

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

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

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 <https://doi.org/10.7717/peerj.2108>

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

Alice Eldridge<sup>1</sup>, Michael Casey<sup>2</sup>, Paola Moscoso<sup>3</sup>, and Mika Peck<sup>4</sup>

<sup>1,3,4</sup>Dept. Evolution, Behaviour and Environment, University of Sussex, UK

<sup>2</sup>Depts. 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 analyse 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 summarise 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

1 Biodiversity assessment is a central and urgent task, not only for research in the biological sciences, but  
2 also in applied conservation biology, including major multi-lateral initiatives for promoting and protecting  
3 biodiversity. At the governmental level biodiversity needs to be incorporated into national accounting by  
4 2020 (Aichi Biodiversity targets A2)<sup>1</sup> and cost effective tools necessary to achieve this remain elusive.  
5 Operating within the conceptual and methodological framework of the burgeoning field of Soundscape  
6 Ecology, (Pijanowski et al., 2011; Farina, 2014) we are interested in the potential use of acoustic indices  
7 for interrogating the interactions between the landscape and the communities which live within it, with  
8 ultimate utility in conservation as a rapid acoustic biodiversity assessment tool. In this ideation paper  
9 we first summarise the foundational premises for the soundscape approach and provide an overview of  
10 existing acoustic indices; a key constraint of current indices is identified and a potential new direction for  
11 research is suggested.

### 1.1 Current Methods for Biodiversity Assessment.

12 Biodiversity refers to the variability among living organisms and the ecological complexes of which  
13 they are a part (Buckland et al., 2005) and is seen as an indicator of the ‘health’ of a habitat. It should  
14 be noted that within traditional ecology, the question of how best to measure and describe biodiversity  
15

<sup>1</sup><http://www.cbd.int/sp/targets/>

16 is contentious. There are no absolute measures, rather indices reflecting difference or change across  
17 or between time or space. A key part of biodiversity assessment then, is based on the estimation and  
18 monitoring of changes in species composition and abundance in animal communities. There are a plethora  
19 of approaches to sampling and analysis (Magurran, 2004; Buckland et al., 2005; Pavoine and Bonsall,  
20 2011) most of which are derived from observations of the richness and abundance of species encountered  
21 in a given area at a specified time. In most strategies, there is a trade-off between data quality and quantity.  
22 All Taxa Biodiversity Indices (ATBI) or multi-species field studies are desirable theoretically, but in  
23 practice surveys are invariably subject to financial constraints that bind decision-makers (Lawton et al.,  
24 1998). In ongoing work, we are exploring cost-effective solutions, including remote sensing (camera traps  
25 and aerial photography of canopy) and identification of 'ecological-disturbance indicator species' (Caro,  
26 2010). Remote sensors are an attractive choice for data collection in that they are noninvasive, scalable in  
27 both space and time and remove the bias and cost associated with programs which require either experts  
28 (ATBIs, Gewin, 2002) or even non-specialists (Rapid Biodiversity Assessment, Oliver and Beattie, 1993),  
29 *in situ*.

30 Various forms of remote visual sensing technologies have been explored. Global satellite imaging  
31 has been investigated to monitor biophysical characteristics of the earth's surface by assessing species  
32 ranges and richness patterns indirectly (e.g. Wang et al., 2010). These methods are attractive, but rely  
33 on expensive equipment, are difficult to adapt to small spatial scales and require a time-consuming  
34 validation step. It is possible, for example, to infer valid species-level identification of canopy trees from  
35 high-resolution aerial imagery, providing a means of remote sensing to assess forest status (Peck et al.,  
36 2012). However, the principal weakness of this and other existing visual remote sensing methods is that  
37 they cannot provide direct information on the status of taxa other than plants: they cannot detect 'silent  
38 forests'. The need for innovative remote sensing methods to monitor the status of wildlife remains and  
39 acoustic, rather than visual, sensors have many attractive characteristics.

## 40 1.2 Acoustic Approaches to Biodiversity Assessment

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

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

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

70 signals due to invasive species.

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

### 81 1.3 Existing Acoustic Indices

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

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

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

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

$$H' = - \sum_{i=1}^R p_i \ln p_i \quad (1)$$

111 The **Acoustic Diversity Index** (Villanueva-Rivera et al., 2011) (ADI) is a spectral entropy measure  
 112 which summarises the distribution of the proportion of signals across the spectrum. The FFT spectrogram  
 113 is divided into a number of bins (default 10) the proportion of the signals in each bin above a threshold  
 114 (default = 50 dBFS) is calculated. The Shannon Index (Eq.1) is then applied, where  $p_i$  is the fraction  
 115 of sound in each  $i_{th}$  of  $R$  frequency bands. An evenness metric, the **Acoustic Evenness Index** (AEI) is  
 116 similarly derived by calculating the Gini index (Gini, 1912) (commonly used by ecologists to estimate  
 117 species evenness) on the spectrum. These relatively simple indices are shown to effectively reflect  
 118 observed distinctions in gross acoustic activity, for example between dawn choruses and night activity, or  
 119 between diverse habitats (mature oak forest, secondary forest, wetland and agricultural land).

120 The **Acoustic Entropy Index**,  $H$  (Sueur et al., 2008) is also calculated as the product of spectral  
121 ( $sh$ ) and temporal ( $th$ ) entropies, calculated on the mean spectrum and Hilbert amplitude envelope of a  
122 time wave respectively.  $H$  ranges from 0 for pure tones to 1 for high-energy, evenly distributed sound.  
123 The index was first tested against simulated choruses, generated by mixing together samples of avian  
124 vocalisations and systematically varying the number of species in each track.  $H$  values increased with  
125 species richness  $S$  following a logarithmic model. Field trials were carried out in pristine and degraded  
126 African coastal forests and  $H$  was shown to reflect assumed variation in species richness. (Sueur et al.,  
127 2008). The study was in an area where animal acoustic activity was high and background noise low.  
128 When background noise (such as traffic) or broadband signals (such as rain, cicada or tropical cricket  
129 choruses) are higher, spectral entropy measures may give counter-intuitive results (values for white noise  
130 or motorway traffic, for example would approach 1).

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

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

147 These initial studies are encouraging: indices have been shown to correlate with aurally identified  
148 changes in bird species richness (Depraetere et al., 2012) and reveal dynamic variation across landscape  
149 [32], [28], however there are many open questions both methodologically and theoretically. Existing  
150 indices are inherently likely to be affected by several factors including transitory or permanent background  
151 noise, variation in distance of the animal to the microphone and relative intensity of particular species  
152 call patterns. Theoretically, we are still far from understanding exactly what aspects of biodiversity these  
153 indices might represent (Pijanowski et al., 2011; Sueur et al., 2008; Servick, 2014). This is highlighted in  
154 a recent temporal study of dissimilarity indices (Lellouch et al., 2014) in which indices were shown to  
155 correlate well with simulated communities, but did not track community composition changes in the wild,  
156 raising the question of what, if any, aspect of compositional diversity such indices represent.

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

#### 164 **1.4 Sparse Coding and Latent Component Analysis**

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

169 The basic idea is fairly simple. For a given set of input signals, a number of atomic functions are  
170 sought, such that each input signal can be approximated sparsely by a linear combination of a relatively  
171 small number of atomic functions. This set of atoms is called a dictionary. In sparse coding, the number  
172 of atomic functions is higher than the dimensionality of the signal such that a subset of them can span the  
173 whole signal space – an overcomplete dictionary (Scholler and Purwins, 2011). Sparse approximations

174 of the signal area are then constructed by finding the “best matching” projections of multidimensional  
 175 data onto an over-complete dictionary, Matching Pursuit (Mallat and Zhang, 1993) (MP) being a popular  
 176 choice.

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

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

$$V \approx WZH = \sum_{k=0}^{K-1} w_k z_k h_k^T \quad (2)$$

190 where each column of  $W$  can be thought of as a recurrent frequency template and each row of  $H$  as  
 191 the excitations in time of the corresponding basis.  $Z = \text{diag}(z)$  is a diagonal matrix of mixing weights  
 192  $z$  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  
 193 probability distributions and are normalized to sum to 1.

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

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

216 In Music Information Retrieval and composition tasks, SI-PLCA provides a tool for accessing  
 217 perceptually relevant musical objects from time-frequency shifted patterns in a dynamic signal. From the  
 218 perspective of Soundscape Ecology, we are not necessarily concerned with the identification of specific  
 219 species, so much as achieving a numerical description of the *qualitative* patterns of interaction between  
 220 them. By way of musical analogy, we don’t care what the specific instruments of the orchestra are, rather  
 221 we wish to assess characteristics of the arrangement and how the voices interact as an ensemble toward a  
 222 coherent global composition through time, timbre and pitch space. Frequency-based indices may succeed  
 223 in tracking species richness in simulated communities by measuring gross changes in frequency band  
 224 occupancy. Perhaps their failure to track variation in species richness in the wild is because the defining

225 feature of acoustic communities are global patterns of interaction across a more complex spectro-temporal  
226 space, rather than frequency band occupancy or amplitude variation alone. Current indices based on  
227 frequency *or* amplitude statistics inherently throw away information crucial to the analysis of spectro-  
228 temporal patterns: SI-PLCA provides a tool for extracting dynamic sound objects grouped by common  
229 frequency-amplitude statistics, even when pitch or time shifted. That it has been demonstrated to be  
230 effective in extracting the perceptually-meaningful but nebulous concept of ‘groove’ (Saroff and Casey,  
231 2013) suggests potential as a tool for beginning to interrogate the concept of the acoustic niche.

232 In this paper we take a first look at how these methods might provide a complementary approach  
233 to current acoustic indices for investigation of soundscape dynamics and ultimately for biodiversity  
234 assessment. Taking a small sample of field recordings across different habitats in an Ecuadorian cloud  
235 forest reserve we compare existing spectral and temporal indices with sample analyses of a number of  
236 approaches to sparse approximation, including dictionaries built using mini-batch gradient descent, Gabor  
237 functions and SI-PLCA2D. The potential value of this approach is illustrated with example reconstructions  
238 from a new variant of SI-PLCA using dual dictionaries.

## 239 **2 METHODS AND MATERIALS**

### 240 **2.1 Data Collection**

#### 241 **2.1.1 Study Area and Acoustic Survey Methods**

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

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

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

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

#### 265 **2.1.2 Species Identification**

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

### 269 **2.2 Acoustic Indices**

270 For the purposes of this illustrative exercise, analyses were carried out on dawn chorus recordings from  
271 just one day for three habitat types sampled. A range of indices described in Section 1.3 were calculated:  
272 NDSI, H (including sh and th components), ADI, AEI, ACI and BI. Indices were calculated for the same  
273 10min periods during which point counts were made at each site. Calculations were made using the  
274 *seewave*<sup>2</sup> and *acousticecology*<sup>3</sup> packages in R.

<sup>2</sup><http://cran.r-project.org/web/packages/seewave/index.html>

<sup>3</sup><http://cran.r-project.org/web/packages/soundecology/index.html>

## 275 **2.3 Audio Spectrum Approximation Methods**

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

283 A potential future direction is illustrated using a SI-PLCA variant (SI-PLCA2) using 2D dual dictio-  
284 naries (Smaragdis and Raj., 2007; Weiss and Bello, 2010; Sarroff and Casey, 2013) based on frequency \*  
285 local time functions and frequency-shift \* global time-activations. The expectation-maximisation (EM)  
286 algorithm (Smaragdis and Raj., 2007) is used to build component reconstructions.

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

## 292 **3 RESULTS AND DISCUSSION**

### 293 **3.1 Species Composition of Acoustic Communities**

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

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

### 311 **3.2 Acoustic Indices**

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

<sup>4</sup><https://github.com/bregmanstudio/audiospectrumpatchapproximation>

<sup>5</sup>1 min excerpts available via PeerJ linked data service



325 be countered by factoring in field-based point count distance measures (recorded, but not included here)  
326 and call frequencies, as well as tallies of individuals, the latter being expedited by the use of automatic  
327 segmentation software (as in Pieretti et al., 2011).

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

### 332 3.3 Sparse-approximation Outputs

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

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

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

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

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

368 The entropies of each distribution are given in the subfigure captions of Figures 4, 5 and 6. Whether  
369 these can provide useful information as a difference measure either between components *within* a particular  
370 reconstruction, or *between* kernels extracted from different soundscapes deserves further investigation.  
371 No conclusions can be drawn from this illustrative analysis, but it raises a number of questions for future  
372 research: (1) Are the component reconstructions meaningful soundscape objects in ecological terms?  
373 Are the structurally distinct characteristics of geophony and biophony separated? Are vocalising species  
374 separated in any meaningful ecological way (species, functionality etc.)? (2) Might the statistics generated  
375 be meaningful? Does the number of components ( $k$ ) returned reflect ‘complexity’ or ‘decomposability’ in  
376 a way which may reflect the status of the acoustic community? Could the entropy summaries of each  
377 component be used as a measure of diversity within or between communities?

378 The ability of PLCA to separate streams of distinct sound objects or voices from background noise

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

#### 387 4 SUMMARY AND FUTURE WORK

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

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

#### 404 ACKNOWLEDGMENTS

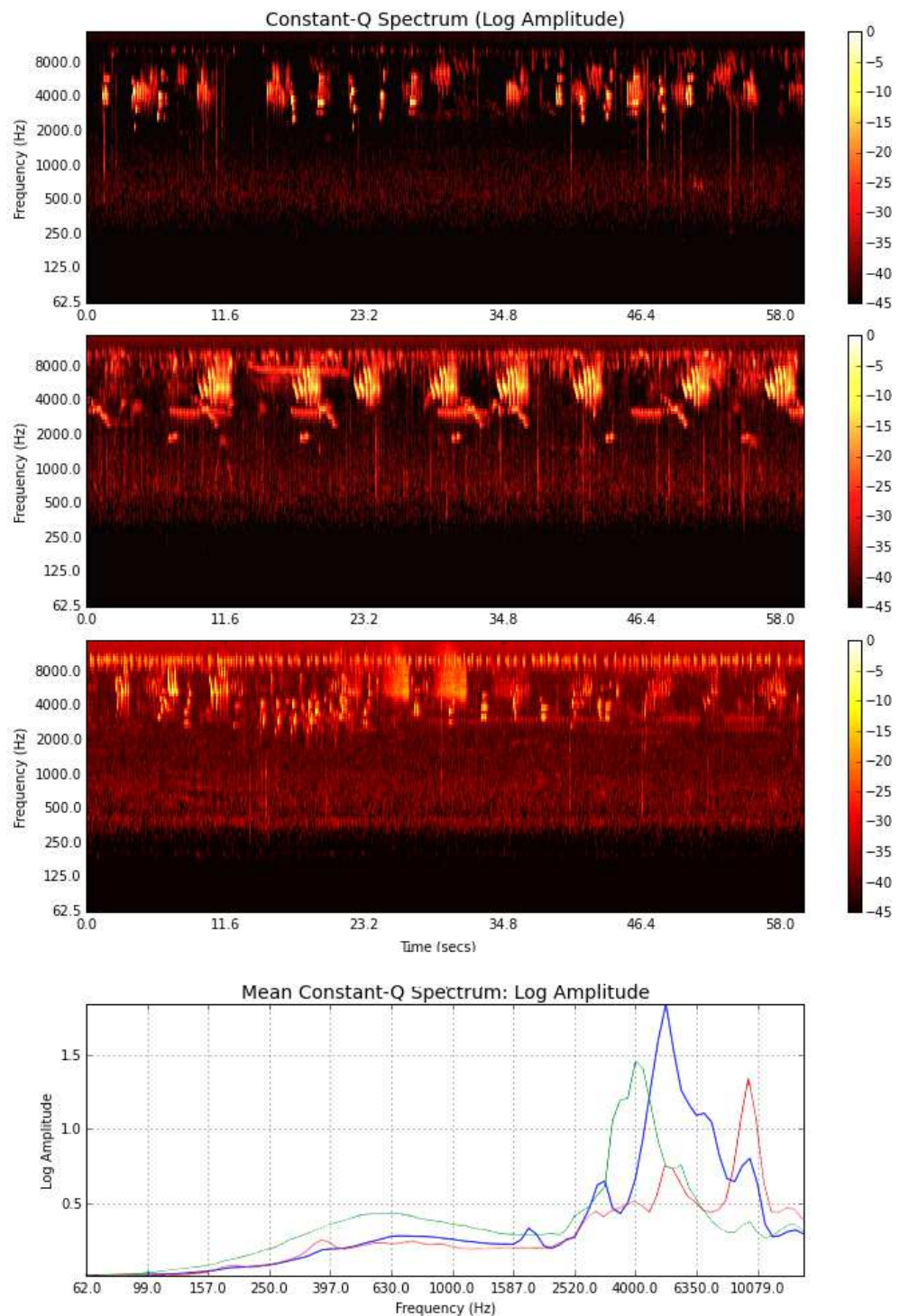
405 Many thanks to Noé Morales of Santa Lucia for carrying out point count surveys.

#### 406 REFERENCES

- 407 Adiloglu, K., Anniés, R., Wahlen, E., Purwins, H., and Obermayer, K. (2012). A graphical representation  
408 and dissimilarity measure for basic everyday sound events. *Audio, Speech, and Language Processing,*  
409 *IEEE Transactions on*, 20(5):1542–1552.
- 410 Boelman, N., Asner, G., Hart, P., and Martin., R. (2007). Multi-trophic invasion resistance in Hawaii:  
411 Bioacoustics, field surveys, and airborne remote sensing,. *Ecological Applications*, 17(8):2137–2144.
- 412 Buckland, S. T., Magurran, A., Green, R., and Fewster, R. M. (2005). Monitoring change in biodiversity  
413 through composite indices. *Philosophical Transactions of the Royal Society of London B*, 360:243–254.
- 414 Caro, T. M. (2010). *Conservation by proxy: Indicator, umbrella, keystone, flagship, and other surrogate*  
415 *species*. Island Press, Washington DC.
- 416 Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., and Sueur, J. (2012). Monitoring animal  
417 diversity using acoustic indices: Implementation in a temperate woodland. *Ecological Indicators*,  
418 13(1):46–54.
- 419 Farina, A. (2014). *Soundscape Ecology: Principles, Patterns, Methods and Applications*. Springer.
- 420 Gewin, V. (2002). All living things, online. *Nature*, 418(18(6896)):362 – 363.
- 421 Gini, C. (1912). Variability and mutability: Contribution to the study of statistical distribution and  
422 relations. *Studi Economico-Giuricici della R*.
- 423 Glotin, H., Sueur, J., Artières, T., Adam, O., and Razik, J. (2013). Sparse coding for scaled bioacoustics:  
424 From humpback whale songs evolution to forest soundscape analyses. *The Journal of the Acoustical*  
425 *Society of America*, 133(5):3311–3311.
- 426 Gregory, R. and Strien, A. (2010). Wild bird indicators: using composite population trends of birds as  
427 measures of environmental health. *Ornithological Science*, 9:3 – 22.
- 428 Kasten, E., Gage, S., Fox, J., and Joo., W. (2012). The remote environmental assessment laboratory's  
429 acoustic library: An archive for studying soundscape ecology. *Ecological Informatics*, 12:50–67.

- 430 Krause, B. (1987). Bioacoustics, habitat ambience in ecological balance. *Whole Earth Review*, 57:14 –  
431 18.
- 432 Lawton, J., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M.,  
433 Holt, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S., and Watt., A. D. (1998).  
434 Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*,  
435 391:72–76.
- 436 Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In Leen, T.,  
437 Dietterich, T., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages  
438 556–562. MIT Press.
- 439 Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., and Sueur, J. (2014). Monitoring temporal change of bird  
440 communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution*, 4:495–505.
- 441 Magurran, A. E. (2004). *Measuring biological diversity*. Blackwell Science Ltd., Malden, USA.
- 442 Mallat, S. G. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *Signal Processing*,  
443 *IEEE Transactions on*, 41(12):3397–3415.
- 444 Oliver, I. and Beattie, A. J. (1993). A possible method for the rapid assessment of biodiversity. *Conserva-*  
445 *tion Biology*.
- 446 Pavoine, S. and Bonsall, M. B. (2011). Measuring biodiversity to explain community assembly: a unified  
447 approach. *Biol Rev Camb Philos Soc*, 86(4):792–812.
- 448 Peck, M., Maddock, S., Morales, J., Peñafiel, V., Torres-Carvajal, O., Pozo-Rivera, W., Cueva-Arroyo,  
449 X., and Tolhurst., A. (2015). Cost-effectiveness of using small vertebrates as indicators of disturbance.  
450 *Conservation Biology*.
- 451 Peck, M., Mariscal, A., Padbury, M., Cane, T., Kniveton, D., M., and A., C. (2012). Identifying tropical  
452 ecuadorian andean trees from inter-crown pixel distributions in hyperspatial aerial imagery. *Applied*  
453 *Vegetation Science*, 15(4):548 – 559.
- 454 Pieretti, N., Farina, A., and Morri., D. (2011). A new methodology to infer the singing activity of an avian  
455 community: The acoustic complexity index (aci). *Ecological Indicators*, 11(3):868–873.
- 456 Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. S., Farina, A., Krause, B., Napoletano, B. M.,  
457 Gage, S. H., and Pieretti, N. (2011). Soundscape ecology: the science of sound in the landscape.  
458 *BioScience*, 61(3):203–216.
- 459 Riede, K. (1993). Monitoring biodiversity: Analysis of amazonian rainforest sounds. *Ambio*, pages  
460 546–548.
- 461 Rodriguez, A., Gasc, A., Pavoine, S., Grandcolas, P., Gaucher, P., and Sueur., J. (2013). Temporal and  
462 spatial variability of animal sound within a neotropical forest. *Ecological Informatics*.
- 463 Sarroff, A. M. and Casey, M. (2013). Groove kernels as rhythmic-acoustic motif descriptors. In  
464 *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*,  
465 pages 299–304.
- 466 Schafer, R. (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books,  
467 Rochester, Vermont.
- 468 Scholler, S. and Purwins, H. (2011). Sparse approximations for drum sound classification sparse  
469 approximations for drum sound classification sparse approximations for drum sound classification.  
470 *Selected Topics in Signal Processing, IEEE Journal of*, 5(5):933 – 940.
- 471 Seddon, N. (2005). Ecological adaptation and species recognition drives vocal evolution in neotropical  
472 suboscine birds. *Evolution*. 2005, 59(1):200–15.
- 473 Servick, K. (2014). Eavesdropping on ecosystems. *Science*, 343(6173):834–837.
- 474 Shannon, C. and Weaver., W. (1949). *The Mathematical Theory of Communication*. Illinois University  
475 Press.
- 476 Skowronski, M. and Harris, J. (2006). Acoustic microchiroptera detection and classification using machine  
477 learning: lessons learned from automatic speech recognition. *Journal of Acoustical Society of America*,  
478 19(3):1817 – 1833.
- 479 Smaragdis, P. and Raj., B. (2007). Shift-invariant probabilistic latent component analysis. Technical  
480 Report TR2007-009, MERL.
- 481 Smaragdis, P., Raj, B., and Shashanka., M. (2008). Sparse and shift-invariant feature extraction from  
482 non-negative data. In *Proceedings of ICASSP*.
- 483 Smith, E. and Lewicki, M. S. (2005). Efficient coding of time-relative structure using spikes. *Neural*  
484 *Computation*, 17(1):19–45.

- 485 Sueur, J., Farina, A., Gasc, A., Pieretti, N., and Pavoine, S. (2014). Acoustic indices for biodiversity  
486 assessment and landscape investigation. *Acta Acoustica United with Acoustica*, 100:772–781.
- 487 Sueur, J., Pavoine, S., Hamerlynck, O., and Duvail, S. (2008). Rapid acoustic survey for biodiversity  
488 appraisal. *PLoS One*, 3(12):e4065.
- 489 Towsey, M., Wimmer, J., Williamson, I. I., and Roe, P. (2013). The use of acoustic indices to determine  
490 avian species richness in audio-recordings of the environment. *Ecological Informatics*.
- 491 Villanueva-Rivera, L., Pijanowski, B., Doucette, J., and Pekin, B. (2011). A primer of acoustic analysis  
492 for landscape ecologists. *Landscape Ecology*, 26:1233–1246.
- 493 Wang, K., F., S., Guo, X., and Cattet, M. (2010). Remote sensing of ecology, biodiversity and conservation:  
494 a review from the perspective of remote sensing specialists. *Sensors*, 10:9647 – 9667.
- 495 Weiss, R. J. and Bello, J. P. (2010). Identifying repeated patterns in music using sparse convolutive  
496 non-negative matrix factorization. In *In Proc. International Conference on Music Information Retrieval*  
497 *(ISMIR)*.



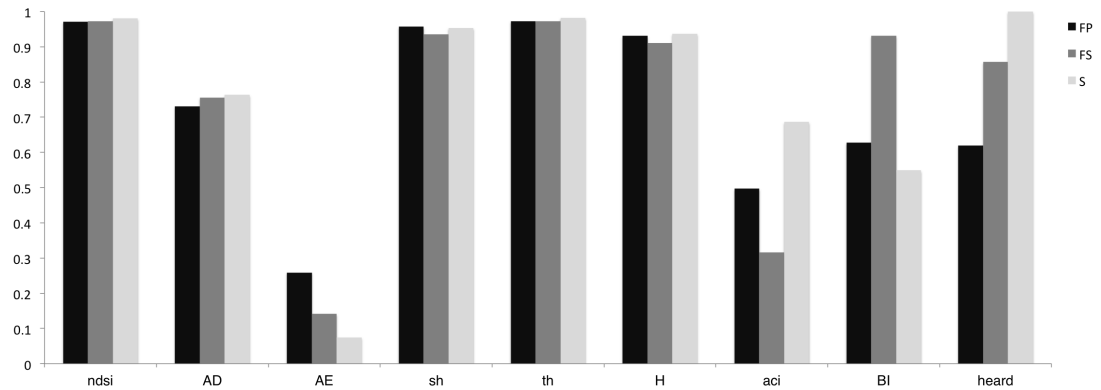
**Figure 1.** Top: Constant-Q spectrograms for 1 min excerpts for each site FP (top) FS (middle) S (bottom). 1 minute resolutions are presented to illustrate the periodic call patterns. These were consistent across the 10 min sampling time in each habitat. Bottom: Mean Constant-Q Spectrum (log amplitude) for FP (green), FS (blue) and S (red)

Common Name	Heard			Seen		
	FP	FS	S	FP	FS	S
Andean Solitaire	1	1	1	1	-	-
Azara's Spinetail	-	-	1	-	-	-
Beryl-spangled Tanager	-	-	-	1	1	-
Blue-winged Mountain-Tanager	-	1	-	-	-	1
Booted Racket-tail	-	1	-	1	-	-
Brown Inca	1	-	-	-	-	-
Brown-capped Vireo	1	-	-	-	-	-
Collared Forest-Falcon	1	-	-	-	-	-
Dusky Bush-Tanager	-	-	1	1	-	-
Empress Brilliant	-	-	1	-	-	1
Flame-faced Tanager	-	-	-	1	1	-
Golden-Crowned Flycatcher	-	-	1	-	-	-
Golden-Headed Quetzal	-	1	1	-	-	-
Gray-breasted Wood-wren	1	1	1	-	-	-
Immaculate Antbird	-	1	-	-	-	-
Lineated Foliage-Gleaner	-	1	1	1	-	-
Long-tailed Antbird	1	-	1	-	-	-
Masked Trogon	1	1	1	-	-	-
Metallic-Green Tanager	-	-	1	-	-	-
Nariño Tapaculo	-	1	-	-	-	-
Orange-bellied Euphonia	1	1	-	1	-	-
Plumbeous Pigeon	1	1	1	-	-	1
Red-faced Spinetail	-	1	1	-	-	-
Roadside Hawk	-	-	1	-	-	-
Rufous-breasted Antthrush	-	1	-	-	-	-
Russet-crowned Warbler	1	1	-	-	-	-
Scale-crested Pygmy-Tyrant	-	1	1	-	-	-
Slate-throated Whitestart	1	-	1	-	-	-
Smoke-colored Pewee	-	-	1	-	-	-
Three-striped Warbler	-	1	-	-	-	-
Toucan Barbet	-	-	1	-	-	-
Tricolored Brush-Finch	-	1	1	-	-	-
Tyrannine Woodcreeper	-	-	-	1	-	-
Uniform Antshrike	-	-	1	-	-	-
Wattled Guan	1	1	1	-	-	-
White-capped Parrot	1	-	-	-	-	-
<b>Total</b>	13	18	21	8	2	3

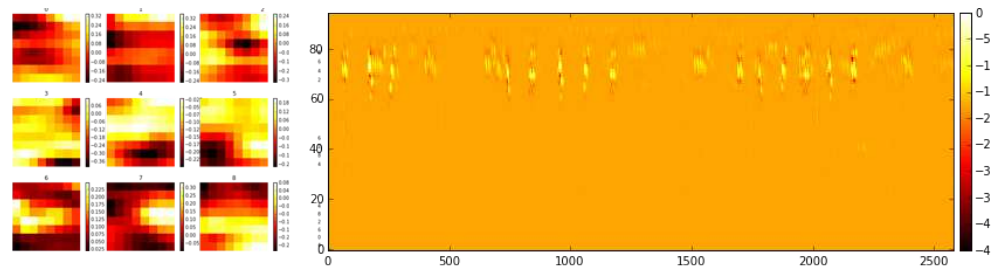
**Table 1.** Field observations for species heard and seen at FP (0600-0610), FS (0619-0629) and S(0639-0649) on June 15th 2014. As only 1 individual of each species was recorded, this data provides a record of species absence-presence as well as abundance for each site.

	NDSI	ADI	AEI	sh	th	H	ACI	BI
<b>FP</b>	0.9716	2.1919	0.2591	0.9567	0.9730	0.9309	18497.14	7.5393
<b>FS</b>	0.9727	2.2684	0.1418	0.9355	0.9729	0.9102	18315.61	11.1780
<b>S</b>	0.9809	2.2909	0.0749	0.9539	0.9825	0.9372	18686.78	6.5867

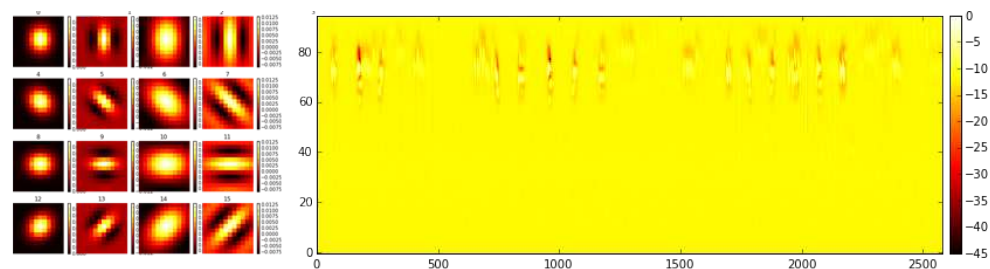
**Table 2.** Acoustic indices values for the three study sites: FP, FS and S shown to 4 decimal places.



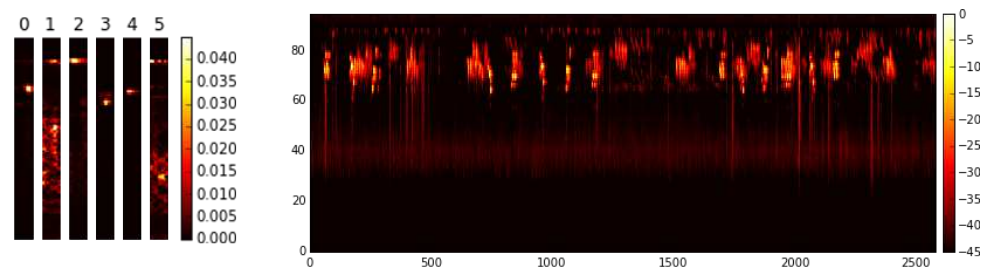
**Figure 2.** Bar plots of indices results for same recording showing indices values 10 minutes of dawn chorus at FP (black), FS (grey) and S (light grey) for NDSI, ADI, AEI, sh, th, H, ACI and BI. Values for AD, ACI and BI are scaled in the ranges 0:3, 1800:1900 and 0:12 respectively. Point count ‘heard’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



(a) Mini-batch Gradient Descent

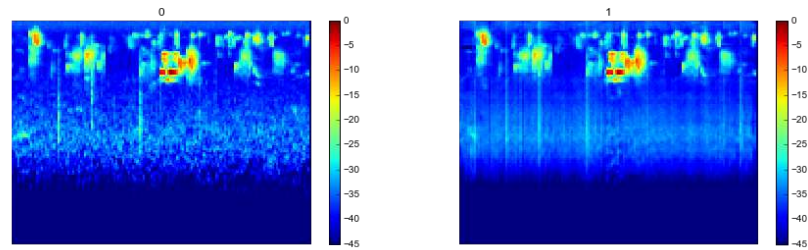


(b) Gabor Field

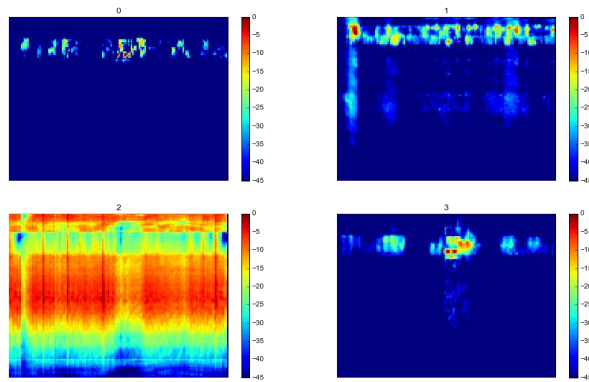


(c) SI-PLCA2D

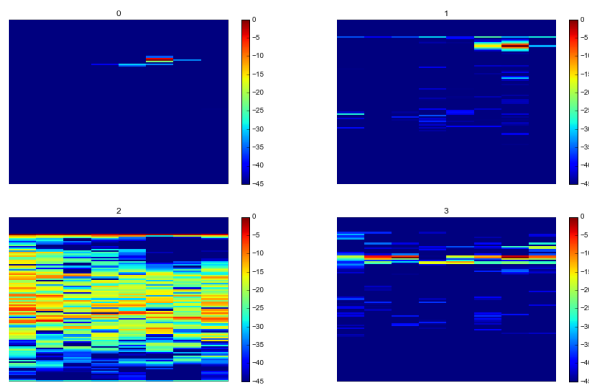
**Figure 3.** Over-complete Dictionaries (left) and sparse-approximations of original spectrogram shown in Figure 1 for (a) Mini-batch gradient learning, (b) Gabor Field Dictionary and (c) SI-PLCA2D Component Dictionary.



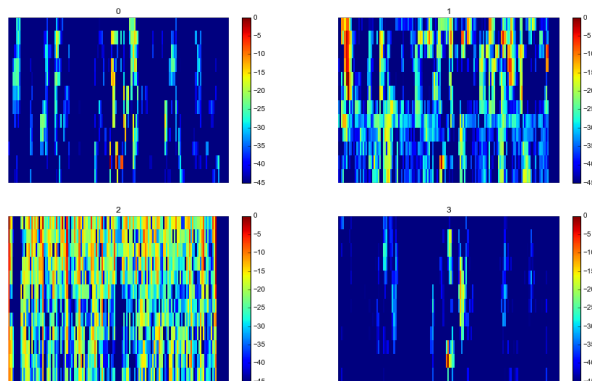
(a) FP: Original Spectrum (left) and Component Reconstructions (right) Original  $S = 4.9439$ ; Reconstruction.  $S = 4.9303$



(b) Individual Component Reconstructions (FP).  $S = (5.3347, 5.5418, 0.0000, 5.4501)$



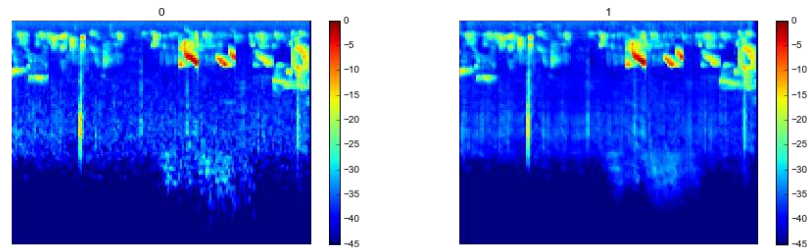
(c) Time-Frequency Kernels (FP).  $S = (1.4749, 3.7474, 5.7675, 3.7552)$



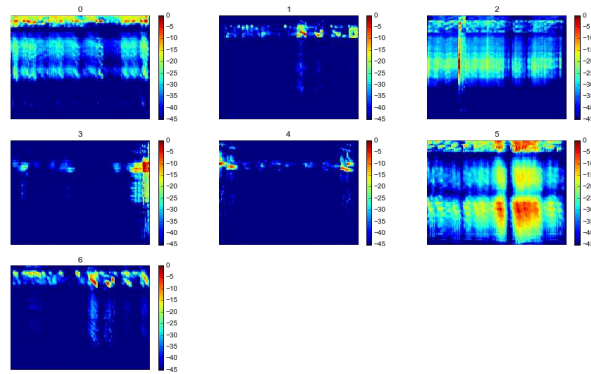
(d) Activation (shift-time) Functions (FP).  $S = (5.0762, 5.7370, 6.2901, 4.6718)$

**Figure 4.** SIPLCA2 outputs for Primary Forest site dawn chorus. Entropy ( $S$ ) values are shown in brackets

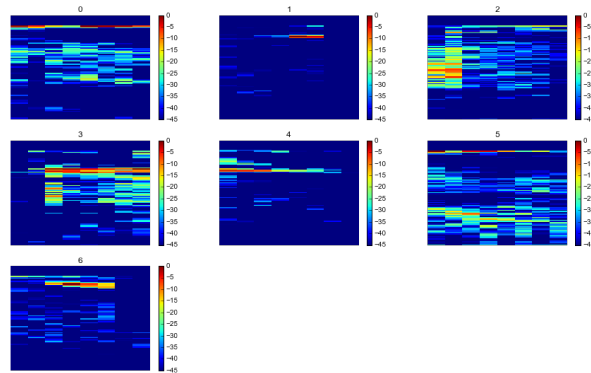




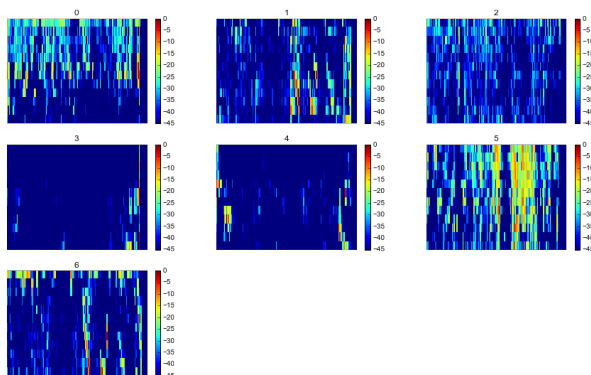
(a) FS: Original Spectrum (left) and Component Reconstructions (right) Original  $S = 5.0291$ ; Reconstruction.  $S = 4.9576$



(b) Individual Component Reconstructions (FS).  $S = (5.4921, 5.4735, 4.3225, 5.5836, 5.1488, 4.7036, 5.6571)$

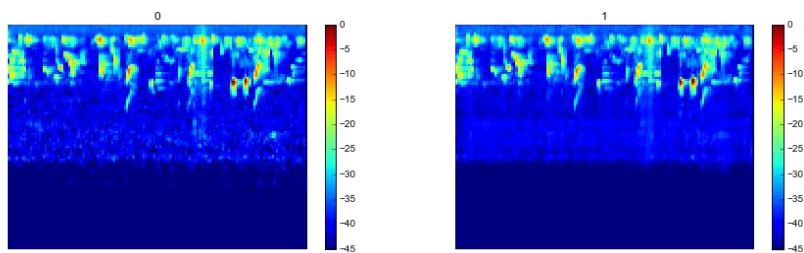


(c) Time-Frequency Kernels (FS).  $S = (4.5131, 2.9622, 4.9541, 4.6132, 3.4991, 5.2196, 3.9248)$

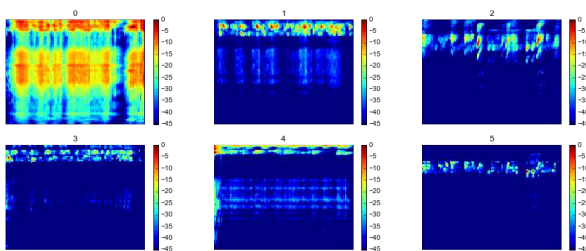


(d) Activation (shift-time) Functions (FS).  $S = (5.7739, 5.1516, 5.9054, 3.6572, 4.2803, 6.0051, 4.9906)$

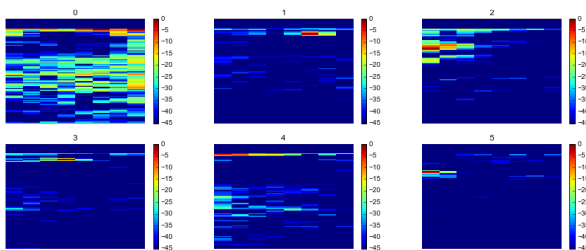
**Figure 5.** SI-PLCA2 outputs for Secondary Forest site dawn chorus. Entropy ( $S$ ) values are shown in brackets.



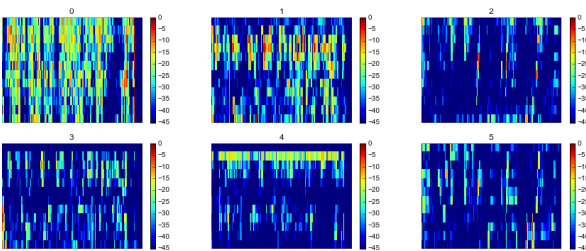
(a)  $S$ : Original Spectrum (left) and Component Reconstructions (right) Original.  $S = 5.1797$ ; Reconstruction  $S = 5.1715$



(b) Individual Component Reconstructions ( $S$ ).  $S = (0.6920, 5.8020, 5.5895, 5.6326, 5.9547, 5.4067)$



(c) Time-Frequency Kernels ( $S$ ).  $S = (5.4939, 3.8297, 3.9353, 3.7362, 4.0415, 3.1377)$



(d) Activation (shift-time) Functions ( $S$ ).  $S = (6.2345, 5.7911, 5.1608, 5.4627, 5.3665, 5.3127)$

**Figure 6.** SI-PLCA2 outputs for Silvopasture site dawn chorus. Entropy ( $S$ ) values are shown in brackets.