Long-duration, false-colour spectrograms for detecting species in large audio data-sets

*Correspondence:
MT: m.towsey@qut.edu.au

Michael Towsey[1,*], Elizabeth Znidersic[3], Julie Broken-Brow[2], Karlina Indraswari[1], David M. Watson[3], Yvonne Phillips[1], Anthony Truskinger[1], Paul Roe[1]

[1] QUT Ecoacoustics Research Group, Queensland University of Technology, George Street, 4001 Brisbane, Queensland, Australia

[2] School of Agriculture and Food Sciences, University of Queensland, St Lucia, Brisbane, Queensland, Australia

[3] Institute for Land, Water and Society, Charles Sturt University, Elizabeth Mitchell Drive, Albury, NSW 2640, Australia

Abstract

Long-duration recordings of the natural environment have many advantages in passive monitoring of animal diversity. Technological advances now enable the collection of far more audio than can be listened to, necessitating the development of scalable approaches for distinguishing signal from noise. Computational methods, using automated species recognisers, have improved in accuracy but require considerable coding expertise. The content of environmental recordings is unconstrained, and the creation of labelled datasets required for machine learning purposes is a time-consuming, expensive enterprise. Here, we describe a visual approach to the analysis of environmental recordings using *long-duration false-colour* (LDFC) spectrograms, prepared from combinations of spectral indices. The technique was originally developed to visualize 24-hour "soundscapes." A soundscape is an ecoacoustics concept that encompasses the totality of sound in an ecosystem. We describe three case studies to demonstrate how LDFC spectrograms can be used, not only to study soundscapes, but also to monitor individual species within them. In the first case, LDFC spectrograms help to solve a "needle in the haystack" problem—to locate vocalisations of the furtive Lewin's Rail (Tasmanian), *Lewinia pectoralis brachipus*. We extend the technique by using a machine learning method to scan multiple days of LDFC spectrograms. In the second case study, we demonstrate that frog choruses are easily identified in LDFC spectrograms because of their extended time-scale. Although calls of individual frogs are lost in the cacophony of sound, spectral indices can distinguish different chorus characteristics. Third, we demonstrate that the method can be extended to the detection of bat echolocation calls. By converting complex acoustic data into readily interpretable images, our practical approach bridges the gap between bioacoustics and ecoacoustics, encompassing temporal scales across three orders of magnitude. Using the one methodology, it is possible to monitor entire soundscapes *and* individual species within those soundscapes.

Introduction

Long-duration acoustic recordings of the environment are increasingly used by ecologists to monitor species diversity in terrestrial ecosystems (Gage and Farina, 2017). Acoustic recordings have several advantages. First, sensors can record continuously for weeks or months (depending on their power source) whereas an observer's ability to access and collect data in the field is limited. Second, multiple sensors can be distributed through a landscape to record simultaneously. Third, multiple people can listen to acoustic recordings multiple times, to facilitate interpretation. Fourth, recordings can be stored indefinitely until more powerful analytical techniques become available. Fifth, acoustic recorders create minimal disturbance and are unlikely to alter the vocalising behaviour of local animals.

The downside to easy acquisition of long-duration recordings is that most of it will not be listened to. Computational methods are required to reveal features of interest, usually by writing automated species recognisers. However, the content of environmental recordings is unconstrained, containing much that is not of interest and confounding to even well-written recognisers. Many papers and international competitions concerning computational bird call recognition depend on data sets that have been extensively curated and cleaned of unwanted sounds (Priyadarshani et al., 2016, 2018) or only include calls recorded at close range. (Note however that the LifeCLEF bird identification challenge now includes several hours of soundscape recording [Goeau et al., 2017]). Consequently, writing recognizers is a time consuming and expensive enterprise. And recognizers reveal nothing about the content of a recording where the species of interest is absent.

An alternative approach is to treat long-duration recordings as soundscapes (Pijanowski et al., 2011). Here the focus is not on individual species, but rather on the broad categories of sound sources that contribute to the recording, typically biophony (sounds due to mammals, birds, frogs, insects, etc), geophony (sounds due to wind, rain, surf), and anthropophony (human-made sounds, whether speech, music or the multitude of machine sounds). To quantify this approach, acoustic indices, such the Acoustic Complexity Index (Pieretti et al., 2010) and the Acoustic Entropy Index (Sueur et al., 2014), are used to estimate the complexity of the biophony in a soundscape, which is, in turn, used as an indirect measure of species diversity and ecosystem health.

There is now a considerable literature on acoustic indices (Gage and Farina, 2017, Chapter 16; Sueur et al., 2014), but single indices do not readily reveal the detailed acoustic structure in long recordings. As an attempt to address this problem, *long-duration false-colour* (hereafter LDFC) spectrograms were developed. These are constructed by calculating three spectral indices at coarse resolution (typically 60 seconds per spectrum) and assigning the three indices to the red, green and blue channels of a coloured spectrogram image (Towsey et al., 2014). The spectral indices employed in this article to construct LDFC spectrograms are described in the next section.

A spectral acoustic index, as incorporated into a LDFC spectrogram, is better understood as a mathematical "filter"—it describes some feature of the distribution of acoustic energy in each of the frequency bins of a one-minute recording segment. The utility of the method depends upon combining three indices that behave as different "filters" providing different "views" into the soundscape (Towsey et al., 2016).

Acoustic indices have previously been used to monitor soundscapes (Gage et al., 2017; Phillips et al., 2018b) and, in addition, to filter long-duration recordings in the search for a single species. See for example, Gage and Farina (2017) who use spectral energy to determine the date of first appearance of the Spring Peeper frog (potentially useful in detecting long-term climate change). Here, we extend their application to incorporate multiple acoustic indices and LDFC spectrograms.

The purpose of LDFC spectrograms, when originally introduced, was to provide soundscape ecologists with a tool to navigate the content of very long recordings (Towsey et al., 2015). However, a surprising amount of detail can be obtained from false-colour spectrograms, even at a resolution of 60 seconds per spectrum, and it has subsequently become apparent that LDFC spectrograms can also assist ecologists in *bioacoustic* studies involving only one or a few species. The purpose of this article is to

describe three case studies that highlight the utility of LDFC spectrograms in the study of individual species from three diverse vocal taxa: birds, frogs and bats.

General methods

Recording acquisition

All recordings described in this article were obtained with Song Meter SM2+, SM3, or SM4 recorders (Wildlife Acoustics, 2017). Each recorder was attached to a tree or metal stake at a height of ≈0.7–1.5 m. Recordings were saved in stereo, 16-bit WAVE format. Sampling rates differed for each study.

All recordings were mixed down to mono and split into one-minute, non-overlapping segments. Each one-minute audio segment was converted to an amplitude spectrogram using non-overlapping frames of 512 or 1,024 samples and a fast-Fourier transform (FFT) with Hamming window. Each spectrum was smoothed using a moving average filter (width = 3).

In this work, seven spectral indices were calculated for each one-minute segment. In our case, a spectral index is a 256 (or 512) element vector, each element of which summarizes some aspect of the distribution of acoustic energy in one frequency bin of the signal spectrogram. The dimension of the vector is half the FFT window size. With the computing resources available to us (a 16-core machine), we are able to process a 24-hour recording (sampled at 22,050 Hz) in 16 minutes. This includes calculation of indices and preparation of spectrogram images.

We present only an abbreviated account of the calculation of the seven indices used in this work. Further detail can be obtained in (Towsey, 2017). A three-letter code is used for ease of identifying each spectral index. The software to perform the analysis is open source and written in C# (Towsey et al., 2018).

Calculation of spectral acoustic indices

Seven spectral indices were calculated as follows:

1. Acoustic Complexity Index (ACI): A measure of the relative change in acoustic intensity (A) in each frequency bin, f, of the amplitude spectrogram:

$$\text{ACI}[f] = \sum_i |A_{if} - A_{i-1,f}| / \sum_i A_i \tag{1}$$

   where i is an index over all frames and f is an index over all frequency bins (Pieretti et al., 2010).

2. Temporal Entropy (ENT): A measure of the dispersal of acoustic energy through the frames of each frequency bin (Sueur et al., 2014, 2008). The squared amplitude values in each frequency bin are normalized to unit area and treated as a probability mass function (*pmf*). The entropy of the *pmf* vector for frequency bin f is a measure of the energy "dispersal" through time and is calculated as:

$$H_t[f] = \sum_i \left(\text{pmf}_{if} \times \log_2 \left(\text{pmf}_{if} \right) \right) / \log_2(N) \tag{2}$$

   where i is an index over frames and N is the number of frames. To obtain a more "intuitive" index, we convert $H_t[f]$ to "energy concentration":

$$\text{ENT}[f] = 1 - H_t[f] \tag{3}$$

3. Background Noise (BGN): The decibel value of background noise in each frequency bin calculated as the modal decibel value in each frequency bin of the decibel spectrogram. The decibel spectrogram was prepared by converting the amplitude values (A) to decibels using dB = $20 \times \log_{10} A$. Noise reduction was achieved by subtracting the modal decibel value in a bin from each value in the bin (Towsey, 2017).

4. Power minus Noise (PMN): The maximum decibel value in each frequency bin of the noise-reduced decibel spectrogram (Towsey, 2017).

5. Event Count (EVN): A measure of the number of acoustic events per minute in each frequency bin, f, of the noise-reduced decibel spectrogram. An event is counted each time the bin's (noise-reduced) decibel value exceeds the 3-dB threshold (Towsey, 2017).

6. Horizontal Ridge Count (RHZ): Also derived from the noise-reduced decibel spectrogram. Many bird songs, consist of whistles with harmonics which appear in the standard grey-scale spectrogram as horizontal ridges. These can be detected using a 5 × 5 ridge mask. Each element of RHZ is the average decibel value of ridge cells identified within the corresponding frequency bin (f). RHZ can be helpful for detecting bird call activity (Towsey, 2017).

7. Spectral Peak Tracks (SPT): The spectral peaks (local maxima) are identified in each spectrum. To qualify as a "peak," the cell's amplitude value must exceed 6 dB. Each element of SPT is the sum of the "peak" decibel values identified within the corresponding frequency bin (f), divided by the number of cells within the bin (Towsey, 2017).

Preparation of false-colour spectrograms

The choice of spectral indices to combine into an LDFC spectrogram depends upon the study. Experience suggests that a combination of ACI, ENT and EVN assigned to the red, green and blue channels respectively satisfies many cases. LDFC spectrograms are most informative where the three indices are not correlated. A 24-hour LDFC spectrogram is 1,440 pixels wide (there are 1,440 minutes in a day) and has a height equal to the number of frequency bins, or half the frame size.

Three case studies

Case study 1: Lewin's Rail (Tasmanian) *Lewinia pectoralis brachipus*

Background

Lewin's Rail, *Lewinia pectoralis*, is a furtive wetland-dependent bird which inhabits thick vegetation, calls rarely and is seldom seen. Of the eight subspecies, two are of particular conservation concern. *L. p. clelandi* is known only from south Western Australia but was last seen in 1932 and is presumed extinct (Garnett et al., 2011). *L. p. brachipus* is restricted to Tasmania but has a patchy distribution which extends to several offshore islands. Fewer than 70 observations of this taxon have been made since 1995 (Department of the Environment, 2015) raising concerns that this ground-nesting species is disproportionately affected by invasive predators (Woinarski et al., 2017). Rails are detected primarily by their vocalizations, and most monitoring relies upon presence/absence estimates using call playback and passive aural surveys (Conway and Gibbs, 2005). Repeated use of call playback (which simulates a territorial intrusion) may negatively affect resident pairs, resulting in territory abandonment or nest failure. Additionally, this methodology also requires a costly extended survey effort to enable high confidence levels inferring absence. Lewin's Rail vocalization repertoire changes temporally from an acoustically simple contact call to a complex call repertoire with harmonic elements that is thought to be associated with breeding. Vocalizations are also sporadic, and of either short or long duration. To establish their current distribution and evaluate their population status, a monitoring approach is needed that can reliably detect small numbers of individuals unobtrusively.

This case study investigates whether the application of LDFC spectrograms can detect Lewin's Rail at sites where its presence had previously only been confirmed by camera traps (Znidersic, 2017). Call detection was at first approached using an automated recognizer for the acoustically simple contact call. However, as the vocalization repertoire shifted to an acoustically more complex call, this recognizer proved ineffective. Here, we present the application of LDFC spectrograms as an alternative method for year-round detection.

Method

The study site was Tasman Island, Tasmania (43° 14′ 12.57′′ S, 148° 00′ 13.49′′ E). A Wildlife Acoustics SM3 sensor was deployed for ten days from 10–19 November 2015 (see Znidersic, 2018, for further details and access to recordings). The sensor was programmed for continuous recording at a sampling rate of 24 kHz, yielding 240 hours of audio. For calculation of spectrograms and indices, the recordings were down-sampled to 22,050 Hz. Spectrograms were prepared using a frame-width of 512. Spectrograms, so obtained, have 2,584 frames (frame duration = 23.2 ms) and 256 frequency bins (each with a bandwidth of ≈43 Hz). Spectral indices and LDFC spectrograms were prepared as described above.

Results

A six-hour sample from a 24-hour LDFC spectrogram is shown on the left side of Figure 1. To the right side are eight seconds of grey-scale spectrogram (within the same time period) illustrating the "grunt" and "wheeze" vocalizations. These vocalizations can be identified in the LDFC spectrogram as green vertical lines in the range 100–3,500 Hz (see white rectangles). Some of the call elements, primarily the "grunt," occupy an available (free) acoustic space in the lower frequencies, readily enabling identification. The three examples in Figure 1 have differing call durations (05:35, ≈40 sec duration; 06:48, ≈8 sec; 08:08, ≈12 sec).

Not only does the LDFC spectrogram enable discrimination of Lewin's Rail calls, it also situates them within the broad-scale soundscape context. For example, the dawn bird chorus is shown in Figure 1 by the green and pink hues that commence at 05:00 in the 1,500–5,000 Hz frequency range and additionally, it illustrates a visual representation of the available acoustic space.

Method to build automated recogniser using spectral indices features

The visibility of Lewin's Rail calls in these LDFC spectrograms suggests that the spectral indices themselves could be used as features to train an automated recogniser. This would be useful to speed the search for calls where there are many days of LDFC spectrograms to be examined.

To explore the feasibility of this approach, we used the PMN, ENT, SPT, RHZ, and ACI spectral indices derived from an entire day of recording. We used only 20 frequency bins (5–25 or 215 Hz to 1,075 Hz) from each spectral index. This yielded a data set of 1,440 instances, consisting of 49 positive and 1,391 negative instances. Each instance consisted of a 5 × 20 = 100 element feature vector and a binary label (0 = not an Lewin's Rail call, 1 = Lewin's Rail call). The feature values were converted to z-scores for each feature independently. The data were used to train a Support Vector
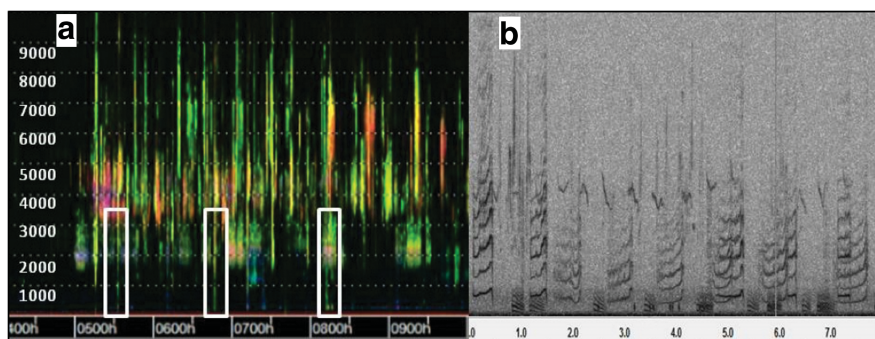

Figure 1. (a) A six-hour sample (04:00 to 10:00) from a 24-hour LDFC spectrogram. (b) An eight-second portion of standard grey-scale spectrogram extracted from the same time period.

The vertical axis (0–10 kHz) is the same for both spectrograms. The grey-scale spectrogram illustrates the "grunt" and "wheeze" of the Lewin's Rail. These can be identified in the LDFC spectrogram within the white rectangles. The dawn chorus is visible at 05:00.

Table 1. Confusion matrix for performance of the Lewin's Rail call recogniser.

Not-Lewin's Rail	Lewin's Rail	Classification
1,383	8	Class: Not Lewin's Rail
16	33	Class: Lewin's Rail

Machine (SVM) in the WEKA Machine Learning package (Frank et al., 2016). Performance was assessed using five-fold cross-validation.

Results of recognizer using spectral index features

A confusion matrix is shown in Table 1. Precision and recall were 80% and 67% respectively.

A better interpretation of this result should take into account that the 49 positive instances consisted of 18 *easy* instances (having an average signal-to-noise ratio [SNR] of 11.6 dB and easy to recognise by eye in a standard grey-scale spectrogram); 12 *difficult* instances (having an average SNR of 4.0 dB and more difficult to recognise by eye); and 19 *very difficult* instances (having an average SNR of 3.2 dB and very difficult to recognise, even for the human eye, because the wheeze part of the call was sometimes missing). 100% of the easy and difficult instances were correctly recognised by the SVM recogniser and three (16%) of the very difficult instances.

Case study 2: Frog communities of the Gulf of Carpentaria

Background

Frog calls are considered an easier target than bird calls for automated recognisers because their structure is simpler, and many species can be distinguished on just three major features, dominant frequency, pulse rate and pulse duration (Savage, 2002). Nevertheless, automated frog call recognition presents difficulties because frog chorusing behaviour involves multiple individuals of the same and different species calling simultaneously, thus obscuring the individual calls on which a recogniser is typically trained, comparable to the cocktail-party effect in speech recognition.

Frog choruses can be hours in duration and consequently soundscape analysis techniques become viable. Furthermore, three of the spectral indices that have proven to be useful for constructing LDFC spectrograms (ACI, ENT, and EVN) respond differentially to important distinguishing features of frog calls. ACI is sensitive to complexly structured calls having multiple levels of amplitude modulation. ENT responds to pulses or croaks having high amplitude, and EVN responds to high pulse rates.

This case study concerns frog communities on Groote Eylandt in the Gulf of Carpentaria, northern Australia. The island has so far remained free of the cane-toad *Rhinella marinus* invasion that is sweeping across northern Australia. The purpose of this study was to build a thorough understanding of frog communities on the island prior to a possible invasion. We originally approached this task by writing traditional call recognisers, but, despite some success (*e.g.*, Xie et al., 2016), this approach provides little insight into the north Australian soundscape and the place of frog chorusing within it. Hence, we investigated the use of LDFC spectrograms.

Methods

The study site was Groote Eylandt (13° 58′ 24.48″ S, 136° 27′ 34.25″ E). A Wildlife Acoustics SM2 sensor was deployed for recording 12 hours at night (19:00–07:00) over consecutive days. The sampling rate was 22,050 Hz. Spectrograms were prepared using a frame-width of 512. Spectral indices and LDFC spectrograms were prepared as described above.

Results

[Figure 2](#) shows four LDFC spectrograms, each of three hours duration, alongside five seconds of standard grey-scale spectrogram taken from the same time-period. The dominant feature in all LDFC spectrograms is an insect chorus at 5,000 Hz, picked up strongly by the ACI index, and hence displayed in a red hue. In the language of soundscape ecology, such dominant features are referred to as *sound-marks* ([Schafer, 1994](#)). Sound-marks are extremely useful in orienting oneself in a visual representation of a soundscape, analogous to landmarks in a landscape.

The LDFC spectrogram in [Figure 2a](#) reveals intermittent chorusing of the ornate burrowing frog *Platyplectrum ornatum* over a two-hour period in the 500–2,500 Hz band. The starts and ends of chorus episodes are clearly delineated in solid colour, although a single call has a two second interval between each brief pulse (see standard grey scale spectrogram). Colour changes reflect the chorus intensity.

The chorusing of the Northern dwarf tree frog *Litoria bicolor* ([Figure 2b](#)) occupies two frequency bands which straddle the insect chorus at 5,000 Hz. The synchronisation of green banding above and below the insect track helps to identify this species. Once again, changes in colour reflect chorus intensity. The call of the flood plain toadlet *Uperoleia inundata* ([Figure 2c](#)) also has a pulse period of two seconds but in the 1,200–2,200 Hz band which helps to distinguish it from *P. ornatum*.
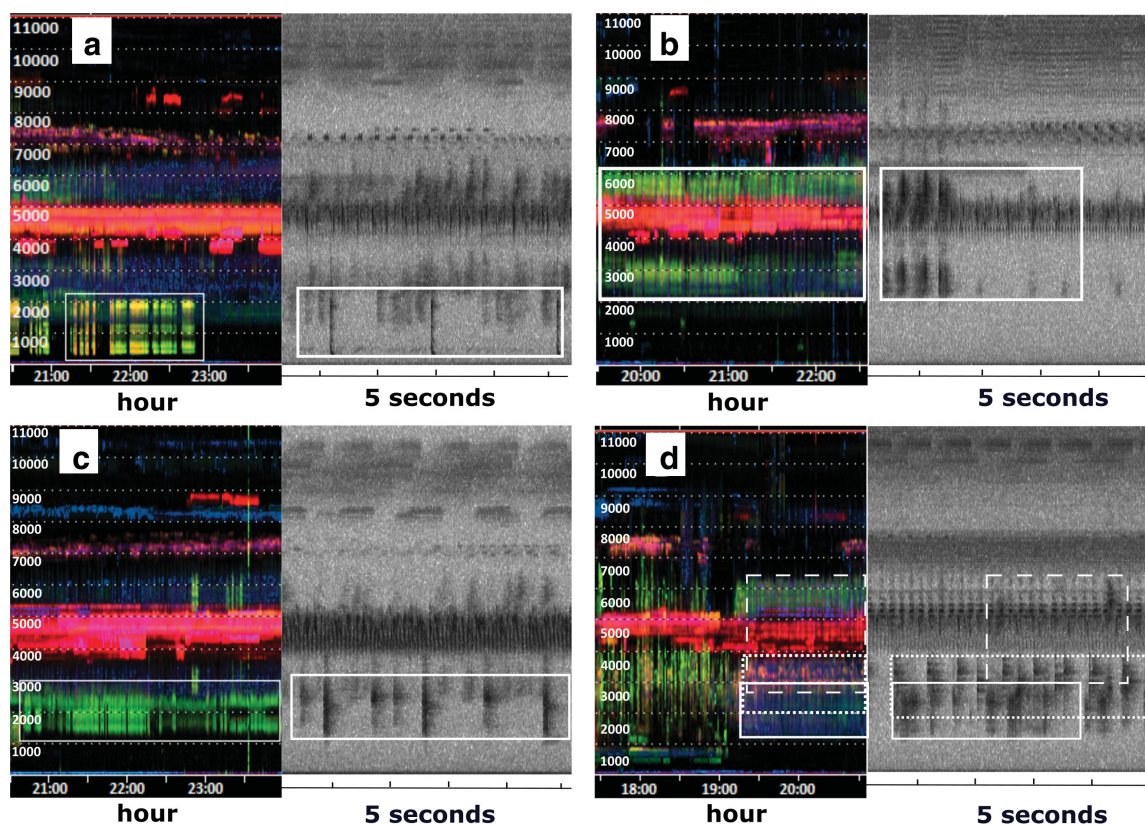

Figure 2. Four LDFC spectrograms, each of three hours duration, alongside five-seconds of standard spectrogram taken from the same time-period.

White rectangles identify frog choruses and calls of interest. The vertical Hertz scale is the same for all spectrograms. (a) Intermittent chorusing of the ornate burrowing frog. On the right are three brief pulses of the same species. (b) Chorusing of the Northern dwarf tree frog. The red band at 5 kHz is an insect chorus. The standard spectrogram illustrates how a single call pulse almost straddles the narrow-band insect chorus. (c) Chorusing of the flood plain toadlet (*left*) and three pulses from one individual (*right*). (d) This evening soundscape (*left*) contains a strong insect chorus and chorusing of three frog species.

The soundscape in Figure 2d spans 1.5 hours either side of dusk. A strong insect chorus persists throughout, although there is a slight drop in frequency coinciding with the onset of colder night-time temperatures. Dusk is preceded by bird calls (green/orange hues) and followed by chorusing of three frog species. Despite spectral and temporal overlap of calls, colour differentiation helps to distinguish the three species more easily in the LDFC spectrogram than in the standard grey-scale spectrogram. Spectral overlap of two choruses leads to a change in colour rendering, thereby helping to identify the bounds of each chorus.

Case study 3: The Gympie bat community

Background

Bats represent approximately 20% of all mammalian biodiversity and the vast majority of bat species globally use ultrasonic echolocation (Hutson et al., 2001). Echolocating bats are typically surveyed by recording their echolocating pulses. Recordings are analysed using a variety of systems, the most common being manual analysis, where field recordings are visually compared to known reference calls and/or published metrics (Kunz and Parsons, 2009). This type of analysis can reveal the species present, and an activity index. As with other acoustic analysis methods, manual analysis can be extremely time-consuming, limiting the ability to analyse long-duration acoustic recordings.

Given the high sampling rates required for bat recordings, it might be supposed that the LDFC spectrogram technique would be inappropriate for bat monitoring. Despite the increasing capability of recording technology, current storage limitations prevent recording for long duration at the high sampling rates required for ultrasonic vocalisations. However, in an acoustic study of Gympie National Park (Phillips et al., 2017), LDFC spectrograms detected calls of the White-striped free-tailed bat, *Austronomous australis*. This species is unusual because its calls are within human hearing range and just below the Nyquist frequency of that study (11,025 Hz). The White-striped free-tailed bat was detected in one month of spectrograms over 13 months of recording (Figure 3). This chance discovery motivated a "proof of concept" investigation into how the LDFC spectrogram technique could be extended to a study of higher frequency bat calls.
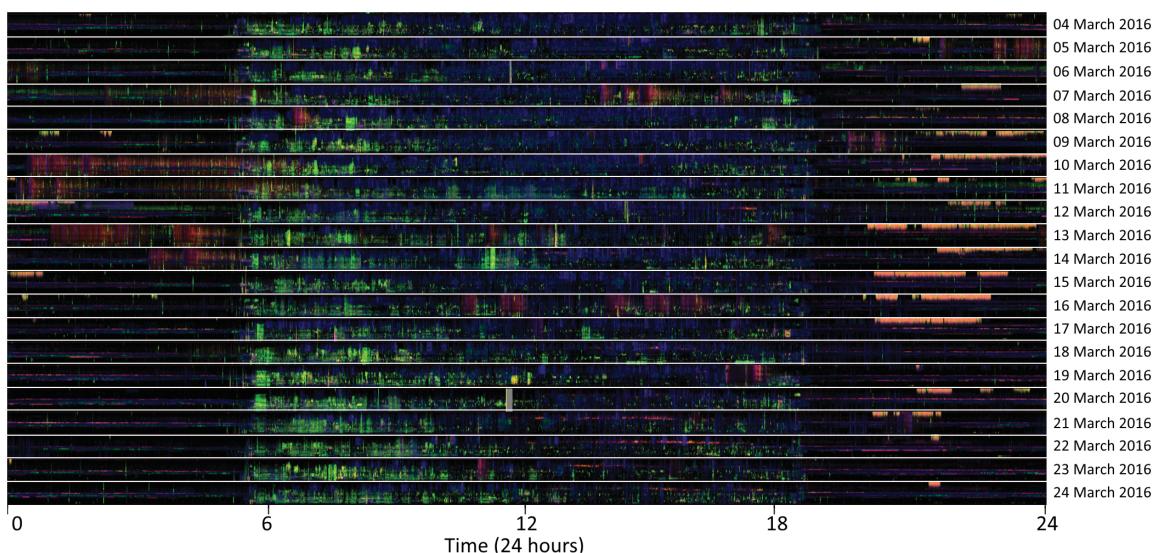

Figure 3. Twenty-one days of continuous recording in Gympie National Park represented by 21 concatenated LDFC spectrogram ribbons.

Dawn and dusk are about 06:00 and 19:00 respectively. The lower portion (around 10 kHz) of echolocating calls of the White-striped free-tailed bat can be seen between 20:00 and 24:00 in many of the spectrogram ribbons (rendered in orange). Note that the temporal resolution is 60 seconds per frame or pixel.

Method

The study site was Gympie National Park (26° 4′27.51′′S, 152°43′2.45′′). Over six consecutive nights, in August 2017, we recorded twelve hours (18:00–06:00) of continuous recording, using an SM4 at a sampling rate of 96 kHz. This is the maximum available on an SM4 and allowed detection of bat calls up to the Nyquist of 48 kHz (see Phillips, 2018a, for further details and access to the recordings). Spectrograms were prepared using a frame-width of 1,024. Spectrograms, so obtained, have 5,625 frames (frame duration = 10.66 ms) and 512 frequency bins (each with a bandwidth of ~93.8 Hz). Spectral indices and LDFC spectrograms were prepared as for case studies 1 and 2, except for one significant difference—the acoustic indices were calculated at 15 second resolution rather than 60 second, to better accommodate the brevity of echolocating pulses. It results in spectrograms that are 4× wider (showing correspondingly more detail) and with slightly changed colour rendering.

Standard grey-scale spectrograms at high temporal resolution were manually viewed to identify bat passes having a minimum of four call pulses. A pass is defined as a series of consecutive pulses (minimum of four characteristic pulses), indicating the bat has "passed" the detector once. Each pass was assigned to a species, given an indication of strength (weak, moderate, strong; based on the dB relative to ambient noise), the number of pulses recorded, and whether it was observable in the LDFC spectrogram. The presence of the pass in the LDFC spectrogram was compared against the strength and the number of pulses.

Results

Over the six nights, 35 strong, 91 moderate, and 184 weak passes were recorded from eight species or species complexes. Of these 310 passes, approximately 54% were observable in the LDFC spectrograms. All of the species present in the dataset were observable in the LDFC spectrogram; including the White-striped free-tailed bat *Austronomous australis*, Chocolate wattled bat *Chalinolobus morio*, Gould's wattled bat *C. gouldii*, Eastern horseshoe bat *Rhinolophus megaphyllus*, Yellow-bellied sheath-tailed bat *Saccolaimus flaviventris*, and three species complexes (broad-nosed bats *Scotorepens* sp./*S. greyii*/*S. orion*; forest bats *Vespadelus darlingtoni*/*V. vulturnus*; and the Inland broad-nosed bat *S. balstoni*/ greater broad-nosed bat *Scoteanax rueppellii*). The number of pulses in a pass varied from four to 613. Figure 4 shows a portion of one LDFC spectrogram from 18:00-10:15. A preliminary scan of this image reveals the presence of at least four bat species. Figure 5 shows an excerpt from the LDFC spectrogram and the corresponding bat passes in a standard grey-scale spectrogram.

Both the length and strength of a pass influenced whether it was observable in the LDFC spectrogram. Typically, passes of moderate strength required at least 20 pulses to be observable; whilst strong passes were observed from as little as eight pulses. Whether a pass is identifiable to species in a LDFC spectrogram depends on the frequency of the echolocation, as there are certain frequencies where multiple species overlap. Even traditional manual analysis methods struggle to tease apart certain species from each other.

Discussion

From an ecological monitoring perspective, LDFC spectrograms are a novel visualisation tool to assist detection of vocal species. With one set of "generic" acoustic features or indices, one can construct an LDFC soundscape spectrogram, with the additional possibility of detecting species of interest within that soundscape. Detection of an individual species (even at the coarse resolution of one spectral index per minute) is possible where its call has clearly defined acoustic properties. Importantly, these distinctive properties refer not just to a frequency band, but to a combination of features that yield an identifiable pattern of colour in a false-colour rendering.

Where the calling behaviour is extended through time, as in the case of frog chorusing, the resultant traces in an LDFC spectrogram are particularly easy to detect. However, it should be noted that there is no simple correlation between a species and the false-colour rendering of its vocalisations. Variations in the intensity and structure of a species' call/chorus behaviour lead to variations in colour renderings.
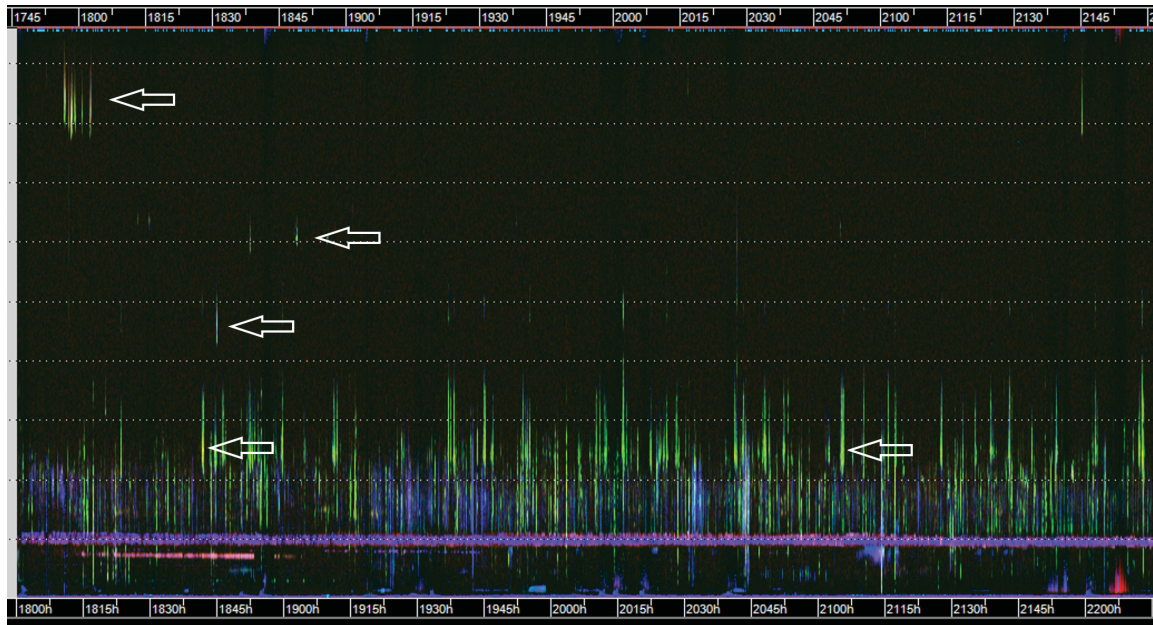
Figure 4. LDFC spectrogram, 4.25 hours. Unlike the previous spectrograms, this one has a temporal resolution of 15-seconds per pixel.

The vertical axis ranges from 0–40 kHz (grid-lines spaced at 5 kHz). The arrows mark four different species of echo-locating bats that are clearly visible at this temporal and spectral resolution.
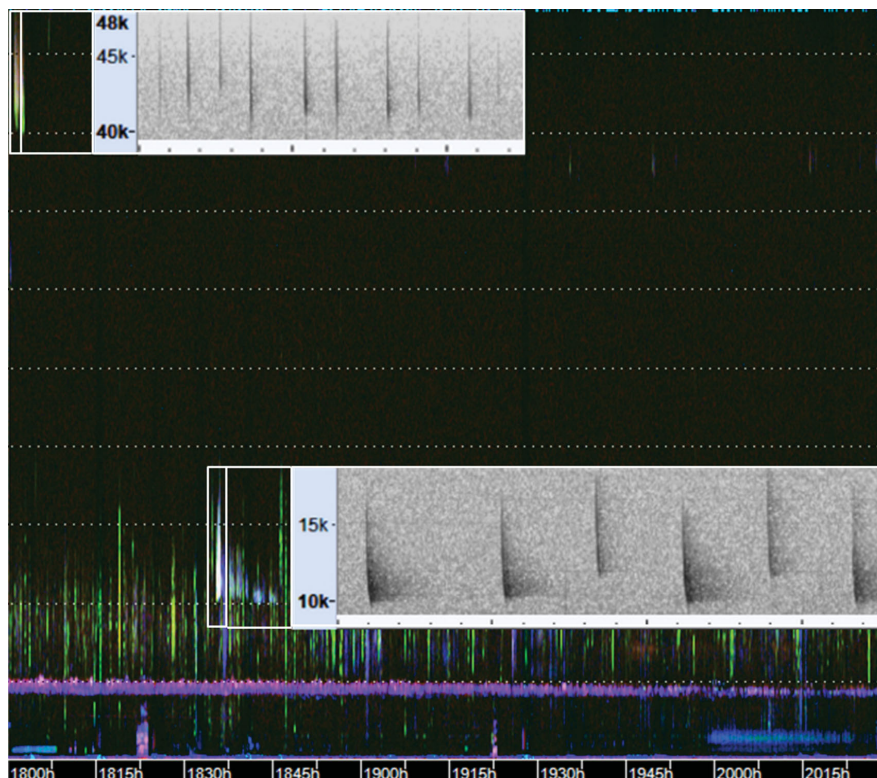

Figure 5. LDFC spectrogram of 2.5 hours of having the same temporal and spectral scale as Figure 4.

The inserts are standard grey-scale spectrograms (interval between minor ticks = 0.1 seconds) illustrating the individual echolocating pulses that contributed to call detection in the LDFC spectrogram.

However, the extent of this variation can be learned with experience. The advantage of rendering three acoustic indices in colour, as opposed to a single index, such as amplitude in grey scale (Gage and Farina, 2017), is that three indices offers more possibilities to detect subtle changes in acoustic behaviour or to detect temporal and spectral overlap of different call types.

The LDFC technique is further enhanced by the ability to prepare spectrograms at different resolutions (Towsey et al., 2015) and by the possibility of constructing call recognisers with data sets derived from the same spectral indices used to construct the soundscape LDFC spectrograms. We believe this article to be the first published example using a (SVM) call recogniser to find bird calls within a LDFC spectrogram. The preparation of an automated recogniser is useful where an ecologist must scan many days of data (one LDFC spectrogram per 12–24 hours) to determine the presence/absence of a species.

We have found that the seven spectral indices described in this study have broad applicability to many soundscape investigations, whether terrestrial, marine, or freshwater. The analysis of the recordings and the production of the false-colour spectrograms is entirely automated. We routinely produce two false-colour spectrograms, the first assigning ACI-ENT-EVN and the second assigning BGN-PMN-RHZ to the RGB channels respectively. After a short period of learning, one quickly recognises the sound-marks in a soundscape and whether calls of the species of interest can be recognised. Of course, further experimentation is required to prepare an automated recogniser. For detecting Lewin's Rail calls using an SVM, we found that a combination of five spectral indices gave the optimum result: ACI, ENT, EVN, PMN, and RHZ. The amount of time required to build this recogniser using generic acoustic indices (six hours to construct the data set, two hours to train and test an SVM recogniser) was small compared to the days of work required to code and incrementally improve the performance of a "traditional" recogniser using features derived from standard spectrograms of the Lewin's Rail "contact" call.

In addition, just to determine recall and precision rates for the "contact call" recognizer required an additional 14 hours of manual work (unpublished data, Znidersic and Towsey). By comparison, using the LDFC images, near immediate visual detection and validation was achieved for the "good" calls, without need for the SVM recogniser. Using the SVM recognizer accelerated visual detection of the difficult and very difficult calls.

Despite the above advantages, the LDFC spectrogram technique is ineffective where the species of interest only calls during the morning chorus or contemporaneously with many other species. In addition, in the case of bat calls, the method does not appear to detect very low amplitude calls compared to a well-developed custom recogniser and would therefore be inappropriate for determining presence/absence of species of "whispering" bats. However, the inability of the technique to detect Lewin's Rail and bat calls to the same level of accuracy as scanning standard grey-scale spectrograms with the human eye, deserves further comment.

Real world datasets derived from environmental recordings, present three difficulties, particularly when used for machine learning purposes. First, the number of negative instances greatly outnumbers the positives (1391:49 in case of the Lewin's Rail dataset). Second, in addition to geographic variation, the calls of most species vary within and between individuals. Third, poor quality calls will almost always outnumber good quality calls. This last difficulty, which has obvious consequences for training a recogniser, is a consequence of two physical laws: where a vocal species is approximately uniformly distributed through the landscape, 1. the number of vocal individuals at distance r from the microphone will increase in proportion to r, but 2. the amplitude of calls arriving at the microphone will decrease with r^2. These two factors conspire to ensure that a majority of positive instances in a truly representative data set will be of low quality. This is a common problem for real-world datasets, as opposed to curated datasets, because every call detected by the human ear is included (Digby et al., 2013). Both the Lewin's Rail and Gympie bat datasets contained a majority of poor quality calls, and standard recognisers or analysis techniques would also have difficulty with these.

Conclusion

The three case studies presented in this article describe the conversion of complex long-duration acoustic data into readily interpreted false-colour spectrogram images. Using the one methodology, it is possible to monitor entire soundscapes *and* some of the individual species that help create that soundscape. In addition, the ability to visualise sound scapes at multiple temporal scales is a distinct advantage and helps pose ecological questions that might not otherwise be obvious. An interesting feature of this approach is that it begins to bridge the gap between bioacoustics and ecoacoustics, two disciplines whose traditional temporal scales of interest have been separated by some three orders of magnitude.

Acknowledgements

Funding sources

Competing interests

All authors, Michael Towsey, Elizabeth Znidersic, Julie Broken-Brow, Karlina Indraswari, David M. Watson, Yvonne Phillips, Anthony Truskinger and Paul Roe declare that they have no conflicts of interest.

References

Conway C. J. and Gibbs J. P. (2005). Effectiveness of call-broadcast surveys for monitoring marsh birds. The Auk. 122 (1): 26–35. https://doi.org/10.1642/0004-8038(2005)122[0026:EOCSFM]2.0.CO;2.

Department of the Environment. (2015). Listing Advice—Lewin's Rail (Tasmanian). Unpublished Report, Canberra, Australia: Australian Government. Available at: http://www.environment.gov.au/biodiversity/threatened/species/lewins-rail-listing-advice. Accessed 20 June 2016.

Digby A., Towsey M., Bell B. D., and Teal P. D. (2013). Practical comparison of manual and autonomous methods for acoustic monitoring. Methods in Ecology and Evolution. 4 (7): 675–683. https://doi.org/10.1111/2041-210X.12060.

Frank E., Hall M. A., and Witten I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, Fourth Edition.

Gage S. H. and Farina A. (2017). The role of sound in terrestrial ecosystems: Three case examples form Michigan, USA. In: Ecoacoustics: The Ecological Role of Sounds, First Edition, edited by Farina A. and Gage S. H. Hoboken, NJ: John Wiley and Sons Ltd. ISBN: 978-1-119-23069-4.

Gage S. H., Towsey M., and Kasten E. P. (2017). Analytical methods in ecoacoustics. In: Ecoacoustics: The Ecological Role of Sounds, First Edition, edited by Farina A. and Gage S. H. Hoboken, NJ: John Wiley and Sons Ltd. ISBN: 978-1-119-23069-4.

Garnett S. T., Szabo J. K., and Dutson G. (2011). The Action Plan for Australian Birds. Collingwood, Victoria, Australia: CSIRO Publishing.

Goeau H., Glotin H., Vellinga W., Planqué R., and Joly A. (2017). LifeCLEF Bird Identification Task 2017. CLEF 2017—Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 2017, 1–9. Available at: http://ceur-ws.org/Vol-1866/invited_paper_8.pdf. Accessed 28 February 2018.

Hutson A. M., Mickleburgh S. P., and Racey P. A. (comp.). (2001). Microchiropteran Bats: Global Status Survey and Conservation Action Plan. IUCN/SSC Chiroptera Specialist Group. Gland, Switzerland and Cambridge, UK: IUCN.

Kunz T. H. and Parsons S. (2009). Ecological and Behavioral Methods for the Study of Bats. Baltimore: Johns Hopkins University Press.

Phillips Y. (2018a). Gympie Bat Communities: Gympie National Park. [Queensland University of Technology]. . https://doi.org/10.4225/09/589d21a7495cd.

Phillips Y., Towsey M., and Roe P. (2017). Visualization of environmental audio using ribbon plots and acoustic state sequences Paper presented at the IEEE International Symposium on Big Data Visual Analytics (BDVA), Adelaide, SA, Australia, 7–10 November. https://doi.org/10.1109/bdva.2017.8114628.

Phillips Y., Towsey M., and Roe P. (2018b). Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation. Plos One. 13 (3): e0193345. https://doi.org/10.1371/journal.pone.0193345.

Pieretti N., Farina A., and Morri D. (2010). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). Ecological Indicators. 11 (3): 868–873. https://doi.org/10.1016/j.ecolind.2010.11.005.

Pijanowski B. C., Farina A., Gage S. H., Dumyahn S. L., and Krause B. L. (2011). What is soundscape ecology? An introduction and overview of an emerging new science. Landscape Ecology. 26 (9): 1213–1232. https://doi.org/10.1007/s10980-011-9600-8.

Priyadarshani N., Marsland S., and Castro I. (2018). Automated birdsong recognition in complex acoustic environments: A review. Journal of Avian Biology Accepted manuscript online: 6 January 2018. https://doi.org/10.1111/jav.01447.

Priyadarshani N., Marsland S., Castro I., and Punchihewa A. (2016). Birdsong denoising using wavelets. Plos One. 11 (1): e0146790. https://doi.org/10.1371/journal.pone.0146790.

Savage J. M. (2002). The Amphibians and Reptiles of Costa Rica. Chicago: University of Chicago Press. 163.

Schafer R. M. (1994). The Soundscape: Our Sonic Environment and the Tuning of the World, 2nd EditionRochester, USA: Destiny Books.

Sueur J., Farina A., Gasc A., Pieretti N., and Pavoine S. (2014). Acoustic indices for biodiversity assessment and landscape investigation. Acta Acustica United with Acustica. 100 (10): 772–781. https://doi.org/10.3813/AAA.918757.

Sueur J., Pavoine S., Hamerlynck O., and Duvail S. (2008). Rapid acoustic survey for biodiversity appraisal. PLoS One. 3 (12): e4065. https://doi.org/10.1371/journal.pone.0004065.

Towsey M. (2017). The Calculation of Acoustic Indices Derived From Long-Duration Recordings of the Natural Environment. Available at: https://eprints.qut.edu.au/110634. Accessed 27 March 2018.

Towsey M., Truskinger A., and Roe P. (2015). The navigation and visualisation of environmental audio using zooming spectrograms. International Conference on Data Mining Workshop (ICDMW), IEEE, Atlantic City, New Jersey, USA, 14–17 November 2015.

Towsey M., Truskinger A., and Roe P. (2016). Long-duration Audio-recordings of the Environment: Visualisation and Analysis. Available at: http://research.ecosounds.org/research/eadm-towsey/long-duration-audio-recordings-of-the-environment. Accessed 27 March 2018.

Towsey M., Truskinger A., Cottman-Fields M., and Roe P. (2018). Ecoacoustics Audio Analysis Software v18.03.0.41 (Version v18.03.0.41). Zenodo. http://doi.org/10.5281/zenodo.1188744.

Towsey M., Zhang L., Cottman-Fields M., Wimmer J., Zhang J., et al. (2014). Visualization of long-duration acoustic recordings of the environment. Procedia Computer Science. 29: 703–712. https://doi.org/10.1016/j.procs.2014.05.063.

Wildlife Acoustics (2017). Wildlife Acoustics—Bioacoustics Monitoring Systems for Bats, Birds, Frogs, Whales, Dolphins and Many Other Species. Available at: http://www.wildlifeacoustics.com. Accessed 27 March 2018.

Woinarski J. C. Z., Woolley L. A., Garnett S. T., Legge S. M., Murphy B. P., et al. (2017). Compilation and traits of Australian bird species killed by cats. Biological Conservation. 216: 1–9. https://doi.org/10.1016/j.biocon.2017.09.017.

Xie J., Towsey M., Zhang J., and Roe P. (2016). Adaptive frequency scaled wavelet packet decomposition for frog call classification. Ecological Informatics. 32: 134–144. https://doi.org/10.1016/j.ecoinf.2016.01.007.

Znidersic E. (2017). Camera traps are an effective tool for monitoring Lewin's Rail (*Lewinia pectoralis brachipus*). Waterbirds. 40 (4): 417–422. https://doi.org/10.1675/063.040.0414.

Znidersic E. (2018). Lewins Rail recordings from Tasman Island Tractor. [Queensland University of Technology]. https://doi.org/10.4225/09/5a72a3a2c33a6.