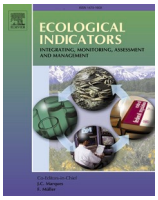
Ecological Indicators 126 (2021) 107316

Contents lists available at ScienceDirect

Ecological Indicators

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



Visualization and categorization of ecological acoustic events based on discriminant features

Liz Maribel Huancapaza Hilasaca [a,*], Lucas Pacciullio Gaspar [b], Milton Cezar Ribeiro [b], Rosane Minghim [c]

[a] Instituto de Ciências Matemáticas e de Computação (ICMC), University of São Paulo, Brazil
[b] Department of Biodiversity, São Paulo State University - UNESP, Brazil
[c] School of Computer Science and Information Technology, University College Cork, Ireland

ARTICLE INFO

Keywords:
Soundscape ecology
Discriminant features
Visualization
Classification
Feature selection

ABSTRACT

Although sound classification in soundscape studies are generally performed by experts, the large growth of acoustic data presents a major challenge for performing such task. At the same time, the identification of more discriminating features becomes crucial when analyzing soundscapes, and this occurs because natural and anthropogenic sounds are very complex, particularly in Neotropical regions, where the biodiversity level is very high. In this scenario, the need for research addressing the discriminatory capability of acoustic features is of utmost importance to work towards automating these processes. In this study we present a method to identify the most discriminant features for categorizing sound events in soundscapes. Such identification is key to classification of sound events. Our experimental findings validate our method, showing high discriminatory capability of certain extracted features from sound data, reaching an accuracy of 89.91% for classification of frogs, birds and insects simultaneously. An extension of these experiments to simulate binary classification reached accuracy of 82.64%, 100.0% and 99.40% for the classification between combinations of frogs-birds, frogs-insects and birds-insects, respectively.

1. Introduction

Habitat loss, fragmentation and degradation are between the main threats to biodiversity and ecosystem function maintenance worldwide (Butchart et al., 2010). Therefore, evaluating the health of environments is of utmost importance in the Anthropocene (Johnson et al., 2017). Acquiring reliable information on occurrence of species in the field is very costly, which many times impairs the simultaneous monitoring of large or multiple areas for longer terms. While autonomous audio recorders generate huge amount of data to support those tasks, identifying the best strategies to handle this data –and rapidly extract information from them– is amongst the top demands for ecologists and stakeholders that use audio data on their studies and decisions.

Ecoacoustics (Sueur and Farina, 2015), which includes soundscape ecology –the study of geophony, anthrophony and biophony (Pijanowski et al., 2011) – is gaining space in research and monitoring arenas, becoming a very promising venue to tackle current challenges in environmental studies. Audio recordings of soundscapes are becoming important tools for measuring biodiversity integrity and environmental health (Servick, 2014), as well as monitoring and understand how different environments respond to human-induced modifications (Hu et al., 2009; Joo et al., 2011; Parks et al., 2014). With the advances of soundscape analysis and autonomous audio recording technologies, with consequent growing interest in these approaches, data sets have increased in size and become more complex over time (Servick, 2014; Towsey et al., 2014c; Sankupellay et al., 2015). As a consequence, a large amount of time is necessary for experts to identify events of interest in recordings. They employ their knowledge about specific sound features to identify acoustic events (defined as sound produced by animals (e.g. bird, dog, frog), nature (e.g. rain, wind, river) and human activity (e.g. speech, cars, steps)). Advancing computational ability to support those users tagging events and phenomena of interest is paramount to advance other technologies such as classification and retrieval of soundscape data. Supporting these tasks is one of our main goals.

In order to employ sound recordings in accomplishing analysis tasks, features are extracted from both the signal and the corresponding

* Corresponding author at: Avenida Trabalhador São-carlense, 400 - São Carlos, SP - Brasil.
E-mail addresses: lizhh@usp.br (L.M. Huancapaza Hilasaca), lucas.pacciullio@unesp.br (L.P. Gaspar), milton.c.ribeiro@unesp.br (M.C. Ribeiro), rosane.minghim@ucc.ie (R. Minghim).

https://doi.org/10.1016/j.ecolind.2020.107316
Received 28 March 2020; Received in revised form 16 December 2020; Accepted 22 December 2020
Available online 26 January 2021

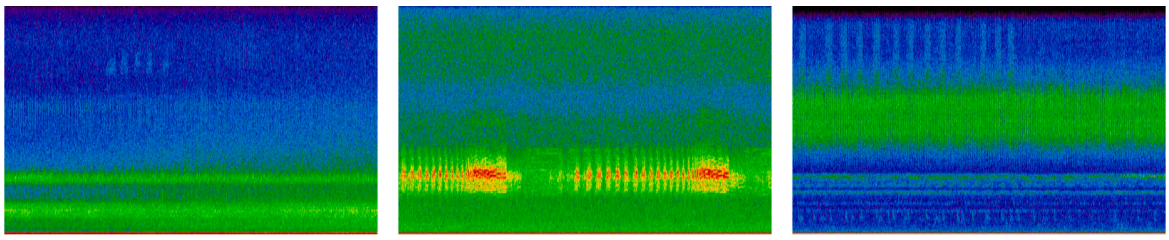


Fig. 1. One minute spectrograms labeled with the dominant events: (a) frogs, (b) birds and (c) insects. In the figures we can observe background noise, which can be a sound another event happening in parallel to an event of interest.

spectrogram. A large number of features can be obtained, aiming to describe different aspects the audio data and summarize their content. The resulting data space has a high number of dimensions or attributes. In this context, one of the main challenges is identifying a subset of adequate and sufficient features to describe events of interest. Finding this best ideal subset implies in targeting a small number of features capable of performing specific classification tasks with high precision (Alpaydin, 2014).

In sound classification tasks (e.g. Phillips et al., 2018; Xie and Towsey, 2016), methods usually extract acoustic metrics or features that are used to train a learning model. Features are therefore employed to summarize and describe a soundscape. However, due to the complexity of environmental sounds (see Figs. 1a, 1b and 1c), characterizing the sound content can become a difficult task. A typical scenario is the combination of time of recording (morning, afternoon and night) and animal habit; for instance frogs and insects usually occur at night and are simultaneous (Pijanowski et al., 2011), adding additional burden to the classification of event categories.

Other research work have focused on the study of features and patterns in soundscapes with ecological perspectives (Fuller et al., 2015). These studies reveal that acoustic peculiarities vary according to specific conditions of each environment. Due the lack of enough studies, for the larger part one does not know what possible features are to be employed to adequately describe a particular acoustic event. However, some work targeted scenarios based on extracted features (Sankupellay et al., 2015; Dias, 2018; Phillips et al., 2018). In a recent review of soundscape study gaps from 1,309 published papers (Scarpelli et al., 2020b) observed that: (a) regarding the scope, 58% focused on anthrophony, and only 8% on biophony; (b) regarding the region, 56% were developed into temperate areas, and only 18% are dedicated to tropical ecosystems; (c) terrestrial ecosystems were 89%, and 11% aquatic ecosystems; (d) regarding the focus of study, 61% were in urban ecosystems, and only 7% in natural areas. However, it is still important to offer tools for the user to understand different environmental situations at a more specific level within a particular recording location, such as distinguish the daytime period (morning, afternoon or night), groups of animals, and categories of events present in the environment, among other elements coded by the sound signals.

In this scenario, the main questions we are interested in contributing to answer are: (a) How classify large amount of environmental audios efficiently? (b) What is the segregation ability of each group of features? Considering that some features vary according to the specific conditions of each environment, (c) What set of features describe a particular set of acoustic events adequately?.

In this work, we propose a method to identify the most discriminant features of sound events based on analysis and evaluation of the accuracy achieved by automatic classifiers. We also employ algorithms for feature rankings and visualizations to support identification of effective features.

The purpose of our study was to examine the power of discrimination of a set of features extracted from natural sounds, which have been categorized into three groups according to its source of extraction: audio, spectrum and image. From the audios, we have extracted acoustic indices, which are usually employed to evaluate the environment (Towsey et al., 2014a); from the spectrogram, we have extracted features with cepstral coefficients (Terasawa et al., 2005); and in the image for the spectrogram, we have applied descriptors provides by image processing (such as texture and color) (Gonzalez and Woods, 2010). The research is focused on the analysis of such features to determine what subsets of them best describe the soundscape under study. The performances were assessed evaluating the accuracy achieved when classification tasks were performed.

We have employed different machine learning and visualization techniques to assist in the exploration, analysis and description of acoustic data, due to the multi-dimensional and exploratory aspects of the data, generating a large set of features that describe each audio instance (Mazza, 2009), from which visualization methods can support identifying and confirming segregation power (Card et al., 1999) of sets of features.

In the Section 2 we describe the uses for feature sets in studying soundscapes. In Section 3 we explain each element of our study. In Section 4 we describe the application of such methods for the discrimination of frogs, birds and insect, and discuss the various results. Our conclusions are presented in Section 5.

2. Audio-based feature extraction and its applications

Studies have analyzed the scope of acoustic indices (for a review see Towsey et al., 2014a; Sueur et al., 2014; Gasc et al., 2015). Gasc et al. (2013) performed an acoustic indices analysis and concluded that there is correlation between the acoustic diversity index (ADI) and phylogenetic diversity in bird communities. Another study with the goal of finding appropriate descriptors used Linear Correlation Coefficients among 14 acoustic indices, to select the least correlated, which were 9: Background noise, Average signal-to-noise ratio, Acoustic event count, High-frequency coverage, Mid-frequency coverage, Low-frequency coverage, Acoustic complexity index (ACI), Entropy of the signal envelope and Spectral Entropy. In this study they have achieved their goal of summarizing the content of 24-h recordings visually in a spectrogram to monitor, describe and compare two distinct acoustic landscapes (Sankupellay et al., 2015). Other researchers (Fuller et al., 2015) also evaluated acoustic indices as indicatives of ecological conditions of a landscape. They studied six types of acoustic indices, identified the relationship between acoustic complexity index (ACI) and bioacoustics (BIO) for bird vocalization, an determined that three acoustic indices known as entropy (H), acoustic evenness (AEI) and the normalized difference soundscape index (NDSI) are the ones that best related soundscape, ecological conditions and bird species.

Dias et al. (2021) aimed at defining a strategy for exploration of soundscapes, capable of revealing similarity level between distinct natural environments, based on features extracted from the image of spectrograms, features of type Cepstral, and acoustic indices. According to the authors, the descriptors Mel-Frequency Cepstrum Coefficients (MFCCs) can distinguish effectively between two different soundscape

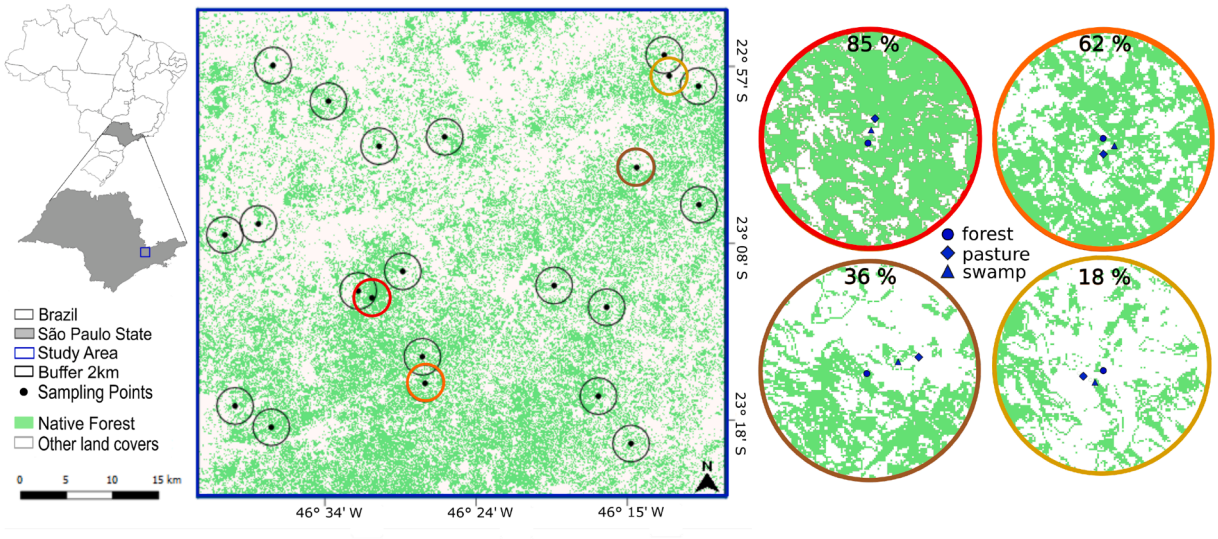


Fig. 2. On the left the location of landscapes where soundscape data were sampled using autonomous audio recorders within the Long Term Ecological Research of Ecological Corridor Cantareira-Mantiqueira (LTER CCM or PELD CCM), São Paulo, Brazil. On the right side we present four of our sampling landscapes varying in forest cover, with the location of each audio recorder. Black blue dots represent forest, rhombus shape are pastures and crosses triangle are swamps.

locations. Also aiming to compare two soundscapes environment visually, Sankupellay et al. (2015) generated false-color spectrograms from three acoustic indices (Towsey et al., 2014c) revealing distinct visual patterns for different habitats. Another analysis strategy (Towsey et al., 2014b) is based on the combination of several acoustic indices that can produce more useful ecological information than individual indices, using ranking of acoustic indices (features) based an "acoustic richness" score. the study concluded that indices such as temporal entropy H[t], spectral entropy H[s] and ACI are useful indicators of bio-acoustic activity and of species diversity.

The classification of soundscapes for multiple species based on Gamatone filter features was addressed in Agrawal et al. (2017). The data set employed was ESC-50[1] with category of events as animals, natural soundscapes and water sounds, human, interior domestic, exterior urban noises; in that case, the data set is not recorded under natural conditions and sound files contain 5 s with one specific event. Also in the literature, more rigorous work that perform sound classification for specific subspecies is presented, for example, for classification of subspecies frogs (Xie et al., 2018; Han et al., 2011; Xie and Towsey, 2016), anurans (Noda et al., 2016) or for classification of bird subspecies (Raghuram et al., 2016; Stowell and Plumbley, 2014; Qian et al., 2015).

Several studies have investigated the relationship between different acoustic indices and biodiversity, mainly for birds (Depraetere et al., 2012; Towsey et al., 2014a; Mammides et al., 2017; Retamosa and Ramírez-Alán, 2018; Machado et al., 2017; Jorge et al., 2018; Zhao et al., 2019; Moreno-Gómez et al., 2019). A consensus is that rarely did investigators identify the best index, since the climate (tropical or temperate), type of environment (forest or savanna), species composition or even incorrect sampling design resulted in inconsistencies when trying to capture differences between different areas (Eldridge et al., 2018). In a recent study, also conducted in the LTER CCM region, target of this paper, several acoustic indices were used to describe the soundscapes and to access how the forest cover influences these indices (Scarpelli et al., 2020a).

In contrast to such studies, here we use the indices not to correlate directly with bird of forest richness or abundance, but combine indices with other extracted features in order to best predict the presence of a particular event (such as bird vocalizations), in a large and diverse database. Being able to answer that is important in preprocessing steps of the database, both for the tagging of species of interest and to automatize the analyzes, such as to describe biodiversity and to tell species apart automatically, allowing their monitoring.

Our work presents a study of such segregation tasks by applying classification of events after visually and numerically assessing the relevance of features for a particular event of interest. The ranking of features can also be used for predicting categories of events. This work takes a different approach from previous ones by employing visualization to support the initial analysis and verification of meaningful attributes regarding particular events (such as the presence of anurans, birds and insects). We analyze a series of attributes from various sources, ranking them and narrowing down the choices to the ones representing each phenomenon of interest. We also use classification to confirm findings. We aim to achieve a stable methodology that can be replicated for characterization in the study of additional events beyond those exemplified in our case study.

3. Material and methods

3.1. Region of study

The region under study hosts the Long-Term Ecological Research project in the Ecological Corridor of Cantareira-Mantiqueira (LTER-CCM or Pesquisa Ecológica de Longa Duração-PELD CCM in Portuguese), localized in the transition between northeastern São Paulo state and south Minas Gerais state - Brazil (Fig. 2). The region is considered as a conservation priority because it connects two important highly forested blocks, the Cantareira and the Mantiqueira (Boscolo et al., 2017). Landscapes are composed by forest in different succession stages, agriculture, forestry plantation (mainly *Eucalyptus spp.*), swamps, water, roads, villages and larger urban areas (Barros et al., 2019). The region is part of Atlantic Forest biodiversity hotspot (Mittermeier et al., 2011), where habitat loss and fragmentation reduced forest cover to about 16%, with remnants reduced in size (84% is <50 ha), highly isolated (average isolation c.a. 1440 *m*) and high edge effect (50% of forests are less than 100 m from any edge) (Ribeiro et al., 2009).

3.2. Experimental design and sound recordings

We have collected sound data in 22 landscapes, which varied in forest cover (16% to 86%) and spatial heterogeneity levels. The same sampling design and audio data was used by Scarpelli et al. (2020a), and

[1] available in: https://github.com/karoldvl/ESC-50/

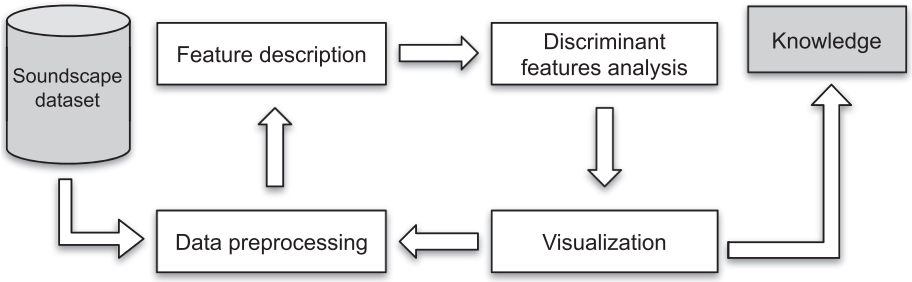


Fig. 3. Schematic process of the Methodology to identify discriminant features.

the recordings occurred between October 2016 and January 2017. In each landscape, we set up autonomous audio recorders Song Meter Digital Field Recorders (SM3) (Wildlife Acoustics, Inc., Massachusetts), which were fixed on trees at 1.5 *m* above ground. Three sites with different environment types were sampled within each landscape: forest, pasture and swamps. However, it is not our focus here compare the differences of soundscape variations between the above environment types. The equipment used two omni-directional microphones (frequency between 20 Hz and 20 kHz), and sounds were recorded at 44.1 kHz with 16 bits, in "mono" mode to save battery and memory card space. The recording schedule of the equipment is presented in Supplementary Material S1. For each site we recorded 34 h, totaling 102 h per landscape, and summing 2,255 h of raw data for the entire project.

3.3. Audio subset and species labeling

For the current study, we employed only a subset of audios per landscape and sites, a selection made in two steps. First, we selected five files of 25-min each in the period of highest activity of birds in the morning (5h00 AM to 8h30 AM), and nine files of 15-min each in the period that amphibians and insects are more active (6h30 PM to 10h45 PM); we adopted this collection strategy for balance of the amount of sampling of the landscape, the animal activity times and the use of hardware battery. In the morning there is a maximum of 3 h of collection in consequence recorded every 25 min; however, at night it is almost 4.5 h of collection, in consequence recorded every 15 min. Second, we randomly selected 2 non-continuous minutes of each file, totaling 18,594 min for the morning, and 22,132 min for the night. These files were selected equally from each type of environment (forest, pasture and swamps) of each landscape. Therefore, our subset has a total of 40,726 min.

From these subsets we proceeded with partial labeling at minute-level, which resulted in 822 min tagged as birds, 615 min tagged as anurans (frogs and tree frogs) and 840 min tagged as insects, in a total of 2,277 min. For the purpose of the current study, the minutes were labeled considering only the presence of acoustic signals of these three groups of species. For birds and anurans, every minute with presence of acoustic signal has been labeled as "bird occurrence" or "frog occurrence". However, for insects, only minutes with predominance of insect vocalizations were labeled as "insect occurrence". The management of the audio recordings was performed in the free programming environment *R* (R Core Team, 2018); labeling was done using the acoustic analysis software Raven Pro 1.5.

3.4. Dataset

For this study our dataset comprised 2,277 sound files of one-minute each, divided into three classes: 615 for frogs, 822 for birds and 840 for insects. However, for the experiments, a total of four new data partitions were created: (a) DS1 (frogs, birds and insect), with 2,277 instances (minutes-files); (b) DS2 (frogs and birds), with 1437 instances (minutes-files); (c) DS3 (frogs and insects), with 1455 instances (minutes-files); and (d) DS4 (birds and insects), with 1,662 instances.

3.5. Method

In Fig. 3 we illustrate the main steps of the methodology. First, data pre-processing step is responsible for collecting and transforming audio signal in spectrograms. Then, feature extraction is performed on the audio signal and the on spectrograms for each selected recording. After that, the proposed feature analysis is performed to determine what features are most discriminating for the target events. Subsequently, each feature configuration is visualized to allow further analysis and confirm accuracy. Each step of the methodology is further described in the next sections.

3.6. Data pre-processing

Audio files were converted to mono channel format employing 40,124 *(samples/s)*. Spectrum creation was generated using the following configuration: (n_fft=512, $hop\_length = 2048$, win_length=512) (McFee et al., 2015; McFee et al., 2019). Them, the spectrogram was obtained using only the real part of the spectrum.

From the spectrograms, color images of dimensions 100×100 were created, in which the spectrogram intensities were normalized in the range of real numbers $(0 - 1)$. Colors for the pixels were given following a particular color lookup table. The *nipy spectral* and *inferno* color maps of the matplotlib library[2] were used.

3.7. Feature description

In this section the descriptors used for feature extraction are briefly presented and were categorized into three groups: (a) based on images; (b) based on spectrum; and (c) based on acoustic indices.

3.7.1. Descriptors based on image

- Gray Level Co-occurrence Matrix (GLCM): In order to describe the texture in the spectrogram images, GLCM features were computed. For this, the color image was converted to gray-scale, then six matrices were computed with the following offsets $d_{(xy)} = \{(0,1),(0,3),(0,5),(1,0),(3,0),(5,0)\}$. Texture features were obtained by computing, for each co-occurrence matrix, the following six Haralick measurements: Energy, Entropy, Contrast, Correlation, Homogeneity and Maximum Probability. In the present study a total of 36 features were extracted for each sound minute file.
- Border Interior classification (BIC): Using the color images of the spectrogram, features of the Border Interior Pixel Classification (BIC) type were extracted. In this case, the colors of the images have been quantized to 64 colors. Thus, two feature vectors were computed to classify the pixels colors located at the edges and the pixels located

[2] available in: https://matplotlib.org/

within regions. A total of 128 features were extracted for each sound minute file.

3.7.2. Descriptors based on cepstral

Cepstral or "Ceptrum" was introduced by Bogert (1963) and is defined as frequency (Fourier transform) of the logarithm of the magnitude of spectrum of the original signal. Cepstral features facilitate similarity comparisons among signals.

- Mels-Scaled Spectrogram (MEL): Directly from the spectrum, 16 MELs were extracted. Mel-Scaled Spectrogram is computed by passing the fourier transformed signal through a set of band-pass filters known as Mel-filter bank. A Mel is a unit of measurement based on the frequency perceived by human ears (Rao and Sarkar, 2014).
- Mel-frequency Cepstral coefficients (MFCC): Employing the spectrum, 16 MFCC were computed. MFCC makes an analysis of spectral features of short time, based on the use of the voice spectrum converted to a frequency Mel-Scale. These coefficients are a representation defined as the cepstrum of a time-winded signal, which has been derived from the application of Discrete Cosine transform (DCT), in non-linear frequency scales, where DCT is used to remove redundant information (Mitrović et al., 2010).
- Gammatone Frequency Cepstral Coefficients (GFCC): Similarly to the MFCC feature, from the spectrum 16 GFCC were computed. GFCC has several advantages compared with MFCC. For example GTCC use Gammatone filter-bank that is a group of filters for the cochlea simulation, which is more accurate than triangle filters employed in MFCC (Liu et al., 2013).
- Linear Predictive Coefficients (LPC): Employing the essentia[3] library directly from the audio signal frames, 11 LPC feature were computed. In LPC, filter coefficients can be computed directly in the time domain. Coefficients can be obtained using an inverse filter to produce a prediction error signal. Then the original signal can be synthesized exactly from the prediction error signal using a synthesis filter (Harma, 2001).

3.7.3. Descriptors based on Acoustic Indices:

- Acoustic indices (AI): Finally, the following 15 Acoustic Indices feature were computed:Average Signal Amplitude (ASA) calculated as the average amplitude of the wave envelope, formula in Towsey et al. (2014b); Background Noise (BGN) calculated from removing acoustic activity of the wave envelope using (Lamel et al., 1981) method (Towsey, 2013); Signal to Noise Ratio (SNR) calculates the difference ratio between the envelope (maximum amplitude) and the background noise (Towsey, 2013); Acoustic Activity (AA) computes a fraction of frames within a one-minute segment, where the signal envelope is more than 3 dB above the background noise level (Towsey, 2013); Number of Acoustic Events (NAE) determined by the number of times the signal envelope overcome the 3 dB limit (Towsey et al., 2014b); Temporal Entropy (Ht) calculated on the amplitude envelope over the time unit, with formula in Sueur et al. (2008); Spectral Entropy (Hs) seeks the concentration of energy on the frequencies, with formula in Sueur et al. (2008); Acoustic Entropy (H) is obtained by multiplying (Ht) and (Hs) (Sueur et al., 2009); Antrhophony (A) is calculated in function of the portion of acoustic components generated by humans (Sueur et al., 2014); Biophony (B) calculated in function of the portion of acoustic components generated by biological entities (Sueur et al., 2014); Normalized Difference Soundscape Index (NDSI) calculated from the proportion between anthropological (human) and biological (animal species) sounds (Kasten et al., 2012); Acoustic Complexity Index (ACI) quantifies the acoustic activities (biological sounds, anthropological sounds) of soundscape, based on differences in the variability of intensities produced by sounds (Pieretti et al., 2011); Shannon Index (H') derives from the entropy calculation, the formula is found in Villanueva-Rivera et al. (2011); Median Of Amplitude Envelope (M) is the median of the amplitude envelope whit formula in Depraetere et al. (2012); and Mid Band Activity (MBA) that is fraction of fragments of the spectrogram between the values 482 Hz and 3500 Hz, where the spectral amplitude exceeds 0.015, this for one minute of audio (Towsey et al., 2014a).

As a result, 238 features were processed for each audio minute file –hereafter named instance. Instances with extracted features are denoted by X^m, where X the instance and m the number of features per instance.

3.8. Discriminant feature analysis

In the following text we describe the steps taken to process the feature space formed by the collection of all previously mentioned.

- *Data cleaning*: similar to how data processing is performed in other tasks in data science, features that had constant values or low variability were removed because they are redundant and less informative (Faceli et al., 2011). In this case when presenting constant values or low variability, the features were eliminated using a standard deviation of less than or equal to 0.015.
- Normalization: the features were normalized using the Min–Max and Z-Score methods. One of the purposes of normalization is to place the variables within a certain numerical range of distribution, increasing reliability of distance calculations performed in subsequent tasks.
- Feature analysis: visual tools such as Box-plot Haemer, 1948 and Histograms Faceli et al., 2011 supported visual analyzes of the variability and behavior of features and comparing the behaviour of values when normalized by Min–Max and Z-Score.
- Feature selection: to determine the most representative features, we proceeded to run feature selection methods. This helps us manage the problem of the high dimensionality of the feature space, while also improving segregation results. In the process, further redundancy and correlation are handled (Kohavi and John, 1997; Hall, 2000). Given the set of instances and features, feature selection can be mathematically denoted as $X^m \Rightarrow X^p$, where m denotes dimension the original set of features, and p the dimension of optimal subset of features, $p < m$. Approaches to feature selection are based on filters, wrappers and embedding (Miao and Niu, 2016). Due to its advantages in classification tasks, we have chosen the embedded method known as Extra Trees Classifier (ETC); this method weighs features according to their importance, where the importance is given specifically by calculating the reduction of impurity in the split of features values when multiple decision trees are built; high values of the decrease in impurity indicate important features.
- Identification of the most discriminating features: Employing the ranking of importance of features, classification tasks are performed for the n first ranked features. Training and testing instances are established by 5-k-folds cross-validation for Training and some percentage of instance for Testing. The classification models used are: Random Forest (RFC), K-Nearest Neighbor Classifier (KNNC), Support Vector Classifier (SVC) and Extreme Gradient Boosting Classifier (XGBC). The most discriminating features are determined by the highest accuracy results of the classification of the n feature combinations.

3.9. Visualization

To validate the quality of identified features visually, multidimensional projections were employed. In this way, the set of instances and

[3] available in: https://essentia.upf.edu/

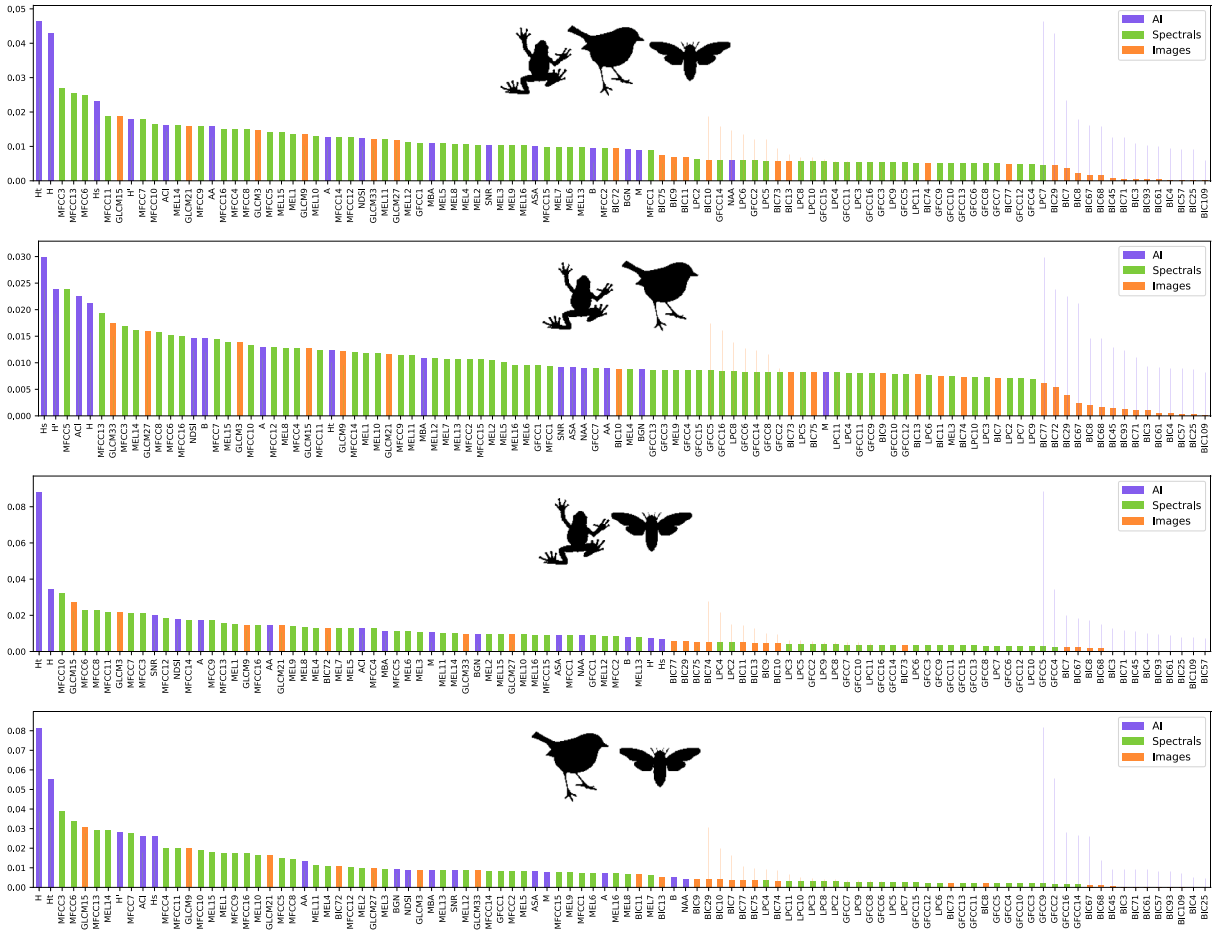


Fig. 4. Ranking of features: from top to bottom results of ranking of features from data partitions DS1 (frogs, birds and insect), DS2 (frogs and birds), DS3 (frogs and insects) and DS4 (birds and insects). Colors are: purple for features based on acoustic indices; green for spectrum; and orange for images. x axis shows features analyzed and y axis shows the level of importance for each feature.

Features X^p is mapped onto two dimensions $p = 2$ so we can visually distinguish sets of similar points by proximity. The projection techniques employed were t-SNE (van der Maaten and Hinton, 2008) and UMAP (McInnes et al., 2018).

It is expected that the employment of best features in multidimensional projections will improve the quality of the visualizations. We use the Stress metric (Kruskal, 1964) to measure such quality. Stress has values in the interval $[0, 1]$ for normalized distances values, with 0 being the highest quality and 1 the lowest, meaning the projection is not capable to reconstruct distance relationships in the data. The Stress that we use is defined in Eq. 1:

$$stress = \frac{\sum_{ij}\left(d_{ij} - \overline{d}_{ij}\right)^2}{\sum_{ij} d_{ij}^2} \tag{1}$$

where d_{ij} is the distance in the original space and $\overline{d}_{ij}$ is the distance in the visual space. We also employ the Silhouette Coefficient (Rousseeuw, 1987) to evaluate the cohesion and separation between grouped instances on visual space, computed by Eq. 3.9:

$$S = \frac{1}{n}\sum_{i=1}^{n}\frac{(b_i - a_i)}{max\{a_i, b_i\}}$$

where n is the number of instances and for each instance i, a_i is the average distance between all instances with the same class of y_i (cohesion), and b_i is the minimum average distance between all other instances in other groups different of y_i (separation). S has values in the interval $[-1, 1]$, with values closer to 1 meaning that the projection is better in terms of cohesion and separability.

3.10. Implementation

The methodology was implemented in Python, Cython and C. For preprocessing data and reading audio files we used the follow libraries: *soundfile*[4] version 0.10.3, *librosa*[5] version 0.8.0. For feature extraction were used *essentia*[6] version 2.1. For analysis and identification of discriminant features we used the algorithm ETC implemented in scikit-learn[7]. The source code of the methodology is available in the repository (https://github.com/hhliz/SoundscapeEcologyFeatures).

4. Results and discussions

In order to predict in a minute-file acoustic database with predominant vocalization of frogs, birds and insects, we performed all the tasks described in the previous section. After feature redundancy detection, the number of features reduced from 238 to 102. The experiments were divided into two steps: the analysis of discriminating features and the

[4] Available in: https://pypi.org/project/SoundFile
[5] Available in: https://librosa.github.io/librosa
[6] Available in: https://essentia.upf.edu
[7] Available in: https://scikit-learn.org/

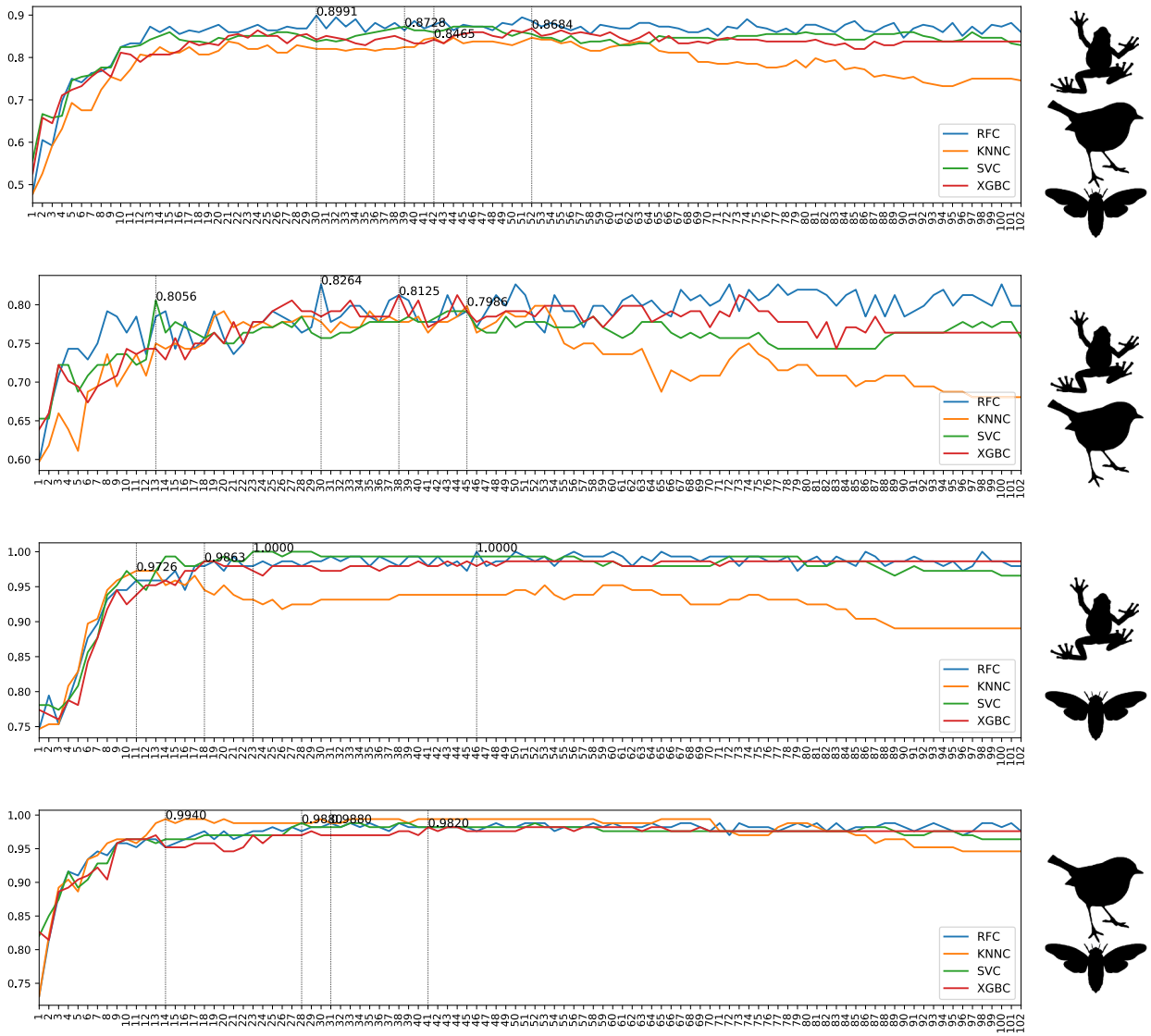


Fig. 5. Best features: from top to bottom results of best features for each data partition, employing four classifiers (Random Forest Classifier (RFC), K-Nearest, Neighbor classifier (KNNC), Support Vector Classifier (SVC) and Extreme-Gradient Boosting Classifier (XGBC)). The first graphic shows DS1 (frogs, birds and insect) with 30 features (89.91% accuracy); the second is DS2 (frogs and birds) with 30 features (82.64% accuracy); the third is DS3 (frogs and insects) with 46 features (100.0% accuracy) and the fourth is DS4 (birds and insects) with 31 features (99.40% accuracy).

visualization, described below.

4.1. Ranking of features

For the four data sets - (recalling DS1 (frogs, birds and insects); DS2 (frogs and birds); DS3 (frogs and insects) and DS4 (birds and insects) - we have followed the methodology for discriminant analysis of features, that is, embedded feature selection by Extra Trees Classifier (ETC). After the relevance for the features was obtained, they were sorted in descending order (highest to lowest) forming a ranking (see Fig. 4). The figure shows the categorization features, in which orange corresponds to image-based features, purple to spectrum-based features and green to acoustic indices.

In this experiment, it can be noted that features based on images such as BIC color descriptors do not perform as well as GLCM features. This may be because the color map does not provide relevant information to describe the spectrograms and, also, the color quantization might not favor them. On the other hand, between the two descriptors, the texture descriptors are more important and can better describe spectrogram information, and 36 texture features are more relevant than most of the 128 BIC color features. With relation to the features based on cepstral, it can be noted that they are within the top 10 features. MFCC and MEL features along with some GLCM texture features are the first in the ranking. Considering the acoustic indices feature set, the most important features are Temporal Entropy (*Ht*), Spectral Entropy (*Hs*), Acoustic Entropy (*H*), Acoustic Complexity Index (*ACI*) and the Shannon Index (*H′*). These features are generally placed in the 5 first positions in the rankings and have better predictive power compared to other descriptors. Entropy (*H*) indices and their variations in addition to *ACI*, were good indicators of bio-acoustic activity in previous work (Towsey et al., 2014a).

4.2. Identification of best features

Following the ranking of relevance of features in each partition, the discriminatory power of the combinations for the *n* first features was evaluated. In this experiment, in each combination of features, the accuracy in the training and test subsets was computed using respectively

Table 1
Classification results.

Dataset	Model	Train 90%	Test 10%	Train 80%	Test 20%	Train 70%	Test 30%	Train 60%	Test 40%	Train 50%	Test 50%
DS1 Frogs BirdsInsects(2277)	RFC	86.43(38)±0.017	89.91(30)	85.94(48)±0.008	87.28(32)	85.12(57)±0.012	85.09(59)	84.48(36)±0.017	84.74(51)	82.60(45)±0.017	83.41(59)
	KNNC	81.89(26)±0.018	84.65(42)	83.14(23)±0.011	83.11(19)	81.48(29)±0.016	82.16(21)	81.99(26)±0.019	80.90(15)	81.02(35)±0.019	81.30(32)
	SVC	85.55(54)±0.011	87.28(39)	85.12(27)±0.015	85.53(22)	84.62(52)±0.016	85.53(44)	83.75(59)±0.031	84.52(57)	84.19(52)±0.022	83.14(49)
	XGBC	84.77(76)±0.007	86.84(52)	84.51(54)±0.003	84.65(89)	83.24(66)±0.009	85.38(76)	83.02(60)±0.021	84.85(67)	82.78(52)±0.018	82.62(34)
DS2 Frogs Birds(1437)	RFC	81.59(46)±0.022	82.64(30)	81.03(55)±0.012	82.64(37)	81.00(47)±0.020	82.41(74)	78.76(42)±0.025	81.91(83)	78.68(31)±0.029	81.36(48)
	KNNC	76.64(28)±0.015	79.86(45)	75.80(26)±0.030	77.43(33)	76.72(26)±0.048	75.69(32)	74.59(39)±0.013	77.91(21)	74.51(43)±0.022	78.03(48)
	SVC	80.82(47)±0.017	80.56(13)	79.20(22)±0.019	80.56(46)	79.90(57)±0.029	80.09(37)	78.77(45)±0.017	80.70(30)	78.13(56)±0.015	79.69(25)
	XGBC	81.05(46)±0.020	81.25(38)	79.72(41)±0.017	78.47(49)	79.40(47)±0.012	78.94(49)	78.30(47)±0.020	79.65(59)	77.58(42)±0.020	79.00(77)
DS3 Frogs Insects (1455)	RFC	98.32(41)±0.005	100.00(46)	98.28(55)±0.013	99.31(32)	97.54(41)±0.018	99.31(67)	97.25(51)±0.017	99.14(53)	96.83(42)±0.014	96.70(27)
	KNNC	96.79(37)±0.011	97.26(11)	97.34(12)±0.008	98.28(13)	96.17(31)±0.013	95.65(39)	96.11(26)±0.017	96.74(13)	96.29(26)±0.011	95.05(17)
	SVC	98.40(38)±0.009	100.00(23)	98.11(27)±0.006	98.63(28)	98.13(44)±0.015	97.25(39)	98.17(48)±0.004	97.59(28)	97.11(53)±0.019	97.39(44)
	XGBC	97.33(44)±0.008	98.63(18)	97.59(60)±0.008	98.28(61)	96.56(64)±0.021	97.03(22)	96.68(51)±0.007	97.25(26)	96.84(53)±0.003	96.15(54)
DS4 Birds Insects (1662)	RFC	98.60(52)±0.004	98.80(31)	98.42(62)±0.008	98.80(34)	98.19(40)±0.005	99.00(77)	98.19(24)±0.004	98.50(33)	98.07(45)±0.013	98.44(78)
	KNNC	98.60(16)±0.007	99.40(14)	98.19(18)±0.006	99.10(18)	98.02(22)±0.011	99.20(17)	98.30(23)±0.007	98.95(20)	97.71(24)±0.005	98.07(16)
	SVC	98.66(27)±0.008	98.80(28)	98.50(31)±0.004	99.10(25)	98.19(58)±0.010	98.20(41)	98.29(38)±0.007	98.65(45)	98.43(45)±0.012	98.56(28)
	XGBC	97.93(56)±0.007	98.20(41)	98.19(73)±0.007	98.80(51)	98.19(41)±0.006	98.40(48)	98.30(46)±0.007	98.50(51)	98.08(27)±0.006	97.95(61)

the 5-k-fold cross validation strategy and 10% of instances per class. Fig. 5 presents the average accuracy results of the 10% Test subsets with the features normalized using Z-Score. We have also computed results with original features (without normalization) and with original features processed by min–max normalization, but Z-Score normalization provides the best results. Results were obtained from four learning models: Random Forest Classifier (RFC), K-Nearest, Neighbor classifier (KNNC), Support Vector Classifier (SVC) and Extreme-Gradient Boosting Classifier (XGBC).

In this experiment, the best results were obtained using RFC learning model, with 30 features (89.91% accuracy) for DS1 (frogs, birds and insect); on the other hand, other models present results similar to RFC, such as the SVC model. In this case, either model may be employed, but, in order to select the best model, we have to evaluate the computational cost also. For this reason, RFC is the best model to learn acoustic data of the categories discussed in this work.

For a deeper analysis of feature quality, the results of an extra study of the previous experiment are presented in Table 1. These results correspond to the 4 Training and Testing data sets, with accuracy of the training and testing, with their best n first score features. In this experiment, the percentage variation (10%, 20%, 30%, 40% and 50%) is taken into account in the formation of Test set; the rest of the instances are employed for training, in which it is also used a 5-k-fold Cross-validation as a strategy to validate learning. Based on the accuracy of most results in the training, we can demonstrate the superiority of the Random Forest classification model. In variations of test percentages, we can also see the preservation of accuracy results.

4.3. Visualizations

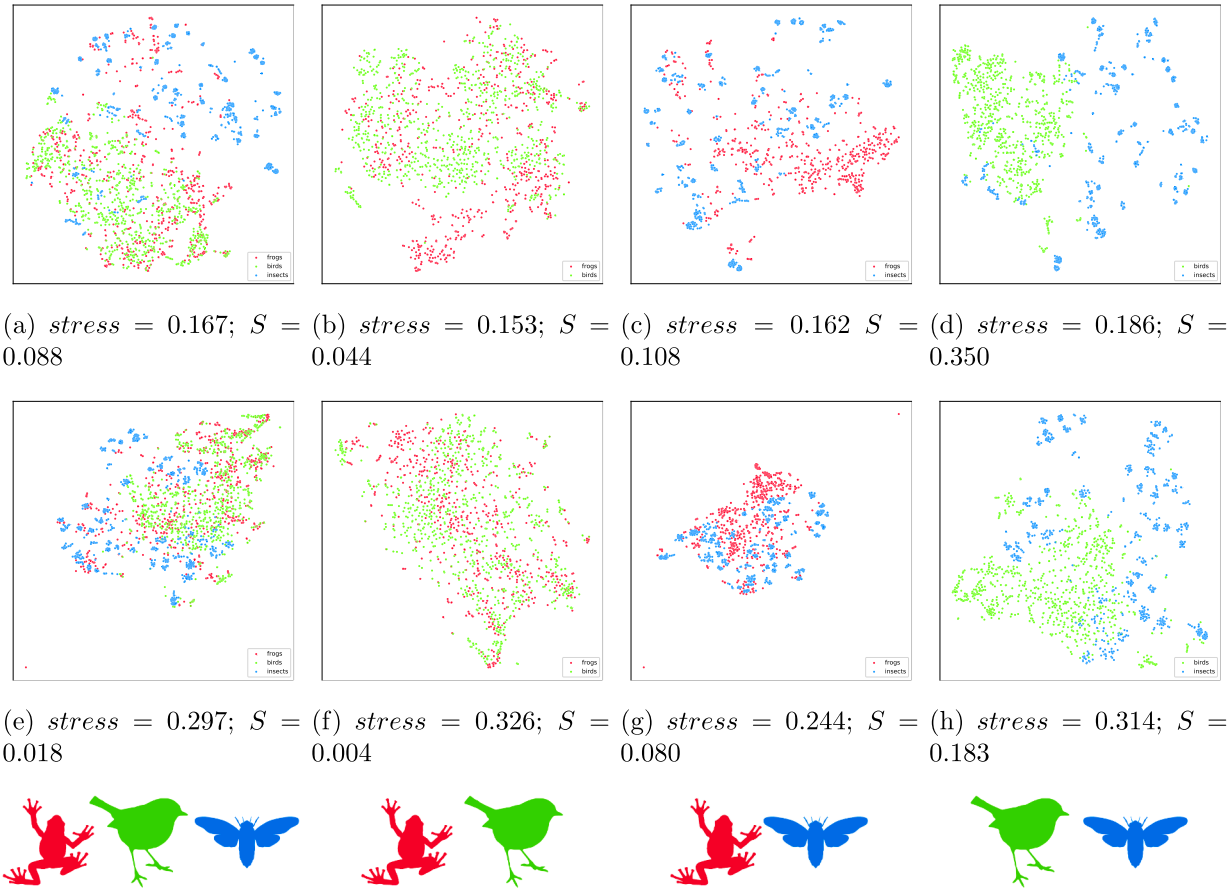
After identifying the best features, multidimensional projection techniques were employed to visualize the dataset, allowing us to confirm the quality of the features. When projection-based visualizations show greater separability, it can be inferred that the corresponding features are more discriminating. Fig. 6 illustrates the visualization results employing t-SNE technique (visually the choice for segregation purposes) for the whole instance set, joining training and test sets. The results are presented comparing with the same visualizations generated with the most discriminating features and with all features (n = 238). In the pictures, the dots represent sound files, with blue for sounds labeled as frogs, pink for birds and green for insects instances. The Stress and Silhouette results also validate the quality of the visualizations.

From the results, at the same time we can consider that a good visualizations are produced from best features extracted and selected.

5. Conclusions

In this study, we were able to automatically classify the audio files in instances of three classes of interest: frogs, birds and insects. The proposed method generated a total of 238 features, indicating a large survey and analysis of features not carried out so far in relation to acoustic landscape audios. The proposed methodology has also allowed us to obtain 102 features as a result of a first visual analysis, after which the level of importance for each of them was assessed in the context of sound classification, issuing a set of relevant features for each of the target events.

The Random Forest Classifier (RFC) learning model allowed the selection of the best features, with an average accuracy of 89.91%in the 10% test subset for the classification of frogs, birds and insects. An extension of the experiments to simulate the binary classification reached the precision of 82.64%, 100.0%and 98.80% for the classification between events of frogs-birds, frogs-insects and birds-insects, respectively. As the RF partitions the space recursively, making orthogonal cuts, this leads us to infer that RF captures information better when the classes of sound events are more spread out, justifying the accuracy values achieved in this study.



(a) $stress = 0.167$; $S = 0.088$ (b) $stress = 0.153$; $S = 0.044$ (c) $stress = 0.162$ $S = 0.108$ (d) $stress = 0.186$; $S = 0.350$

(e) $stress = 0.297$; $S = 0.018$ (f) $stress = 0.326$; $S = 0.004$ (g) $stress = 0.244$; $S = 0.080$ (h) $stress = 0.314$; $S = 0.183$

Fig. 6. Visualization of data sets: columns from left to right show results for DS1, DS2, DS3 and DS4. In the 1st row (a-d) t-SNE projections are shown for the data sets with 30, 30, 46 and 41 best features respectively. 2nd row (e-h) shows t-SNE projections for the same data sets employing the complete set of 238 features.

The learning models used also provide useful knowledge for future research. For example the SVC model when cutting the space by hyperplanes, this indicates that if the space of features is very well formed and does not present overlap, as a result it generates accurate results. Therefore, it can be concluded that the classes are linearly separable. Therefore, in the DS3 dataset we can conclude that the most discriminating features (first 23) identified with SVC provide relevant information when classifying anurans and insects, without resorting to more advanced methods such as RF. Additionally, we can infer that the discriminating features identified with the KNNC model can also be used in tasks of sound event retrieval, judged by the way in which the model performs the task of classification, as can be seen for the DS4 set.

The analyzes allowed us to identify in a general way the most discriminatory features of the sound events under study, among them acoustic indices such as: Temporal Entropy (*Ht*), Spectral Entropy (*Hs*), Acoustic Entropy (*H*), Acoustic Complexity Index (*ACI*) and the Shannon Index (H'). We also highlight the MFCC, MEL cepstral features (coefficient). Regarding the features extracted from the spectrogram image, those based on GLCM texture stand out. Identifying a set of more discriminating features is of great importance because it offers larger efficiency when carrying out classification tasks and we can infer that the discriminating features can provide better results in tasks of automatic labeling of sound events.

This methodology can be used in future studies to verify accuracy of feature in locating events of interest. In particular one could verify whether classification methods are capable of differentiating between insect minutes and minutes of rain or strong wind, since these events have a high entropy index. Due to the challenge that is to deal with large audio datasets on the different phases of data handling, being able to pre-identify where in the raw data we have greater probability of vocalizations of events of interest can significantly contribute with the advances of ecoacoustics research field. In the future, the proposed method could be applied to data sets with more categories of events and with features extracted directly from the audio signal through other methods, such as deep convolutional networks.

As the identification problems becomes more specific, it is very important to be able to consider user's knowledge in segregation tasks. We have recently developed visual tools for feature analysis and selection based on correlation with target categorical attributes (Minghim et al., 2020; Artur and Minghim, 2019). In them, correlation as well as relevance of features are displayed in a similarity layout allowing the experts to select sub-set of features associated with particular occurrences in the data. These tools will soon be adapted specifically to the soundscape ecology case, as a follow-up to the work presented in this paper.

CRediT authorship contribution statement

Liz Maribel Huancapaza Hilasaca: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Visualization. **Lucas Pacciullio Gaspar:** Resources, Data curation, Writing - review & editing. **Milton Cezar Ribeiro:** Resources, Data curation, Writing - review & editing, Investigation. **Rosane Minghim:** Supervision, Investigation, Writing - review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) [Grant No. 307411/2016-8] and the Coordination for the Improvement of Higher Education Personnel (CAPES) [Grant No. 133718/2018-2] - Finance Code 001. MCR thanks to FAPESP (processes #2013/50421-2; #2020/01779-5), CNPq (processes # 312045/2013-1; #312292/2016-3; #442147/2020-1) and PROCAD/CAPES (project # 88881.068425/2014–01) for their financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.ecolind.2020.107316.

References

Agrawal, D.M., Sailor, H.B., Soni, M.H., Patil, H.A., 2017. Novel teo-based gammatone features for environmental sound classification. In: 2017 25th European Signal Processing Conference (EUSIPCO), pp. 1809–1813. https://doi.org/10.23919/EUSIPCO.2017.8081521.

Alpaydin, E., 2014. Introduction to machine learning. MIT press.

Artur, E., Minghim, R., 2019. A novel visual approach for enhanced attribute analysis and selection. Computers Graphics 84, 160–172. https://doi.org/10.1016/j.cag.2019.08.015.

Barros, F., Martello, F., Peres, C., Pizo, M., Ribeiro, M., 2019. Matrix type and landscape attributes modulate avian taxonomic and functional spillover across habitat boundaries in the brazilian atlantic forest. Oikos 128. https://doi.org/10.1111/oik.05910.

Bogert, B.P., 1963. The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. Proc. Symposium on Time Series Analysis, 1963, https://ci.nii.ac.jp/naid/10022304707/en/.

Boscolo, D., Tokumoto, P.M., Ferreira, P.A., Ribeiro, J.W., dos Santos, J.S., 2017. Positive responses of flower visiting bees to landscape heterogeneity depend on functional connectivity levels. Perspectives Ecol. Conser. 15, 18–24. https://doi.org/10.1016/j.pecon.2017.03.002.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M. H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.-C., Watson, R., 2010. Global biodiversity: Indicators of recent declines. Science, 328, 1164–1168. https://science.sciencemag.org/content/328/5982/1164. DOI: 10.1126/science.1187512. http://arxiv.org/abs/https://science.sciencemag.org/content/328/5982/1164.full.pdfarXiv:https://science.sciencemag.org/content/328/5982/1164.full.pdf.

Card, S.K., Mackinlay, J.D., Shneiderman, B. (Eds.), 1999. Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

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

Dias, F.F., 2018. Uma estratégia para análise visual de Paisagens Acústicas com base em seleçáo de caracteristicas discriminantes. In: Master's dissertation in ciências de computaçáo e matemática computacional Instituto de Ciências Matemáticas e de Computaçáo. University of Sáo Paulo.

Dias, F.F., Pedrini, H., Minghim, R., 2021. Soundscape segregation based on visual analysis and discriminating features. Ecol. Informatics 61, 101184. https://doi.org/10.1016/j.ecoinf.2020.101184.

Eldridge, A., Guyot, P., Moscoso, P., Johnston, A., Eyre-Walker, Y., Peck, M., 2018. Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats. Ecol. Ind. 95, 939–952. https://doi.org/10.1016/j.ecolind.2018.06.012.

Faceli, K., Lorena, A.C., Gama, J., Carvalho, A.C.P. d. L.F. d., 2011. Inteligência artificial: uma abordagem de aprendizado de máquina. LTC.

Fuller, S., Axel, A.C., Tucker, D., Gage, S.H., 2015. Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? Ecol. Ind. 58, 207–215. https://doi.org/10.1016/j.ecolind.2015.05.057.

Gasc, A., Pavoine, S., Lellouch, L., Grandcolas, P., Sueur, J., 2015. Acoustic indices for biodiversity assessments: Analyses of bias based on simulated bird assemblages and recommendations for field surveys. Biol. Conserv. 191, 306–312. https://doi.org/10.1016/j.biocon.2015.06.018.

Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., Depraetere, M., Pavoine, S., 2013. Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? Ecol. Ind. 25, 279–287. https://doi.org/10.1016/j.ecolind.2012.10.009.

Gonzalez, R.C., Woods, R.E., 2010. Digital Image Processing, 3rd Edition. Prentice-Hall Inc., Upper Saddle River, NJ, USA.

Haemer, K.W., 1948. Range-bar charts. Am Statist 2, 23.

Hall, M.A., 2000. Correlation-based feature selection for discrete and numeric class machine learning. In: Proceedings of the Seventeenth International Conference on Machine Learning ICML, '00. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 359–366.

Han, N.C., Muniandy, S.V., Dayou, J., 2011. Acoustic classification of australian anurans based on hybrid spectral-entropy approach. Appl. Acoust. 72, 639–645. https://doi.org/10.1016/j.apacoust.2011.02.002.

Harma, A., 2001. Linear predictive coding with modified filter structures. IEEE Trans. Speech Audio Processing 9, 769–777. https://doi.org/10.1109/89.966080.

Hu, W., Bulusu, N., Chou, C.T., Jha, S., Taylor, A., Tran, V.N., 2009. Design and evaluation of a hybrid sensor network for cane toad monitoring. ACM Trans. Sen. Netw. 5 https://doi.org/10.1145/1464420.1464424, 4:1–4:28.

Johnson, C.N., Balmford, A., Brook, B.W., Buettel, J.C., Galetti, M., Guangchun, L., Wilmshurst, J.M., 2017. Biodiversity losses and conservation responses in the anthropocene. Science, 356, 270–275. https://science.sciencemag.org/content/356/6335/270. DOI: 10.1126/science.aam9317. http://arxiv.org/abs/https://science.sciencemag.org/content/356/6335/270.full.pdf.

Joo, W., Gage, S.H., Kasten, E.P., 2011. Analysis and interpretation of variability in soundscapes along an urban–rural gradient. Landscape Urban Planning 103, 259–276. https://doi.org/10.1016/j.landurbplan.2011.08.001.

Jorge, F.C., Machado, C.G., da Cunha Nogueira, S.S., Nogueira-Filho, S.L.G., 2018. The effectiveness of acoustic indices for forest monitoring in atlantic rainforest fragments. Ecol. Ind. 91, 71–76. https://doi.org/10.1016/j.ecolind.2018.04.001.

Kasten, E.P., Gage, S.H., Fox, J., Joo, W., 2012. The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. Ecol. Inform. 12, 50–67. https://doi.org/10.1016/j.ecoinf.2012.08.001.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intell. 97, 273–324.

Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1–27. https://doi.org/10.1007/BF02289565.

Lamel, L., Rabiner, L., Rosenberg, A., Wilpon, J., 1981. An improved endpoint detector for isolated word recognition. IEEE Trans. Acoust. Speech Signal Process. 29, 777–785.

Liu, J., You, M., Li, G., Wang, Z., Xu, X., Qiu, Z., Xie, W., An, C., Chen, S., 2013. Cough signal recognition with gammatone cepstral coefficients. In: 2013 IEEE China Summit and International Conference on Signal and Information Processing, pp. 160–164. https://doi.org/10.1109/ChinaSIP.2013.6625319.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 http://www.jmlr.org/papers/v9/vandermaaten08a.html.

Machado, R.B., Aguiar, L., Jones, G., 2017. Do acoustic indices reflect the characteristics of bird communities in the savannas of central brazil? Landscape Urban Planning 162, 36–43. https://doi.org/10.1016/j.landurbplan.2017.01.014.

Mammides, C., Goodale, E., Dayananda, S.K., Kang, L., Chen, J., 2017. Do acoustic indices correlate with bird diversity? insights from two biodiverse regions in yunnan province, south china. Ecol. Ind. 82, 470–477. https://doi.org/10.1016/j.ecolind.2017.07.017.

Mazza, R., 2009. Introduction to Information Visualization, 1st ed. Springer Publishing Company. Incorporated.

Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg, Oriol Nieto, 2015. librosa: Audio and Music Signal Analysis in Python. In Kathryn Huff, & James Bergstra (Eds.), Proceedings of the 14th Python in Science Conference (pp. 18–24). 10.25080/Majora-7b98e3ed-003.

McFee, B., Lostanlen, V., McVicar, M., Metsai, A., Balke, S., Thomé, C., Raffel, C., Lee, D., Zalkow, F., Lee, K., Nieto, O., Mason, J., Ellis, D., Yamamoto, R., Battenberg, E., Bittner, R., Choi, K., Moore, J., Wei, Z., Seyfarth, S., nullmightybofo, Friesch, P., Stöter, F.-R., Thassilo, Kim, T., Vollrath, M., Weiss, A., Weiss, A., 2019. librosa/librosa: 0.7.1. doi: 10.5281/zenodo.3478579. DOI: 10.5281/zenodo.3478579.

McInnes, L., Healy, J., Melville, J., 2018. Umap: Uniform manifold approximation and projection for dimension reduction. http://arxiv.org/abs/1802.03426 cite arxiv:1802.03426Comment: Reference implementation available at http://github.com/lmcinnes/umap.

Miao, J., Niu, L., 2016. A survey on feature selection. Procedia Computer Science, 91, 919–926. http://www.sciencedirect.com/science/article/pii/S1877050916313047. doi: 10.1016/j.procs.2016.07.111. Promoting Business Analytics and Quantitative Management of Technology: 4th International Conference on Information Technology and Quantitative Management (ITQM 2016).

Minghim, R., Huancapaza, L., Artur, E., Telles, G.P., Belizario, I.V., 2020. Graphs from features: Tree-based graph layout for feature analysis. Algorithms 13. https://doi.org/10.3390/a13110302.

Mitrović, D., Zeppelzauer, M., Breiteneder, C., 2010. Chapter 3 - features for content-based audio retrieval. In Advances in Computers: Improving the Web (pp. 71–150). Elsevier volume 78 of Advances in Computers. http://www.sciencedirect.com/science/article/pii/S0065245810780037. doi: 10.1016/S0065-2458(10)78003-7.

Mittermeier, R., Turner, W., Larsen, F., Brooks, T., Gascon, C., 2011. Global biodiversity conservation: The critical role of hotspots. In Biodiversity Hotspots (pp. 3–22). DOI: 10.1007/978-3-642-20992-5_1.

Moreno-Gómez, F.N., Bartheld, J., Silva-Escobar, A.A., Briones, R., Márquez, R., Penna, M., 2019. Evaluating acoustic indices in the valdivian rainforest, a biodiversity hotspot in south america. Ecol. Ind. 103, 1–8. https://doi.org/10.1016/j.ecolind.2019.03.024.

Noda, J.J., Travieso, C.M., Sánchez-Rodríguez, D., 2016. Methodology for automatic bioacoustic classification of anurans based on feature fusion. Expert Syst. Appl. 50, 100–106. https://doi.org/10.1016/j.eswa.2015.12.020.

Parks, S.E., Miksis-Olds, J.L., Denes, S.L., 2014. Assessing marine ecosystem acoustic diversity across ocean basins. Ecological Informatics, 21, 81–88. http://www.sciencedirect.com/science/article/pii/S1574954113001167. doi: 10.1016/j.ecoinf.2013.11.003. Ecological Acoustics.

Phillips, Y.F., Towsey, M., Roe, P., 2018. Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation. PLOS ONE 13, 1–27. https://doi.org/10.1371/journal.pone.0193345.

Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: The acoustic complexity index (aci). Ecol. Ind. 11, 868–873. https://doi.org/10.1016/j.ecolind.2010.11.005.

Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011. What is soundscape ecology? an introduction and overview of an emerging new science. Landscape Ecol. 26, 1213–1232. https://doi.org/10.1007/s10980-011-9600-8.

Qian, K., Zhang, Z., Ringeval, F., Schuller, B., 2015. Bird sounds classification by large scale acoustic features and extreme learning machine. In: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 1317–1321. https://doi.org/10.1109/GlobalSIP.2015.7418412.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria http://www.R-project.org/.

Raghuram, M.A., Chavan, N.R., Belur, R., Koolagudi, S.G., 2016. Bird classification based on their sound patterns. Int. J. Speech Technol. 19, 791–804. https://doi.org/10.1007/s10772-016-9372-2.

Rao, K.S., Sarkar, S., 2014. Robust Speaker Recognition in Noisy Environments. Springer Publishing Company. Incorporated.

Retamosa Izaguirre, M.I.R.I., Ramírez-Alán, O.R.-A., 2018. Acoustic indices applied to biodiversity monitoring in a costa rica dry tropical forest. Journal of Ecoacoustics, 2, 1–1. https://jea.jams.pub/article/2/1/40. 10.22261/jea.tnw2np.

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The brazilian atlantic forest: How much is left, and how is the remaining forest distributed? implications for conservation. Biological Conservation, 142, 1141–1153. http://www.sciencedirect.com/science/article/pii/S0006320709000974. doi: 10.1016/j.biocon.2009.02.021. Conservation Issues in the Brazilian Atlantic Forest.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Sankupellay, M., Towsey, M., Truskinger, A., Roe, P., 2015. Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. In 2015 Big Data Visual Analytics (BDVA) (pp. 1–8). DOI: 10.1109/BDVA.2015.7314306.

Scarpelli, M.D., Ribeiro, M.C., Teixeira, C.P., 2020a. What does atlantic forest soundscapes can tell us about landscape? Ecol. Ind. 121, 107050 https://doi.org/10.1016/j.ecolind.2020.107050.

Scarpelli, M.D., Ribeiro, M.C., Teixeira, F.Z., Young, R.J., Teixeira, C.P., 2020b. Gaps in terrestrial soundscape research: It's time to focus on tropical wildlife. Sci. Total Environ. 707, 135403 https://doi.org/10.1016/j.scitotenv.2019.135403.

Servick, K., 2014. Eavesdropping on ecosystems. Science, 343, 834–837. http://science.sciencemag.org/content/343/6173/834. DOI: 10.1126/science.343.6173.834. http://arxiv.org/abs/http://science.sciencemag.org/content/343/6173/834.full.pdf.

Stowell, D., Plumbley, M.D., 2014. Audio-only bird classification using unsupervised feature learning. In CLEF.

Sueur, J., Farina, A., 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound. Biosemiotics 8, 493–502.

Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acustica United With Acustica 100, 772–781.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid acoustic survey for biodiversity appraisal. PLoS ONE 3.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2009. Rapid acoustic survey for biodiversity appraisal. PLOS ONE 3, 1–9. https://doi.org/10.1371/journal.pone.0004065.

Terasawa, H., Slaney, M., Berger, J., 2005. Perceptual distance in timbre space. In: ICAD2005.

Towsey, M., 2013. Noise removal from wave-forms and spectrograms derived from natural recordings of the environment. In http://eprints.qut.edu.au/41131/.

Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014a. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecological Informatics, 21, 110–119. http://www.sciencedirect.com/science/article/pii/S1574954113001209. doi: 10.1016/j.ecoinf.2013.11.007. Ecological Acoustics.

Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014b. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecol. Inform. 21, 110–119. https://doi.org/10.1016/j.ecoinf.2013.11.007.

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., Roe, P., 2014c. Visualization of long-duration acoustic recordings of the environment. Procedia Computer Science, 29, 703–712. http://www.sciencedirect.com/science/article/pii/S1877050914002403. doi: 10.1016/j.procs.2014.05.063. 2014 International Conference on Computational Science.

Villanueva-Rivera, L.J., Pijanowski, B.C., Doucette, J., Pekin, B.K., 2011. A primer of acoustic analysis for landscape ecologists. Landscape Ecol. 26, 1233–1246.

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

Xie, J., Towsey, M., Zhang, J., Roe, P., 2016. Acoustic classification of australian frogs based on enhanced features and machine learning algorithms. Applied Acoustics, 113, 193 – 201. http://www.sciencedirect.com/science/article/pii/S0003682X16301864. https://doi.org/10.1016/j.apacoust.2016.06.029.

Zhao, Z., yong Xu, Z., Bellisario, K., wen Zeng, R., Li, N., yang Zhou, W., Pijanowski, B.C., 2019. How well do acoustic indices measure biodiversity? computational experiments to determine effect of sound unit shape, vocalization intensity, and frequency of vocalization occurrence on performance of acoustic indices. Ecological Indicators, 107, 105588. http://www.sciencedirect.com/science/article/pii/S1470160X19305801. https://doi.org/10.1016/j.ecolind.2019.105588.