2020.113390

Xie, J., Indraswari, K., Schwarzkopf, L., Towsey, M., Zhang, J., & Roe, P. (2018). Acoustic classification of frog within-species and species-specific calls. *Applied Acoustics*, *131*, 79–86. https://doi.org/10.1016/j.apacoust.2017.10.024

Xie, J., Towsey, M., Zhu, M., Zhang, J., & Roe, P. (2017). An intelligent system for estimating frog community calling activity and species richness. *Ecological Indicators*, *82*, 13–22. https://doi.org/10.1016/j.ecolind.2017.06.015

Xue, J., Feng, Z., & Zhang, P. (2013). Spectrum occupancy measurements and analysis in Beijing. *IERI Procedia*, *4*, 295–302. https://doi.org/10.1016/j.ieri.2013.11.042

Zhao, Z., Zhang, S., Xu, Z., Bellisario, K., Dai, N., Omrani, H., & Pijanowski, B. C. (2017). Automated bird acoustic event detection and robust species classification. *Ecological Informatics*, *39*, 99–108. https://doi.org/10.1016/j.ecoinf.2017.04.003

SUPPORTING INFORMATION

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

Appendix A. Ground truth (labels).

Appendix B. Datasets summary.

Appendix C. Segmentation test.

Appendix D. Segmentation Algorithms.

Appendix E. Low-confidence species.

Appendix F. Species performance in each analysed methodology.

Appendix G. Low-intensity vocalizations regarding background noise.

Appendix H. Low-intensity vocalizations.

Appendix I. Particularities in cluster analysis.

How to cite this article: Guerrero, M. J., Bedoya, C. L., López, J. D., Daza, J. M., & Isaza, C. (2023). Acoustic animal identification using unsupervised learning. *Methods in Ecology and Evolution*, *14*, 1500–1514. https://doi.org/10.1111/2041-210X.14103