



HAL
open science

Contributions of MIR to Soundscape Ecology. Part 2: Spectral timbral analysis for discriminating soundscape components

Kristen M. Bellisario, Jack Vanschaik, Zhao Zhao, Amandine Gasc, Hichem Omrani, Bryan Pijanowski

► To cite this version:

Kristen M. Bellisario, Jack Vanschaik, Zhao Zhao, Amandine Gasc, Hichem Omrani, et al.. Contributions of MIR to Soundscape Ecology. Part 2: Spectral timbral analysis for discriminating soundscape components. *Ecological Informatics*, 2019, 51, pp.1-14. 10.1016/j.ecoinf.2019.01.008 . hal-02080755

HAL Id: hal-02080755

<https://hal.science/hal-02080755>

Submitted on 10 Apr 2019

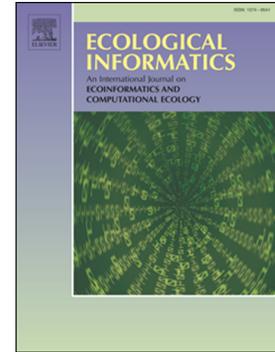
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Contributions of MIR to soundscape ecology. Part 2: Spectral timbral analysis for discriminating soundscape components

Kristen M. Bellisario, Jack VanSchaik, Zhao Zhao, Amandine Gasc, Hichem Omrani, Bryan C. Pijanowski



PII: S1574-9541(18)30190-0
DOI: <https://doi.org/10.1016/j.ecoinf.2019.01.008>
Reference: ECOINF 923
To appear in: *Ecological Informatics*
Received date: 15 July 2018
Revised date: 8 January 2019
Accepted date: 14 January 2019

Please cite this article as: K.M. Bellisario, J. VanSchaik, Z. Zhao, et al., Contributions of MIR to soundscape ecology. Part 2: Spectral timbral analysis for discriminating soundscape components, *Ecological Informatics*, <https://doi.org/10.1016/j.ecoinf.2019.01.008>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Title: Contributions of MIR to Soundscape Ecology. Part 2: Spectral timbral analysis for discriminating soundscape components

Running Title: Spectral timbral analysis

Kristen M. Bellisario¹, Jack VanSchaik¹, Zhao Zhao², Amandine Gasc^{1,4}, Hichem Omrani³, Bryan C. Pijanowski¹

¹ Purdue University, Dept of Forestry & Natural Resources
203 S. Martin Jischke Drive, Mann Hall Ste B066
Discovery Park Center for Global Soundscapes
West Lafayette, Indiana 47906

² Nanjing University of Science and Technology, School of Electronic and Optical Engineering,
Nanjing 210094, China

³ Luxembourg Institute of Research

⁴ Institut Méditerranéen de Biodiversité et d'Écologie marine et continentale, France

Corresponding author: Kristen Bellisario (kbellisa@purdue.edu)

Contributions of MIR to Soundscape Ecology. Part 2: Spectral timbral analysis for discrimination of soundscape components

by Kristen M. Bellisario, Jack VanSchaik, Zhao Zhao,
Amandine Gasc, Hichem Omrani, and Bryan C. Pijanowski

Abstract

Soundscape ecology evaluates biodiversity and environmental disturbances by investigating the interaction among soundscape components (biological, geophysical, and human-produced sounds) using data collected with autonomous recording units. Current analyses consider the acoustic properties of frequency and amplitude resulting in varied metrics, but rarely focus on the discrimination of soundscape components. Computational musicologists analyze similar data but consider a third acoustic property, timbre.

Here, we investigated the effectiveness of spectral timbral analysis to distinguish among dominant soundscape components. This process included manually labeling and extracting spectral timbral features for each recording. Then, we tested classification accuracy with linear and quadratic discriminant analyses on combinations of spectral timbral features.

Different spectral timbral feature groups distinguished between biological, geophysical, and manmade sounds in a single field recording. Furthermore, as we tested different combinations of spectral timbral features that resulted in both high and very low accuracy results, we found that they could be ordered to “sift” out field recordings by individual dominant soundscape component.

By using timbre as a new acoustic property in soundscape analyses, we could classify dominant soundscape components effectively. We propose further investigation into a sifting scheme that may allow researchers to focus on more specific research questions such as understanding changes in biodiversity, discriminating by taxonomic class, or to inspect weather-related events.

1. Introduction

Soundscape ecology is an emerging science that provides a new paradigm to address a variety of global environmental change issues (Pijanowski, Farina, Gage, Dumyahn, & Krause, 2011a; Pijanowski, Villanueva-Rivera, Dumyahn, Farina, Krause, Napoletano, ... & Pieretti, 2011b; *in sensu* Vitousek, Mooney, Lubchenco, & Melillo 1997). Soundscapes — defined as the combination of biological (e.g., animal vocalizations), geophysical (e.g., thunder, rain) and anthropogenic (e.g., transportation) sounds in a landscape (Pijanowski et al., 2011b) — contain information that can be used to study environmental patterns and processes. Soundscape analyses have been successfully applied to evaluate patterns of biodiversity in the context of environmental disturbances (e.g., Gasc et al., 2018; Farina & Pieretti, 2014; Sueur, Aubin, & Simonis, 2008).

Autonomous recording units (ARUs) are widely used survey instruments deployed to collect sound in the field for large temporal and/or spatial scale ecological research projects. As audio files contain a vast amount of data per unit, long-term landscape level studies tend to amass large volumes of recordings that are challenging to analyze with traditional “listen and tag” approaches (Towsey, Parsons, & Sueur, 2014; Zhao, Zhang, Xu, Bellisario, Dai, Omrani, & Pijanowski, 2017). Thus, researchers need efficient, automated analysis methods for conducting a majority of soundscape ecological research.

Two classes of approaches are currently being used by soundscape ecologists using automated techniques. This first approach is applied to species presence-absence surveys; these methods use segmentation of a sound file into species calls which are then matched to a known template of spectral characteristics using machine learning algorithms (e.g., Keen, Ross, Griffiths, Lanzone, & Farnsworth, 2014; Zhao et al., 2017). Automating these tools has been successfully demonstrated by a variety of researchers and is useful for studies designed to determine if a rare species is present at a location.

The second approach includes the application of metrics of broad spectral features in a recording (Villanueva et al., 2011; Sueur et al., 2014). These acoustic indices calculate the sonic complexity across a recording, integrating frequency and amplitude properties through such features as the entropy, evenness and dominance of frequencies, frequency band ratios, the number of frequency peaks, or the complexity of all signals across temporal and spectral ranges of the recordings. In one comparison of indices in varying landscapes, the acoustic complexity index (ACI) (Pieretti & Farina, 2013) and bioacoustics index (Boelman, Asner, Hart, & Martin, 2007) were linked to bird acoustic activity while the acoustic diversity index (ADI) (Villanueva-Rivera, Pijanowski, Doucette & Pekin, 2011) was sensitive to nocturnal activity (Fuller, Axel, Tucker, & Gage, 2015). Pekin et al. (2014) also showed that specific frequency bands as well as the ADI were positively correlated with the structural complexity of vegetation in a Neotropical rainforest that varied from old growth to young, secondary forests.

However, both current approaches to developing measures of acoustic patterns across space and time have their shortcomings. For the first approach, automated species detection requires the knowledge of the calls of every species present in a recording. Although much is known of bird

calls, species specific sounds of other animal calls such as amphibians, insects, fish and aquatic invertebrates are widely unknown. The call of many terrestrial mammals are extremely complex (e.g., primates) and species detection methods applied to those species have been limited (Aide, Corrada-Bravo, Campos-Cerqueira, Milan, Vega & Alvarez, 2013). The second approach, although widely used currently, make numerous assumptions about the nature of sound and their sound sources limiting their use in natural resource management. For instance, the Normalized Difference Soundscape Index (NDSI) (Kasten, Gage, Fox, & Joo, 2012) provides a ratio of animal and human soundscape contribution but its anthrophony frequency boundaries may include animal species that vocalize within the anthrophony frequency range (Eldridge, Casey, Moscoso, & Peck, 2016; Towsey et al., 2014). Thus, the reliability of acoustic indices is influenced by several factors that require substantial interpretation and may affect the quality of the result (Fairbrass, Rennett, Williams, Titheridge, & Jones, 2017; Mammides, Goodale, Dayananda, Kang, & Chen, 2017). Often, reliability of these indices is affected by geophonic and anthropogenic sounds captured in soundscape field recordings. As many researchers focus exclusively on the analysis of biophonic sounds, a filtering method of these competing sounds might aid researchers who are limited in the ability to interpret the biophony component and potentially improve index reliability.

In this paper, we present a novel approach for classifying the dominant soundscape composition of biophony, geophony, and/or anthrophony using a method classically used in musicology, spectral timbral analysis. Musicians have recognized “timbre” as a significant property of a sound signal influenced by composition (e.g., wood versus metal), shape (e.g., bell size), and sound production mechanism (e.g., string versus reed). The Acoustical Society of America defines timbre as the “attribute of auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness and pitch, are dissimilar” (Kitamura 1988). Recently, a group of computational musicologists has developed approaches that quantify timbral qualities, for instance using spectral features, to classify musical instruments (Peeters, Giordano, Susini, Misdariis, & McAdams, 2011). This group refers to this approach as Music Information Retrieval (MIR), a system that relies on models to solve audio classification problems using temporal and spectral features (Downie, 2003).

Spectral timbral analysis has been applied to tonal and noise-like signals in controlled tests using labeled samples (e.g., bird, cow, thunder, specific music instrument) and effective classification results have been obtained (e.g., Mitrovic, Zeppelzauer, & Eidenberger, 2007). As sound production mechanisms of terrestrial species (Marler, 1967; Wiley & Richards, 1978) are similar to those of idiophone (e.g. stringed) and aerophone (e.g., air) instruments (von Hornbostel & Sachs, 1914), we propose that MIR classification methods (Herrera, Amatriain, Batlle, & Serra, 2000) may be an adaptable analysis for soundscape ecologists. For instance, some birds and most insects may be considered idiophones, while more melodic birds and many amphibians may be aerophones.

Limited research has been conducted on how spectral timbral features classify field recordings that contain biological and non-biological sound sources (Bormpoudakis, Sueur, & Pantis, 2013) and to date, MIR methods have not been applied to soundscape ecology research (see Bellisario and Pijanowski, in review, for additional information). Considering that musical sound production properties are similar to those in the ecology domain, we were interested in whether these features

could distinguish between soundscape classes of anthrophony, geophony, and biophony, or in what capacity this is possible. Methods that could assist in further deconstructing sound sources to major animal taxonomic classes (e.g., taxonomic groups of birds, insects, amphibians, etc.) would be a great improvement over the approaches that employ acoustic metrics. We argue that spectral timbral features have the potential to characterize soundscape field recordings by dominant soundscape composition group (biophony, geophony, and anthrophony) and should complement current analysis approaches. The use of timbral features in bioacoustics and soundscape analyses has benefited previous research. For instance, spectral centroid was a favorable feature in the classification of anurans (Han et al, 2011). And, Sueur et al (2008) used spectral evenness as the basis for the H(f) entropy index. Another timbral feature used in current bioacoustic classification analyses is MFCC (Mel-frequency cepstral coefficients). This timbral feature has measurable success at classification of anurans (Noda et al 2015) and insects (Noda et al 2016) but is less robust in classifying birds with urban noise (Ricard and Glotin, 2016).

Here, we use MIR methods to demonstrate the efficacy of using a spectral timbral analysis approach to the field of soundscape ecology. Recordings from three contrasting ecosystems: tropical forest, an arid grassland, and an open field in an estuary system, were analyzed and the performance of spectral timbral features in classifying sound samples to ecologically meaningful classes was evaluated. In the rest of this paper, Section 2 briefly describes timbre as an acoustic feature from the perspective of the musicology field. Section 3 describes field recordings and spectral timbral analysis process. Experimental results and discussion are provided in Section 4 and 5.

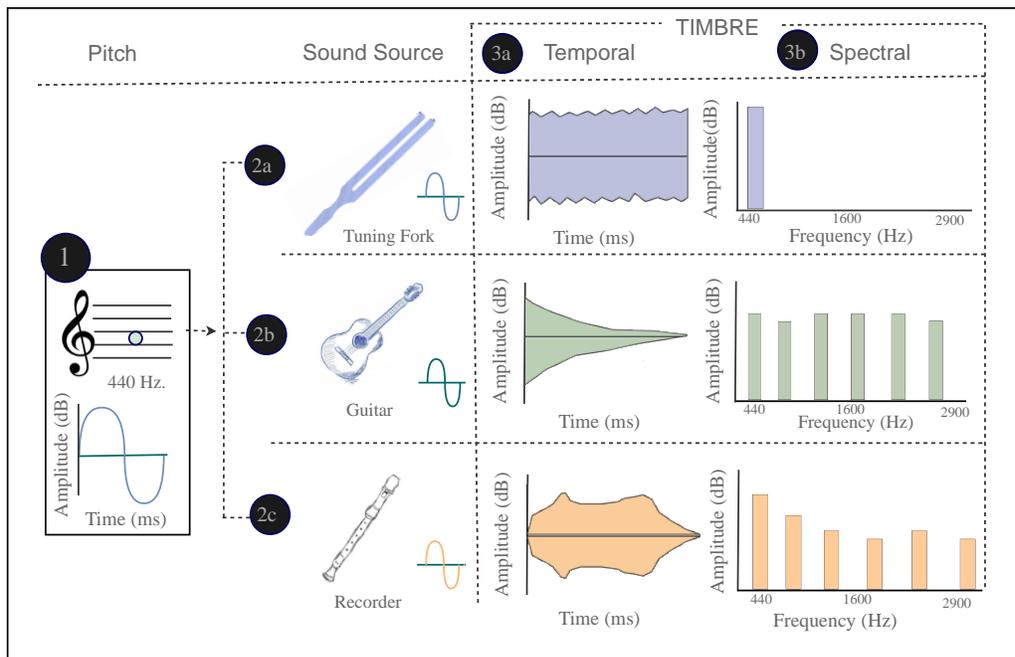


Figure 1: Timbre Description

The pitch 440 Hz (1) is played by three different sound sources (2a, 2b, and 2c) at the same loudness, or amplitude. Even though pitch and amplitude are the same for each sound source, the resulting temporal envelope (3a) and spectral envelope (3b) are different. These differences are features of timbre and can help discern between a played tuning fork, guitar, or recorder. Ultimately, these features provide unique patterns that can be used in machine learning applications.

2. Background: timbre as an acoustic feature

One of the first musicologists to quantify timbre was Schouten (1968) who suggested that timbre has several acoustic dimensions including temporal and spectral envelopes (Figure 1). The temporal envelope describes the waveform by overall shape of the signal (attack, sustain, decay, and release) over time while the characteristics of the spectral envelope are determined by the size of its sound source and the composition of its resonator, e.g., wood or metal. In this example, a single pitch (440 Hz) (1) is notated by three different sound sources, a tuning fork (2a), guitar (2b), and recorder (2c) that results in a waveform. The temporal envelope (3a) is the shape of this waveform and described as the amplitude of the pitch over time. The shape of the recorder's envelope differs from the guitar because of amplitude variation caused by vibrato and air pressure to produce sound. The differences between these sounds are evident in the spectral envelope (3b). The pure tone emitted from the tuning fork has a single harmonic peak at 440 Hz, while the recorder and guitar have a harmonic peak at 440 Hz but also harmonic peaks at different simultaneous frequencies that resonate when the single pitch is played by the instrument. It is this type of difference between the spectral envelopes of the tuning fork, guitar, and recorder that provide a unique set of descriptors that can identify one instrument from another instrument. Ultimately, the full set of timbral features exhibited in temporal and spectral envelopes can be quantified.

Instrument timbral profiles result in unique patterns including those generated by a vibration from hitting, shaking, and/or scraping (idiophone), and by the modulation of vibrations of air or other mediums (aerophone). These sound-production mechanisms are similar to those of sources in soundscapes. Although MIR models are not expressly designed to analyze soundscape recordings, some non-musical sounds have been identified in MIR analysis that share similar features with soundscape ecology recordings, such as a dog bark or rain (Mitrovic, 2006).

In particular, one MIR feature set used for instrument classification consists of spectral timbral features, a type of frequency-domain measurement resulting in unique spectral envelopes. More specifically, *spectral centroid* describes the brightness of sound and is also known to detect the difference among sounds, for instance those produced by various techniques of a violin bow; and *spectral slope* indicates how the sound tapers off toward high frequencies, with higher values indicating a stronger low frequency component (Bullock & Conservatoire, 2007). Both spectral centroid and spectral slope are highly correlated for harmonic instruments ($r=0.96$), but may be unique classifiers in a soundscape. Three other spectral features measure differences in the bandwidth and include *spectral skewness* — relationship to the mean and spread of the spectrum; *spectral spread* — the distribution of energy among frequencies; and *spectral variance* — a weighted version of spectral spread. These measurements are influenced by the beginning of a sound, or articulation features, and offer unique characteristics that MIR technologists (Agostini, Longari, & Pollastri, 2003; Gutierrez, 2006; Peeters et al., 2011; Tzanetakis & Cook, 2002) and audiologists (Krimphoff, McAdams, Winsberg, Petit, Bakchine, Dubois, ... & Vellas, 1994) have successfully used for sound classification tasks. Soundscape recordings often include overlapping,

or polyphonic, sounds. Recent MIR work has shown that instrument families (e.g., flute, violin, french horn) can be classified with the correct complimentary set of features even within a polyphonic texture (Muller, Ellis, Klapuri, & Richard, 2011). Based on the aforementioned studies, spectral centroid, spectral slope, spectral skewness, spectral spread, and spectral variance, as well as their different combinations, were investigated in this paper.

TIMBRAL ANALYSIS PROCESS

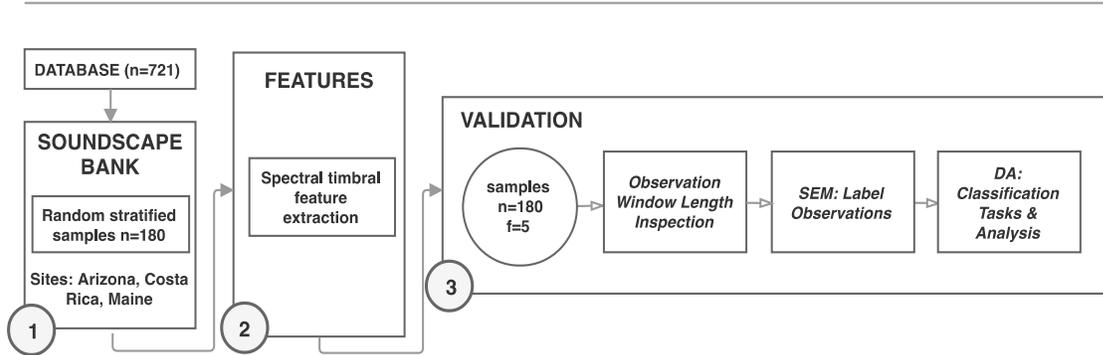


Figure 2: Block diagram of the proposed timbral analysis process.

3. Materials and methods

The overall workflow of our proposed spectral timbral analysis procedure consists of three main phases: 1) soundscape bank selection, 2) feature extraction, and 3) validation (Figure 2). The stratified random samples of sound recordings used in this work, i.e. the soundscape bank, are field recordings collected from three different ecosystems. The feature extraction step is a crucial step that computes spectral timbral measurements for each recording in the sound bank. Validation focused on identifying ecologically meaningful classes using an adapted sensory evaluation method (SEM) after an observation window length was determined. Then, based on five individual spectral timbral features, feature groups were assessed in terms of soundscape composition classification accuracy using discriminant analysis (DA).

3.1 Soundscape bank selection

We selected soundscape recordings from the Center for Global Soundscapes (CGS) soundscape library located at Purdue University (West Lafayette, Indiana, USA). These stereo recordings were collected in the field with ARU SM2+ (Wildlife Acoustics Inc., Maynard, MA, USA). All field recordings had a duration of ten minutes and a sampling rate of 44,100 Hz with 16-bit depth. We considered field recordings for four discrete time periods (dawn – 6-8:30 am, midday – 11 am-1:30 pm, dusk – 6-8:30 pm, and late evening – 11 pm-1:30 am) and from three different ecosystems: a tropical forest located in La Selva Biological Station (Costa Rica; $n = 239$), an arid grassland located in the Chiricahua National Monument located in the Sonoran desert (Arizona, USA; $n = 241$), and an estuarine open field located in the Wells Marine Reserve (Maine, USA; $n = 241$). The samples from the La Selva Biological Station were collected from the old growth and mature secondary forest dominated by sounds from over 700 species of mammals, birds, amphibians and insects (Pekin et al, 2012). Most common sounds are from rain, howler monkeys (*Alouatta*

species), toucans (Ramphastidae), tinamous (Tinamidae), cicadas (Cicadidae), and a variety of tropical frog species (Anura). The samples from the study in Chiricahua National Monument focused on arid environments with soundscapes of insects, birds, and wind. Samples from the Wells Estuary Reserve study focused on anthropogenic influences on a coastline reserve that serves as a fragile interface between the land and ocean systems. Common sounds in the terrestrial estuary were comprised of shorebirds and forest breeding passerine birds, along with those sounds from human activities (motors), rain, wind, and insects (Appendix A). The testing dataset used in this work consisted of the first 60s of 180 recordings with 60 recordings per site that were randomly selected via stratified sampling per site and time period from the 721 soundscape bank recordings.

3.2 Spectral timbral feature extraction

Spectral timbral features were extracted using a short-time Fourier Transform (STFT) that was computed with a window size of 1024. As the sample rate of the recordings were 44100 Hz, this resulted in a discrete frame length of 23ms. We used a Blackman-Harris window for the transform because it minimizes side-lobe levels, or undesired artifacts (spectral components that are present during the transform process), and is also a preferred window in polyphonic pitch detection (Harris, 1978; Herrera, Amatriain, Batlle, & Serra 2000; Masri & Bateman, 1995). Five spectral timbral features—spectral centroid, spectral spread, spectral skewness, spectral slope, and spectral variance—were extracted per frame of 23 ms from each recording in the dataset using Sonic Annotator (Cannam, Landone, & Sandler, 2010) with libxtract algorithms (Bullock & Conservatoire, 2007). A brief description of the spectral timbral features selected for this experiment is presented in Appendix B.

3.3 Validation

The validation phase aimed to determine how ecologically meaningful classes, identified using an adapted sensory evaluation method (SEM), match different combinations of the spectral timbral features. This phase included 1) determining the best window length of spectral timbral features for the DA; 2) describing the soundscape composition class for each recording using SEM by expert coders with assistance of visually inspecting the corresponding spectrogram (full bandwidth) or melodic-range spectrogram (a type of zoomed spectrogram used to inspect information in the 1.2 kHz bandwidth range) (Cannam et al., 2010); and 3) evaluating how the timbral features classify soundscape classes using DA. All analyses of this section were computed in the R program (initially Rx86_64 3.0.2; 2014 and optimized for Rx86_64 3.5.0, R Core Team, 2018) using packages ‘vegan’ (Oksanen et al, 2018), ‘MASS’ (Venables and Ripley, 2002), ‘pROC’ (Robin et al, 2011) and ‘caret’ (Kuhn, 2016). Visualizations used packages ‘ggplot2’ (Wickham, 2016), ‘gplot’(Wasserman & Faustm 1994), ‘dplyr’ (Wickham & Francois, 2016) , ‘colorspace’ (Ihaka et al, 2016), and ‘alluvial’ (Bojanowski & Edwards, 2016), here on after referred to by the package name (“package name,” R package).

3.3.1 Observation window segment inspection

Spectral timbral features were first extracted on a randomly selected subsample from the testing database in a frame-by-frame manner. However, considering the acoustic interpretation of a sound by humans for SEM labeling requires a longer duration, we calculated the centered mean of different window segment durations (1s, 3s, 10s, and 60s) for each feature and compared the

resulting mean value with those single frame counterparts using the absolute value of a Pearson correlation (mean value is the sum of all discrete 23ms for a length duration divided by the number of frames centered around the compared 23ms frame). The window length that had the high and consistent Pearson correlation values across features was considered the best performing choice, in our case (3s). Then, this window length segment inspection was used to establish the duration of the recordings used in the SEM (subsection 3.3.2) as well as to compute each mean feature value in our analysis (subsection 3.3.3) for each recording. Thus, we trimmed the testing dataset recordings from 60s to 3s segments (1s to 4s) for all further phases of this experiment.

3.3.2 Sensory evaluation method (SEM)

SEM is a method whereby humans use a sensory response and rate results according to that sense (Lokki, Pätynen, Kuusinen, & Tervo, 2012; Meilgaard, Carr, & Civille, 2006). In our adaptation of this method, we used both sight and hearing to label soundscape recordings in the dataset (n=180). First, KB and JV listened to 1s-4s of each recording in the testing dataset and coded it as containing biophony, geophony, and/or anthrophony. We assigned a single label to each recording using a hierarchical approach when two or more classes were present in a single sound file (Appendix A). Here, a class means anthrophony, geophony, or biophony. Any sound file that did not result in agreement was reviewed using a spectrogram and/or melodic-range spectrogram by KB and ZZ. The final label of the soundscape was determined in a single-label manner using the following coding rule: a recording was considered anthrophony if any motor sound (vehicle, motor boat, airplane) or coastal noise (n=12) (identified using a melodic-range spectrum as a constant band of low frequency with or without harmonic structure) was present (Pieretti & Farina 2013); then, geophony if any geophysical sounds of rain, wind or wind-like noise (identified using a melodic-range spectrum as different than anthrophony or coastal noise by fluctuations in the low frequency bands) were present (Digby, Towsey, Bell, & Teal, 2013; Hofmann, Wolff, Buck, Haulick, & Kellermann, 2012; Soares & Cherneva, 2005); and, then biophony when the soundscape consisted exclusively of biological sounds (unless minimal rain drops) from insects, amphibians, and/or some birds. This coding rule was developed using an exploratory data analysis of a random sample (n=60, 23ms frame) drawn from the original dataset to see how spectral timbral features measure sound characteristics of the soundscape composition classes (Borcard, Gillet, & Legendre, 2011).

Furthermore, considering that the alluvial diagram is a powerful visualization tool that compares categorical data within groups (Liu, Derudder, & Taylor, 2013), we used an alluvial diagram to compare the distribution of the SEM-derived classes within and between sites and time of the day ('alluvial', R package).

3.3.3 Discriminant analysis (DA)

Discriminant analyses searches feature combinations that best separate classes (Legendre and Legendre 1998). The soundscape recordings from the testing dataset were classified using a combinatorial computation of all possible feature combinations excluding single spectral timbral feature. We did not evaluate single features, as our preliminary work found that the individual features were not suitable for discrimination tasks. Additionally, considering that the unbalanced conditions of the SEM-derived classes could bias the DA results, the use of our small sample size

per class reduced potential bias when working with an unbalanced dataset (Blagus & Lusa, 2010; Xue & Titterington, 2008).

Table 1: Feature Groups

Listing of possible feature groups [FG] combinations of spectral timbral features used in the discrimination analysis.

Feature Group	Spectral Timbral Features
FG1	centroid, skewness, slope, spread, variance
FG2	centroid, skewness, slope, spread
FG3	centroid, skewness, slope
FG4	centroid, skewness
FG5	centroid, skewness, spread
FG6	centroid, skewness, variance
FG7	centroid, skewness, slope, variance
FG8	centroid, skewness, spread, variance
FG9	centroid, slope, spread, variance
FG10	centroid, slope, variance
FG11	centroid, slope, spread
FG12	centroid, slope
FG13	centroid, spread, variance
FG14	centroid, spread
FG15	centroid, variance
FG16	skewness, slope, spread, variance
FG17	skewness, slope, spread
FG18	skewness, slope, variance
FG19	skewness, slope
FG20	skewness, variance
FG21	skewness, spread, variance
FG22	skewness, spread
FG23	slope, spread, variance
FG24	slope, spread
FG25	slope, variance
FG26	spread, variance

We assessed conformance of the data to assumptions of multivariate normality, homogeneity of variance, and non-collinearity (Anderson, 2006). We tested multivariate normality using a Bonferroni-corrected permutational ANOVA (permutations=10,000) (Oksanen, 2016; Team,

2013; Venables & Ripley, 2002) which assessed if mean distances from group centroids in the multivariate space were similar among each group (Anderson, 2006; McArdle & Anderson, 2001), examined resulting covariance matrices for homogeneity of variance ('vegan,' R package), and checked collinearity ('MASS,' R package). Specifically, we tested the 26 possible combinations of the five features (Table 1), i.e. 26 different feature groups (FG), with two discriminant analysis classification algorithms: linear (LDA) and quadratic discriminant analysis (QDA) ('MASS,' R package).

These classifications were validated by comparing them to SEM results using a jackknife cross-validation (Oksanen, 2016; Venables & Ripley, 2002), which produced a confusion matrix ('vegan,' R package). This confusion matrix was used to compute performance metrics, i.e. accuracy, sensitivity (true positive rate, TPR) specificity (1-specificity is false positive rate, FPR) ('caret,' R package), and a multi-class area under the curve (AUC) (using TPR and TFR) ('pROC,' R package). AUC was the ultimate factor in deciding the highest performing feature groups if sensitivity / specificity was similar as it validates performance metrics with the lowest occurrence of misclassification (Hand & Till 2001). Overall, we considered high precision for a single feature group to be $TPR > 70\%$ and $AUC > 70\%$ following Powers (2011).

4. Results

First, the suitable window length duration was determined. Then, the recordings labeled by SEM were provided and the corresponding alluvial diagram was depicted. Finally, the performance comparison among different feature groups using two DA algorithms: LDA and QDA, was illustrated. In the following subsections, we present each experimental result in detail.

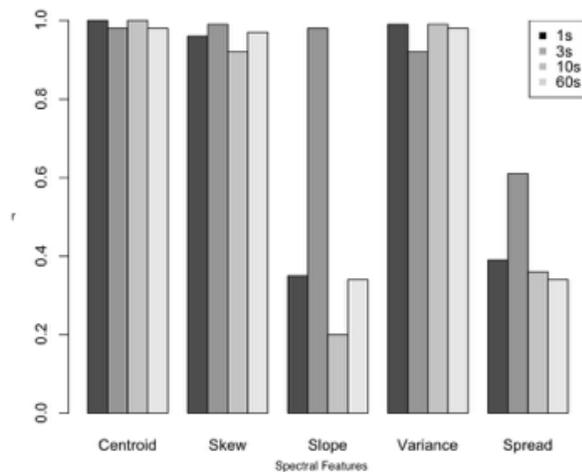


Figure 3: Observation Window Length Inspection Results

The absolute value of Pearson correlation (r) values of the discrete spectral feature (23 ms frame) compared with mean values of window lengths (1s, 3s, 10s, 60s) of individual spectral timbral features.

4.1 Observation window length inspection

The plot of Pearson correlation coefficients (r) of the mean feature values for each window length duration (1s, 3s, 10s, and 60s) is compared with corresponding single frame counterparts (Figure

3). The r values were similar for all window lengths for the spectral timbral features centroid, skew, and variance. However, the r values for slope and spread were not similar across window lengths. This plot indicated that slope and spread were measurements that are influenced by quick timbral changes in the sound event while the other features were more stable over longer durations. Due to this variation, as both slope and spread had highest r value at the 3s window, we selected a 3s window as the most suitable for all tested features.

4.2 Sensory evaluation method

Our observations, those dataset recordings labeled by SEM, were organized by class composition of biophony ($n=59$), geophony ($n=74$), and anthrophony ($n=47$). Figure 4 shows the resulting soundscape composition classes from SEM. More specifically, the alluvial diagram shows the density of class membership and its distribution across daily time periods (1-dawn: 6-8:30 am), 2-midday: 11am-1:30pm; 3-dusk: 6-8:30pm, and 4-late evening: 11pm-1:30am), and sites (Arizona, Costa Rica, and Maine). Each class was color coded with ribbons extending from the time period node to the site node. Here, we interpret the results of the alluvial diagram with respect to SEM class. Geophony had the greatest presence in the midday as represented by the width of the blue ribbon. From here, we can see that the geophony presence in midday is comprised of sounds from each site, with the widest band present in Arizona. Biophony had the greatest presence in the dawn, dusk, and late evening time periods as represented by the width of the green ribbon with the widest band present in Costa Rica. Anthrophony had the greatest presence in the dawn, midday, and dusk time periods as represented by the width of the red ribbon with the widest band present in Maine. Overall, the soundscape component classes, biophony, geophony, and anthrophony confirmed that SEM classes were distributed across each time period and the three sites.

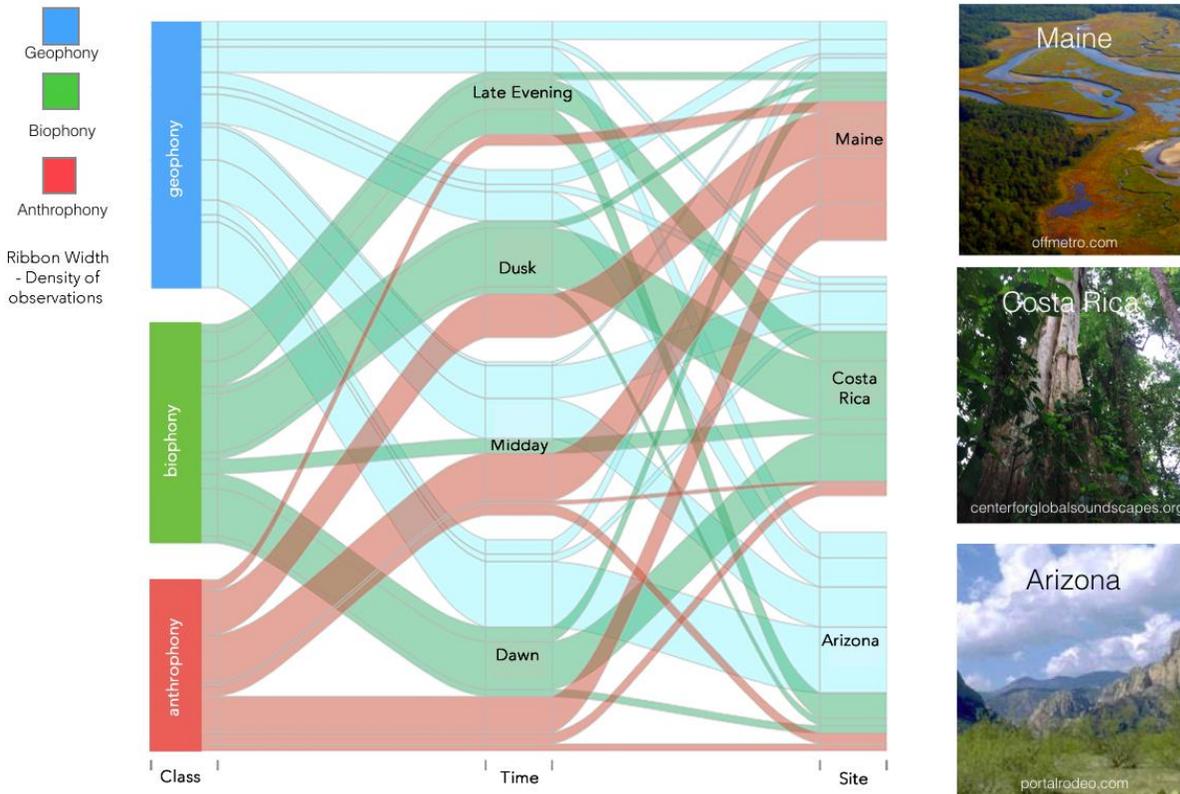


Figure 4: Alluvial Diagram of SEM Classes

The alluvial diagram displays the SEM class by density (ribbon width) and distribution across time period and site. The ribbon color is associated with the specific soundscape composition class (e.g., geophony is blue, biophony is green, and anthrophony is red). This alluvial diagram displays non-correlation between each node (class, time, site) representing an unbiased sample.

4.3 Discriminant analysis

To begin our comparison of DA results, we used a confusion matrix to calculate accuracy, sensitivity, and specificity using the observations (SEM-derived classes) and the DA-classified classes (Table 2 and 3). A performance measure of 0-100% is calculated using multiple algorithms (resulting in accuracy, sensitivity, specificity, or AUC) that use the number of correctly identified DA-classified classes as compared to the labeled observation, with 50% or less performance relative to low performance or a null hypothesis, and 70% or higher performance relative to high performance (Powers, 2011). We assessed different combinations of feature groups and DA algorithms that had the lowest and highest performances in terms of accuracy. As this analysis was exploratory in nature, different feature groups yielded a wide range of classification performance, suggesting that certain feature groups were better at classifying sonic characteristics for a particular soundscape class – and thus, have discriminatory value. For instance, when considering the metric accuracy, the lowest performing feature group for anthrophony had 0% (LDA) while the highest performing feature group had 87.2% (QDA). The lowest performing feature group for biophony was 0% (LDA) while the highest performing feature group was 86.7% (QDA). The performance range for geophony was similar having a range of 0% (QDA) to 75.3% (LDA). After comparing

all the results, though, we found there was not a single feature group that could classify the three classes with accuracy metrics all higher than 70%.

In this context, considering that those feature groups with the best performance on predicting a certain single class might be further utilized in a synergistic way, we selected a separate feature group to classify each soundscape class. Specifically, according to Tables 3 and 4, we determined the best performing feature groups for each class considering the complementary set of sensitivity, R^2 value, and AUC as listed below (complete set of metrics on Tables 3 and 4).

1) Biophony [FG12-plus-QDA]:

sensitivity = 86.7%, $R^2=0.21$, AUC=75.5%

2) Geophony [FG4-plus-LDA]:

sensitivity = 74%, $R^2=0.07$, AUC=71.4%

3) Anthrophony [FG11-plus-QDA]:

sensitivity = 87.2%, $R^2=0.21$, AUC=71.8%

Table 2: Results of Linear Discriminant Analysis

LDA: Feature Group Combination	Permutational Anova Results	AUC	Biophony : accuracy	Geophony : accuracy	Anthrophony : accuracy	Biophony : sensitivity	Biophony : specificity	Geophony : sensitivity	Geophony : specificity	Anthrophony : sensitivity	Anthrophony : specificity
FG1	F(2,179)=9.27, $R^2=0.09$, $p<0.001$	79.87%	55%	57.5%	83%	55%	86.7%	56.2%	73.8%	83%	83%
FG2	F(2,179)=6.45, $R^2=0.0$	70.1%	13.3%	64.4%	78.7%	13.3%	89.2%	64.4%	54.2%	78.7%	80.5%

	7, p<0.00 1										
FG3	F(2,179)=6.45, R ² =0.07, p<0.001	70.03%	8.3%	67.1%	78.7%	8.3%	90.8%	67.1%	50.5%	78.7%	81.2%
FG4	F(2,179)=6.45, R²=0.07, p<0.001	71.8%	0%	74%	78.7%	0%	97.5%	74%	43%	78.7%	81.2%
FG5	F(2,179)=6.45, R ² =0.07, p<0.001	72.2%	0%	72.6%	78.7%	0%	99.2%	71.2%	41.1%	78.7%	80.0%
FG6	F(2,179)=9.27, R ² =0.09, p<0.001	80.02%	56.7%	54.8%	80.9%	56.7%	86.7%	55%	73%	80.9%	82.7%
FG7	F(2,179)=9.27, R ² =0.09, p<0.001	80.07%	53.3%	60.3%	83%	53.3%	88.3%	58.9%	73%	83%	82.7%
FG8	F(2,179)=9.27, R ² =0.09, p<0.001	79.69%	56.7%	53.4%	80.9%	56.7%	85.8%	52.1%	72.9%	80.9%	82%
FG9	F(2,179)=9.27, R ² =0.09, p<0.001	78.75%	58.3%	38.4%	83%	58.3%	81.7%	38.4%	74.8%	83%	78.2%
FG10	F(2,179)=9.27, R ² =0.09, p<0.001	79.21%	56.7%	38.4%	83%	56.7%	80.8%	37%	72.9%	83%	79%
FG11	F(2,179)=24.03, R ² =0.2	70.07%	0%	71.2%	70.2%	0%	99.2%	72.6%	36.4%	70.2%	81.2%

	1, p<0.00 1										
FG12	F(2,179))=24.03 , R ² =0.2 1, p<0.00 1	69. 23 %	0%	72.6 %	68.1%	0%	99.2 %	72.6 %	35.5 %	68.1%	81.2%
FG13	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	79. 25 %	58.3 %	42.5 %	80.9%	58.3 %	80.8 %	42.5 %	72.9 %	80.9%	82%
FG14	F(2,179))=24.03 , R ² =0.2 1, p<0.00 1	68. 68 %	0%	71.2 %	66%	0%	97.5 %	69.8 %	33.6 %	66%	82%
FG15	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	78. 99 %	56.7 %	41.1 %	80.9%	56.7 %	80.0 %	41.1 %	72%	80.9%	82%
FG16	F(2,179))=9.27, R ² =0.9 4, p<0.00 1	79. 55 %	55%	57.5 %	80.9%	55%	88.3 %	57.5 %	72.9 %	80.9%	82%
FG17	*F(2,17 9)=1.95 , R ² =0.0 2, p=0.15	70. 03 %	8.3%	68.5 %	78.7%	8.3%	90.8 %	67.1 %	50.4 %	78.7%	81.2%
FG18	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	79. 61 %	53.3 %	60.3 %	80.9%	53.3 %	89.2 %	60.3 %	72%	80.9%	82.7%
FG19	*F(2,17 9)=1.95 , R ² =0.0 2, p=0.13	70. 03 %	8.3%	68.5 %	78.7%	8.3%	91.7 %	68.5 %	50.5 %	78.7%	81.2%

FG20	F(2,179)=9.27, R ² =0.09, p<0.001	66.78%	61.7%	65.8%	0%	61.7%	79.2%	65.7%	39.3%	0%	96.2%
FG21	F(2,179)=9.27, R ² =0.09, p<0.001	67.33%	63.3%	64.4%	0%	63.3%	79.2%	65.8%	40.2%	0%	96.2%
FG22	*F(2,179)=1.95, R ² =0.02, p=0.12	50.0%	0%	98.7%	0%	0%	99.2%	98.6%	0%	0%	100%
FG23	F(2,179)=9.27, R ² =0.94, p<0.001	79.26%	58.3%	42.5%	83%	58.3%	80.8%	42.5%	74.8%	83%	81.2%
FG24	*F(2,179)=1.17, R ² =0.01, p=0.32	69.36%	0%	74%	68.1%	0%	99.2%	74%	35.5%	68.1%	82%
FG25	F(2,179)=9.27, R ² =0.09, p<0.001	79.2%	56.7%	42.5%	80.9%	56.7%	81.7%	42.5%	72%	80.9%	81.2%
FG26	F(2,179)=9.27, R ² =0.09, p<0.001	59.92%	46.7%	75.3%	0%	46.7%	81.7%	74%	33.6%	0%	96.2%

Results for linear discriminant analysis show feature groups with high geophony classification accuracy. Performance measures include a permutational ANOVA, area under the curve, and sensitivity for the class. Feature groups with an “*” indicate a p-value greater than the Bonferroni-corrected significance of $p < 0.002$ and were not considered for the “sifting” workflow (gray) (Figure 5). Feature group 4 [FG4] had the overall highest performance for classification of geophony (black).

Table 3: Results for Quadratic Discriminant Analysis

QDA: Feature Group Combinations	Permutational Anova Results	AUC	Biophony : accuracy	Geophony : accuracy	Anthrophony : accuracy	Biophony : sensitivity	Biophony : specificity	Geophony : sensitivity	Geophony : specificity	Anthrophony : sensitivity	Anthrophony : specificity
FG1	F(2,179)=9.27, R ² =0.09, p<0.001	75.17%	81.20%	8.30%	87.20%	81.70%	58.00%	8.30%	97.20%	87.20%	77.30%
FG2	F(2,179)=6.45, R ² =0.07, p<0.001	74.38%	80.00%	8.30%	87.20%	80.00%	57.10%	8.30%	97.20%	87.20%	77.30%
FG3	F(2,179)=6.45, R ² =0.07, p<0.001	75.91%	83.30%	9.70%	87.20%	83.00%	54.60%	9.70%	97.20%	87.20%	81.80%
FG4	F(2,179)=6.45, R ² =0.07, p<0.001	73.93%	80.00%	11.00%	83.00%	80.00%	56.70%	11.00%	96.30%	83.00%	78.20%
FG5	F(2,179)=6.45, R ² =0.07, p<0.001	74.28%	78.30%	11.00%	85.10%	78.30%	60.80%	11.00%	95.30%	85.10%	75.20%
FG6	F(2,179)=9.27, R ² =0.09, p<0.001	74.81%	78.30%	12.30%	87.20%	78.30%	58.30%	12.30%	96.30%	87.20%	78.20%
FG7	F(2,179)=9.27, R ² =0.09, p<0.001	76.91%	86.70%	0.00%	78.70%	85.00%	58.00%	9.70%	98.10%	89.40%	79.50%
FG8	F(2,179)=9.27, R ² =0.09, p<0.001	74.85%	80.00%	11.00%	85.10%	80.00%	59.20%	11.00%	95.30%	85.10%	77.40%

	p<0.00 1										
FG9	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	74. 75 %	81.7 0%	0.00 %	91.50 %	81.7 0%	60.5 0%	0.00 %	99.10 %	91.50 %	70.50 %
FG10	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	75. 50 %	85.0 0%	0.00 %	91.50 %	85.0 0%	58.8 0%	0.00 %	100.0 0%	91.50 %	72.70 %
FG11	F(2,179))=24.03 , R²=0.2 1, p<0.00 1	72. 21 %	78.3 0%	0.00 %	87.20 %	78.3 0%	58.8 0%	0.00 %	99.10 %	87.20 %	68.90 %
FG12	F(2,179))=24.03 , R²=0.2 1, p<0.00 1	71. 80 %	86.7 0%	0.00 %	78.70 %	86.7 0%	51.3 0%	0.00 %	100.0 0%	78.70 %	75.80 %
FG13	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	74. 17 %	81.7 0%	4.10 %	87.20 %	81.7 0%	60.8 0%	4.10 %	98.10 %	87.20 %	71.40 %
FG14	F(2,179))=24.03 , R ² =0.2 1, p<0.00 1	73. 59 %	80.0 0%	4.10 %	87.20 %	80.0 0%	60.0 0%	4.10 %	98.10 %	87.20 %	71.40 %
FG15	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	79. 07 %	56.7 0%	58.90 %	83.00 %	56.7 0%	92.5 0%	58.90 %	77.60 %	83.00 %	76.70 %
FG16	F(2,179))=9.27, R ² =0.9 4, p<0.00 1	74. 70 %	80.0 0%	9.60 %	85.10 %	80.0 0%	59.2 0%	9.60 %	95.30 %	85.10 %	76.70 %

FG17	*F(2,179)=1.95 , R ² =0.02, p=0.15	74.12 %	78.30 %	9.60 %	85.10 %	78.30 %	60.00 %	9.60 %	95.30 %	85.10 %	75.20 %
FG18	F(2,179)=9.27, R ² =0.09, p<0.001	74.65 %	78.30 %	11.00 %	87.20 %	78.30 %	58.30 %	11.00 %	96.30 %	87.20 %	77.40 %
FG19	*F(2,179)=1.95 , R ² =0.02, p=0.13	73.77 %	80.00 %	9.60 %	83.00 %	80.00 %	55.80 %	9.60 %	96.30 %	83.00 %	78.20 %
FG20	F(2,179)=9.27, R ² =0.09, p<0.001	71.51 %	75.00 %	9.60 %	78.70 %	75.00 %	59.20 %	9.60 %	94.40 %	78.70 %	72.90 %
FG21	F(2,179)=9.27, R ² =0.09, p<0.001	75.14 %	76.70 %	9.60 %	89.40 %	76.70 %	67.50 %	9.60 %	95.30 %	89.40 %	69.20 %
FG22	*F(2,179)=1.95 , R ² =0.02, p=0.12	63.66 %	48.30 %	9.60 %	85.10 %	48.30 %	92.50 %	9.60 %	95.30 %	85.10 %	32.30 %
FG23	F(2,179)=9.27, R ² =0.94, p<0.001	74.01 %	81.70 %	2.70 %	87.20 %	81.70 %	60.80 %	2.70 %	98.10 %	87.20 %	70.70 %
FG24	*F(2,179)=1.17 , R ² =0.01, p=0.32	73.44 %	80.00 %	2.70 %	87.20 %	80.00 %	60.00 %	2.70 %	98.10 %	87.10 %	70.70 %
FG25	F(2,179)=9.27, R ² =0.09, p<0.001	78.87 %	55.00 %	58.90 %	83.00 %	55.00 %	93.30 %	58.90 %	76.60 %	83.00 %	75.90 %

FG26	F(2,179))=9.27, R ² =0.0 9, p<0.00 1	65. 07 %	58.3 0%	1.40 %	83.00 %	58.3 0%	83.3 3%	1.37 %	97.20 %	83.00 %	38.30 %
------	---	----------------	------------	-----------	------------	------------	------------	-----------	------------	------------	------------

Results for quadratic discriminant analysis show feature groups with high biophony and anthrophony classification accuracy. Performance measures include a permutational ANOVA, area under the curve, and sensitivity for the class. Feature groups with an “*” indicate a p-value greater than the Bonferroni-corrected significance of $p < 0.002$ and were not considered for the “sifting” workflow (gray) (Figure 5). Feature group 11 had the overall highest performance for classification of anthrophony while feature group 12 had the overall highest performance for classification of biophony (black).

QDA: Feature Group Combinations	Permutational Anova Results	AUC	Biophony: accuracy	Geophony: accuracy	Anthrophony: accuracy	Biophony: sensitivity	Biophony: specificity	Geophony: sensitivity	Geophony: specificity	Anthrophony: sensitivity	Anthrophony: specificity
FG1	F(2,179)=9.27, R ² =0.09, p<0.001	75.17%	81.20%	8.30%	87.20%	81.70%	58.00%	8.30%	97.20%	87.20%	77.30%
FG2	F(2,179)=6.45, R ² =0.07, p<0.001	74.38%	80.00%	8.30%	87.20%	80.00%	57.10%	8.30%	97.20%	87.20%	77.30%
FG3	F(2,179)=6.45, R ² =0.07, p<0.001	75.91%	83.30%	9.70%	87.20%	83.00%	54.80%	9.70%	97.20%	87.20%	81.80%
FG4	F(2,179)=6.45, R ² =0.07, p<0.001	73.93%	80.00%	11.00%	83.00%	80.00%	56.70%	11.00%	96.30%	85.00%	78.20%
FG5	F(2,179)=6.45, R ² =0.07, p<0.001	74.28%	78.30%	11.00%	85.10%	78.30%	60.80%	11.00%	95.30%	85.10%	75.20%
FG6	F(2,179)=9.27, R ² =0.09, p<0.001	74.81%	78.30%	12.30%	87.20%	78.30%	58.30%	12.30%	96.30%	87.20%	78.20%
FG7	F(2,179)=9.27, R ² =0.09, p<0.001	76.91%	86.70%	0.00%	78.70%	85.00%	58.00%	9.70%	98.10%	89.40%	79.50%
FG8	F(2,179)=9.27, R ² =0.09, p<0.001	74.85%	80.00%	11.00%	85.10%	80.00%	59.20%	11.00%	95.30%	85.10%	77.40%
FG9	F(2,179)=9.27, R ² =0.09, p<0.001	74.75%	81.70%	0.00%	91.50%	81.70%	60.50%	0.00%	99.10%	91.50%	70.50%
FG10	F(2,179)=9.27, R ² =0.09, p<0.001	75.50%	85.00%	0.00%	91.50%	85.00%	58.80%	0.00%	100.00%	91.50%	72.70%
FG11	F(2,179)=24.03, R ² =0.21, p<0.001	72.21%	78.30%	0.00%	87.20%	78.30%	58.80%	0.00%	99.10%	87.20%	68.90%
FG12	F(2,179)=24.03, R ² =0.21, p<0.001	71.90%	86.70%	0.00%	78.70%	86.70%	51.90%	0.00%	100.00%	78.70%	75.80%
FG13	F(2,179)=9.27, R ² =0.09, p<0.001	74.17%	81.70%	4.10%	87.20%	81.70%	60.80%	4.10%	98.10%	87.20%	71.40%
FG14	F(2,179)=24.03, R ² =0.21, p<0.001	73.59%	80.00%	4.10%	87.20%	80.00%	60.00%	4.10%	98.10%	87.20%	71.40%
FG15	F(2,179)=9.27, R ² =0.09, p<0.001	79.07%	56.70%	58.90%	83.00%	56.70%	92.50%	58.90%	77.80%	85.00%	76.70%
FG16	F(2,179)=9.27, R ² =0.09, p<0.001	74.70%	80.00%	9.60%	85.10%	80.00%	59.20%	9.60%	95.30%	85.10%	76.70%
FG17	*F(2,179)=1.95, R ² =0.02, p=0.15	74.12%	78.30%	9.60%	85.10%	78.30%	60.00%	9.60%	95.30%	85.10%	75.20%
FG18	F(2,179)=9.27, R ² =0.09, p<0.001	74.85%	78.30%	11.00%	87.20%	78.30%	58.30%	11.00%	96.30%	87.20%	77.40%
FG19	*F(2,179)=1.95, R ² =0.02, p=0.13	73.77%	80.00%	9.60%	83.00%	80.00%	55.80%	9.60%	96.30%	83.00%	78.20%
FG20	F(2,179)=9.27, R ² =0.09, p<0.001	71.51%	75.00%	9.60%	78.70%	75.00%	59.20%	9.60%	94.40%	78.70%	72.90%
FG21	F(2,179)=9.27, R ² =0.09, p<0.001	75.14%	76.70%	9.60%	89.40%	76.70%	67.50%	9.60%	95.30%	89.40%	69.20%
FG22	*F(2,179)=1.95, R ² =0.02, p=0.12	63.86%	48.30%	9.60%	85.10%	48.30%	92.50%	9.60%	95.30%	85.10%	32.30%
FG23	F(2,179)=9.27, R ² =0.09, p<0.001	74.01%	81.70%	2.70%	87.20%	81.70%	60.80%	2.70%	98.10%	87.20%	70.70%
FG24	*F(2,179)=1.17, R ² =0.01, p=0.32	73.44%	80.00%	2.70%	87.20%	80.00%	60.00%	2.70%	98.10%	87.10%	70.70%
FG25	F(2,179)=9.27, R ² =0.09, p<0.001	78.87%	55.00%	58.90%	83.00%	55.00%	93.30%	58.90%	76.60%	83.00%	75.90%
FG26	F(2,179)=9.27, R ² =0.09, p<0.001	65.07%	58.30%	1.40%	83.00%	58.30%	83.33%	1.37%	97.20%	83.00%	38.30%

5. Discussion

The classification of soundscape composition classes using field recordings is challenging due to the inherent diversity of sounds within a class. As soundscape ecologists often consider biological events in relation to other sound events, the biophony class needs to include acoustically different sounds (e.g., a howler monkey vocalization and a cricket stridulation), while separating them from geophony or anthrophony. Furthermore, the identification of characteristics that group these sounds within the boundaries of each soundscape class has not been widely studied or implemented.

We investigated spectral timbral feature analysis to address this fundamental problem in soundscape ecology due to its high performance capabilities in classifying musical instruments. We found in preliminary work that, although in musical instruments centroid and slope are highly correlated, in field recordings this was rarely the case. This could be due to the nature of an instrument's harmonic qualities whereby a fundamental frequency is present unlike most sounds found in a soundscape bank (Agostini et al., 2003; Beauchamp, 2007; Marozeau, de Cheveigné, McAdams, & Winsberg, 2003). In fact, the slope and spread were more sensitive to immediate changes in our soundscape bank and responsive to nuances in sound such as wind pitch rising slightly, while the other features were relatively stable over a span of time. Evaluating feature groups in timbral space was not dependent on template matching or species identification, but evaluated a recording using the full spectrum in three-second windows. This approach is useful as template matching and species-identification of sounds within a large collection of soundscapes is challenging due to unknown species and competing signals (Towsey et al., 2014; Zhao et al., 2017).

The set of tested feature groups had a range of classification accuracy performance for each class. We had suspected that one feature group might be able to classify the three classes. However, there was not one feature group that had strong performance for all soundscape composition classes. The most difficult soundscape class to classify was geophony. Upon listening to the entire group of geophonic sounds, several consisted of thunder and heavy rain, a natural ecological process that is often present in unfiltered soundscape banks. One possible improvement for geophonic classification is a pre-filtering method that removes extreme weather events and inclusion of a new class for inconclusive sounds.

Another interesting result from this experiment is that certain feature groups resulted in a single class with low accuracy. For instance, the classification result for [FG10-plus-QDA] was 0% accuracy for the geophony class. We can infer that this combination does not measure the timbral qualities present in that feature group and could be used to potentially “sift” out sounds that do not have geophony, a typical source of noise that is difficult to determine without manually searching sound files. Additionally, there were several high performing feature groups for an individual soundscape class with accuracy and AUC higher than 70%. Indeed, many feature groups had high performance for biophony and anthrophony, especially using the QDA algorithm. As for [FG14-plus-QDA], this option had a higher R^2 value ($R^2 = 0.21$) from the permutational ANOVA and resulted in 80% accuracy for biophony, 87.2% accuracy for anthrophony, with 0% for geophony. So, although we did not include this feature group for inclusion in the proposed “sifting” workflow,

this single workflow may be useful for ecologists with questions regarding biological and anthropogenic sounds without influence by weather-related occurrences.

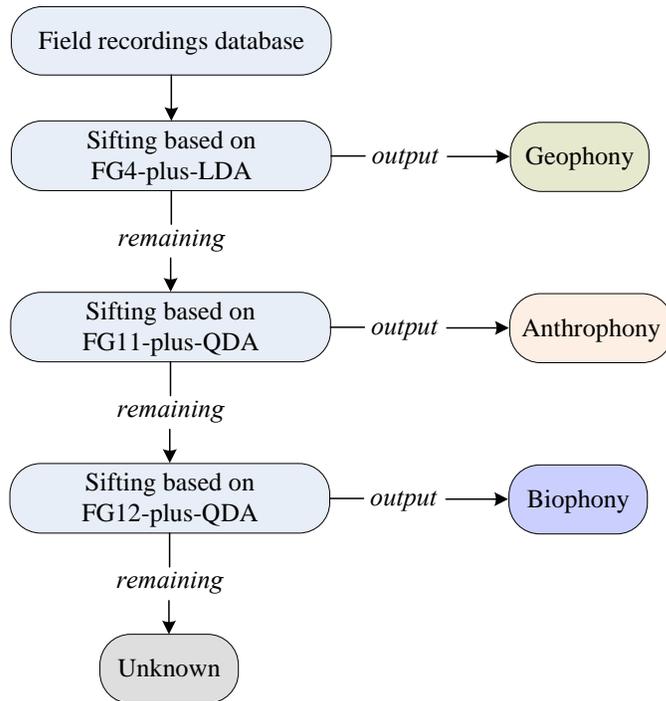


Figure 5: Proposed Sifting Workflow

A “sifting” workflow for soundscape ecology composition classification. FG refers to feature group, LDA is linear discriminant analysis, and QDA is quadratic discriminant analysis.

Many relationships exist amongst the feature groups, but those feature groups with higher F statistics and R^2 values, as well as sensitivity and multi-class AUC, should be considered for deeper analysis. In this study, utilizing this criterion, we compiled a list of potential “sifting” feature groups, geophony [FG-4 plus-LDA], biophony [FG12-plus-QDA], and anthrophony [FG11-plus-QDA], that could benefit soundscape analyses (Figure 5). In this proposed method, we demonstrate a series of steps that can output a single soundscape composition class. In Figure 5, we illustrate a proposed “sifting” process to output all biophonic sounds could benefit research studies that relate to an underlying theme in soundscape analysis, such as the acoustic niche hypothesis, whereby species are thought to have evolved to produce sound in their own acoustic bandwidth to avoid masking (Brumm & Slabbekorn, 2005). This “sifting” process could be expanded to include ecological “genres” that become more specialized resulting in a species community profile. This paper explored only one type of MIR feature, but other features may offer novel solutions such as algorithms that utilize tempos of a location (tempograms), concentration of frequencies (chromagrams), and other temporal and spectral timbral features (timbregrams) (Grosche, Müller, & Kurth, 2010). As we have investigated one type of feature analyses, a low-level audio descriptor in MIR, ecologists might consider using analyses of other audio descriptor scales. Certainly, the gamut of computational musicological analyses is a useful exploratory tool for the soundscape ecologist.

Overall, this work demonstrates that we can use different spectral timbral feature groups to classify sounds into soundscape composition classes, and offers the potential for a “sifting” process that may retrieve recordings that contain or do not contain a soundscape class. It is worth remarking that this hierarchical scheme has been recently suggested by soundscape researchers as a needed tool to reduce analysis time (Towsey et al., 2014). A similar multi-step “sifting” approach has been used in recent visualization problems (Lahoulou, Larabi, Beghdadi, Viennet, & Bouridane, 2016) and as a tool on the progress of classifying musical genres (Li & Ogihara, 2006).

Our method worked in a single-label manner, which allowed discrimination of a single, dominant soundscape composition class. Additionally, since determining the SEM class was difficult due to interpreting the difference between distant noise and wind, visual cues were involved in the inspection of the sound. Further improvements to this experiment include the use of multi-label concepts (Omrani, Abdallah, Charif, & Longford, 2015; Omrani, Tayyebi, & Pijanowski, 2017; Zhang et al., 2016) and a larger sample size.

6. Appendix A

Spectral Timbral Feature Definitions (Bullock & Conservatoire, 2007; Cannam et al., 2010; Krimphoff et al., 1994; Peeters et al., 2011)

Spectral Centroid

Measures spectral center of gravity of the magnitude spectrum at time window t , defined as

$$\mu(t) = \sum_{n=1}^N f_n p(t, n)$$

Where n is the spectral index resultant from the Fourier transform and f_n is the n th frequency. And

$$p(t, n) = \frac{v(t, n)}{\sum_{n=1}^N v(t, n)}$$

Where $v(t, n)$ is the amplitude spectrum at frequency index n and time window t .

Spectral Spread

Measures the spread of the spectrum around it's mean value at time window t , defined as

$$\sigma^2(t) = \sum_{n=1}^N (f_n - \mu(t))^2 p(t, n)$$

Spectral Skewness

Measures the asymmetry of the spectrum by using the frequency distribution around its main value at time window t . A value of $??_3$ indicates a symmetric distribution, $??_3 < 0$ indicates more spectral

energy at frequencies lower than the mean value, and $\gamma_3 > 0$ indicates more energy at higher frequencies, defined as

$$\gamma_3(t) = \frac{m_3(t)}{\sigma^3(t)}$$

where

$$m_3(t) = \sum_{n=1}^N (f_n - \mu(t))^3 p(t, n)$$

Spectral Slope

Measures the amount of decrease of spectral amplitude with respect to frequency at time window t . It is calculated using the slope coefficient of a simple linear regression, defined as

$$\frac{1}{\sum_{n=1}^N v(t, n)} * \frac{N \sum_{n=1}^N f_n v(t, n) - (\sum_{n=1}^N f_n)(\sum_{n=1}^N v(t, n))}{N \sum_{n=1}^N f_n^2 - (\sum_{n=1}^N f_n)^2}$$

Spectral Variance

Measures spectral spread weighted by the inverse total amplitude.

$$v(t) = \frac{1}{\sum_{n=1}^N p(t, n)} \sum_{n=1}^N [(f_n - \mu(t))^2 p(t, n)]$$

7. Appendix B

This is the master list for soundscape recordings used in this experiment where B represents biophony, G represents geophony, A represents anthrophony, C represents observation code, SEM class represents the class assignment using the sensory evaluation method (SEM), Time represents the time period of the random sample draw, and Site is the location where the recording occurred.

No.	File Name	B	G	A	C	SEM class	Time	Site
1	Arizona_H1_U_T1_H1-NO-R3_20140313_060000	1	1	0	6	geophony	1-dawn	Arizona
2	Arizona_H1_U_T1_H1-NO-R3_20140316_080000	1	1	0	6	geophony	1-dawn	Arizona
3	Arizona_H1_U_T1_H1-NO-R3_20140320_063000	1	1	0	6	geophony	1-dawn	Arizona
4	Arizona_H1_U_T1_H1-NO-R3_20140421_080000	1	1	0	6	geophony	1-dawn	Arizona
5	Arizona_H1_U_T1_H1-NO-R3_20140422_060000	1	1	0	6	geophony	1-dawn	Arizona
6	Arizona_H1_U_T1_H1-NO-R3_20140427_060000	1	1	0	6	geophony	1-dawn	Arizona
7	Arizona_H1_U_T1_H1-NO-R3_20140430_060000	1	1	0	6	geophony	1-dawn	Arizona
8	Arizona_H1_U_T1_H1-NO-R3_20140430_080000	1	1	0	6	geophony	1-dawn	Arizona
9	Arizona_H1_U_T1_H1-NO-R3_20140502_060000	0	1	0	7	anthrophony	1-dawn	Arizona
10	Arizona_H1_U_T1_H1-NO-R3_20140502_063000	1	1	0	6	geophony	1-dawn	Arizona
11	Arizona_H1_U_T1_H1-NO-R3_20140506_070000	1	1	0	6	geophony	1-dawn	Arizona

12	Arizona_HI_U_T1_H1-NO-R3_20140603_060000	1	1	0	6	geophony	1-dawn	Arizona
13	Arizona_HI_U_T1_H1-NO-R3_20140603_063000	1	1	0	6	geophony	1-dawn	Arizona
14	Arizona_HI_U_T1_H1-NO-R3_20140606_080000	0	0	0	4	biophony	1-dawn	Arizona
15	Arizona_HI_U_T1_H1-NO-R3_20140608_060000	1	1	0	6	geophony	1-dawn	Arizona
16	Arizona_HI_U_T1_H1-NO-R3_20140610_063000	1	0	1	7	anthrophony	1-dawn	Arizona
17	Arizona_HI_U_T1_H1-NO-R3_20140612_063000	1	1	0	6	geophony	1-dawn	Arizona
18	Arizona_HI_U_T1_H1-NO-R3_20140901_060000	0	0	0	4	biophony	1-dawn	Arizona
19	Arizona_HI_U_T1_H1-NO-R3_20140905_070000	1	1	0	6	geophony	1-dawn	Arizona
20	Arizona_HI_U_T1_H1-NO-R3_20140914_080000	1	1	0	6	geophony	1-dawn	Arizona
21	Arizona_HI_U_T1_H1-NO-R3_20141003_060000	1	1	0	6	geophony	1-dawn	Arizona
22	Arizona_HI_U_T1_H1-NO-R3_20141003_063000	1	1	0	6	geophony	1-dawn	Arizona
23	Arizona_HI_U_T2_H1-NO-R3_20140306_120000	1	1	1	7	anthrophony	2-midday	Arizona
24	Arizona_HI_U_T2_H1-NO-R3_20140306_130000	1	1	0	6	geophony	2-midday	Arizona
25	Arizona_HI_U_T2_H1-NO-R3_20140310_120000	0	1	0	6	geophony	2-midday	Arizona
26	Arizona_HI_U_T2_H1-NO-R3_20140316_120000	1	1	0	6	geophony	2-midday	Arizona
27	Arizona_HI_U_T2_H1-NO-R3_20140419_110000	0	1	0	6	geophony	2-midday	Arizona
28	Arizona_HI_U_T2_H1-NO-R3_20140420_130000	0	1	0	6	geophony	2-midday	Arizona
29	Arizona_HI_U_T2_H1-NO-R3_20140422_130000	1	1	0	6	geophony	2-midday	Arizona
30	Arizona_HI_U_T2_H1-NO-R3_20140502_130000	1	1	0	6	geophony	2-midday	Arizona
31	Arizona_HI_U_T2_H1-NO-R3_20140504_130000	1	1	0	6	geophony	2-midday	Arizona
32	Arizona_HI_U_T2_H1-NO-R3_20140905_110000	1	1	1	7	anthrophony	2-midday	Arizona
33	Arizona_HI_U_T2_H1-NO-R3_20140906_110000	0	1	1	7	anthrophony	2-midday	Arizona
34	Arizona_HI_U_T2_H1-NO-R3_20140914_120000	1	1	0	6	geophony	2-midday	Arizona
35	Arizona_HI_U_T2_H1-NO-R3_20140918_110000	1	1	0	6	geophony	2-midday	Arizona
36	Arizona_HI_U_T2_H1-NO-R3_20141003_130000	0	1	0	6	geophony	2-midday	Arizona
37	Arizona_HI_U_T3_H1-NO-R3_20140320_180000	1	1	0	6	geophony	3-dusk	Arizona
38	Arizona_HI_U_T3_H1-NO-R3_20140425_180000	1	1	0	6	geophony	3-dusk	Arizona
39	Arizona_HI_U_T3_H1-NO-R3_20140505_190000	1	1	0	6	geophony	3-dusk	Arizona
40	Arizona_HI_U_T3_H1-NO-R3_20140601_180000	1	1	0	6	geophony	3-dusk	Arizona
41	Arizona_HI_U_T3_H1-NO-R3_20140603_180000	1	1	0	6	geophony	3-dusk	Arizona
42	Arizona_HI_U_T3_H1-NO-R3_20140906_190000	0	0	0	4	biophony	3-dusk	Arizona
43	Arizona_HI_U_T3_H1-NO-R3_20140908_200000	0	0	0	4	biophony	3-dusk	Arizona
44	Arizona_HI_U_T3_H1-NO-R3_20140914_180000	1	1	0	6	geophony	3-dusk	Arizona
45	Arizona_HI_U_T3_H1-NO-R3_20140917_180000	1	1	0	6	geophony	3-dusk	Arizona
46	Arizona_HI_U_T3_H1-NO-R3_20140917_200000	1	1	0	6	geophony	3-dusk	Arizona
47	Arizona_HI_U_T4_H1-NO-R3_20140320_010000	1	1	0	6	geophony	4-late evening	Arizona
48	Arizona_HI_U_T4_H1-NO-R3_20140419_010000	1	1	0	6	geophony	4-late evening	Arizona
49	Arizona_HI_U_T4_H1-NO-R3_20140419_230000	1	1	0	6	geophony	4-late evening	Arizona
50	Arizona_HI_U_T4_H1-NO-R3_20140425_000000	1	1	0	6	geophony	4-late evening	Arizona
51	Arizona_HI_U_T4_H1-NO-R3_20140502_230000	1	1	0	6	geophony	4-late evening	Arizona
52	Arizona_HI_U_T4_H1-NO-R3_20140601_000000	1	1	0	6	geophony	4-late evening	Arizona
53	Arizona_HI_U_T4_H1-NO-R3_20140605_010000	1	0	0	4	biophony	4-late evening	Arizona
54	Arizona_HI_U_T4_H1-NO-R3_20140609_230000	0	1	0	6	geophony	4-late evening	Arizona
55	Arizona_HI_U_T4_H1-NO-R3_20140612_230000	0	0	0	4	biophony	4-late evening	Arizona
56	Arizona_HI_U_T4_H1-NO-R3_20140619_000000	0	0	0	4	biophony	4-late evening	Arizona
57	Arizona_HI_U_T4_H1-NO-R3_20140916_010000	0	0	0	4	biophony	4-late evening	Arizona
58	Arizona_HI_U_T4_H1-NO-R3_20141002_010000	1	0	0	4	biophony	4-late evening	Arizona
59	Arizona_HI_U_T4_H1-NO-R3_20141006_000000	1	0	0	4	biophony	4-late evening	Arizona
60	Arizona_HI_U_T4_H1-NO-R3_20141007_010000	0	0	0	4	biophony	4-late evening	Arizona

61	Laselva_exswamp_2011_T1_015087_20150118_083000	1	0	0	6	geophony	1-dawn	Costa Rica
62	Laselva_exswamp_2011_T1_EXPSWAMP_20110813_074523	1	0	0	4	biophony	1-dawn	Costa Rica
63	Laselva_exswamp_2011_T1_EXPSWAMP_20110813_083023	0	0	0	4	biophony	1-dawn	Costa Rica
64	Laselva_exswamp_2011_T1_EXPSWAMP_20110817_060000	1	0	0	4	biophony	1-dawn	Costa Rica
65	Laselva_exswamp_2011_T1_EXPSWAMP_20110825_080000	0	0	0	4	biophony	1-dawn	Costa Rica
66	Laselva_exswamp_2011_T1_EXPSWAMP_20110826_061500	0	0	0	4	biophony	1-dawn	Costa Rica
67	Laselva_exswamp_2011_T1_EXPSWAMP_20110827_064500	1	0	0	4	biophony	1-dawn	Costa Rica
68	Laselva_exswamp_2011_T1_EXPSWAMP_20110831_070000	1	0	0	4	biophony	1-dawn	Costa Rica
69	Laselva_exswamp_2011_T1_EXPSWAMP_20110904_071508	1	0	0	4	biophony	1-dawn	Costa Rica
70	Laselva_exswamp_2011_T1_EXPSWAMP_20110908_073000	1	0	0	7	anthrophony	1-dawn	Costa Rica
71	Laselva_exswamp_2011_T1_EXPSWAMP_20110909_073000	0	0	0	4	biophony	1-dawn	Costa Rica
72	Laselva_exswamp_2011_T1_EXPSWAMP_20110922_080000	0	0	0	4	biophony	1-dawn	Costa Rica
73	Laselva_exswamp_2011_T1_EXPSWAMP_20110926_071501	0	0	1	7	anthrophony	1-dawn	Costa Rica
74	Laselva_exswamp_2011_T1_EXPSWAMP_20110927_074502	0	0	0	4	biophony	1-dawn	Costa Rica
75	Laselva_exswamp_2011_T1_EXPSWAMP_20110927_080002	0	0	0	4	biophony	1-dawn	Costa Rica
76	Laselva_exswamp_2011_T1_EXPSWAMP_20110929_074500	1	0	0	4	biophony	1-dawn	Costa Rica
77	Laselva_exswamp_2011_T1_EXPSWAMP_20111003_060000	0	1	1	7	anthrophony	1-dawn	Costa Rica
78	Laselva_exswamp_2011_T1_EXPSWAMP_20111003_073000	1	1	0	6	geophony	1-dawn	Costa Rica
79	Laselva_exswamp_2011_T2_015087_20150113_123000	1	1	0	6	geophony	2-midday	Costa Rica
80	Laselva_exswamp_2011_T2_015087_20150115_113000	1	0	1	6	geophony	2-midday	Costa Rica
81	Laselva_exswamp_2011_T2_015087_20150115_130000	1	0	1	6	geophony	2-midday	Costa Rica
82	Laselva_exswamp_2011_T2_015087_20150117_123000	1	1	0	6	geophony	2-midday	Costa Rica
83	Laselva_exswamp_2011_T2_EXPSWAMP_20110812_120020	1	1	0	6	geophony	2-midday	Costa Rica
84	Laselva_exswamp_2011_T2_EXPSWAMP_20110831_124500	0	0	0	4	biophony	2-midday	Costa Rica
85	Laselva_exswamp_2011_T2_EXPSWAMP_20110902_134503	1	1	0	6	geophony	2-midday	Costa Rica
86	Laselva_exswamp_2011_T2_EXPSWAMP_20110903_121506	0	0	1	7	anthrophony	2-midday	Costa Rica
87	Laselva_exswamp_2011_T2_EXPSWAMP_20110904_113009	0	0	0	4	biophony	2-midday	Costa Rica
88	Laselva_exswamp_2011_T2_EXPSWAMP_20110909_130000	1	1	0	6	geophony	2-midday	Costa Rica
89	Laselva_exswamp_2011_T2_EXPSWAMP_20110923_110000	1	0	0	6	geophony	2-midday	Costa Rica
90	Laselva_exswamp_2011_T2_EXPSWAMP_20110930_110000	1	1	0	6	geophony	2-midday	Costa Rica
91	Laselva_exswamp_2011_T2_EXPSWAMP_20111001_110000	0	0	0	4	biophony	2-midday	Costa Rica
92	Laselva_exswamp_2011_T2_EXPSWAMP_20111004_134500	0	0	0	4	biophony	2-midday	Costa Rica
93	Laselva_exswamp_2011_T3_015087_20150115_193000	0	1	1	4	biophony	3-dusk	Costa Rica
94	Laselva_exswamp_2011_T3_015087_20150118_180000	0	1	0	4	biophony	3-dusk	Costa Rica
95	Laselva_exswamp_2011_T3_EXPSWAMP_20110810_191515	0	0	0	4	biophony	3-dusk	Costa Rica
96	Laselva_exswamp_2011_T3_EXPSWAMP_20110811_191518	0	0	0	4	biophony	3-dusk	Costa Rica
97	Laselva_exswamp_2011_T3_EXPSWAMP_20110812_203021	0	0	0	4	biophony	3-dusk	Costa Rica
98	Laselva_exswamp_2011_T3_EXPSWAMP_20110813_204526	1	1	0	6	geophony	3-dusk	Costa Rica
99	Laselva_exswamp_2011_T3_EXPSWAMP_20110818_180004	0	0	0	4	biophony	3-dusk	Costa Rica
100	Laselva_exswamp_2011_T3_EXPSWAMP_20110820_191500	0	0	0	4	biophony	3-dusk	Costa Rica
101	Laselva_exswamp_2011_T3_EXPSWAMP_20110821_183000	0	0	0	4	biophony	3-dusk	Costa Rica
102	Laselva_exswamp_2011_T3_EXPSWAMP_20110821_204500	0	0	0	4	biophony	3-dusk	Costa Rica
103	Laselva_exswamp_2011_T3_EXPSWAMP_20110823_180006	1	1	0	6	geophony	3-dusk	Costa Rica
104	Laselva_exswamp_2011_T3_EXPSWAMP_20110827_194501	0	0	0	4	biophony	3-dusk	Costa Rica
105	Laselva_exswamp_2011_T3_EXPSWAMP_20110828_184503	0	0	0	4	biophony	3-dusk	Costa Rica
106	Laselva_exswamp_2011_T3_EXPSWAMP_20110828_203004	0	0	0	4	biophony	3-dusk	Costa Rica
107	Laselva_exswamp_2011_T3_EXPSWAMP_20110831_183000	1	0	0	4	biophony	3-dusk	Costa Rica
108	Laselva_exswamp_2011_T3_EXPSWAMP_20110904_180010	1	0	0	4	biophony	3-dusk	Costa Rica
109	Laselva_exswamp_2011_T3_EXPSWAMP_20110926_194502	1	0	0	4	biophony	3-dusk	Costa Rica

110	Laselva_exswamp_2011_T3_EXPSWAMP_20110929_190000	0	0	1	4	biophony	3-dusk	Costa Rica
111	Laselva_exswamp_2011_T4_015087_20150119_013000	1	1	0	6	geophony	4-late evening	Costa Rica
112	Laselva_exswamp_2011_T4_EXPSWAMP_20110811_004516	1	0	0	4	biophony	4-late evening	Costa Rica
113	Laselva_exswamp_2011_T4_EXPSWAMP_20110811_230019	1	0	0	4	biophony	4-late evening	Costa Rica
114	Laselva_exswamp_2011_T4_EXPSWAMP_20110815_004500	0	1	0	4	biophony	4-late evening	Costa Rica
115	Laselva_exswamp_2011_T4_EXPSWAMP_20110820_010000	1	0	0	4	biophony	4-late evening	Costa Rica
116	Laselva_exswamp_2011_T4_EXPSWAMP_20110824_010007	0	0	0	4	biophony	4-late evening	Costa Rica
117	Laselva_exswamp_2011_T4_EXPSWAMP_20110904_013007	1	1	0	6	geophony	4-late evening	Costa Rica
118	Laselva_exswamp_2011_T4_EXPSWAMP_20110922_004500	0	0	0	4	biophony	4-late evening	Costa Rica
119	Laselva_exswamp_2011_T4_EXPSWAMP_20110923_011500	0	0	0	4	biophony	4-late evening	Costa Rica
120	Laselva_exswamp_2011_T4_EXPSWAMP_20110925_231500	0	0	0	4	biophony	4-late evening	Costa Rica
121	Wells_PilgerPoint_U_T1_015100_20140505_073000	1	0	1	7	anthrophony	1-dawn	Maine
122	Wells_PilgerPoint_U_T1_015100_20140530_080000	1	0	0	4	biophony	1-dawn	Maine
123	Wells_PilgerPoint_U_T1_015100_20140531_060000	1	0	0	4	biophony	1-dawn	Maine
124	Wells_PilgerPoint_U_T1_015100_20140625_083000	1	0	0	4	biophony	1-dawn	Maine
125	Wells_PilgerPoint_U_T1_015100_20140812_080000	0	0	1	7	anthrophony	1-dawn	Maine
126	Wells_PilgerPoint_U_T1_015100_20140819_073000	1	1	1	7	anthrophony	1-dawn	Maine
127	Wells_PilgerPoint_U_T1_015100_20140827_073000	1	0	1	7	anthrophony	1-dawn	Maine
128	Wells_PilgerPoint_U_T1_015100_20140901_073000	1	0	0	4	biophony	1-dawn	Maine
129	Wells_PilgerPoint_U_T1_015100_20140911_083000	1	1	0	6	geophony	1-dawn	Maine
130	Wells_PilgerPoint_U_T1_015100_20140918_073000	1	0	1	7	anthrophony	1-dawn	Maine
131	Wells_PilgerPoint_U_T1_015100_20141001_070000	1	1	0	6	geophony	1-dawn	Maine
132	Wells_PilgerPoint_U_T1_015100_20141001_083000	1	1	1	6	geophony	1-dawn	Maine
133	Wells_PilgerPoint_U_T1_015100_20141009_073000	0	1	1	7	anthrophony	1-dawn	Maine
134	Wells_PilgerPoint_U_T1_015100_20141010_070000	0	1	1	7	anthrophony	1-dawn	Maine
135	Wells_PilgerPoint_U_T1_015100_20141013_070000	0	1	0	7	anthrophony	1-dawn	Maine
136	Wells_PilgerPoint_U_T1_015100_20141014_063000	0	0	1	7	anthrophony	1-dawn	Maine
137	Wells_PilgerPoint_U_T1_015100_20141101_080000	1	1	0	6	geophony	1-dawn	Maine
138	Wells_PilgerPoint_U_T1_015100_20141103_083000	1	0	1	7	anthrophony	1-dawn	Maine
139	Wells_PilgerPoint_U_T2_015100_20140505_120000	0	1	1	7	anthrophony	2-midday	Maine
140	Wells_PilgerPoint_U_T2_015100_20140507_123000	0	1	1	7	anthrophony	2-midday	Maine
141	Wells_PilgerPoint_U_T2_015100_20140508_123000	1	0	1	7	anthrophony	2-midday	Maine
142	Wells_PilgerPoint_U_T2_015100_20140512_120000	1	0	1	7	anthrophony	2-midday	Maine
143	Wells_PilgerPoint_U_T2_015100_20140607_120000	1	0	1	7	anthrophony	2-midday	Maine
144	Wells_PilgerPoint_U_T2_015100_20140826_133000	0	0	1	7	anthrophony	2-midday	Maine
145	Wells_PilgerPoint_U_T2_015100_20140903_110000	0	0	0	7	anthrophony	2-midday	Maine
146	Wells_PilgerPoint_U_T2_015100_20140904_113000	1	0	0	7	anthrophony	2-midday	Maine
147	Wells_PilgerPoint_U_T2_015100_20140905_113000	0	0	1	7	anthrophony	2-midday	Maine
148	Wells_PilgerPoint_U_T2_015100_20140912_113000	1	0	1	7	anthrophony	2-midday	Maine
149	Wells_PilgerPoint_U_T2_015100_20141004_130000	0	1	1	7	anthrophony	2-midday	Maine
150	Wells_PilgerPoint_U_T2_015100_20141010_110000	0	0	1	7	anthrophony	2-midday	Maine
151	Wells_PilgerPoint_U_T2_015100_20141029_113000	0	0	1	7	anthrophony	2-midday	Maine
152	Wells_PilgerPoint_U_T2_015100_20141101_130000	1	1	1	6	geophony	2-midday	Maine
153	Wells_PilgerPoint_U_T3_015100_20140505_180000	1	1	1	7	anthrophony	3-dusk	Maine
154	Wells_PilgerPoint_U_T3_015100_20140507_200000	0	0	1	7	anthrophony	3-dusk	Maine
155	Wells_PilgerPoint_U_T3_015100_20140512_203000	1	0	0	4	biophony	3-dusk	Maine
156	Wells_PilgerPoint_U_T3_015100_20140607_203000	1	1	0	7	anthrophony	3-dusk	Maine
157	Wells_PilgerPoint_U_T3_015100_20140620_193000	0	1	0	7	anthrophony	3-dusk	Maine
158	Wells_PilgerPoint_U_T3_015100_20140626_183000	0	1	1	7	anthrophony	3-dusk	Maine

159	Wells_PilgerPoint_U_T3_015100_20140810_183000	1	1	0	6	geophony	3-dusk	Maine
160	Wells_PilgerPoint_U_T3_015100_20140813_190000	0	1	0	6	geophony	3-dusk	Maine
161	Wells_PilgerPoint_U_T3_015100_20140815_193000	1	1	0	6	geophony	3-dusk	Maine
162	Wells_PilgerPoint_U_T3_015100_20140819_183000	0	1	0	6	geophony	3-dusk	Maine
163	Wells_PilgerPoint_U_T3_015100_20140824_193000	0	1	1	7	anthrophony	3-dusk	Maine
164	Wells_PilgerPoint_U_T3_015100_20140826_203000	0	0	1	7	anthrophony	3-dusk	Maine
165	Wells_PilgerPoint_U_T3_015100_20140829_193000	0	1	1	7	anthrophony	3-dusk	Maine
166	Wells_PilgerPoint_U_T3_015100_20140830_193000	0	1	1	7	anthrophony	3-dusk	Maine
167	Wells_PilgerPoint_U_T3_015100_20140906_203000	1	0	0	4	biophony	3-dusk	Maine
168	Wells_PilgerPoint_U_T3_015100_20140916_190000	0	1	1	7	anthrophony	3-dusk	Maine
169	Wells_PilgerPoint_U_T3_015100_20141007_183000	0	1	0	7	anthrophony	3-dusk	Maine
170	Wells_PilgerPoint_U_T3_015100_20141013_203000	1	1	1	7	anthrophony	3-dusk	Maine
171	Wells_PilgerPoint_U_T4_015100_20140606_233000	1	1	0	6	geophony	4-late evening	Maine
172	Wells_PilgerPoint_U_T4_015100_20140626_003000	1	1	0	6	geophony	4-late evening	Maine
173	Wells_PilgerPoint_U_T4_015100_20140721_230000	0	1	0	6	geophony	4-late evening	Maine
174	Wells_PilgerPoint_U_T4_015100_20140802_000000	0	1	1	7	anthrophony	4-late evening	Maine
175	Wells_PilgerPoint_U_T4_015100_20140806_013000	0	1	0	6	geophony	4-late evening	Maine
176	Wells_PilgerPoint_U_T4_015100_20140822_010000	1	0	1	4	biophony	4-late evening	Maine
177	Wells_PilgerPoint_U_T4_015100_20140914_003000	0	1	0	6	geophony	4-late evening	Maine
178	Wells_PilgerPoint_U_T4_015100_20140921_233000	1	0	0	4	biophony	4-late evening	Maine
179	Wells_PilgerPoint_U_T4_015100_20140925_233000	1	0	1	7	anthrophony	4-late evening	Maine
180	Wells_PilgerPoint_U_T4_015100_20140928_003000	0	0	1	7	anthrophony	4-late evening	Maine

8. Acknowledgements

KMB was supported by NSF Advancing Informal STEM Learning (1323615) grant to Pijanowski (PI) and a Purdue University Graduate School's Bilslund Dissertation Fellowship. BCP was supported by Purdue University's Executive Office of the Vice President for Research and Engagement, the Department of Forestry and Natural Resources Wright Fund and the USDA NIFA Hatch/McIntire Stennis Accession Number: 1016730. AG was partially funded by the NSF AISL program (#1323615). ZZ's work was supported by the National Natural Science Foundation of P.R. China [#61401203]; and the State Scholarship Fund of China [#201606840023]. JV was supported by an NSF undergraduate research fellowship. HO was funded by the Luxembourg Institute of Socio-Economic Research (LISER) Institute.

9. References

- Agostini, G., Longari, M., & Pollastri, E. (2003). Musical instrument timbres classification with spectral features. *EURASIP Journal on Advances in Signal Processing*, 2003(1)943279. doi:10.1155/S1110865703210118
- Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, Alvarez R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1:e103 <https://doi.org/10.7717/peerj.103>.
- Anderson, M. J. (2006). Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics*, 62(1), 245–253. doi:10.1111/j.1541-0420.2005.00440.x
- Beauchamp, J. W. (2007). *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 1–89). New York, NY: Springer.
- Blagus, R., & Lusa, L. (2010). Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 11(1), 523. doi:10.1186/1471-2105-11-523
- Boelman, N. T., Asner, G. P., Hart, P. J., & Martin, R. E. (2007). Multi-Trophic invasion resistance in Hawaii: bioacoustics, field surveys, and airborne remote sensing. *Ecological Applications*, 17(8), 2137–2144. doi:10.1890/07-0004.1
- Bojanowski M and Edwards R (2016). *alluvial: R Package for Creating Alluvial Diagrams*. R package version: 0.1-2, <https://github.com/mbojan/alluvial>.
- Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology in R*. Springer, New York, NY.
- Bormpoudakis, D., Sueur, J., & Pantis, J. D. (2013). Spatial heterogeneity of ambient sound at the habitat type level: ecological implications and applications. *Landscape Ecology*, 28(3), 495–506. doi:10.1007/s10980-013-9849-1
- Brumm, H., & Slabbekoorn, H. (2005). Acoustic Communication in Noise. *Advances in the Study of Behavior* 35, 151-209. doi:10.1016/S0065-3454(05)350004-2
- Bullock, J., & Conservatoire, U. C. E. B. (2007, August). Libxtract: a Lightweight Library for audio Feature Extraction. *In ICMC*.
- Cannam, C., Landone, C., & Sandler, M. (2010, October). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1467–1468). ACM. doi:10.1145/1873951.1874248
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4(7), 675–683. doi:10.1111/2041-210X.12060
- Downie, J.S. (2003). Music information retrieval. *Annual review of information science and technology*, 37(1), 295-340. doi:10.1002/aris.1440370108
- Eldridge, A., Casey, M., Moscoso, P., & Peck, M. (2016). A new method for ecoacoustics? Toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. *PeerJ*, 4, e2108. doi:10.7717/peerj.2108

- Fairbrass, A. J., Rennett, P., Williams, C., Titheridge, H., & Jones, K. E. (2017). Biases of acoustic indices measuring biodiversity in urban areas. *Ecological Indicators*, *83*, 169–177. doi:10.1016/j.ecolind.2017.07.064
- Farina, A., & Pieretti, N. (2014). Sonic environment and vegetation structure: a methodological approach for a soundscape analysis of a Mediterranean maqui. *Ecological Informatics*, *21*, 120–132. doi:10.1016/j.ecoinf.2013.10.008
- Fuller, S., Axel, A. C., Tucker, D., & Gage, S. H. (2015). Connecting soundscape to landscape: Which acoustic index best describes landscape configuration?. *Ecological Indicators*, *58*, 207–215. doi:10.1016/j.ecolind.2015.05.057
- Grosche, P., Müller, M., & Kurth, F. (2010, March). Cyclic tempogram—a mid-level tempo representation for music signals. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (pp. 5522-5525). IEEE. doi:10.1109/ACASSP.2010.5495219
- Gutiérrez, E. G. (2006). *Tonal description of music audio signals*. Universitat Pompeu Fabra.
- Hand, D.J. & Till, R.J. (2001). A simple generalization of the area under the ROC curve for multiple class classification problems. *Machine learning*, *45*(2), 171-186. doi.org/10.1023/A:1010920819831
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, *66*(1), 51–83. doi:10.1109/PROC.1978.10837
- Herrera, P., Amatriain, X., Batlle, E., & Serra, X. (2000). Towards Instrument Segmentation for Music Content Description: a Critical Review of Instrument Classification Techniques. In *ISMIR* (Vol. 1, pp. 115-119).
- Hofmann, C., Wolff, T., Buck, M., Haulick, T., & Kellermann, W. (2012, September). A morphological approach to single-channel wind-noise suppression. In *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, pp. 1-4. VDE.
- 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. *Ecological Informatics*, *12*, 50–67. doi:10.1016/j.ecoinf.2012.08.001
- Keen, S., Ross, J. C., Griffiths, E. T., Lanzone, M., & Farnsworth, A. (2014). A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). *Ecological Informatics*, *21*, 25–33. doi:10.1016/j.ecoinf.2014.01.001
- Kitamura, O. (1988). The content of “timbre.” *The Journal of the Acoustical Society of America*, *84*(S1), S143–S143. doi:10.1121/1.2025835
- Krimphoff, J., McAdams, S., Winsberg, S., Petit, H., Bakchine, S., Dubois, B., ... Vellas, B. (1994). Characterization of the timbre of complex sounds. 2. Acoustic analysis and psychophysical quantification. *J. de Physique*, *4*, 625–628.
- Kuhn, M. (2016). Contributions from J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, & C. Candan. Caret: Classification and Regression Training. R package version 6.0-71. <https://CRAN.R-project.org/package=caret>.

- Lahoulou, A., Larabi, M. C., Beghdadi, A., Viennet, E., & Bouridane, A. (2016). Knowledge-based Taxonomic Scheme for Full-Reference Objective Image Quality Measurement Models. *Journal of Imaging Science and Technology*, 60(6), 60406-1. doi:10.2352/J.ImagingSci.Technol.2016.60.6.060406
- Li, T., & Ogihara, M. (2006). Toward intelligent music information retrieval. *IEEE Transactions on Multimedia*, 8(3), 564–574. doi:10.1109/TMM.2006.870730
- Liu, X, Derudder, B., & Taylor, O. (2013). Mapping the evolution of hierarchical and regional tendencies in the world city network, 2000-2010. *Computers, environment and urban systems*, 43, 51-66.
- Lokki, T., Pätynen, J., Kuusinen, A., & Tervo, S. (2012). Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *The Journal of the Acoustical Society of America*, 132(5), 3148–3161. doi:10.1121/1.4756826
- 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. *Ecological Indicators*, 82, 470–477. doi:10.1016/j.ecolind.2017.07.017
- Marler, P. (1967). Animal Communication Signals: We are beginning to understand how the structure of animal signals relates to the function they serve. *Science*, 157(3790), 769–774. doi:10.1126/science.157.3790.769
- Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, 114(5), 2946–2957. doi:10.1121/1.1618239
- McArdle, B. H., & Anderson, M. J. (2001). Fitting Multivariate Models to Community Data: a comment on distance-based redundancy analysis. *Ecology*, 82(1), 290–297. doi:10.1890/0012-9658(2001)082[0290:FMMTCD] 2.0.CO;2
- Meilgaard, M. C., Carr, B. T., & Civille, G. V. (2006). Pairwise ranking test: Friedman analysis-comparing several samples in all possible pairs. *Sensory evaluation techniques*, 109.
- Mitrovic, D., Zeppelzauer, M., & Eidenberger, H. (2007). Analysis of the data quality of audio descriptions of environmental sounds. *Journal of Digital Information Management*, 5(2), 48.
- Muller, M., Ellis, D. P. W., Klapuri, A., & Richard, G. (2011). Signal Processing for Music Analysis. *IEEE Journal of Selected Topics in Signal Processing*, 5(6), 1088–1110. doi:10.1109/JSTSP.2011.2112333
- Oksanen, Jari. (2015). Vegan: an introduction to ordination. URL <http://cran.r-project.org/web/packages/vegan/vignettes/introvegan.pdf>
- Omrani, H., Abdallah, F., Charif, O., & Longford, N. T. (2015). Multi-label class assignment in land-use modelling. *International Journal of Geographical Information Science*, 29(6), 1023-1041. doi.org:10.1080/13658816.2015.1008004
- Omrani, H., Tayyebi, A., & Pijanowski, B. (2017). Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: an integrated ML-CA-LTM modeling framework. *GIScience & Remote Sensing*, 54(3), 283–304. doi:10.1080/15481603.2016.1265706

- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902–2916. doi:10.1121/1.3642604
- Pekin, B. K., Jung, J., Villanueva-Rivera, L. J., Pijanowski, B. C., & Ahumada, J. A. (2012). Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. *Landscape Ecology*, *27*(10), 1513–1522. doi:10.1007/s10980-012-9806-4
- Pieretti, N., & Farina, A. (2013). Application of a recently introduced index for acoustic complexity to an avian soundscape with traffic noise. *The Journal of the Acoustical Society of America*, *134*(1), 891–900. doi:10.1121/1.4807812
- 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 Ecology*, *26*(9), 1213–1232. doi:10.1007/s10980-011-9600-8
- Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., ...Pieretti, N. (2011). Soundscape Ecology: the science of sound in the landscape. *BioScience*, *61*(3), 203–216. doi:10.1525/bio.2011.61.3.6
- Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Retrieved from <https://dspace.flinders.edu.au/xmlui/handle/2328/27165>
- Schouten, Jan Frederik. (1968). The perception of timbre.
- Soares, C. G., & Cherneva, Z. (2005). Spectrogram analysis of the time–frequency characteristics of ocean wind waves. *Ocean Engineering*, *32*(14), 1643–1663. doi:10.1016/j.oceaneng.2005.02.008
- Sueur J., Aubin T, & Simonis C. (2008). Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*(2), 213–226. doi:10.1080/09524622.2008.9753600
- 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*(4), 772–781. doi:10.3813/AAA.918757
- Team, R. C. (2013). R: A language and environment for statistical computing.
- Towsey, M., Parsons, S., & Sueur, J. (2014). Ecology and acoustics at a large scale. *Ecological Informatics*, *21*, 1–3. doi:10.1016/j.ecoinf.2014.02.002
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. In *IEEE Transactions on Speech and Audio Processing* *10.5*: 293–302
- Venables, W. N. & Ripley. B.D., 2002. Modern Applied Statistics with S. *New York: Springer Science & Business Media* *200*, 183–206.
- Villanueva-Rivera, L. J., Pijanowski, B.C., Doucette, J., & Pekin, B. (2011). A primer of acoustic analysis for landscape ecologists. *Landscape Ecology*, *26*(9), 1233. doi:10.1007/s10980-011-9636-9
- Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human Domination of Earth's Ecosystems. *Science*, *277*(5325), 494–499. doi:10.1126/science.277.5325.494

- von Hornbostel, E. M., & Sachs, C. (1914). Systematik der Musikinstrumente. Ein Versuch. *Zeitschrift Für Ethnologie*, 46(H. 4/5), 553–590
- Wickham H (2016). **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York. ISBN 978-3-319-24277-4, <http://ggplot2.org>.
- Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wickham H and Francois R (2016). dplyr: A Grammar of Data Manipulation.
- Wiley, R. H., & Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology*, 3(1), 69–94. doi:10.1007/BF00300047
- Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, p. 77. doi:10.1186/1471-2105-12-77. <http://www.biomedcentral.com/1471-2105/12/77/>.
- Xue, J.-H., & Titterton, D. M. (2008). Do unbalanced data have a negative effect on LDA? *Pattern Recognition*, 41(5), 1558–1571. doi:10.1016/j.patcog.2007.11.008
- Zhang, L., Towsey, M., Xie, J., Zhang, J., & Roe, P. (2016). Using multi-label classification for acoustic pattern detection and assisting bird species surveys. *Applied Acoustics*, 110, 91–98. doi:10.1016/j.apacoust.2016.03.027
- Zhao, Z., Zhang, S. H., Xu, Z. Y., Bellisario, K., Dai, N. H., Omrani, H., & Pijanowski, B. C. (2017). Automated bird acoustic event detection and robust species classification. *Ecological Informatics*, 39, 99–108. doi:10.1016/j.ecoinf.2017.04.003

Contributions of MIR to Soundscape Ecology. Part 2: Spectral timbral analysis for discrimination of soundscape components

Highlights

- Spectral timbral features used in combination together can differentiate between major soundscape classes
- Classification method distinguishes between classes without template matching or species identification
- Sifting method from resulting discriminant analysis feature groups provides new direction for big data sorting