Ecological Indicators 90 (2018) 346–355

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

# Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis

Juan Sebastian Ulloa[a,b,*], Thierry Aubin[b], Diego Llusia[a,b], Charles Bouveyron[c], Jérôme Sueur[a]

[a] Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, Sorbonne Université CNRS-MNHN-UPMC-EPHE, 57 rue Cuvier, CP 50, F-75005, Paris, France
[b] Equipe Communications Acoustiques, UMR 9197 Neuro-PSI-CNRS, Université Paris Sud, bat. 446, F-91405 Orsay, France
[c] Laboratoire MAP5, UMR CNRS 8145, Université Paris Descartes & Sorbonne Paris Cité, Paris, France

ARTICLE INFO

*Keywords:*
Ecoacoustic monitoring
Acoustic community
Unsupervised machine learning
Wavelets
Nocturnal soundscape

ABSTRACT

Ecoacoustic monitoring has proved to be a viable approach to capture ecological data related to animal communities. While experts can manually annotate audio samples, the analysis of large datasets can be significantly facilitated by automatic pattern recognition methods. Unsupervised learning methods, which do not require labelled data, are particularly well suited to analyse poorly documented habitats, such as tropical environments. Here we propose a new method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the detection of relevant structure in audio data. MAAD was designed to decompose the acoustic community into few elementary components (soundtypes) based on their time–frequency attributes. First, we used the short-time Fourier transform to detect regions of interest (ROIs) in the time–frequency domain. Then, we characterised these ROIs by (1) estimating the median frequency and (2) by running a 2D wavelet analysis at multiple scales and angles. Finally, we grouped the ROIs using a model-based subspace clustering technique so that ROIs were automatically annotated and clustered into soundtypes. To test the performance of the automatic method, we applied MAAD to two distinct tropical environments in French Guiana, a lowland high rainforest and a rock savanna, and we compared manual and automatic annotations using the adjusted Rand index. The similarity between the manual and automated partitions was high and consistent, indicating that the clusters found are intelligible and can be used for further analysis. Moreover, the weight of the features estimated by the clustering process revealed important information about the structure of the acoustic communities. In particular, the median frequency had the strongest effect on modelling the clusters and on classification performance, suggesting a role in community organisation. The number of clusters found in MAAD can be regarded as an estimation of the soundtype richness in a given environment. MAAD is a comprehensive and promising method to automatically analyse passive acoustic recordings. Combining MAAD and manual analysis would maximally exploit the strengths of both human reasoning and computer algorithms. Thereby, the composition of the acoustic community could be estimated accurately, quickly and at large scale.

## 1. Introduction

The diversity of life forms is an invaluable biological resource threatened by anthropogenic environmental change (Pimm et al., 1995; Thomas et al., 2004). Given the pace of this change, there is an imperative need to develop quantitative indicators that provide specific information on the state of biodiversity (Pereira et al., 2013). With the advent of new sensor technology it is possible to remotely collect environmental data, assisting to determine, and eventually buffer, the pressures on biological diversity and ecosystem services (Petrou et al., 2015). In particular, the use of passive acoustic sensors in ecological research, or ecoacoustics (Sueur and Farina, 2015), has proved to be a viable method for biodiversity assessment that can be scaled up at multiple spatial and temporal scales (Towsey et al., 2014). The environmental sounds collected by these automated sensors usually include a large combination of both biotic and abiotic sounds, which are mixed down into a single time series. Such interlaced audio data needs to be unravelled in order to extract and to decipher ecological meaningful information, which represents to date a prominent bottleneck for the application of acoustic sensors in biodiversity monitoring.

A significant proportion of animal species produce sounds for social interaction, navigation or predator-prey encounters (Fletcher, 2014).

* Corresponding author at: Institut de Systématique, Évolution, Biodiversité, ISYEB, UMR 7205 CNRS-MNHN-UPMC-EPHE, Muséum national d'Histoire naturelle, Sorbonne Universités, 57 rue Cuvier, CP 50, F-75005 Paris, France.
*E-mail address:* juan.ulloa@mnhn.fr (J.S. Ulloa).

https://doi.org/10.1016/j.ecolind.2018.03.026
Received 10 November 2017; Received in revised form 7 March 2018; Accepted 12 March 2018
1470-160X/ 

Most of these acoustic signals have a species-specific signature that can be exploited for the remote identification of species. The use of these signatures is a direct way to retrieve ecological data about species presence, abundance, status and distribution. Manual species identification by experts can be carried on audio datasets, but for large collections, the analysis can be facilitated by automatic pattern recognition methods such as supervised learning (Kershenbaum et al., 2016). Supervised learning is a method to build a statistical classifier based on labelled training data (Webb and Copsey, 2011). An increasing number of supervised learning tools have been adapted to identify automatically single species (Dugan et al., 2013; Ganchev et al., 2015; Ulloa et al., 2016) or several species (Briggs et al., 2012; Potamitis, 2014; Heinicke et al., 2015; Dong et al., 2015; Xie et al., 2016; Ruiz-Muñoz et al., 2016). The application of supervised learning is limited by the large reference datasets required to 'train' the classifiers and the high acoustic similarity sometimes observed between closely related taxa. The available sound libraries, even if providing thousands of samples, still cover only a small fraction of the animal sound diversity, at both population and species scales.

An alternative to species identification consists in characterising the acoustic community or the soundscape with the use of acoustic indices (Sueur et al., 2014). Rather than focusing on target species, acoustic indices aim to describe the global structure of the soundscape. A variety of indices have been proposed and applied to terrestrial (Lellouch et al., 2014; Farina et al., 2015; Fuller et al., 2015) and underwater habitats (Parks et al., 2014; Desjonquères et al., 2015; Harris et al., 2016; Buscaino et al., 2016). These indices revealed, for example, changes in bird species richness among woodland habitats (Depraetere et al., 2012) or dynamics of the soundscape across different temporal scales (Rodriguez et al., 2014). However, they also showed to be sensitive to transitory or permanent background noise, variation in the distance of the animals to the sensor, and the relative sound amplitude or the calling rate of the signalling animal (Gasc et al., 2015; Kendrick et al., 2016).

More recently, methods based on unsupervised learning have been adapted to audio recordings achieved in natural environments. Unsupervised learning searches for structures or patterns in a dataset without using labels. This approach has been extensively used to draw inferences in areas where labelled data is inaccessible or too expensive, such as astronomy (Way, 2012), genetics and genomics (Libbrecht and Noble, 2015). In an innovative work, Eldridge et al. (2016) adapted sparse-coding and source separation algorithms to extract shift-invariant spectro-temporal "atoms" from environmental recordings. However, the authors did not establish a clear link between the spectro-temporal "atoms" and ecological or biological processes. Unsupervised learning has also been used as a pre-processing step for the classification task, significantly improving the classification performance on species recognition (Stowell and Plumbley, 2014). In their approach, Stowell and Plumbley (2014) first decomposed the sounds into "atoms" with spherical k-means, and then used the "atoms" as features for the supervised learning framework. Thus, unsupervised learning offers new means to characterise sounds and may provide insights on the acoustic communities of diverse and threatened ecosystems, such as those of tropical regions (Pekin et al., 2012; Rodriguez et al., 2014).

The present work emerges from the question: how to best measure, quantify and characterise environmental sounds (from biotic and abiotic sources) in passive acoustic recordings to get valuable ecological indicators? We propose a new data-driven method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the discovery of plausible and interpretable patterns in passive acoustic recordings. To build a generalized method for multiple conditions and environments, we adapted methods from the unsupervised learning field. We estimated acoustic diversity by detecting regions of interest in sound recordings and grouping them into soundtypes based on the value of their time-frequency attributes. To test the flexibility and robustness of the method, we applied MAAD to two distinct night tropical environments in French Guiana, a lowland high rainforest (HF) and a rock savanna (RS). The RS is inhabited by a distinct and likely less diverse animal community in comparison with the HF (Bongers et al., 2001) so that it was expected to find contrasting acoustic communities between these two tropical environments. We compared manual and automated annotations to (1) evaluate the model selection procedure; (2) assess the relevance of different features in the clustering process; and (3) quantify the overall similarity between manual and MAAD soundtypes. To conclude, we give practical advices and discuss how MAAD can potentially be transferred to other environments in order to track the state and dynamics of animal communities for biodiversity studies.

## 2. Material and methods

The workflow of the proposed method (MAAD) followed four main steps: (1) passive acoustic recordings were transformed into the time-frequency domain using the windowed short-time Fourier transform and the Fourier coefficients were filtered to remove noise and to highlight sounds that can be delimited in time and frequency, here defined as regions of interest (ROIs); (2) each ROI was then characterised by features in the time-frequency domain using 2D wavelets; (3) the ROIs with their attributes were used to automatically estimate clustering hyper-parameters; and (4) the hyper-parameters and the attributes of the ROIs were passed to a clustering algorithm that formed homogenous groups of ROIs, namely soundtypes (Fig. 1). This led to an automatic partitioning and characterization of soundtypes, which can be used to determine their presence, relative abundance and diversity within acoustic communities. To validate the proposed approach, the automatic partitioning provided by MAAD was compared to expert manual annotations using the adjusted Rand index (ARI).

### 2.1. Audio dataset

Audio data were collected in French Guiana at the CNRS Nouragues Research Station (4°05′N; 54°40′W). The station is mainly occupied by a lowland high rainforest (HF) and a rock savanna (RS), among other ecosystems. The HF dominates on lower parts of the station (40–100 metres above sea level), has a fairly open understory and is closed on top by a dense canopy elevating at 25–35 m. The tree density in HF is

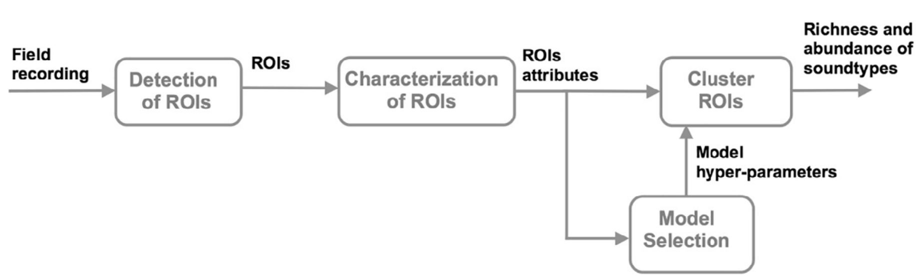


**Fig. 1.** Block diagram of MAAD. Each step of the workflow is depicted as a grey box. Input and output after each step are indicated in black. Model selection is an optional step. Model hyper-parameters can also be set based on prior information about the acoustic community.