A peer-reviewed version of this preprint was published in PeerJ on 30 June 2016.

View the peer-reviewed version (peerj.com/articles/2108), which is the preferred citable publication unless you specifically need to cite this preprint.

A New Direction for Soundscape Ecology? 
Toward the Extraction and Evaluation of Ecologically-Meaningful Soundscape Objects using Sparse Coding Methods

Alice Eldridge¹, Michael Casey², Paola Moscoso³, and Mika Peck⁴

¹,³,⁴Dept. Evolution, Behaviour and Environment, University of Sussex, UK
²Depts. of Music and Computer Science, Dartmouth College, US

ABSTRACT

Efficient methods of biodiversity assessment and monitoring are central to ecological research and crucial in conservation management; technological advances in remote acoustic sensing inspire new approaches. In line with the emerging field of Soundscape Ecology [Pijanowski et al., 2011], the acoustic approach is based on the rationale that the ecological processes occurring within a landscape are tightly linked to and reflected in the high-level structure of the patterns of sounds emanating from those landscapes – the soundscape. Rather than attempting to recognise species-specific calls, either manually or automatically, analysis of the high-level structure of the soundscape tackles the problem of diversity assessment at the community (rather than species) level (Pijanowski et al., 2011; Farina, 2014). Preliminary work has attempted to make a case for community-level acoustic indices (e.g. Pieretti et al., 2011; Farina, 2014; Sueur et al., 2008), existing indices provide simple statistical summaries of the frequency or time domain signal. We suggest that under this approach, the opportunity to analyze spectro-temporal structural information is diminished, limiting power both as monitoring and investigative tools. In this paper we consider sparse-coding and source separation algorithms (specifically, shift-invariant probabilistic latent component analysis in 2D) as a means to access and summarize ecologically-meaningful sound objects. In doing so we highlight a possible new approach for understanding and assessing ecologically relevant interactions within the conceptual framework of Soundscape Ecology.

Keywords: Soundscape Ecology, Biodiversity, Sparse Coding, Probabilistic Latent Component Analysis, Unsupervised Learning, Acoustic Niche Hypothesis

1 INTRODUCTION

Biodiversity assessment is a central and urgent task, not only for research in the biological sciences, but also in applied conservation biology, including major multi-lateral initiatives for promoting and protecting biodiversity. At the governmental level biodiversity needs to be incorporated into national accounting by 2020 (Aichi Biodiversity targets A2[1]), and cost effective tools necessary to achieve this remain elusive. Operating within the conceptual and methodological framework of the burgeoning field of Soundscape Ecology, [Pijanowski et al., 2011] Farina [2014] we are interested in the potential use of acoustic indices for interrogating the interactions between the landscape and the communities which live within it, with ultimate utility in conservation as a rapid acoustic biodiversity assessment tool. In this ideation paper we first summarise the foundational premises for the soundscape approach and provide an overview of existing acoustic indices; a key constraint of current indices is identified and a potential new direction for research is suggested.

1.1 Current Methods for Biodiversity Assessment.

Biodiversity refers to the variability among living organisms and the ecological complexes of which they are a part (Buckland et al., 2005) and is seen as an indicator of the ‘health’ of a habitat. It should be noted that within traditional ecology, the question of how best to measure and describe biodiversity

¹http://www.cbd.int/sp/targets/
is contentious. There are no absolute measures, rather indices reflecting difference or change across or between time or space. A key part of biodiversity assessment then, is based on the estimation and monitoring of changes in species composition and abundance in animal communities. There are a plethora of approaches to sampling and analysis (Magurran, 2004; Buckland et al., 2005; Pavoine and Bonsall, 2011), most of which are derived from observations of the richness and abundance of species encountered in a given area at a specified time. In most strategies, there is a trade-off between data quality and quantity. All Taxa Biodiversity Indices (ATBI) or multi-species field studies are desirable theoretically, but in practice surveys are invariably subject to financial constraints that bind decision-makers (Lawton et al., 1998). In ongoing work, we are exploring cost-effective solutions, including remote sensing (camera traps and aerial photography of canopy) and identification of ‘ecological-disturbance indicator species’ (Caro, 2010). Remote sensors are an attractive choice for data collection in that they are noninvasive, scalable in both space and time and remove the bias and cost associated with programs which require either experts (ATBIs, Gewin, 2002) or even non-specialists (Rapid Biodiversity Assessment, Oliver and Beattie, 1993), in situ.

Various forms of remote visual sensing technologies have been explored. Global satellite imaging has been investigated to monitor biophysical characteristics of the earth’s surface by assessing species ranges and richness patterns indirectly (e.g. Wang et al., 2010). These methods are attractive, but rely on expensive equipment, are difficult to adapt to small spatial scales and require a time-consuming validation step. It is possible, for example, to infer valid species-level identification of canopy trees from high-resolution aerial imagery, providing a means of remote sensing to assess forest status (Peck et al., 2012). However, the principal weakness of this and other existing visual remote sensing methods is that they cannot provide direct information on the status of taxa other than plants: they cannot detect ‘silent forests’. The need for innovative remote sensing methods to monitor the status of wildlife remains and acoustic, rather than visual, sensors have many attractive characteristics.

### 1.2 Acoustic Approaches to Biodiversity Assessment

Acoustic surveys have most obvious relevance for the identification of vocal animals. Bird species in particular are of interest as their importance as indicator species of environmental health has been demonstrated in temperate (Gregory and Strien, 2010) and tropical (Peck et al., 2015) climates. One approach is to focus on automatic species call identification, but current methods are far from reliable (e.g. Skowronski and Harris, 2006) for bats), increasingly difficult in complex environments such as tropical forest soundscapes, where tens of signals mix and many species still remain unknown (Riede, 1993) and notoriously difficult to generalize across locations due to natural geographic variation in species’ calls (Towsey et al., 2013).

Rather than focusing on individual species, there is a growing interest in monitoring community structure within the emerging field of Soundscape Ecology (Pijanowski et al., 2011) in which systematic interactions between animals, humans and their environment are studied at the landscape level. From this emerging perspective, the landscape’s acoustic signature – the soundscape – is seen as a unique component in the evaluation of its function, and therefore potential indicator of its status (Krause, 1987; Schafer, 1977). By extension, vocalising species establish an acoustic community when they sing at the same time at a particular place. The potential for estimation of acoustic community dynamics as key to understanding what drives changes in community composition and species abundance is being recognised (Lellouch et al., 2014).

This relationship can be understood in evolutionary terms: the same competitive forces which drive organisms to partition and therefore structure dimensions of their shared biophysical environment (food supply, nesting locations etc.) apply in the shared sonic environment; the soundscape is seen as a finite resource in which organisms (including humans) compete for spectro-temporal space. These ideas were first explicitly captured in Krause’s Acoustic Niche Hypotheses (ANH) (Krause, 1987) which suggests that vocalising organisms have evolved to occupy unique spectro-temporal ‘niches’, minimising competition and optimising intraspecific communication mechanisms. Formulated following countless hours recording in pristine habitats, Krause goes so far as to posit that this spectro-temporal partitioning structures the global soundscape, such that the global compositional structure is indicative of the ‘health’ of a habitat. Crudely put, in ancient, stable ecosystems, the soundscape will comprise a complex of non-overlapping signals well dispersed across spectro-temporal niches; a newly devastated area would be characterised by gaps in the spectro-temporal structure; and an area of regrowth may comprise competing, overlapping
signals due to invasive species.

Krause’s ANH can be understood in terms of several theories of the evolution of bird species, which are supported by field studies. Avian mating signals are thought to diverge via several processes: (1) as a by-product of morphological adaptation, the Morphological Adaptation Hypothesis; (2) through direct adaptation to physical features of the signaling environment, the Acoustic Adaptation Hypothesis; and (3) to facilitate species recognition, the Species Recognition Hypothesis. Field studies of the Neotropical suboscine antbird (Thamnophilidae) provide direct evidence that species recognition and ecological adaptation operate in tandem, and that the interplay between these factors drives the evolution of mating signals in suboscine birds (Seddon 2005). The ANH is tenable in evolutionary terms, but to date we have lacked the tools for any serious experimental investigation of exactly which dimensions of acoustic ecospace niches may occupy.

1.3 Existing Acoustic Indices

This emerging framework, coupled with the technical feasibility of remote acoustic sensing and pressure to meet strategic biodiversity targets, fuels a growing research interest in ecological applications of acoustic indices; several dozen have been proposed over the last 6 years (see Sueur et al. 2014; Towsey et al. 2013; Lellouch et al. 2014, for good overviews). These are predominantly derived from statistical summaries of Fast Fourier Transforms (FFTs) or time wave representations of soundscape recordings and are motivated by slightly different approaches to measuring the ‘health’ of a habitat or species diversity or abundance.

The simplest indices provide summaries of the Sound Pressure Level (e.g. peaks, or specific times of day). In Rodriguez et al. (2013), for example, root mean square values of raw signals from a network of recorders are used to create maps of amplitude variation to reveal spatiotemporal dynamics in a neotropical forest. The predominant approach, however, is to consider amplitude variation in time or magnitude differences between frequency bands of a spectrogram; a selection of these indices are described below.

Under the assumption that anthropogenic noise contribution is band-limited to a frequency range (anthrophony: 2 – 8 kHz) below that of the biological world (biophony: 0.2 – 2 kHz), the Normalized Difference Soundscape Index (NDSI) (Kasten et al. 2012) seeks to describe the ‘health’ of the habitat in terms of the level of anthropogenic disturbance by calculating the ratio (biophony - anthrophony) / (biophony + anthrophony). In long term studies, the NDSI has been shown to reflect assumed seasonal and diurnal variation in a landscape and may prove useful for observing high level, long term interactions between animals and human populations (Kasten et al. 2012). However, it does not give an estimation of local diversity within the range of biophony, or provide a means to investigate short term interactions in detail. Further, assumptions about frequency ranges may not generalize. For example in non-industrialized tropical climes (arguably the most precious in ecological terms) animals vocalize outside the 2-8 kHz range (Sueur et al. 2014), and industrial anthrophony is minimal.

A range of entropy indices are based on the assumption that the acoustic output of a community will increase in complexity with the number of singing individuals and species. A summary of the complexity of the sound is assumed to give a proxy of animal acoustic activity. Complexity here is used as a synonym of heterogeneity and many indices derive from classical ecological biodiversity indices. Shannon Entropy (Shannon and Weaver, 1949) (Equation 1) is favoured by ecologists as a measure of species diversity, where \( p_i \) is the proportion of individuals belonging to the \( i \)th species in the data set of interest; it quantifies the uncertainty in predicting the species identity of an individual that is taken at random from the dataset.

\[
H' = - \sum_{i=1}^{R} p_i \ln p_i
\]

The Acoustic Diversity Index (Villanueva-Rivera et al. 2011) (ADI) is a spectral entropy measure which summarises the distribution of the proportion of signals across the spectrum. The FFT spectrogram is divided into a number of bins (default 10) the proportion of the signals in each bin above a threshold (default = 50 dBFS) is calculated. The Shannon Index (Eq 1) is then applied, where \( p_i \) is the fraction of sound in each \( i \)th of \( R \) frequency bands. An evenness metric, the Acoustic Evenness Index (AEI) is similarly derived by calculating the Gini index (Gini 1912) (commonly used by ecologists to estimate species evenness) on the spectrum. These relatively simple indices are shown to effectively reflect observed distinctions in gross acoustic activity, for example between dawn choruses and night activity, or between diverse habitats (mature oak forest, secondary forest, wetland and agricultural land).
The Acoustic Entropy Index, $H$ [Sueur et al., 2008] is also calculated as the product of spectral ($s_h$) and temporal ($t_h$) entropies, calculated on the mean spectrum and Hilbert amplitude envelope of a time wave respectively. $H$ ranges from 0 for pure tones to 1 for high-energy, evenly distributed sound. The index was first tested against simulated choruses, generated by mixing together samples of avian vocalisations and systematically varying the number of species in each track. $H$ values increased with species richness $S$ following a logarithmic model. Field trials were carried out in pristine and degraded African coastal forests and $H$ was shown to reflect assumed variation in species richness. [Sueur et al., 2008]. The study was in an area where animal acoustic activity was high and background noise low.

When background noise (such as traffic) or broadband signals (such as rain, cicada or tropical cricket choruses) are higher, spectral entropy measures may give counter-intuitive results (values for white noise or motorway traffic, for example would approach 1).

The spectral indices provide a statistical summary of the distribution of energy across the sample, typically 1-10 mins are analysed at a time. These prove useful in long term studies or for observing gross changes in time or space. Seeking to capture subtler changes in behaviour and composition of vocalising communities, and to counter the noise-sensitivity of the entropy indices, the Acoustic Complexity Index (ACI) was developed specifically to capture the dynamic changes in the soundscape: “many biotic sounds, such as bird songs, are characterised by an intrinsic variability of intensities, while some types of human generated noise (such as car passing or airplane transit) present very constant intensity values” [Pieretti et al., 2011]. The (ACI) is derived from measures of absolute difference in adjacent bins in a spectrogram and was shown to correlate with the number of bird vocalisations in a small scale spatial study in an Apennine National Park, Italy [Pieretti et al., 2011].

The Bioacoustic Index [Boelman et al., 2007] is presented as measure of avian abundance and is calculated simply as the area under the mean frequency spectrum (minus the value of the lowest bin), providing a measure of both the sound level and the number of frequency bands used by the avifauna. It was used to investigate differences between exotic and native species in Hawaii and shown to be strongly correlated with counts from direct ornithological survey when calculated for single samples taken across a 6 week period.

These initial studies are encouraging: indices have been shown to correlate with aurally identified changes in bird species richness [Depraeter et al., 2012] and reveal dynamic variation across landscape [32], [28], however there are many open questions both methodologically and theoretically. Existing indices are inherently likely to be affected by several factors including transitory or permanent background noise, variation in distance of the animal to the microphone and relative intensity of particular species call patterns. Theoretically, we are still far from understanding exactly what aspects of biodiversity these indices might represent [Pijanowski et al., 2011] [Sueur et al., 2008] [Servick, 2014]. This is highlighted in a recent temporal study of dissimilarity indices [Lellouch et al., 2014] in which indices were shown to correlate well with simulated communities, but did not track community composition changes in the wild, raising the question of what, if any, aspect of compositional diversity such indices represent.

By virtue of being based on either time-averaged spectrograms or amplitude changes in the time domain, indices under this approach are fundamentally limited in their ability to detect spectro-temporal patterns, which may be key to evaluating the acoustic dynamics of specific communities. Frequency-based indices can pick up on crude differences in gross frequency range, but are inherently constrained in their ability to detect global spectro-temporal patterns created by cohabiting species interacting in an acoustic community. As the motivational premise of the community level approach assumes that spectro-temporal partitioning is responsible for structuring the soundscape, this constraint may be relevant.

### 1.4 Sparse Coding and Latent Component Analysis

Time-frequency tradeoffs are an important issue in all signal processing tasks. Sparse coding is gaining popularity in brain imaging, image analysis and audio classification tasks as an alternative to vector-based feature representations. Depending on the dictionary used, sparse representations using overcomplete dictionaries may have more time-frequency flexibility than standard Fourier transform representations.

The basic idea is fairly simple. For a given set of input signals, a number of atomic functions are sought, such that each input signal can be approximated sparsely by a linear combination of a relatively small number of atomic functions. This set of atoms is called a dictionary. In sparse coding, the number of atomic functions is higher than the dimensionality of the signal such that a subset of them can span the whole signal space – an overcomplete dictionary [Scholler and Purwins, 2011]. Sparse approximations...
we wish to assess characteristics of the arrangement and how the voices interact as an ensemble toward a
occupancy. Perhaps their failure to track variation in species richness in the wild is because the defining
in tracking species richness in simulated communities by measuring gross changes in frequency band
coherent global composition through time, timbre and pitch space. Frequency-based indices may succeed
qualitative species, so much as achieving a numerical description of the
perspective of Soundscape Ecology, we are not necessarily concerned with the identification of specific
perceptually relevant musical objects from time-frequency shifted patterns in a dynamic signal. From the
soundscape analysis, might this provide an ecologically-relevant indicator of the compositional complexity
of an acoustic community?

Sparse decomposition using dictionaries of atoms based on biologically informed time-frequency
atoms such as Gabor and Gammatone functions – which are seen to resemble characteristics of cochlea
filters – are intuitively attractive as they can provide an oriented feature set with which to approximate the
original signal. This has been shown to be more efficient than Fourier or wavelet representations (Smith
and Lewicki, 2005) and to provide effective and efficient input features in a range of audio discrimination
tasks in everyday sounds (Adiloglu et al., 2012), drum samples (Scholler and Purwins, 2011) and similarity
matching of bioacoustic data (Glotin et al., 2013).

Probabilistic Latent Component Analysis (PLCA) is one of a family of techniques used for source
separation, which similarly provides a tool for extracting sound objects according to common frequency-
amplitude statistics. PLCA is a probabilistic variant of non-negative matrix factorization (NMF) (Lee and
Seung, 2001). It decomposes a non-negative matrix V into the product of two multinomial probability
distributions, W and H, and a mixing weight, Z. In the auditory domain, V would be a matrix representing
the time-frequency content of an audio signal:

\[ V \approx WZH = \sum_{k=0}^{K-1} w_k z_k h_k^T \]  

where each column of W can be thought of as a recurrent frequency template and each row of H as
the excitations in time of the corresponding basis. Z = diag(z) is a diagonal matrix of mixing weights
z and K is the number of bases in W (Weiss and Bello, 2010). Each of V, w_k, z_k, and h_k correspond to
probability distributions and are normalized to sum to 1.

Sparse and shift-invariant PLCA (SI-PLCA) extends PLCA to enable the extraction of multiple
shift-invariant features from analysis of non-negative data of arbitrary dimensionality and was first
demonstrated as an effective unsupervised tool for extracting shift-invariant features in images, audio
and video (Smaragdis et al., 2008). The algorithm provides a very precise and perceptually meaningful
description of content. A series of piano notes, for example, is automatically decomposed into a kernel
distribution representing the harmonic series common to all notes, the peaks of the impulse distribution
representing the fundamental frequency of each note and its location in time (Smaragdis et al., 2008). Weiss
and Bello (2010) demonstrated application in segmentation task, showing SI-PLCA to be competitive with
Hidden Markov Models and self-similarity matrices. More recently, Sarroff and Casey (2013) developed
a shift and time-scale invariant PLCA which performed well against results of a human ‘groove-similarity’
judgement task.

A common strategy used throughout the NMF literature is to favour sparse settings in order to
learn parsimonious, parts-based decompositions of the data. Sparse solutions can be encouraged when
estimating the parameters of the convolution matrix by imposing constraints using an appropriate prior
distribution (Smaragdis et al., 2008). Under the Dirichlet distribution for example, hyper-parameters
can be set to favour a sparse distribution. In these cases, the algorithm will attempt to use as few bases
as possible, providing an ‘automatic relevance determination strategy’ (Weiss and Bello, 2010): The
algorithm can be initialised to use many bases; the sparse prior then prunes out those that do not contribute
significantly to the reconstruction of the original signal. In the context of pop song segmentation, this
enables the algorithm to automatically learn the number and length of repeated patterns in a song. In
soundscape analysis, might this provide an ecologically-relevant indicator of the compositional complexity
of an acoustic community?

In Music Information Retrieval and composition tasks, SI-PLCA provides a tool for accessing
perceptually relevant musical objects from time-frequency shifted patterns in a dynamic signal. From the
perspective of Soundscape Ecology, we are not necessarily concerned with the identification of specific
species, so much as achieving a numerical description of the qualitative patterns of interaction between
them. By way of musical analogy, we don’t care what the specific instruments of the orchestra are, rather
we wish to assess characteristics of the arrangement and how the voices interact as an ensemble toward a
coherent global composition through time, timbre and pitch space. Frequency-based indices may succeed
in tracking species richness in simulated communities by measuring gross changes in frequency band
occupancy. Perhaps their failure to track variation in species richness in the wild is because the defining
feature of acoustic communities are global patterns of interaction across a more complex spectro-temporal space, rather than frequency band occupancy or amplitude variation alone. Current indices based on frequency or amplitude statistics inherently throw away information crucial to the analysis of spectro-temporal patterns: SI-PLCA provides a tool for extracting dynamic sound objects grouped by common frequency-amplitude statistics, even when pitch or time shifted. That it has been demonstrated to be effective in extracting the perceptually-meaningful but nebulous concept of ‘groove’ (Sarroff and Casey 2013) suggests potential as a tool for beginning to interrogate the concept of the acoustic niche. In this paper we take a first look at how these methods might provide a complementary approach to current acoustic indices for investigation of soundscape dynamics and ultimately for biodiversity assessment. Taking a small sample of field recordings across different habitats in an Ecuadorian cloud forest reserve we compare existing spectral and temporal indices with sample analyses of a number of approaches to sparse approximation, including dictionaries built using mini-batch gradient descent, Gabor functions and SI-PLCA2D. The potential value of this approach is illustrated with example reconstructions from a new variant of SI-PLCA using dual dictionaries.

2 METHODS AND MATERIALS

2.1 Data Collection

2.1.1 Study Area and Acoustic Survey Methods

The data reported here was collected during an 8 week field survey (June - August 2014) in the Ecuadorian Andean cloud forest at the Santa Lucia Cloud Forest Reserve (SLR). The SLR (0°07'30” N, 78°40'3” W) is situated on the Western (Pacific) slopes of the Andes in northwestern Ecuador and spans an elevational range of 1400 - 2560 m. The forest is lower montane rain forest (cloud forest). The area has a humid subtropical climate and is composed of fragmented forest reserves surrounded by a matrix of cultivation and pasture lands. It lies within the Tropical Andes biodiversity hotspot and exhibits high plant species endemism and diversity. Topography is defined by steep-sloping valley systems of varying aspect.

The SLR was awarded reserve status 20 years ago, prior to which areas of Primary Forest had cleared for fruit farming. The SLR therefore consists of a complex mosaic of habitat types: Ancient Primary Forest (FP) punctuated by small areas of secondary regrowth of around 20 years (FS) and silvopasture (S), typically elephant grass pastures used as grazing paddocks for the mules, which provide local transport. These areas are less than 5 ha. In contrast to other studies where dramatically different sites have been used to validate indices, this complex patchy habitat provides subtle habitat gradients.

Acoustic data was collected using nine digital audio field recorders Song Meter SM2+ (Wildlife Acoustics), giving three replicates of each of the three habitat types. Minimum distance between recorders was 300m to avoid pseudo sampling. Altitudinal range was minimised. Recording schedules captured the full dawn (150 min), dusk choruses (90 min) plus midday (60 min) activity; throughout the rest of the period 3min recordings were taken every 15 min.

The SM2+ is a schedulable, off-line, weatherproof recorder, with two channels of omni-directional microphone (flat frequency response between 20 Hz and 20 kHz). Gains were set experimentally at 36 dB and recordings made at 16 bit with a sampling rate of 44.1 kHz. All recordings were pre-processed with a high pass filter at 500 Hz (12 dB) to attenuate the impact of the occasional aircraft and local generator noise.

2.1.2 Species Identification

Point counts were carried out by a local expert ornithologist and were made in situ in order to record species seen as well as heard. A record was made for each individual, rather than individual vocalisations, providing species presence-absence and abundance.

2.2 Acoustic Indices

For the purposes of this illustrative exercise, analyses were carried out on dawn chorus recordings from just one day for three habitat types sampled. A range of indices described in Section 1.3 were calculated: NDSI, H (including sh and th components), ADI, AEI, ACI and BI. Indices were calculated for the same 10min periods during which point counts were made at each site. Calculations were made using the seewave and acousticecology packages in R.

http://cran.r-project.org/web/packages/seewave/index.html
http://cran.r-project.org/web/packages/acousticecology/index.html
2.3 Audio Spectrum Approximation Methods

Three approaches to audio decomposition are illustrated using the Bregman Media Labs Audio Patch Approximation Python package[^4]. Dictionary learning using mini-batch gradient learning, a Gabor field dictionary and, shift invariant 2D Principle Latent Component Analysis (SI-PLCA2D). Each uses Orthogonal Matching Pursuit (OMP) to build the component reconstructions. Samples were extracted from analyses of 1 minute sections of the field recordings. These examples are aimed at illustrating the potential of an atomic rather than vector approach in general, rather than experimental validation of any particular algorithm. Default parameters were used in all cases.

A potential future direction is illustrated using a SI-PLCA variant (SI-PLCA2) using 2D dual dictionaries[^Smaragdis and Raj (2007), Weiss and Bello (2010), Sarroff and Casey (2013)] based on frequency × local time functions and frequency-shift × global time-activations. The expectation-maximisation (EM) algorithm[^Smaragdis and Raj (2007)] is used to build component reconstructions.

As described in Section 1.4, the algorithm returns a set of \( K \) from a \( K_{\text{max}} = 16 \): independent component reconstructions, time-frequency kernels and shift-time activation functions. Entropies of each are also returned. The technical details of this approach are not addressed here, but example analyses are used to illustrate SI-PLCA as a potentially rich tool for future research in investigating the complex quasi-periodic signals of wild soundscapes.

3 RESULTS AND DISCUSSION

3.1 Species Composition of Acoustic Communities

The species observations for each site, shown in Table 1, reveal little variation in overall abundance or species number between the sites when seen and heard records are considered together. Several species are observed in all sites; others are observed only in one habitat type. Discounting the seen-only counts, the highest number of species, and individuals, was recorded at S, with least heard FP. The spectrograms and mean spectrum profiles (Figure 1) for these recordings suggest that this information is present in the soundscape. The number of shared species between sites results in acoustic communities with an overall similar overlap, differentiated by calls of ‘keynote’ species. Each acoustic community occupies a broadly similar frequency range, with variation in the peaks of spectral profiles according to the prevalence of calls of habitat-specific species. FP appears to have lowest overall activity, in line with the relatively fewer number of species observed.

Despite occupying an overall similar frequency range and not differing dramatically in abundance, each site is distinctly characterised by differing quasi-periodic patterns of calls. The same patterns observed at the 1 minute shown continued for the full 10 minute sample[^5]. The soundscape is structured, not just by repetitions of specific species calls, but by turn taking, i.e. interactions between species. This is most evident in listening, and can be observed visually as an interplay of periodic gestures in the spectrogram. It is precisely this complex of interacting periodic structures which we wish to evaluate under the soundscape approach, but which are impervious to analyses by current indices.

3.2 Acoustic Indices

Values for each of the acoustic indices calculated for the three habitats are given in Table 2 and shown as bar plots in Figure 2. As we might expect given the minimal anthropogenic noise and broadly similar spectral profile, the NDSI reports near maximum values for each site. The global complexity of each scene is high; it is no surprise then that entropy indices approach 1 and differences between sites are minimal. The ADI reports a small variation, following the rank-order pattern of species heard at each site. Differences between Sueur’s spectral, temporal and therefore overall, H entropy are minimal. ACI similarly shows small variation between sites. This index in particular is very sensitive to the size of the analysis window and requires further exploration to establish which aspects of community composition may be being assessed. BI values report the differences in overall acoustic energy, observable in mean spectrum plot (Figure 1, bottom), with the highest value at FS, FP being slightly higher than site S. These basic features of the acoustic recordings are at odds with the field observations of abundance and species numbers. An increase in overall energy could be due to certain individuals having intrinsically louder calls, calling more frequently, or simply being closer to the microphone. In validation studies the latter could

[^4]: https://github.com/bregmanstudio/audiospectrumpatchapproximation
[^5]: 1 min excerpts available via PeerJ linked data service
be countered by factoring in field-based point count distance measures (recorded, but not included here) and call frequencies, as well as tallies of individuals, the latter being expedited by the use of automatic segmentation software (as in [Pieretti et al., 2011]).

The key issue raised here, however is that in providing summaries of frequency or temporal amplitude profile and magnitude differences, these current indices are not only sensitive to these largely irrelevant variations in overall amplitude changes, but are all insensitive to the periodic structures which uniquely characterise the three soundscapes.

3.3 Sparse-approximation Outputs

Dictionaries and sparse-approximations of recordings of each site using mini-batch gradient descent, Gabor atoms and SI-PLCA2D are shown in Figure 3. The input for each is a log-frequency spectrogram (constant-Q transform) of samples from the field recordings, as shown in Figure 1 Example dictionaries (left) and sparse approximations of the input spectrogram (right) for site FP are shown for each method (component reconstructions not shown). Comparing the sparse-approximation of the original spectrogram for FP (see Figure 1 top image), the superior performance of SI-PLCA2D over the other two methods is evident.

A detailed technical discussion of the methods is beyond the scope of the current paper, however of key interest in this context are the qualitative differences in the dictionaries. The Gabor field dictionary has an intuitive advantage over vector descriptors in representing features oriented in time-frequency space. The dictionary learned under mini-batch gradient descent similarly exhibits time-frequency atoms differing subtly in orientation. The SI-PLCA2D dictionary however, comprises a collection of spectrum patches with a variety of micro-structures across a range of orientation and spread. In terms of the filter model which motivates the use of Gabor atoms, the Gabor and mini-batch dictionaries could be described as having relatively homogenous widths across the dictionary; the SI-PLCA2D dictionary by contrast contains points not only differing in time-frequency orientation, but in spectral width, atoms 0, 2, 3 and 4 being considerably more focused than 1 and 5 (Figure 3[c], left). This is an appealing property for the analysis of broad-spectrum versus pitched soundscape elements.

Full outputs for all three sites using the SI-PLCA2 algorithm with dual 2D dictionaries are shown in Figures 4, 5 and 6. Each 10 min site recording is sampled, taking 16 time windows from across the file of around 4 seconds each, arranged in order. The input is the log-frequency spectrogram of these samples, as before. Extensive analysis of larger data sets across more diverse soundscapes is needed before we can begin to evaluate the ecological significance or application of this approach, but a number of promising observations can be made.

As can be seen in Figures [4][a],[5][a],[6][a], the component reconstructions appear faithful to the original spectrogram. The individual component reconstructions (Figures [4][b],[5][b],[6][b]) pull out clearly distinct components. This is clearest in S3 (Figure [4][b]) where the first component is broadband ambient noise, and each of components 1 - 5 appear as distinct ‘voices’ grouped according to both spectral range and spectro-temporal periodic gesture.

The time-frequency kernels provide a lower dimension representation of components with apparently similar characteristics: compare each component in Figures [4][b] and [5][b] for example. The automatic relevance determination feature deserves further investigation as a quick and dirty proxy for community composition assessment. In this example in FP $k = 4$, FS $k = 7$ and S $k = 6$. Does K increase with the number of vocalising species? Might it reflect the complexity or ‘decomposability’ of a scene in some way?

The entropies of each distribution are given in the subfigure captions of Figures 4, 5 and 6. Whether these can provide useful information as a difference measure either between components within a particular reconstruction, or between kernels extracted from different soundscapes deserves further investigation.

No conclusions can be drawn from this illustrative analysis, but it raises a number of questions for future research: (1) Are the component reconstructions meaningful soundscape objects in ecological terms? Are the structurally distinct characteristics of geophony and biophony separated? Are vocalising species separated in any meaningful ecological way (species, functionality etc.)? (2) Might the statistics generated be meaningful? Does the number of components ($k$) returned reflect ‘complexity’ or ‘decomposability’ in a way which may reflect the status of the acoustic community? Could the entropy summaries of each component be used as a measure of diversity within or between communities?

The ability of PLCA to separate streams of distinct sound objects or voices from background noise
is well recognised. Within the conceptual framework of Soundscape Ecology, such techniques promise utility at two levels. Within the acoustic community, as a possible means to investigate the composition of the local soundscape in terms of dynamic interactions between spectro-temporal patterns of vocalising component species, providing a new tool to begin to experimentally interrogate the concept of acoustic niche. Within the conceptual framework of Soundscape Ecology, geophony (the noises created by wind, rain etc. in interaction with the local landscape) is not noise to be removed from the signal, but a crucial active component: the potential to separate geophony from biophony and anthrophony therefore provides a means to further investigate the patterns of interaction between them.

4 SUMMARY AND FUTURE WORK

Monitoring subtle changes in complex ecosystems is crucial for ecological research and conservation but far from straightforward. Acoustic indices hold promise as a rapid assessment tool, but are subject to the same trade-offs as traditional ecological research of quality versus quantity: any metric necessarily throws away some information. In this paper we have provided an overview of the motivational premises of Soundscape Ecology, including the concept that acoustic communities may be structured according to competition across acoustic niches through spectro-temporal partitioning. We suggest that existing indices operating in time or frequency domain may be insensitive to the dynamic patterns of interaction in the soundscape which characterise specific acoustic communities and propose SI-PLCA2D as a promising new tool for research. This was illustrated with example analyses of tropical dawn chorus recordings along a gradient of habitat degradation.

It seems likely that if acoustic niches exist that they do not lie neatly along 1D vectors in the frequency or time domain but dance dynamically across pitch–timbre–time space. SI-PLCA2D is computationally expensive, but provides a tool for extracting shift-invariant patterns in a dynamic soundscape. In future work we plan to investigate the potential for these tools to help us begin to experimentally investigate the acoustic niche concept, seeking to understand what dimensions of acoustic ecospace it may occupy toward the development of effective tools for rapid acoustic biodiversity assessment.

ACKNOWLEDGMENTS

Many thanks to Noé Morales of Santa Lucia for carrying out point count surveys.

REFERENCES


Figure 1. Top: Constant-Q spectrograms for 1 min excerpts for each site FP (top) FS (middle) S (bottom). 1 minute resolutions are presented to illustrate the periodic call patterns. These were consistent across the 10 min sampling time in each habitat. Bottom: Mean Constant-Q Spectrum (log amplitude) for FP (green), FS (blue) and S (red)
<table>
<thead>
<tr>
<th>Common Name</th>
<th>Heard</th>
<th></th>
<th></th>
<th>Seen</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FP</td>
<td>FS</td>
<td>S</td>
<td>FP</td>
<td>FS</td>
<td>S</td>
</tr>
<tr>
<td>Andean Solitaire</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Azara's Spinetail</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Beryl-spangled Tanager</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Blue-winged Mountain-Tanager</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Booted Racket-tail</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Brown Inca</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Brown-capped Vireo</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Collared Forest-Falcon</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dusky Bush-Tanager</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Empress Brilliant</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Flame-faced Tanager</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Golden-Crowned Flycatcher</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Golden-Headed Quetzal</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gray-breasted Wood-wren</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Immaculate Antbird</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Lineated Foliage-Gleaner</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Long-tailed Antbird</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Masked Trogon</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Metallic-Green Tanager</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nariño Tapaculo</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Orange-bellied Euphonia</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Plumbeous Pigeon</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Red-faced Spinetail</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Roadside Hawk</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rufous-breasted Antthrush</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Russet-crowned Warbler</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scale-crested Pygmy-Tyrant</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Slate-throated Whitestart</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smoke-colored Pewee</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Three-striped Warbler</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Toucan Barbet</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tricolored Brush-Finch</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyrannine Woodcreep</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Uniform Antshrike</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wattled Guan</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>White-capped Parrot</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

| Total | 13 | 18 | 21 | 8 | 2 | 3 |

**Table 1.** Field observations for species heard and seen at FP (0600-0610), FS (0619-0629) and S(0639-0649) on June 15th 2014. As only 1 individual of each species was recorded, this data provides a record of species absence-presence as well as abundance for each site.

<table>
<thead>
<tr>
<th>NDSI</th>
<th>ADI</th>
<th>AEI</th>
<th>sh</th>
<th>th</th>
<th>H</th>
<th>ACI</th>
<th>BI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP</td>
<td>0.9716</td>
<td>2.1919</td>
<td>0.2591</td>
<td>0.9567</td>
<td>0.9730</td>
<td>0.9309</td>
<td>18497.14</td>
</tr>
<tr>
<td>FS</td>
<td>0.9727</td>
<td>2.2684</td>
<td>0.1418</td>
<td>0.9355</td>
<td>0.9729</td>
<td>0.9102</td>
<td>18315.61</td>
</tr>
<tr>
<td>S</td>
<td>0.9809</td>
<td>2.2909</td>
<td>0.0749</td>
<td>0.9539</td>
<td>0.9825</td>
<td>0.9372</td>
<td>18686.78</td>
</tr>
</tbody>
</table>

**Table 2.** Acoustic indices values for the three study sites: FP, FS and S shown to 4 decimal places.
Figure 2. Bar plots of indices results for same recording showing indices values 10 minutes of dawn chorus at FP (black), FS (grey) and S (light grey) for NDSI, ADI, AEI, sh, th, H, ACI and BI. Values for AD, ACI and BI are scaled in the ranges 0:3, 1800:1900 and 0:12 respectively. Point count ‘heard’ data values for each site are given in the end column for comparison.

Figure 3. Over-complete Dictionaries (left) and sparse-approximations of original spectrogram shown in Figure 1 for (a) Mini-batch gradient learning, (b) Gabor Field Dictionary and (c) SI-PLCA2D Component Dictionary.
Figure 4. SIPLCA2 outputs for Primary Forest site dawn chorus. Entropy ($S$) values are shown in brackets.
Figure 5. SI-PLCA2 outputs for Secondary Forest site dawn chorus. Entropy (S) values are shown in brackets.
Figure 6. SI-PLCA2 outputs for Silvopasture site dawn chorus. Entropy (S) values are shown in brackets.