Ecological Informatics 21 (2014) 110–119

Contents lists available at ScienceDirect

Ecological Informatics

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

The use of acoustic indices to determine avian species richness in audio-recordings of the environment

Michael Towsey *, Jason Wimmer, Ian Williamson, Paul Roe

Science and Engineering Faculty, Queensland University of Technology, QLD, Australia

ARTICLE INFO

Article history:
Received 21 June 2013
Received in revised form 30 October 2013
Accepted 14 November 2013
Available online 22 November 2013

Keywords:
Acoustic sensing
Biodiversity monitoring
Species richness
Acoustic sampling
Automated bird call analysis
Ecological indices

ABSTRACT

Interpreting acoustic recordings of the natural environment is an increasingly important technique for ecologists wishing to monitor terrestrial ecosystems. Technological advances make it possible to accumulate many more recordings than can be listened to or interpreted, thereby necessitating automated assistance to identify elements in the soundscape.

In this paper we examine the problem of estimating avian species richness by sampling from very long acoustic recordings. We work with data recorded under natural conditions and with all the attendant problems of undefined and unconstrained acoustic content (such as wind, rain, traffic, etc.) which can mask content of interest (in our case, bird calls).

We describe 14 acoustic indices calculated at one minute resolution for the duration of a 24 hour recording. An acoustic index is a statistic that summarizes some aspect of the structure and distribution of acoustic energy and information in a recording. Some of the indices we calculate are standard (e.g. signal-to-noise ratio), some have been reported useful for the detection of bioacoustic activity (e.g. temporal and spectral entropies) and some are directed to avian sources (spectral persistence of whistles). We rank the one minute segments of a 24 hour recording in descending order according to an "acoustic richness" score which is derived from a single index or a weighted combination of two or more. We describe combinations of indices which lead to more efficient estimates of species richness than random sampling from the same recording, where efficiency is defined as total species identified for given listening effort. Using random sampling, we achieve a 53% increase in species recognized over traditional field surveys and an increase of 87% using combinations of indices to direct the sampling. We also demonstrate how combinations of the same indices can be used to detect long duration acoustic events (such as heavy rain and cicada chorus) and to construct long duration (24 h) spectrograms.

1. Introduction

The analysis of acoustic recordings is an increasingly important technique for ecologists wishing to monitor the terrestrial and aquatic environments. Rapid advances in electronic hardware and computing power now make it possible to leave unattended acoustic sensors in exposed locations for several weeks of continuous recording. It is clearly impossible for ecologists to listen to even a small fraction of this audio data. Some degree of automated assistance is essential.

Recorded audio data can contribute to a number of ecological investigations, most obviously the identification of vocal animals. Bird species in particular are regularly surveyed because of their importance as indicator species of environmental health (Gregory and Strien, 2010). There is now a considerable body of published work on the detection of bird vocalizations (Acevedo et al., 2009; Agranat, 2009; Anderson et al., 1996; Brandes, 2008; Chen and Maher, 2006; Digby et al., 2013; Juang and Chen, 2007; McIlraith and Card, 1997; Somervuo et al., 2006). However vocal frog and insect species are also of interest (Brandes et al., 2006) and, in the Australian context, the koala (*Phascolarctos cinereus*, Ellis et al., 2010, 2011) and the cane toad (*Bufo marinus*, Hu et al., 2010) have received particular attention.

In contrast to the bioacoustic interest in individual species, there is a growing interest in *soundscape ecology*, that is, the study of the temporal and spatial distribution of sound through a landscape, reflecting important ecosystem processes and human activities (Kasten et al., 2012; Pijanowski et al., 2011a, 2011b). From this perspective, the soundscape is a finite resource in which organisms (including humans) compete for spectral space (Krause, 2008).

Although this work does not depend on the theoretical perspective of soundscape ecology, it does address the ecological problem of estimating species richness using acoustic recordings. In theory it might be possible to automate this task by preparing individual recognizers for the expected vocal species (which could number 100 or more) but

* Corresponding author at: Science and Engineering Faculty, QUT, GPO Box 2434, Brisbane, QLD 4000, Australia. Tel.: +61 7 3138 9340; fax: +61 7 3138 4438.
E-mail addresses: m.towsey@qut.edu.au (M. Towsey), j.wimmer@qut.edu.au (J. Wimmer), i.williamson@qut.edu.au (I. Williamson), p.roe@qut.edu.au (P. Roe).

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.11.007

the preparation of call recognizers is not an easy task. Lack of suitable training data can be a significant constraint and, even if a recognizer is successfully trained for one species in one locality, the natural geographic variation of calls may render it less effective in a new locality. Our research group has previously addressed the problem of recognizing vocal species by steering a middle path between "one-recognizer-fits-all-species" and "one-recognizer-for-each-species". The former strategy can sacrifice accuracy for generality but the latter is cumbersome and difficult to maintain. We have built a number of recognizers for generic features shared by many bird calls (Towsey et al., 2012).

This paper investigates the problem of determining species richness by approaching it as a problem of computer assisted sampling from long duration audio recordings. We illustrate our approach by focusing on bird species. The traditional method to determine avian richness at a specific location is the *point count* — one or more appropriately skilled persons count all species heard and/or seen within a specified area over a fixed period of time. Clearly this is a time consuming task where sampling effort is constrained by cost. A typical protocol is to visit a site for 20 min each at morning, noon and dusk over several days (Wimmer et al., 2013) but many other protocols are in use (Bibby et al., 1992).

Automated and semi-automated methods offer the advantage that recording devices can be deployed in the field for days or weeks obviating the need for regular field visits by a trained ecologist. However, the use of acoustic recordings to determine avian species richness is a relatively new technology and there are few well-established protocols or even comparisons of automated methods with traditional (Acevedo and Villanueva-Rivera, 2006). Our research group is investigating protocols for the use of environmental recordings (Digby et al., 2013; Wimmer et al., 2010, 2013). Wimmer et al. (2013) have compared a number of acoustic sampling protocols and demonstrated that they can be significantly more efficient than traditional point counts, where efficiency is defined as the number of species identified for equivalent listening effort. They also found that an effective sampling strategy is to select one minute audio samples at random from the 3 h after civil-dawn which encompasses the morning chorus when most birds are most likely to sing.

In this paper we investigate the use of a variety of acoustic indices to direct sampling from recordings of the environment. An acoustic index is a statistic that summarizes some aspect of the distribution of acoustic energy and information in a recording. We present one minute sound segments to a person skilled in bird identification, in an order ranked by indices that describe the acoustic content of the segments. Success is achieved if an estimate of avian species richness is obtained more efficiently (number of species identified for a given listening effort) than using *either* traditional on-site point-counts *or* random sampling from the recordings.

There is a growing body of work on the ecological uses of acoustic indices. It is convenient to divide the indices into three categories: waveform indices, spectral indices and second order indices. Waveform indices include traditional measures such as signal amplitude and signal-to-noise ratio. More recently, temporal entropy ($H[t]$) was introduced to characterize the temporal dispersal of acoustic energy within a recording (Sueur et al., 2008).

Spectral indices include spectral entropy ($H[s]$), a measure of acoustic energy dispersal through the spectrum (Sueur et al., 2008), and spectral peak count (NP), a measure of the average number of peaks in the spectra of the frames through a recording (Gasc et al., 2013). NP was shown to reflect acoustic activity as determined by ear. Pieretti et al. (2011) have introduced the *acoustic complexity index* (ACI), which is a measure of the average absolute fractional change in signal amplitude from one frame to the next through a recording.

The above indices show varying degrees of correlation with bioacoustic activity. To obtain better correlations, a number of second order indices have been proposed. Sueur et al. (2008) demonstrated that $H[t] * H[s]$ is weakly correlated with "acoustic heterogeneity", and that an *acoustic dissimilarity* index, D_f, between two spectra $S1$ and $S2$, where:

$$D_f = \Sigma_f |S1(f) - S2(f)|/2,$$

correlates with *differences* in "acoustic heterogeneity" between recordings.

A convenient property of $H[t]$, $H[s]$ and ACI is that their values are naturally normalized in [0, 1] and can therefore be used to compare recordings of quite different content and amplitude. It is possible to combine non-normalized indices, such as amplitude, by first converting them to a ranked index. For example, Depraetere et al. (2012) calculate the index *Acoustic Richness* (AR) given by:

$$AR = ((\text{rank}(H[t]) \times \text{rank}(M)))/n^2, \text{ with } 0 \leq AR \leq 1,$$

which combines $H[t]$ and M (median of the recording's amplitude envelope) by combining their ranks rather than their values. AR correlates with avian species richness.

Working on the assumption that acoustic activity in the 1–2 kHz and 2–11 kHz bands is likely to be technophony (sound due to machine sources) and biophony (sound due to animal sources) respectively, Joo et al. (2011) have proposed an *acoustic health quality index* (AHQI), more recently called the *normalized difference soundscape index* (NDSI):

$$\text{NDSI} = (\text{biophony} - \text{technophony}) \ / \ (\text{biophony} + \text{technophony}),$$

where biophony and technophony are the summed power spectral densities (PSD) in the appropriate bands (McLaren, 2012).

In this work we investigate the hypothesis that *combinations* of indices will be more useful than single indices to characterize the acoustic content of one minute recordings. Our hypothesis is that a single index cannot capture all that is acoustically relevant in a recording. For example, $H[t]$ is not sensitive to frequency content and none of $H[t]$, $H[s]$, NP and ACI is sensitive to signal amplitude since their calculation 'normalizes' amplitude information. We apply combinations of acoustic indices to two tasks: 1. the efficient estimation of avian species richness and; 2. the detection of common acoustic "regimes" in Australian sub-tropical environmental recordings, namely rain and cicada choruses. A particular feature of our work is that we directly analyze real field-data recorded under normal environmental conditions and with all the attendant problems of unconstrained and undefined acoustic content. In particular, we do not remove audio segments containing wind and rain "noise" prior to analysis.

In this context, the issue of what constitutes "noise" in recordings of the environment requires some clarification. In a non-technical sense, "noise" is a sound where it is not wanted (adopting the classical definition of a *weed*). Because our focus is bird vocalizations, geophony (sounds due to wind, rain, leaf rustle, etc.), anthrophony (sounds due to human sources, traffic etc.) and biophony (sounds due to other animal vocalizations) can be considered noise. However in this study, we use the term "noise" in a technical sense to mean that acoustic energy which remains constant through the duration of a one-minute audio segment regardless of its source. Thus it is possible that the same acoustic source may contribute to both "noise" and "signal". For example, if we assume that crickets are evenly distributed in the landscape around a sensor, there will be a background "murmur" of crickets but the chirps of those crickets closest to the microphone will register as specific acoustic events within the background. Likewise, wind gusts will stand out as specific acoustic events within the constant noise generated by a background of moving air.

2. Materials and methods

2.1. Hardware

Recordings were obtained using custom-developed acoustic sensors (Wimmer et al., 2010). The recording equipment consisted of Olympus DM-420 (Olympus, Pennsylvania, USA) digital recorders and external omni-directional electret microphones. Data were stored internally in stereo MP3 format (128 kbit/s, 22.05 kHz) on high capacity 32 GB Secure Digital memory cards. The units were stored in weatherproof cases and powered by four D cell batteries, providing up to 20 days of continuous recording. Although MP3 is a lossy format, it is designed to reproduce sound accurately for the human ear and has been found suitable for identifying bird calls (Rempel et al., 2005). However we have not investigated the effect that MP3 compression might (or might not) have on the indices described in this paper.

2.2. Data sets

Our data were derived from five days of continuous audio recording, 13th to 17th October 2010 inclusive, obtained at the Samford Ecological Research Facility (SERF) in bush land on the outskirts of Brisbane city, Australia. The dominant vegetation is open-forest to woodland comprised primarily of *Eucalyptus tereticornis*, *Eucalyptus crebra* (and sometimes *Eucalyptus siderophloia*) and *Melaleuca quinquenervia* in moist drainage. There are also small areas of gallery rainforest (with *Waterhousea floribunda* predominantly fringing the Samford Creek to the west of the property) and areas of open pasture along the southern boundary.

This is the same data set described by Wimmer et al. (2013) and see that paper for more detail about the sites and recording methodology. The recording boxes were attached at chest height to a tree near the centre of the survey site. For this paper, we used recordings from only one of the four sites, Site 3 (south-east). A road (some 100 m distant) meant that recordings contained traffic noise, overflying airplanes, dog barks, human speech and severe wind gusts. Recordings were divided into one minute segments. Three expert 'bird observers' (working in collaboration) identified all the audible bird species in each one minute segment as described in Wimmer et al. (2013). The end result was five days of continuous recording with all bird calls tagged at one minute resolution.

2.3. Signal processing

To reduce subsequent computational burden, the *mp3* recordings were re-sampled at 17,640 samples per second (after filtering to remove content above the Nyquist of 8820 Hz) and divided into non-overlapping frames of 512 samples each. Thus there were approximately 4140 frames per one minute of recording (the exact number depending on how the mp3 recording was split). The final fractional frame in each minute was discarded. It should be noted that almost all of the acoustic activity of interest to us is below the Nyquist.

We calculated fourteen acoustic indices for each minute audio segment. Eight indices were derived from a *wave envelope* which was, in turn, derived from the *maximum absolute* value in each frame. Note that the number of values in the wave envelope therefore equals the number of complete frames in a one-minute audio segment.

The remaining six indices were derived directly or indirectly from one minute spectrograms. FFTs were calculated using a Hamming window. The spectrum derived from each frame has 256 frequency bins, spanning 8820 Hz (34.45 Hz per bin). Each spectrum was smoothed with a moving average filter (window width = 3). We removed from further consideration the lowest 14 bins (0–482 Hz) in order to avoid traffic noise that contaminated recordings. Non-removal of these low frequency bands meant that the extracted indices were dominated by non-avian acoustic sources. In consequence, a one minute spectrogram had $242 * 2067 \approx 500{,}000$ cells. Spectrograms were "noise reduced" using a modification of *adaptive level equalization* (Lamel et al., 1981) applied to every frequency bin independently (Towsey, 2013b). Adaptive level equalization has the effect of removing continuous background acoustic activity and setting that level to zero amplitude. Thus it becomes possible to define a single absolute threshold for the detection of an acoustic event that spans multiple frequency bins.

Spectrograms were not converted to decibels in order to preserve values appropriate for subsequent calculations of ACI and spectral entropy.

2.4. Acoustic indices

This section describes the derivation of each of the fourteen acoustic indices. Some of them are standard (such as signal-to-noise ratio) and some are specifically directed to features expected of biological sources (such as persistence of whistles). Some are real quantities (such as dB values) and some are naturally normalized (such as the different measures of acoustic entropy). Note that we did not calculate 'second-order' indices such as $H[t] * H[s]$, *Acoustic Richness* and *NDSI* because it was our intention to derive 'second-order' indices as weighted combinations of 'raw' indices. To obtain weighted combinations of indices it was necessary to normalize all indices in [0,1]. The minimum and maximum values for normalizing are shown in brackets at the end of each description. Values below the minimum or above the maximum are truncated to 0 and 1 respectively.

1. *Average signal amplitude*: calculated as the average amplitude of the wave envelope. This average value is converted to decibels using:

   $$\mathrm{dB} = 20.\log_{10}(A_{av}),$$

   where A_{av} is the average amplitude of the envelope samples. Decibels have a negative value because A_{av} is in the range [0, 1] (normalizing min/max = −50 dB/−3 dB).
2. *Background noise*: estimated from the wave envelope using the method of Lamel et al. (1981) as described in Towsey (2013b). The value is given in decibels. Note once again that the term *background noise* has a technical definition. It is that acoustic activity removed using the method of Lamel. Background noise over the five days ranged from −47 dB to −9 dB (normalizing min/max = −50 dB/−3 dB).
3. *Signal-to-noise ratio* (SNR): the decibel difference between the maximum envelope amplitude in any minute segment and the background noise (Index 2) (normalizing min/max = 3 dB/50 dB).
4. *Acoustic activity*: the fraction of frames within a one minute segment where the signal envelope is more than 3 dB above the level of background noise (Index 2) (normalizing min/max = 0.0/1.0).
5. *Count of acoustic events*: the number of times that the signal envelope crosses the 3 dB threshold (used to calculate acoustic activity, Index 4) from below to above (normalizing min/max = 0/140).
6. *Average duration of acoustic events*: an acoustic event is a portion of recording which starts when the signal envelope crosses above the 3 dB threshold and ends when it crosses below the 3 dB threshold. The average duration of the acoustic events so identified is measured in milliseconds (normalizing min/max = 0/500 ms).
7. *Entropy of the signal envelope* (henceforth *temporal entropy*, $H[t]$): The squared amplitude values of the wave envelope were normalized to unit area and treated as a probability mass function (*pmf*). The entropy (H) of the signal was calculated as:

   $$H(t) = -\sum_i \log_2(pmf_i) \;/\; \log_2(N),$$

   where i is an index over all integers $0 - N - 1$ and N is the number of values in the wave envelope (normalizing min/max = 0.5/1.0).
8. *Acoustic complexity index* (ACI): for each frequency bin over the entire one minute recording, calculate the average absolute fractional

change in spectral amplitude from one spectrum to the next. The ACI for the entire recording is the average over all frequency bins. See Pieretti et al. (2011) for more detail (normalizing min/max = 0.2/0.8).

9. *Mid-band activity*: the fraction of spectrogram cells in the mid-band (482 Hz–3500 Hz) where the spectral amplitude exceeds 0.015. The suitability of this threshold was determined by trial and error. It is low because background noise has already been removed. Background noise over the five days of recording was typically between −45 and −35 dB. Thus an amplitude threshold of 0.015 corresponds to approximately 6 dB above background (normalizing min/max = 0.0/1.0).
10. *Entropy of the average spectrum* (henceforth *spectral entropy*, *H*[s]): Calculated from the 482–8820 Hz portion of the spectrogram.
    a. Calculate the average of all the spectra.
    b. Calculate the entropy of the average spectrum as described for *Index 7*, except that here *N* = spectral length = 242 (normalizing min/max = 0.5/1.0).
11. *Entropy of spectral maxima* (*H*[m]): calculated from the 482–8820 Hz portion of the spectrogram (normalizing min/max = 0.0/1.0).
    a. Determine the frequency bin in each spectrum having maximum amplitude.
    b. Prepare a histogram of bin IDs having maximum spectral amplitude.
    c. Calculate the entropy of the resulting histogram (of bin counts) as described for *spectral entropy* (*Index 10*).
12. *Entropy of the spectral variance* (*H*[v]): calculated at the same time as *spectral entropy* (*Index 10*) but replacing the average of each frequency bin (over all frames) by its variance (normalizing min/max = 0.0/1.0).
13. *Spectral diversity*: measured as the number of distinct spectral clusters in a one minute recording segment. Like *spectral indices 10*, *11* and *12*, spectral diversity was expected to be a helpful indicator of spectral richness and therefore of species richness. We implement a modified version of the ART1 unsupervised iterative learning algorithm designed to cluster binary input vectors (Grossberg and Carpenter, 2002). The modifications are primarily to speed convergence of clustering (Towsey, 2013a).
    a. Reduce the length of each spectrum to one-third (from 242 to 80) by averaging values in consecutive groups of three. (Last two values ignored.) This step is to reduce computational burden and to reduce spectral detail.
    b. Convert each spectrum (length = 80) to a binary vector using an amplitude threshold = 0.07.
    c. Determine the number of spectral categories using a clustering algorithm. To reduce computational burden, parameters are adjusted to achieve fast convergence.
    d. Prune the resulting list of spectral clusters by removing clusters that contain only one member and clusters whose prototype contains only one non-zero value.
    The final cluster count is sensitive to the choice of spectra that seed the clustering process and to other parameter choices. Nevertheless, it is generally indicative of the spectral diversity in a one minute recording. The threshold of 0.07 is relatively high (~16 dB above typical background noise) to limit detection to birds close to the microphone and to reduce the number of resulting spectral clusters. The maximum cluster count in any minute over the five days of recording was 16. Reducing the amplitude threshold to 0.03 (~10 dB) increased the maximum cluster count to 50 (normalizing min/max = 0/20).
14. *Spectral persistence*: each frame in a one minute segment is assigned to its nearest spectral cluster as determined in *Index 13*, step c. Spectral persistence occurs when consecutive frames are assigned to the same spectral cluster and is defined as the average duration (in milliseconds) of those clusters which persist for longer than one frame (normalizing min/max = 0/200 ms).

2.5. Measuring performance

We use four methods to describe the efficiency of determining species richness. The first is to plot a graph of cumulative species identified versus sample number (Fig. 4). This is a simple way to compare the effectiveness of a small number of sampling protocols. However when making many comparisons, it is more useful to compare *either* the number of one minute samples required to identify some fixed percent of the species known to be present (equivalent to a horizontal line cutting a set of species accumulation curves) *or* the number of species identified for a fixed number of one minute samples (equivalent to a vertical line cutting a set of species accumulation curves). For the former approach, we employ the nomenclature *S75%* (or *S50%*) to mean the number of samples required to identify 75% (or 50%) of species known to be present in a 24 hour period. 75% was taken to be a satisfactory compromise between making the task unnecessarily difficult and producing an inadequate result. For the latter approach, we chose to compare protocols over 60 processed samples because this is equivalent to the three 20 minute listening sessions employed in our standard point-count protocol.

In seeking a sampling protocol that efficiently determines species richness, we desire one that not only selects the most species in 60 samples but one which is also consistent over many days in different kinds of weather. To this end, we calculate a fourth measure of performance, the *inverse coefficient of variance* over the five days of the study, that is, the mean number of unique species recognized in 60 samples per day over five days divided by the standard deviation over the five days. The inverse coefficient of variance (1/CV) is sometimes interpreted as a signal-to-noise ratio, where high values are desirable.

2.6. Training a recognizer for the cicada chorus

As a feature set for recognition of the cicada chorus, we calculated six of the above indices (background noise, SNR, acoustic activity, mid-band activity, *H*[t] and *H*[s]) at 10 second resolution as opposed to 60 second resolution. The data set consisted of 113 segments (each 10 s long): 12 containing cicada chorus, 43 segments of heavy rain and 58 consisting of a mixture of day/night recordings with varying levels of acoustic activity due to avian sources. Although no rain fell in the days reported in this study, we wished to use the recognizer for other recordings where rain was present. Because heavy rain has a variable appearance in the spectrogram (depending on the variable nature of the nearby ground cover, leaf size and the resonant response of flat surfaces near the microphone, including the reverberant qualities of the sensor box itself), four times more rain segments were included in the data set than cicada segments. Light rain was not included as a category because it is too difficult to determine from 10 second recordings alone. The data were used to train a *See5* decision tree (Quinlan, 1993), which adds branches (nodes) using the information gain of the different features and finally prunes branches to reduce estimated classification error.

Decision trees were trained to classify the 10 second audio segments as *cicada chorus*, *heavy rain* or *other*. Cicada noise and rain were incorporated in the same classifier because both, when they occur, are dominant acoustic events in a 24 hour recording that might be confused on the basis of amplitude and signal-to-noise ratio. A significant advantage of decision trees for our purpose is that they output rules which can be inserted as C# code into our analysis software. We used multiple runs of 10-fold cross-validation to build a variety of trees and selected one having a combination of high accuracy and few nodes.

3. Results

3.1. Species counts (ground truth)

The number of unique species calling minute by minute over the five days is shown in Fig. 1. During the morning chorus of a typical day at this site, more than ten unique call types can be identified within a minute period. It should be noted that the month of October is the dominant avian breeding season when their vocalizations can be expected to be most numerous. It is estimated that some 80% of species encountered have multiple call types and consequently it is more accurate to refer to *call types* than to *species*. However for convenience we will use the terms interchangeably. As is to be expected, the number of species calling at night is much reduced.

Over the five days, 77 unique call types were identified but the number of call types in a typical day is about three-quarters of this number (Table 1) because not all birds call every day. An obvious feature in both Fig. 1 and Table 1 is the reduced number of calls on October 16th. This was due to strong gusting winds which developed on the afternoon of the 15th October and persisted to early morning of 17th October. Likewise the number of 'avian-active' minutes (containing at least one identifiable bird call) was nearly halved on 16th October. The combined effect of fewer calling species and reduced calling rate was to reduce *call density* by a factor of 2–3. Call density is defined as the average number of different species calling per 'avian-active' minute (see bottom row of Table 1).

The effects of wind could be detected in measurements of background noise (Fig. 2). Recall that the term "noise" is used here in the technical sense of continuous acoustic activity without reference to the kind or source of the activity. For the first two days, background noise shows a typical 24 hour profile that we have observed at other sites when the weather is calm. In particular, noise decreases steadily during the night as insect and animal activity declines and then increases sharply with the onset of the morning chorus. However during extended periods of high wind (e.g., 16th October), *both* the absolute level of background noise *and* its variability over time increase.

Table 1
Counts and density of unique bird calls over the five days of the study.

	13th Oct	14th Oct	15th Oct	16th Oct	17th Oct
# unique species	62	58	62	45	62
# active minutes	877	872	850	449	884
Call density	5.9	4.7	4.6	2.2	4.2

3.2. Acoustic indices

A graphical presentation of the indices is a useful way to recognize important acoustic episodes in a 24 hour recording. Fig. 3 displays 16 tracks of indices for the morning (0300 to 0900 h) and evening (1700 to 1900 h) of October 13th. Abbreviated names of the indices are on the right. The first 14 tracks are of those indices described in Section 2.4. The *Cicada* and *WeightedIndex* tracks will be explained in later sections. Time of day in hours is indicated at top and bottom of the image. 0300 to 0400 h illustrates the typical nighttime pattern with high temporal entropy and low ACI. An obvious morning chorus starts around 0440 h with a rise in SNR, ACI and mid-band cover and a drop in temporal entropy and spectral entropy. The most obvious feature in the evening is a cicada chorus after 1800 h characterized by an increase in background noise, a decline in SNR and a marked decline in the three spectral entropies due to a dominance of acoustic energy in a relatively narrow bandwidth.

3.3. Sampling recordings with prior knowledge

In order to interpret the effectiveness of different sampling protocols designed to maximize species identified, we first determined some performance benchmarks. The upper (optimum) performance limit is set by the known distribution of call types through the 1440 min of each day. Row 1 of Table 2 shows the minimum number of samples required to detect 25%, 50%, 75% and 100% of the species calling on 13th October 2010 given complete prior knowledge of call distributions. Employing a 'greedy' algorithm, which selects as its next sample the one minute recording containing the *most so-far unidentified* species, only 23 samples are required to identify all species. (Note: this greedy algorithm is *not* guaranteed to find the global minimum values of *S25%–S100%*.)

Another, lesser degree of prior knowledge is the observed counts of unique species or call types per minute as shown in Fig. 1. In other words, we know the species counts but not what those species are. When one-minute segments are selected in descending order of known species counts, the resulting accumulation curve falls below that of the greedy algorithm but above that of random sampling (Fig. 4 and Table 2). We propose that this curve is a realistic performance target for methods without prior knowledge. Note that accumulation curves tend to converge with increasing sample number because, if one samples often enough, every species will be found.

3.4. Random sampling

Another appropriate benchmark is the performance of random sampling over the 1440 minute segments in a 24 hour period. Average values and standard deviations for *S25%*, *S50%*, *S75%* and *S100%* (over 5000 trials) are shown for 13th October in Table 2, bottom row. Note that with the greedy algorithm, 2.5 times as many samples are required to move from *S75%* to *S100%*. By comparison, 10.3 times as many samples are required with random sampling. It becomes increasingly difficult to identify the low calling-rate species when sampling is random. Five species in the study area were heard only once in the five days. In fact, calling frequency displays the "20–80" relationship, that is, many species call infrequently while a few species call frequently. Hence random sampling is an inefficient way to detect low calling rate species.

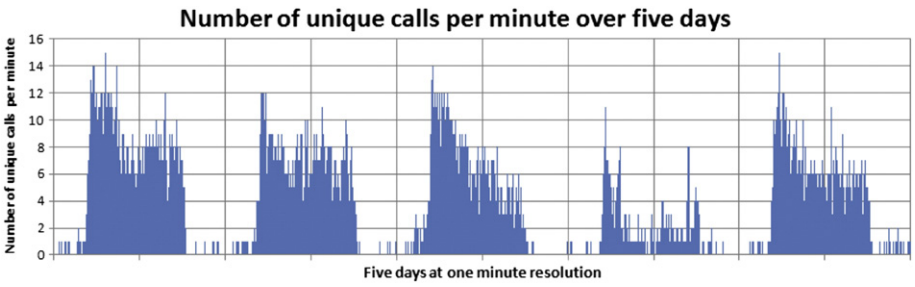

Fig. 1. Number of unique bird call types per minute from 13th to 17th October, 2010. The vertical gridlines are at 12 hour intervals.

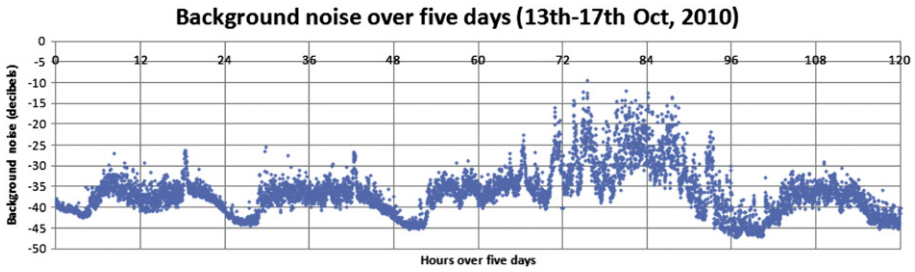

Fig. 2. Background noise (dB) over five consecutive days. Recording dates are 13–17 Oct, 2010. The vertical gridlines are at 12 hour intervals.

The value of S75% varied significantly over the five days (Table 3), the lowest value being for the first day of the study (13th October) and the highest value being for the day of gusting winds (16th October). The question arises as to what factor most determines day to day variations in the efficiency of random sampling. Fig. 5 illustrates that *S75%* (over the five days of the study) is strongly correlated with call density ($r^2 = 0.93$). This is to be expected because increasing call density means that any 'bio-acoustically active' one minute sample is likely to include a larger number of unique call types or different species. The effect of wind is to reduce calling frequency and consequently to reduce the joint probability of two or more species calling in the same minute (assuming minimal interactions between species).

3.5. Sampling based on single indices

In the previous section, we established benchmarks for sampling performance. In this section we examine sampling protocols informed by the acoustic indices described in Section 2.4. In this experiment, the 1440 minute segments of each day were ranked by the value of a single index and sampled in descending order (ascending order in case of H[t]). The percent of call-types (known to be audible in the 24 hour period) identified in the first 60 samples was compared across the five days for each of the indices. To illustrate the method, the results for just four indices are shown in Table 4 and compared with two benchmarks: random sampling (first row in Table 4) and sampling in descending order of the already known unique call count (second row in Table 4). The averages and standard deviations for random sampling are the outcome of 5000 trials. The results for other protocols can then be expressed as a z-score with respect to the mean and standard deviation of random sampling. Note that sampling based on average signal amplitude performs worse than random sampling on all days (row 3 of Table 4 and Fig. 4), particularly on those days which experienced gusty winds. The indices which identified most species in 60 samples were *H*[v], ACI and spectral diversity (or spectral cluster count). Note that the ACI protocol yielded a lower 1/CV score than the other two protocols, presumably because it was more responsive to the presence of gusting wind. (Recall that having a higher value of 1/CV over the five days is a desirable property for an index.)

The average z-score and the inverse coefficient of variation for all the individual indices over the five days are shown in Table 5. A confidence of *N%* can be interpreted as follows: the total species identified in 60 samples using the protocol is expected to exceed that of *N%* of random sampling trials over the same 24 hour recording. A higher z-score and confidence implies a more efficient sampling protocol. A confidence of <50% implies worse performance than random sampling. This could happen for two reasons: 1. the ranking index does not reflect avian sources; or 2. the ranking index selects consecutive samples containing the *same* calling species. Recall that sampling efficiently for species richness requires consecutive samples to contain *diverse* and *different* species.

The best performing index was spectral diversity with a confidence of 92.8% with respect to random sampling. However, in terms of consistency over multiple days (1/CV values), the best index was *H*[v].

3.6. Sampling based on combinations of weighted indices

It was our original intention to find optimally weighted combinations of acoustic indices using the 13th October recording as training

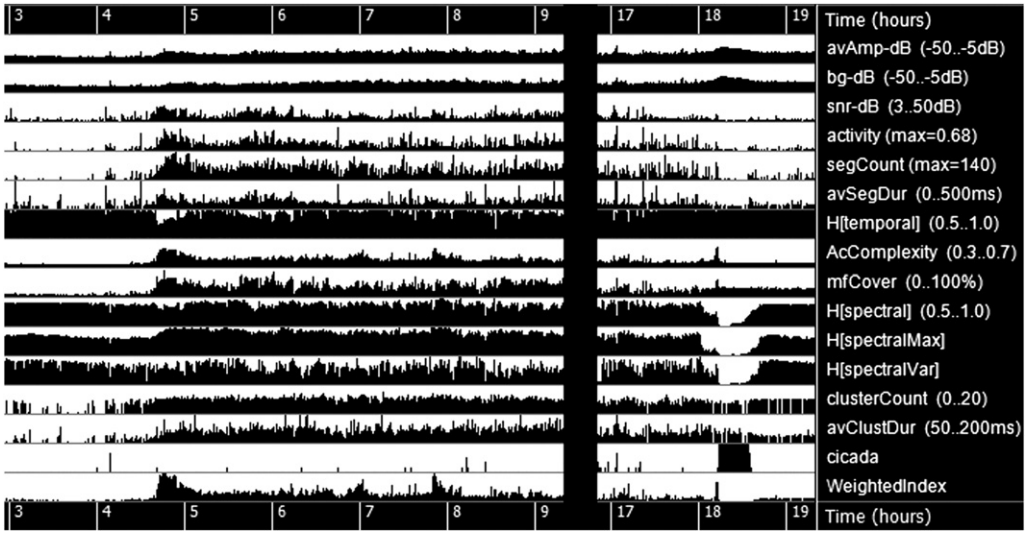

Fig. 3. Tracks of acoustic indices for the morning (0300 to 0900 h) and evening (1700 to 1900 h) of the 13th October 2010. Hour of day is marked at top and bottom of image. Index names in right hand panel refer to the indices described in Section 2.4. AcComplexity = ACI.

Table 2
The number of one minute samples required to identify 25%, 50%, 75% and 100% of calling bird species on 13th October 2010.

Sampling protocol	*S25%*	*S50%*	*S75%*	*S100%*
Using prior knowledge of species distributions [a]	1	4	9	23
Using prior knowledge of species counts per minute	2	5	54	706
Random sampling (no prior knowledge) [b]	7.8 ± 3.4	28.9 ± 8.3	109 ± 28.5	1224 ± 176

[a] See text for a definition of the greedy sampling algorithm.
[b] The averages and standard deviations for random sampling are the outcome of 5000 trials.

Table 3
S75% averages and standard deviations over five consecutive days for the random sampling protocol (see Section 2.5 for definition of *S75%*).

Sampling protocol	13th Oct	14th Oct	15th Oct	16th Oct	17th Oct
Random sampling[a]	109 ± 29	162 ± 41	128 ± 32	290 ± 73	189 ± 65

[a] The averages and standard deviations for random sampling are the outcome of 5000 trials.

data and the remaining days as test data. It quickly became apparent that this approach would not work. For example, by using exhaustive search through weight-space (at 0.1 resolution), we obtained four indices (SNR, ACI, H[v], H[s] combined in the weight ratios of 0.8:0.7:1.0:1.0) that could identify 81% of 13th October species in the first 60 ranked samples. However when this ranking protocol was applied to the following four days, the first 60 samples yielded 57%, 8%, 9% and 58% of species present on those days. This combination of indices, optimized for 13th October, performed particularly poorly on the high wind days of 15th and 16th October. This is an extreme example of the machine learning phenomenon known as *over-learning* — the training data are well learned but performance on the test data is poor. It arises because the composition of the training data is not representative of the test data. In our case, the training data did not contain acoustic noise due to wind.

We were not able to easily incorporate the objective function for this task (a mapping from weighted indices to the number of unique species calls in the first 60 of 1440 ranked samples) into a standard machine learning algorithm. Consequently we report results for weighted combinations of indices that performed well, on average, over all five days. Despite this limitation of our methodology and the relatively small number of study days, we believe that the results provide useful insights in how to approach the problem of combining acoustic indices for determining species richness.

Spectral diversity was the single best performing index and also resistant to the acoustic effects of strong gusting winds (Table 5). We combined spectral diversity with other high performing indices, but kept spectral diversity as the highest weighted index. We report the best performing weighted combinations in Table 6. To aid comparisons, the best performing *single* index is included (row 1 of Table 6). It is apparent that the weighted combinations of indices (rows 2, 3, 4, 5 of Table 6) have higher z-scores (with respect to random sampling) and higher 1/CV scores than the best single index. A combination of five indices yielded the highest z-score and a combination of three indices yielded the highest 1/CV score.

An additional performance benchmark (using prior knowledge) is that obtained by selecting the 60 samples having the highest species counts per minute (bottom row, Table 6 and second curve from top in Fig. 4). Recall that we proposed this as a practical performance target for methods which do not use prior knowledge. The best performing combination of five indices (FS5) achieves a lower z-score than this benchmark ($1.9 < 3.3$) but a similar 1/CV value ($2.25 \approx 2.20$). The highest value of 1/CV (2.67) was achieved by combining three indices.

3.7. Use of a dawn bias

Wimmer et al. (2013) report that restricting random sampling to the three hour period after civil dawn is an efficient way to detect bird species because call density and species number are highest during the dawn chorus. In order to replicate this approach, we multiplied the index scores of the 180 one-minute segments following civil dawn by a *dawn-bias* prior to ranking the 1440 one-minute segments. This had the effect of pushing the dawn chorus segments slightly up the final ranked list. We did this for various values of dawn-bias and for various combinations of indices. Because the general trend was clear and consistent, we show only one result (Table 6) using a dawn bias = 1.05. Incorporating the dawn-bias had little effect on average z-score but reduced consistency (1/CV). We therefore did not incorporate a dawn-bias in any subsequent investigations.

3.8. Combining random sampling with index ranking

We wanted to consider the possibility that the various ranking protocols were assigning high scores to one-minute segments containing the *same* calling species. We approached this possibility by ranking the one-minute segments using the best combination of five indices (FS5 in Table 6) but then random sampling with a probability inversely proportional to the rank of the segment. This biased samples towards the highest ranked but allowed for deviation from the strict ranked order. This protocol did not perform as well as strictly ranked sampling

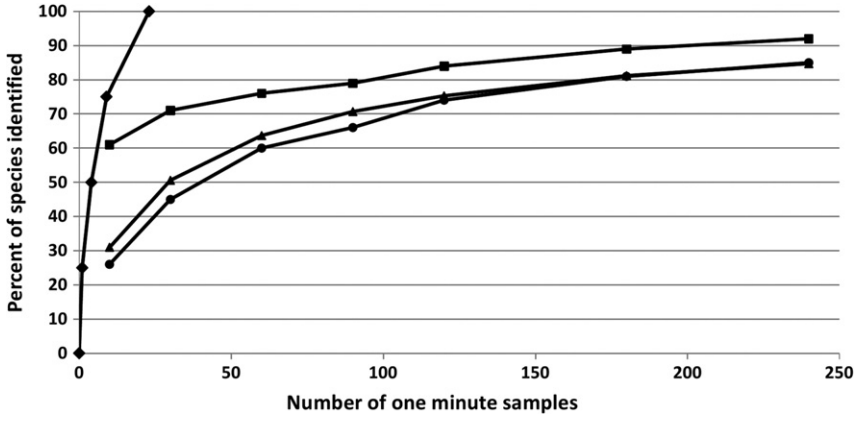

Fig. 4. Species accumulation curves for four sampling protocols on 13th October 2010. -◆- greedy sampling with prior knowledge of species distributions; -■- sampling with prior knowledge of unique species counts but not what those species are; -▲- random sampling; -●- sampling in descending order of average signal amplitude.

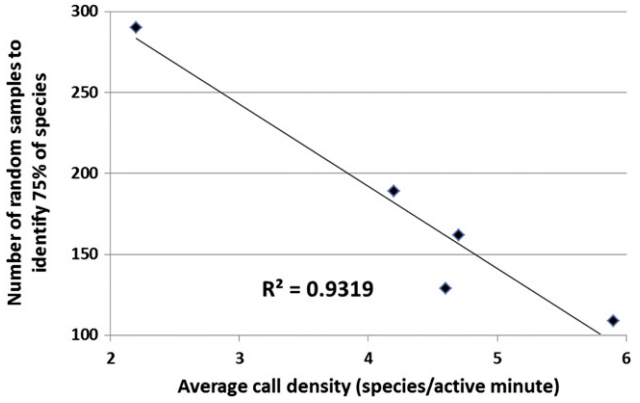

Fig. 5. The dependence of the *S75* sample number on call density. There is one data point for each day of recording.

(second bottom row, Table 6). We explored various sampling distributions but none of them performed as well as strictly ranked sampling.

3.9. Comparison with field survey

Finally we compared the best performing protocols with field survey (Table 7). Recall that we used *species identified in 60 samples* as a standard measure so that we could make valid comparisons with field survey methods. Wimmer et al. (2013) have determined that 60 min of listening time in a field survey is equivalent effort to processing 60 one-minute audio samples (ignoring travel time). Although, on average, it takes 2 min to listen to a one minute recording (due to inevitable play-backs), in the standard field survey two people are listening in concert. The summary averages (rightmost column, Table 7) demonstrate that random sampling will detect ~53% more species than field survey and five weighted indices will detect ~87% more species than field survey.

The question arises as to which one-minute segments are selected using these more efficient sampling techniques. The bottom track in Fig. 3 (labeled *WeightedIndex*) displays the score obtained using a weighted combination of five indices (FS5 in Table 6). It is immediately apparent that the morning chorus segments are most highly ranked followed by the minutes before 8 am which also contain many bird calls. By contrast the one-minute segments in the evening, coinciding with a deafening cicada chorus, are lowest ranked.

3.10. Recognizing the cicada chorus

Ten-fold cross-validation on the data set described in Section 2.5 yielded an average accuracy of 76%. However achieving high accuracy on the data set was not the primary objective of this work because accuracy depends too much on the difficulty of the examples included in the data set. Rather our purpose was to select one of the ten trained trees that combined moderately high accuracy with parsimonious structure (few nodes). Because no rain occurred in the five days of our study, we extracted the following rule applicable to cicadas:

if((*H*[*s*]<0.6) && (background noise>−24 dB)) then *cicada chorus*; else *none*.

This rule is consistent with the observation that a cicada chorus produces continuous high amplitude noise in a relatively narrow part of the spectrum. A low value of *H*[s] indicates spectral concentration of acoustic energy. By contrast, heavy rain produces a high *H*[s] value (broadband acoustic energy) due to percussive effects on surfaces near the microphone. The classifications for six consecutive 10 second segments were averaged to give a 'cicada score' for each minute. The resulting scores for the morning and evening of 13th October are shown in the second track from bottom of Fig. 3. Note that the classifier clearly identifies the cicada chorus at 6:15 pm. It also produces some false positives which could be filtered using a thresholding rule which requires a high and persistent cicada score.

Table 5
The average z-score (over 5 days) and corresponding confidence and 1/CV values for different sampling protocols. Protocols are ordered by their average z-score. Only four protocols perform better than random sampling.

Ranking index	A comparison of 60 ranked samples compared with 60 random samples		
	Av. z-score	Conf.	1/CV
13. Spectral diversity	1.46	92.8%	1.53
8. ACI	1.14	87.3%	0.74
12. *H*[v]	1.01	84.6%	1.64
10. *H*[s]	0.22	58.5%	0.28
9. % mid-band activity	−0.15	43.9%	−0.18
11. *H*[m]	−0.27	39.4%	−0.08
3. SNR	−0.3	37%	−0.24
6. Av event duration	−0.43	33.4%	−0.37
7. Temporal entropy	−0.43	33.4%	−0.39
14. Spectral persistence	−0.57	28.5%	−0.25
4. % frame activity	−0.69	25%	−0.62
5. Number of acoustic events	−1.03	15.2%	−0.50
2. Background noise	−2.9	0.0%	−0.75
1. Av. amplitude	−3.3	0.0%	−0.97

4. Discussion

The increasing use of acoustic sensors to monitor biodiversity presents many technical and procedural problems. Sensor reliability has improved and costs have decreased to the point where many more hours of recording are available than can be listened to. In this work we have approached the "data deluge" problem as one of sampling, an approach with which ecologists are already familiar. We were motivated by the hypothesis that sampling for species richness could be made more efficient by using combinations of acoustic indices to order sampling rather than single indices. The hypothesis has been confirmed. Ranking one-minute segments by scores obtained from combinations of weighted indices yielded more species than ranking with single indices or random sampling (Table 6). Selecting 60 samples ranked by five weighted indices identified 22% more species than 60 random samples and 87% more than a 60 minute field survey (Table 7).

Table 4
The percent of known species identified and z-scores (in brackets) calculated with respect to random sampling for that day.

Sampling protocol	13th Oct.	14th Oct.	15th Oct.	16th Oct.	17th Oct.	5 day average	1/CV[a] (av/sd)
60 random samples [b]	63.7% ± 4.8	57.7% ± 5.0	59.0% ± 5.4	40.6% ± 7.6	55.8% ± 10.6	55.4% ± 6.7	–
Rank on known call count/minute	76% (2.56)	76% (3.66)	79% (3.70)	82% (5.44)	66% (0.96)	75.8% (3.3 ± 1.5)	2.2
Rank on av. signal amplitude	60% (−0.8)	52% (−1.1)	11% (−9.3)	2% (−5.1)	52% (−0.4)	35.5% (−3.3 ± 34)	−1.0
Rank on *H*[v]	73% (1.94)	62% (0.86)	58% (0.06)	51% (1.37)	65% (0.87)	61.8% (1.0 ± 0.6)	1.6
Rank on *ACI*	63% (−0.1)	60% (0.46)	65% (1.46)	71% (4.00)	55% (−0.1)	62.8% (1.1 ± 1.5)	0.7
Rank on spectral diversity	69% (1.1)	72% (2.86)	61% (0.66)	58% (2.29)	60% (0.40)	64.0% (1.5 ± 1.0)	1.5

[a] 1/CV = inverse coefficient of variation which can be interpreted as a signal-to-noise ratio.
[b] The averages and standard deviations for random sampling are the outcome of 5000 trials.

Table 6
The average z-score (with respect to random sampling) and corresponding confidence and 1/CV values for different sampling protocols. Average z-scores are obtained by averaging performance over 5 days.

Number of indices	60 samples ranked by combinations of weighted indices			
	Feature set & feature weights	Av. z-score	Conf.	1/CV
1 index	FS1: Spectral diversity	1.46	92.8%	1.53
2 indices	FS2 = (Spectral diversity, ACI) Weights = (1.0, 0.1)	1.78	96.3%	2.13
3 indices	FS3 = (Spectral diversity., ACI, H[v]) Weights = (1.0, 0.1, 0.1)	1.74	95.9%	2.67
4 indices	FS4 = (Spectral div., ACI, H[v], H[s]) Weights = (1.0, 0.1, 0.1, 0.1)	1.82	96.5%	2.49
5 indices	FS5 = (Sp.div., ACI, H[v], H[m], H[t]) Weights = (1.0, 0.1, 0.1, 0.1, 0.1)	1.89	97.1%	2.25
5 indices	FS5 + dawn bias	1.81	96.5%	1.69
5 indices	Random sampling from FS5 ranked indices	1.66	95.1%	2.22
–	Prior knowledge of call counts/min	3.30	99.9%	2.20

Our work confirms other reports that indices such as *H*[t], *H*[s] and ACI are useful indicators of bio-acoustic activity generally and of species diversity in particular. In this paper we have introduced several new indices, the entropy of spectral variance (*H*[v]), the entropy of spectral maxima (*H*[m]) and spectral diversity (count of spectral clusters). The last proved to be useful because it yielded the most consistent results over days of fine and windy weather, either alone or in combination with other indices (Tables 5 and 6). Like ACI, the various entropy measures have the advantage of being naturally normalized. Contrary to expectations, the use of a dawn bias did not improve sampling efficiency — in fact performance *consistency* over the five days declined.

The question arises as to how well the results of this study can be applied to other ecosystems and weather conditions. Although it is not possible to make claims about performance for all locations in all weathers, we believe we can make a strong claim that sampling guided by weighted combinations of relevant indices will be more robust (yield efficient sampling more consistently under different conditions) than either field survey, random sampling or ranking protocols dependent on single indices. Relevant indices are those which reflect *both* the temporal *and* spectral distribution of acoustic energy at the location, for example, *H*[t], *H*[s] and ACI. Clearly there is more work that could be done here in finding other acoustic indices that are of particular relevance to bird vocalizations, such as, indices to detect whistles, chirps, whips, stacked harmonics and repeated syllables. The recognition of such bird song features is an on-going work in our lab.

We have also used the same acoustic indices to identify dominant "acoustic regimes" in recordings of the Australian environment. In sampling for avian species richness, segments accurately identified as containing cicada choruses, heavy rain and wind gusts, could either be avoided entirely or given a discounted ranking. The bypassing of "irrelevant" audio would be expected to increase the probability of finding species of interest. Combinations of indices could be used to recognize frog choruses and other natural phenomena depending on the environment under study.

A particular feature of this study is the use of 'real-world' recordings whose acoustic content is unconstrained and undefined. We endeavored to develop a protocol which can be applied directly to 'raw' recordings typically obtained by ecologists. Thus to sample one-minute segments containing bird calls, it was necessary to ignore acoustic content below 500 Hz band (primarily due to traffic and airplane noise) and it was necessary to sacrifice some temporal and spectral specificity in order to reduce processing time. Coping with large temporal and spatial scale is one of the distinctions between bioacoustics and ecological acoustics. A parallel implementation (CPU: Intel Xenon E5-2665 0, 2.4 GHz, 16 cores; RAM: 256 GB, DDR) can process a 24 hour recording in 8–10 min.

4.1. Long duration spectrograms

Synchronous fluctuations of acoustic indices can be used to identify acoustic episodes through a 24 hour period (Fig. 3). However a more intuitive visualization technique is to construct a 24 hour spectrogram by combining indices that produce a value for each frequency bin in each time frame. We selected three indices (normalized average power, ACI and *H*[t]) because these were expected to be maximally independent of one another. *H*[t] was 'reversed' to yield an acoustic *concentration* index (1 − *H*[t]) rather than an acoustic *dispersal* index. The resulting three spectra were averaged for each one minute recording segment to produce a 24 hour spectrogram consisting of 1440 spectra each containing 256 values (Fig. 6). The morning and evening portions of the spectrogram correspond to the left and right sides of Fig. 3. The morning chorus is clearly visible as are tracks around 1830 h corresponding to the evening cicada chorus.

A surprising amount of information can be gleaned from these long duration spectrograms and can be used to navigate a 24 hour recording which would otherwise be opaque and impenetrable. In some cases, actual species can be identified just by observation. For example crows (whose calls have a distinctive set of stacked harmonics) can be clearly identified from 1000 to 1010 h because their individual calls leave a visible trace in consecutive one-minute spectra. One can observe the arrival and departure of other species (such as the Grey Fantail (*Rhipidura albiscapa*) calling in the 5–7 kHz band during the 5 h after dawn; the Yellow-faced Honeyeater (*Lichenostomus chrysops*) calling in the 2–4 kHz band around 0938 h; the Striated Pardalote (*Pardalotus striatus*) calling around 1719 h; the Olive-backed Oriole, *Oriolus sagittatus*, calling around 1745 h) because they leave visible traces in consecutive spectra. At night-time, the tracks of different orthoptera species leave a more obvious trace in long duration spectrograms than they do in one-minute spectrograms. The use of long-duration spectrograms to navigate recordings in excess of 24 h is an on-going research project in our lab.

Acknowledgments

This research was conducted with the support of the QUT Science and Engineering Faculty, the QUT Institute of Future Environments, the QUT Samford Ecological Research Facility (SERF) and the Microsoft QUT eResearch Centre (MQUTER). The authors wish to thank Julie Sarna, Tom Tarrant and Rebecca Ryan for identifying the birds in the five days of recording; Mark Cottman-Fields and Anthony Truskinger for IT support and stimulating discussion.

Table 7
The percent of known species identified using five different sampling protocols (each taking 60 one-minute samples per day) over five consecutive days. Percentages are of the total species known to call on that day. Values are rounded to integers.

Sampling protocol	13th Oct	14th Oct	15th Oct	16th Oct	17th Oct	Average over 5 days
Field survey	35% (22/62)	45% (26/58)	32% (20/62)	33% (15/45)	35% (22/62)	36.0%
Random sampling	64 ± 5%	58 ± 5%	59 ± 5%	41 ± 8%	56 ± 11%	55 ± 7%
Five weighted indices	69%	71%	68%	64%	66%	67.4%[a]

[a] Confidence = 97% (with respect to random sampling).

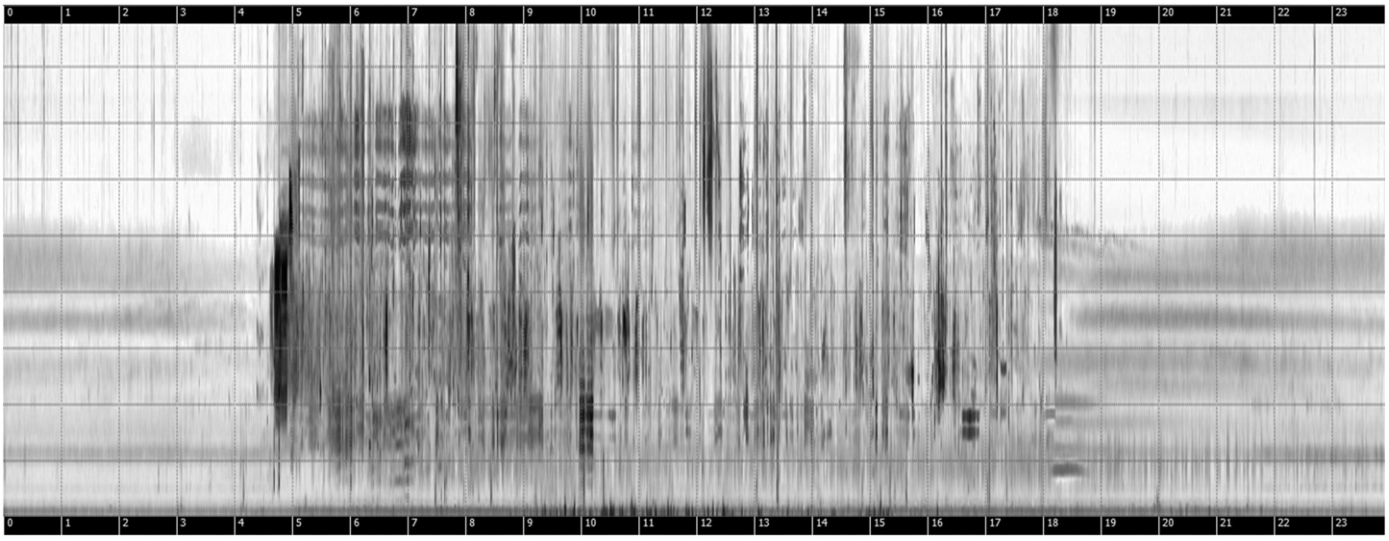

Fig. 6. A 24 hour spectrogram derived from an equal-weighted combination of the spectrograms for average power, ACI and *H*[t]. The vertical gridlines are at one hour intervals, starting and ending at midnight. The horizontal gridlines are at 1 kHz intervals with 0 kHz at bottom of image.

References

Acevedo, M.A., Villanueva-Rivera, L.J., 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildl. Soc. Bull. 34, 211–214.

Acevedo, M.A., Corrada-Bravo, C.J., Corrada-Bravo, H., Villanueva-Rivera, L.J., Aide, T.M., 2009. Automated classification of bird and amphibian calls using machine learning: a comparison of methods. Ecol. Inform. 4, 206–214.

Agranat, I., 2009. Automatically identifying animal species from their vocalizations. Fifth International Conference on Bio-Acoustics (http://bioacoustics2009.lboro.ac.uk/abstract.php?viewabstract=57, Holywell Park).

Anderson, S., Dave, A., Margoliash, D., 1996. Template-based automatic recognition of birdsong syllables from continuous recordings. J. Acoust. Soc. Am. 100, 1209–1219.

Bibby, C.J., Burgess, N.D., Hill, D.A., 1992. Bird Census Techniques, 2nd ed. Academic Press, London.

Brandes, S.T., 2008. Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conserv. Int. 18, S163–S173.

Brandes, S.T., Naskrecki, P., Figueroa, H.K., 2006. Using image processing to detect and classify narrow-band cricket and frog calls. J. Acoust. Soc. Am. 120, 2950–2957.

Chen, Z., Maher, R., 2006. Semi-automatic classification of bird vocalizations using spectral peak tracks. J. Acoust. Soc. Am. 120, 2974.

Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012. Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. Ecol. Indic. 13, 46–54.

Digby, A., Towsey, M., Bell, B., Teal, P., 2013. A practical comparison of manual, semi-automatic and automatic methods for acoustic monitoring. Methods Ecol. Evol. 4, 675–683.

Ellis, W., FitzGibbon, S., Roe, P., Bercovitch, F.B., Wilson, R., 2010. Unraveling the mystery of koala vocalisations: acoustic sensor network and GPS technology reveals males bellow to serenade females. Integr. Comp. Biol. 50, E49-E49.

Ellis, W., Bercovitch, F., FitzGibbon, S., Roe, P., Wimmer, J., 2011. Koala bellows and their association with the spatial dynamics of free-ranging koalas. Behav. Ecol. 22, 372–377.

Gasc, A., Sueur, J., Pavoine, S., Pellens, R., Grandcolas, P., 2013. Biodiversity sampling using a global acoustic approach: contrasting sites with microendemics in New Caledonia. PLoS ONE 8 (5), e65311.

Gregory, R.D., Strien, A.V., 2010. Wild bird indicators: using composite population trends of birds as measures of environmental health. Ornithol. Sci. 9, 3–22.

Grossberg, S., Carpenter, G., 2002. Adaptive Resonance Theory. (Grossberg, S., Carpenter, G.(Grossberg, S., Carpenter, G.s) Second edition. MIT Press, Cambridge, Massachusetts (http://cns.bu.edu/Profiles/Grossberg/CarGro2003HBTNN2.pdf).

Hu, W., Bulusu, N., Dang, T., Taylor, A., Chou, C.T., Jha, S., Tran, V.N., 2010. Canetoad monitoring: Data reduction in a high rate application. In: Al, E.G.E. (Ed.), Google eBooks. Springer Science, New York (al., E.G.e.)al., E.G.e.s).

Joo, W., Gage, S.H., Kasten, E.P., 2011. Analysis and interpretation of variability in soundscapes along an urban–rural gradient. Landsc. Urban Plan. 103, 259–276.

Juang, C., Chen, T., 2007. Birdsong recognition using prediction-based recurrent neural fuzzy networks. Neurocomputing 71, 121–130.

Kasten, E.P., Gage, S.H., Fox, J., Joo, W., 2012. The remote environmental assessment laboratory's acoustic library: an archive for studying soundscape ecology. Ecological Informatics 12, 50–67.

Krause, B., 2008. Anatomy of the soundscape. J. Audio Eng. Soc. 56.

Lamel, L.F., Rabiner, L.R., Rosenberg, A.E., Wilpon, J.G., 1981. An improved endpoint detector for isolated word recognition. IEEE Trans. ASSP ASSP-29, 777–785.

McIlraith, A.L., Card, H.C., 1997. Birdsong recognition using backpropagation and multivariate statistics. IEEE Trans. Signal Process. 45, 2740–2748.

McLaren, J., 2012. Monitoring Techniques for Temperate Bird Diversity: Uncovering Relationships between Soundscape Analysis and Point Counts, Department of Biological Sciences. University of Notre Dame Environmental Research Center, Land O'Lakes, WI 54540.

Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: the acoustic complexity index (ACI). Ecol. Indices 11, 868–873.

Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011a. What is soundscape ecology? An introduction and overview of an emerging new science. Landsc. Ecol. 26, 1213.

Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gage, S.H., Pieretti, N., 2011b. Soundscape ecology: the science of sound in the landscape. Bioscience 61, 203.

Quinlan, R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann publishers.

Rempel, R.S., Hobson, K.A., Holborn, G., Wilgenburg, S.L.V., Elliott, J., 2005. Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory. J. Field Ornithol. 76, 1–11.

Somervuo, P., Harma, A., Fagerlund, S., 2006. Parametric representations of bird sounds for automatic species recognition. IEEE Trans. Audio Speech Lang. Processing 14, 2252–2263.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid acoustic survey for biodiversity appraisal. PLoS ONE 3 (12), e4065.

Towsey, M., 2013a. An Algorithm to Cluster the Spectra in a Spectrogram. Queensland University of Technology, Brisbane (QUT ePrints, http://eprints.qut.edu.au/61509/).

Towsey, M., 2013b. Noise Removal from Wave-forms and Spectrograms Derived from Natural Recordings of the Environment. Queensland University of Technology, Brisbane (QUT ePrints, http://eprints.qut.edu.au/61399/).

Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, P., 2012. A toolbox for animal call recognition. Bioacoustics 21, 107–125.

Wimmer, J., Towsey, M., Planitz, B., Roe, P., Williamson, I., 2010. Scaling Acoustic Data Analysis through Collaboration and Automation, e-Science (e-Science). 2010 IEEE Sixth International Conference on. IEEE, pp. 308–315.

Wimmer, J., Towsey, M., Roe, P., Williamson, I., 2013. Sampling environmental acoustic recordings to determine bird species richness. Ecol. Appl. 23, 1419–1428.