



**Queensland University of Technology**  
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Sankupellay, Mangalam, Towsey, Michael W., Truskinger, Anthony, & Roe, Paul

(2015)

Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. In

*IEEE International Symposium on Big Data Visual Analytics (BDVA 2015)*, 22-25 September 2015, Hobart, Tas.

This file was downloaded from: <https://eprints.qut.edu.au/88899/>

**© Copyright 2015 IEEE**

**Notice:** *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

# Visual fingerprints of the acoustic environment

The use of acoustic indices to characterise natural habitats

M. Sankupellay, M. Towsey, A. Truskinger and P. Roe

Electrical Engineering and Computer Science School

Queensland University of Queensland,

Brisbane, Australia

m.sankupellay; m.towsey; a.truskinger; p.roe@qut.edu.au

**Abstract** — Acoustic recordings play an increasingly important role in monitoring terrestrial environments. However, due to rapid advances in technology, ecologists are accumulating more audio than they can listen to. Our approach to this big-data challenge is to visualize the content of long-duration audio-recordings by calculating acoustic indices. These are statistics which describe the temporal-spectral distribution of acoustic energy and reflect content of ecological interest. We combine spectral indices to produce false-color spectrogram images. These not only reveal acoustic content but also facilitate navigation. An additional analytic challenge is to find appropriate descriptors to summarize the content of 24-hour recordings, so that it becomes possible to monitor long-term changes in the acoustic environment at a single location and to compare the acoustic environments of different locations. We describe a 24-hour ‘acoustic-fingerprint’ which shows some preliminary promise.

**Keywords**— *visualisation of acoustic data; soundscape ecology; self-organising maps; acoustic environment.*

## I. INTRODUCTION

Acoustic recordings play an increasingly important role in monitoring terrestrial environments and they contribute to several kinds of ecological study concerning biodiversity [1], environmental health [2], threatened species, invasive species [3] and climate change [4]. In the last few years, increasing awareness of sound in the environment has given rise to a new discipline, *soundscape ecology*, which investigates the temporal, spatial and spectral distribution of sound through a landscape and its relationship to important ecosystem processes and to human activity [5].

Because recording devices can be deployed in the field for days or weeks on end, audio recordings are a useful methodology for large-scale environmental monitoring. Expert ecologists need not waste time with repeated visits to distant field sites. Rapid advances in recording and computing technology have made it possible to leave unattended acoustic sensors in exposed locations for weeks, even months, of continuous recording. However, it is clearly impossible to listen to all that is collected. Therefore, much research has been devoted to automated and semi-automated methods of acoustic analysis, a difficult task because the content of environmental recordings is unconstrained. Three analysis tasks have received particular attention: automated species recognition [6-8]; protocols for sampling from audio [9] and the extraction of acoustic indices as surrogates for biological activity [10-12].

Despite the development of automated techniques, the accumulation of environmental recordings can rapidly become overwhelming. This is the classical big-data problem – data acquisition is easy but data curation, search, analysis and visualization is difficult. A single 24-hour recording, even when compressed as MP3, is over a gigabyte in size. After only seven years of collecting recordings from around Australia, our lab is now managing the equivalent of 16 continuous years of audio in 259,000 recordings. At 35 TB of data, this would *not* be considered a big-data problem if it were *text*. But long-duration audio recordings (up to 24 hours) are opaque and standard audio software, if it can work with them at all, is clumsy.

### A. The Visualisation of Sound

The challenge addressed in this paper is how to search and visualize long-duration audio recordings and how to make quantitative comparisons of their content on the scale of hours and days. Visualization should facilitate navigation and present meaningful information about acoustic content *prior* to listening to the underlying audio. Furthermore we do not wish to view recordings in isolation but rather to compare daily cycles and to make quantitative comparisons. This approach is based on the intuition that the brain can integrate visual representations of long duration audio more easily and quickly than by listening.

The standard method to visualize sound is the spectrogram, a grey-scale image whose pixel columns are spectra (obtained from signal frames) and whose rows are frequency bins. Pixel shading encodes acoustic intensity. At a typical time scale of fifty frames per second, visual features are easily linked to acoustic content. However the standard spectrogram does not scale to long duration recordings. A 24-hour recording at a typical temporal scale (0.02s/frame, one frame per pixel column) on a current desktop monitor (35.7px/cm density) would produce an image that occupies about 1.2 km of monitor width.

It has been demonstrated [13] that long-duration (24 hour) recordings of the environment can be meaningfully represented on a standard computer monitor, by first extracting acoustic indices – statistics that summarize the temporal and spectral distribution of acoustic energy in a one-minute segment of recording. There is a growing literature on the ecological uses of acoustic indices. Of particular interest are the acoustic complexity index [11] and entropy (a measure of the temporal

and/or spectral dispersal of acoustic energy) [14] because these have been shown to be sensitive to biological sources of acoustic activity, especially birds calls.

After calculating a variety of indices at coarse resolution (one spectrum of index values per minute of audio) and preparing 24-hour spectrograms for each index, false-color spectrograms can be prepared by mapping three indices to the red-green-blue (RGB) color scale [13]. The result is a false-color, long-duration spectrogram, such as that in Fig. 1b, having a temporal scale of 60s/pixel column (hereafter abbreviated to 60s/px).

An important use for long duration spectrograms of the type shown in Fig. 1b is to verify data integrity. Acoustic sensors in ecological studies are exposed to all kinds of weather and, in practice, management and visualization of long-duration audio must accommodate corrupted and noisy data. Typical practice in bioacoustic studies is to manually ‘weed out’ unwanted audio prior to analysis but such methods cannot scale. In this work, we do not remove audio segments that contain clipping or ‘noise’ due to wind, rain, aircraft, traffic and other human activity. In the context of soundscape ecology, such sounds can be ‘signal’ and not ‘noise’.

Another use of 24-hour spectrograms is that they enable the investigation of *acoustic regimes*.

### B. Acoustic Regimes

The key insight explored in this paper is that natural soundscapes can be approached statistically using acoustic indices and that it is possible to abstract from long duration recordings the concept of an acoustic regime. This is analogous to the concept of climatic regimes which are abstracted from long term statistical measures of weather. The Köppen System [15] recognizes climatic regimes (for example, a Mediterranean Climate and a Sub-tropical Climate) arising from seasonal fluctuations in incident solar radiation and precipitation. Note that the concept of a climatic regime, such as the Mediterranean Climate, only has meaning over an annual cycle (hot dry summers, cold wet winters). In the same way, we expect acoustic regimes to be dominated by the local vegetation structure, the resources which local vegetation offers vocal species and the interaction of the whole with daily and annual cycles.

In this work we only investigate 24-hour acoustic cycles, which we consider to be the smallest sensible unit of time over which to define an acoustic regime. The recognition of acoustic regimes requires a visual analytics approach because the 24-hour temporal scale is too large to integrate aurally. We look for acoustic patterns that are characteristic for the habitat and we derive an acoustic fingerprint which shows some preliminary promise in characterizing the 24-hour acoustic regime at a location.

## II. METHODS

### A. The Audio Recordings

The audio recordings used in this study were obtained from two locations:

1. The 51 ha, QUT Samford Ecological Research Facility (SERF) located in the Samford Valley, some 19 km north-west of Brisbane City, Australia; and
2. In the vicinity of Mount Byron Creek, just outside the western boundary of D’Aguilar National Park, some 50 km north-west of Brisbane City.

The dominant vegetation at both locations is open-forest to woodland comprised primarily of *Eucalyptus tereticornis*, *E. crebra* (and sometimes *E. siderophloia*) and *Melaleuca quinquenervia* in moist drainage. Both locations have similar species but in different proportions. Both locations have a sub-tropical climate and receive approximately 1000mm of rainfall per year. Recordings at both locations were made in the usually dry spring month of October and no rain fell at either location in the days of this study.

Four sites were selected within the SERF location, labelled NE, SE, NW and SW according to their relative compass positions and some 200-300 m distant from one another. Two sites were selected at Mount Byron as described in Table 1. Note that a creek site was chosen at both locations, both of which contained more dense vegetation than the non-creek sites. An ‘edge’ site was selected at SERF between woodland and grass land. The principal difference between the sensor locations is that the Mount Byron sensors were placed in corridors of woodland vegetation running through pasture. All the SERF sensors (except SE) were located in habitat interiors.

TABLE I. SUMMARY OF TWELVE AUDIO RECORDINGS USED IN THIS STUDY

Location	Site	Sensor habitat	Dates
SERF	NW	Creek site - densely vegetated	13, 14 <sup>th</sup> Oct 2010
SERF	SW	Open eucalypt woodland	13, 14 <sup>th</sup> Oct 2010
SERF	NE	Open eucalypt woodland	13, 14 <sup>th</sup> Oct 2010
SERF	SE	Edge site between open eucalypt woodland and grassland	13, 14 <sup>th</sup> Oct 2010
Mt. Byron	2	Creek site - within strip of dense vegetation running through pasture	16, 17 <sup>th</sup> Oct 2013
Mt. Byron	4	Strip of more exposed open woodland running through pasture.	16, 17 <sup>th</sup> Oct 2013

Two consecutive days of continuous recording (48 hours) were obtained from each of the six sites. These were treated as twelve 24-hour recordings. In addition to bird calls, the recordings contain distant sounds due to muffled rural traffic, airplanes, dog barks and moderate wind gusts. There were no

sounds due to rain in any of the recordings. All sensor boxes were attached to a tree at chest height.

### B. The Recording Hardware

Recordings from SERF were obtained using Olympus DM-420 digital recorders housed in custom built weatherproof cases. Two external omnidirectional electret microphones were attached to the recorder which was powered by four D-cell batteries, providing up to 20 days of continuous recording. Data were stored internally in stereo MP3 format (128 kbps, 22050 Hz) on 32GB Secure Digital memory cards [9].

Recordings from Mount Byron were obtained with a battery-powered, weatherproof Song Meter (SM2) box (Wildlife Acoustics, <http://www.wildlifeacoustics.com/products/song-meter-sm2-birds>). Recordings were two-channel, sampled at 22.05 kHz and saved in WAC4 format. WAC4 compresses 16 bit samples to 12 bit by removing the least significant 4 bits. This reduces the dynamic range from 87.3 dB to 63.2 dB. Since environmental background noise does not typically fall below -60 dB, even on the coldest winter nights, WAC4 compression does not compromise our recordings despite being a lossy format.

### C. Signal Processing

For ease of data processing, each 24-hour stereo-recording was divided into 1435 one-minute segments and mixed to mono. (Note: The last five minutes of each 24-hour period were not recorded while the audio data was written to file.) FFTs were calculated over non-overlapping 512 windows using a Hamming window (2584 frames / one-minute of audio). The spectrum derived from each frame has 256 frequency bins, spanning 11025 Hz (43.06 Hz per bin). The spectrum was smoothed with a moving average filter (window width = 3). It should be noted that all the bio-acoustic activity of interest in this study is below 11025 Hz.

### D. Spectral Indices

Two kinds of acoustic indices were calculated: 1. spectral indices, each consisting of a vector of values, one for each frequency bin; 2. summary indices, each being a single value to describe one-minute of recording. In this section we describe the derivation of the spectral indices. We describe calculation of summary indices in Section II-F.

We calculated three spectral indices:

1. **ACI spectrum:** Calculated from the amplitude spectrogram. 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. See Pieretti et al. (2011) for more detail.
2. **ENT spectrum:** Calculate the temporal entropy ( $H_t$ ) of all the values in each frequency bin of the amplitude spectrogram. The squared amplitude values were normalized to unit area and treated as a probability mass function whose entropy was calculated. Entropy is a measure of energy dispersal. We converted this to 'energy concentration' by calculating  $ENT[f] = 1 - H_t(f)$ , where  $f$  is an index over frequency bins.

3. **EVN spectrum:** A vector of counts of acoustic events in each frequency bin. Calculated from the noise-reduced dB spectrogram. An event occurs when the spectrogram cell value crosses the 3 dB threshold from below to above.

### E. False-colour long-duration spectrograms

Each of the above indices (Section II-D: Spectral Indices) was prepared as a 24-hour spectrogram with 256 frequency bins. At the 60s/spectrum scale, each 24-hour spectrogram contains 1435 spectral columns. False-colour images were prepared by mapping the three indices, ACI, ENT, EVN, to RGB colours respectively. In order to produce informative images that utilize the full colour range, all index values must be normalized as described in [13].

### F. Summary Indices

Fourteen different summary indices were calculated for each minute of the 12 days. Linear correlation coefficients were calculated pairwise for all 14 indices and nine indices with lowest coefficients were selected for further processing.

1. **Background noise:** Estimated from the wave envelope of a one-minute recording using the method of Lamel et al. [16] as described in [12]. The value is in decibels.
2. **Average signal-to-noise ratio:** The average dB value of those frames having an SNR greater than 3 dB.
3. **Acoustic event count:** The number of times that the signal envelope crosses the 3 dB threshold from below to above.
4. **High-frequency coverage:** The fraction of spectrogram cells in the high-frequency band (>4000 Hz) where the spectral amplitude exceeds 0.015. This index is calculated from the amplitude spectrogram after removal of background noise. The suitability of this threshold was determined by trial and error. Assuming background noise in the recording ranges around -45 to -35 dB (typical values), this threshold corresponds to approximately 6 dB above background.
5. **Mid-frequency coverage:** Calculated as for high-frequency coverage but in the mid-frequency band (482Hz–4000 Hz).
6. **Low-frequency coverage:** Calculated as for high-frequency coverage but in the low-frequency band (<482 Hz).
7. **Acoustic complexity index (ACI):** Calculate the ACI spectrum as described in Section D.1. Take the average ACI value in the frequency band 482Hz–8820Hz to avoid low frequency energy due to wind and anthropogenic sources and high frequency artefacts due to mp3 compression.
8. **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 whose entropy was calculated. We converted this to 'energy concentration' by calculating  $1 - H_t$ .

9. **Spectral Entropy** ( $H_s$ ): Calculated from the 482–8820 Hz band of the spectrogram in order to avoid low frequency energy due to wind and anthropogenic sources and high frequency artefacts due to mp3 compression. The average spectrum was calculated from all the frames in a one-minute recording and the spectral entropy was calculated as per summary index 8.

#### G. The Self-organising Map (SOM)

Three clustering techniques were used to cluster the summary indices (Section II-F) so as to reduce the dimensionality of the acoustic data: self-organizing maps [17], Fuzzy C-Mean clustering [18] and hierarchical clustering[19].

The self-organizing map (SOM) was prepared using the Matlab Neural Networks Toolbox R2014a. The SOM consisted of a 10x10 rectangular grid of 100 neurons, each with nine inputs. The SOM was trained for 75,000 epochs of the 17,220 (12 \* 1435) input vectors. The initial neighbourhood size was 3. The SOM weights were batch updated, that is, at the end of each entire pass of the input data.

#### H. Fuzzy C-Mean Clustering

The weight vectors of the trained SOM were clustered using Fuzzy C-Mean clustering (Matlab Fuzzy Logic Toolbox R2014a). The optimal cluster number was determined from the fractional decline in quantization error as a function of cluster number.

#### I. Hierarchical Clustering

For each of the twelve 24-hour recordings, each minute of recording was assigned to a fuzzy cluster based on its maximum membership value. This produced twelve vectors, each vector describing the cluster occupancy for one day of recording. Agglomerative (bottom-up) hierarchical clustering (using the Matlab Statistical Toolbox R2014a) was performed on the twelve vectors to compute a hierarchical clustering tree.

### III. RESULTS AND DISCUSSION

#### A. The 24-hour Spectrograms

A 24-hour acoustic recording, even in mp3 format, cannot usually be opened by standard desktop software. However, we obtained a 24-hour spectrogram of the SERF-SE 13-Oct-2010 recording using the well-known open-source audio-editor, Audacity <<http://sourceforge.net/projects/audacity/>>, on a high-performance computer (Figure 1 (a)). The Audacity spectrogram reveals little acoustic structure because it is compressed from the full-width spectrogram by averaging. Averaging effectively smooths out most of the acoustic structure.

By comparison, a 24-hour false-colour long-duration spectrogram of the same recording (Figure 1 (b)) reveals a surprising amount of temporal-spectral acoustic structure that is ecologically meaningful. The morning chorus is clearly visible starting around 0440h. The sounds made by Orthoptera species (grasshoppers, crickets and katydids) appear as tracks during night-time hours, decreasing in frequency in the early morning due to temperature sensitivity. Even at this temporal scale (a greater than 3000 fold compression of the standard spectrogram), some bird species can be identified because their

calls leave traces in consecutive one-minute spectra. Two obvious examples are crow calls (*Corvus torres*, stacked harmonics, 1000h) and the Grey Fantail (*Rhipidura albiscapa*, green harmonics 5-7 kHz, 0600h-0800h).

The value of this visual approach to compressing long duration acoustic data becomes apparent when we wish to compare recordings from different locations. The long-duration spectrogram for the Mount Byron creek site (Figure 1 (c)) is different from the SERF-SE spectrogram in at least three respects. 1. The morning chorus at SERF is more pronounced indicating a greater activity of different species closer to the microphone; 2. The frequency band with the maximum daytime activity is 3kHz at SERF-SE and 2kHz at Mount Byron Creek; 3. There is more night-time acoustic activity at Mount Byron than at SERF. In particular, one or more owls (*Ninox boobook*) are calling at Mount Byron (activity at 600Hz, early morning hours), and there is a greater variety of orthopteran vocalisation. Both locations have crows during daylight hours and a strong cicada chorus from 1800h - 1845h. The bright red colouration in the high frequency band of the SERF-SE spectrogram is due to the sensitivity of the ACI index to artefacts introduced by mp3 compression and does not reflect biological activity. Note that this band was excluded from the calculation of the ACI and  $H_s$  summary indices.

The acoustic differences between these locations can likely be attributed to the presence of permanent water at the Mount Byron Creek site compared to the dry location of the SERF-SE site. However we do not wish to draw ecological conclusions here but rather to emphasize that such acoustic comparisons could not be easily made if one depended on listening to excised one-minute samples.

#### B. The Self-organising Map (SOM) & Acoustic Clusters

A 10x10 SOM was prepared using the nine acoustic indices derived from each minute of all twelve 24-hour recordings – for a total of 17,220 nine-element vectors. After training the SOM, the 100 weight vectors were clustered using Fuzzy C-Means Clustering. The fractional decline in quantization error as a function of cluster number indicated that 27 clusters was optimal for the 100 nodes. The clusters are identified with letters in Figure 2 (a).

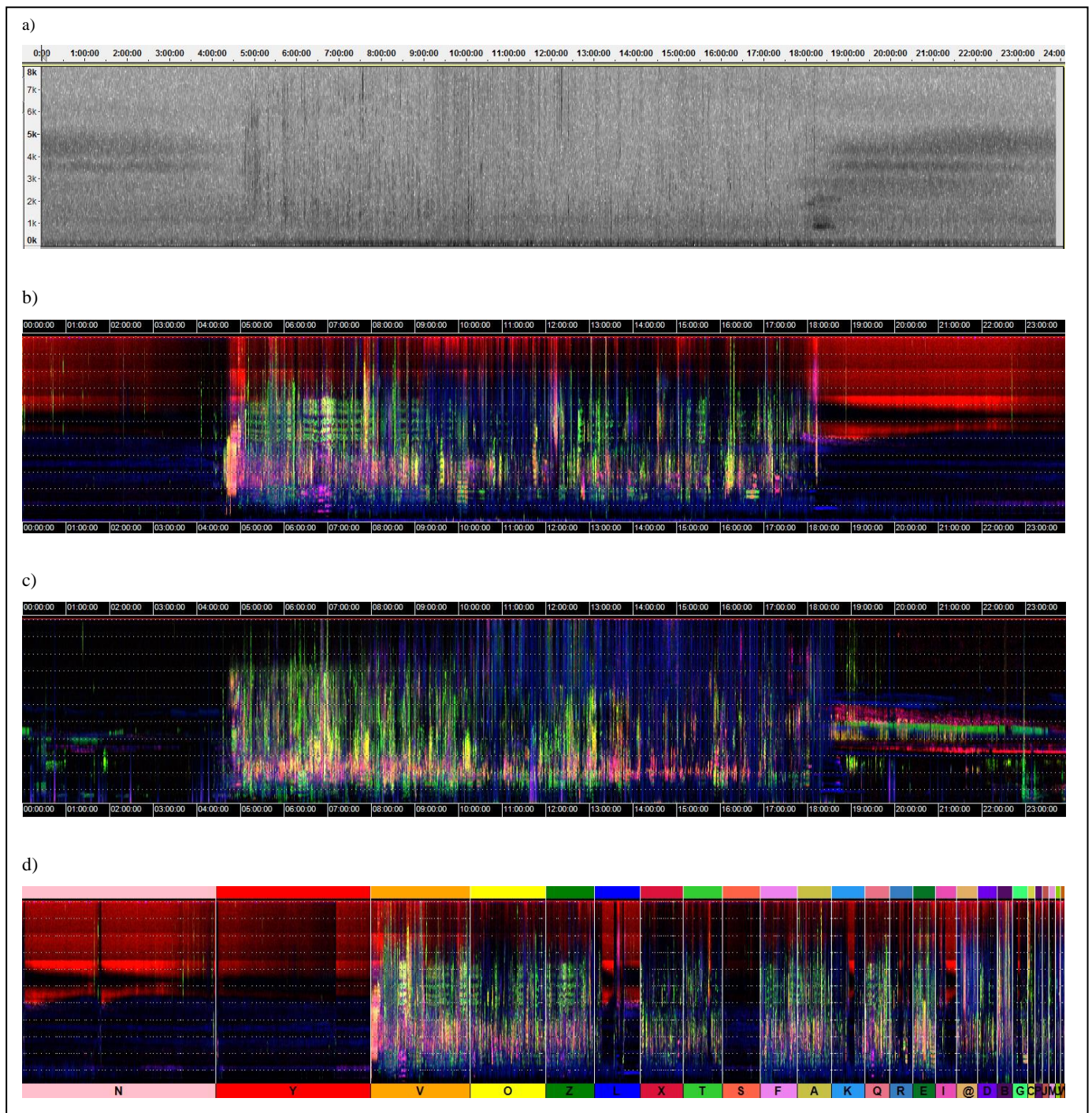
In order to visualize what had been learned by the SOM, each acoustic index vector for the SERF-SE recording (13-Oct-2010) was input to the SOM and assigned to the node with maximal activity. The count of inputs assigned to each node is shown in Figure 2 (c). It can be seen that some nodes received many ‘hits’ and some received none. The median minute (rounded to the nearest 10 minute) of the range of minutes assigned to each node is shown in Figure 2 (b). It is apparent that night-time minutes tend to map to the upper part of the SOM map whereas daytime minutes map to the lower part of the SOM. It is also apparent that the nodes with the greatest number of hits are night-time nodes.

In order to verify the validity of the 27 clusters, we prepared a cluster spectrogram for the SERF-SE recording (Figure 1(d)).



Clusters N, Y, L and S all contain night-time minutes and map to the top centre of the SOM. Clusters V, O, Z, Q, E, A and F all contain day-time minutes and map to the bottom of the

SOM. In particular, cluster V contains the morning chorus and subsequent minutes.



(a) A 24-hour spectrogram of the SERF-SE 13-Oct-2010 recording using the well-known open-source audio-editor, Audacity (y-axis: 0-8kHz). (b) A 24-hour false-colour long-duration spectrogram of the same recording in (a). This image was constructed by combining spectrograms of three different acoustic indices in RGB color (see text). Vertical gridlines are at one hour intervals, starting and ending on midnight. The horizontal gridlines are at 1 kHz intervals. Note: This image is better viewed when enlarged 200% or see the annotated image here: <http://www.ecosounds.org/VIS15/1> (c) The long-duration spectrogram for the Mount Byron creek site. (d) A 'cluster spectrogram' showing how the minutes in the 24-hour spectrogram (Figure 1b) were assigned to clusters using the SOM and Fuzzy-C-Means clustering (the y-axis for figure (b), (c) and (d) is 0 – 11kHz).

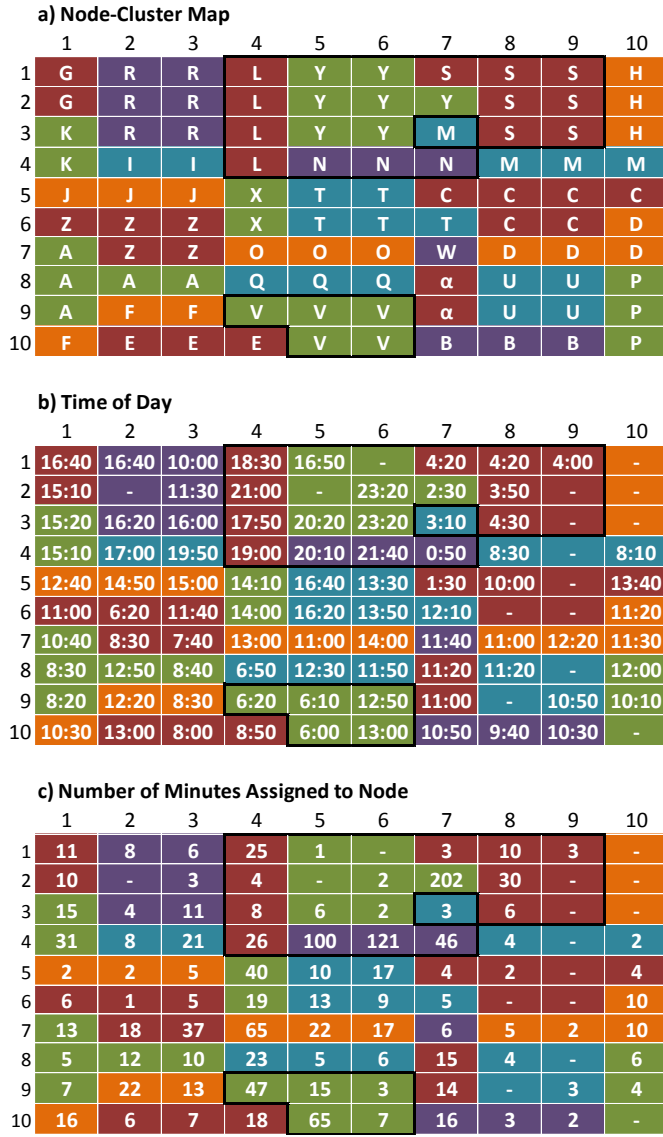


Fig. 1. The 10×10 SOM after training on the 12 recordings (a total of 17,220 nine-element vectors). (a) The result of clustering the 100 weight vectors using fuzzy C-means clustering. The clusters are identified with letters. (b) The median minute (rounded to nearest 10 minute) of all the minutes assigned to each node for the SERF-SE recording (13-Oct-2010). Clusters N, Y, L and S all contain night-time minutes and they map together in the top centre of the SOM. Clusters V, O, Z, Q, E, A and F all contain day-time minutes. Cluster V contains the morning chorus and subsequent minutes. (c) The number of input vectors (minute recordings) assigned to each node for the SERF-SE recording (13-Oct-2010).

### C. 24-Hour Acoustic Fingerprints

As indicated in the introduction, one of the objectives in this work was to derive an acoustic ‘fingerprint’ that would characterize a 24-hour cycle of acoustic activity for a given habitat/location. Given such a fingerprint, it would then become possible to investigate the ‘acoustic regime’ hypothesis, that persistent diurnal patterns of sound are associated with the underlying ecosystem processes in a habitat.

The 27 clusters derived from the SOM nodes can be treated as 27 *acoustic states* (using the language of finite state

automata). The patterns for which we are searching should be expected to occur within the diurnal sequence of acoustic states. Patterns may lie in the state frequencies (fraction of day spent in each state), state transitions or higher order transitions. To demonstrate the possibilities of this approach, we calculated a *state occurrence histogram* for each of our 12 recordings. (Note: in order to avoid ambiguity in the meaning of the word *frequency*, we will use the term *state occurrence histogram* rather than state frequency histogram.)

Figure 3 compares the 24-hour state occurrence histograms for SERF-SE, 13 Oct 2010 and Mt. Byron-2, 17 Oct 2013. The states/clusters have been arranged in approximate temporal sequence (difficult because some states reoccur over an extended period) with day on the left and night on the right. It is apparent that night-time states have the highest occurrence scores because there is less change in the pattern of acoustic activity at night. However the dominant night-time states are different for the two recordings (states N and Y in case of SERF-SE; states H and G in case of Mt. Byron-2). These differences arise for two reasons:

1. The Mt Byron-2 site has more varied night-time acoustic activity due both to birds and orthopteran species;
2. The SERF-SE recording contains mp3 artefacts to which the ACI index is sensitive.

The morning chorus state V is common to both recordings. However, the overall occurrence distribution for the day-time states is different for the two locations. For example, the SERF-SE recording has only two states with over 100 ‘hits’, compared with six states for the Mt. Byron-2 recording, which implies a more even distribution of state occupancy for the latter. It is such kinds of comparison that we believe will reveal interesting insights into the underlying ecosystem processes.

To determine whether the state occurrence histograms provide a ‘meaningful’ acoustic fingerprint for a 24-hour recording, we subjected the twelve fingerprints (one for each recording) to hierarchical clustering (Figure 4). It is apparent that the between-day distances at each site are smaller than the between-site distances, although this conclusion is somewhat compromised by the different recording techniques used at the two sites, a factor that was beyond our control. In addition, the within-location distances are smaller than the between-location distances. Given that there was no significant weather transition between the pairs of days, the acoustic fingerprint has determined that there is a greater acoustic difference between sites and locations than between consecutive days. We consider this to be a promising indication that the 24-hour acoustic fingerprint should be investigated further.

## IV. CONCLUSION

Long-duration acoustic recordings of the environment are useful in ecological studies but they present a number of challenges. First and most obvious is the big-data challenge – more recordings can be accumulated than listened to. We have described a method to visualise long-duration recordings that both facilitates navigation to acoustic events of interest and also enables a big-picture appreciation of an entire 24-hour recording.

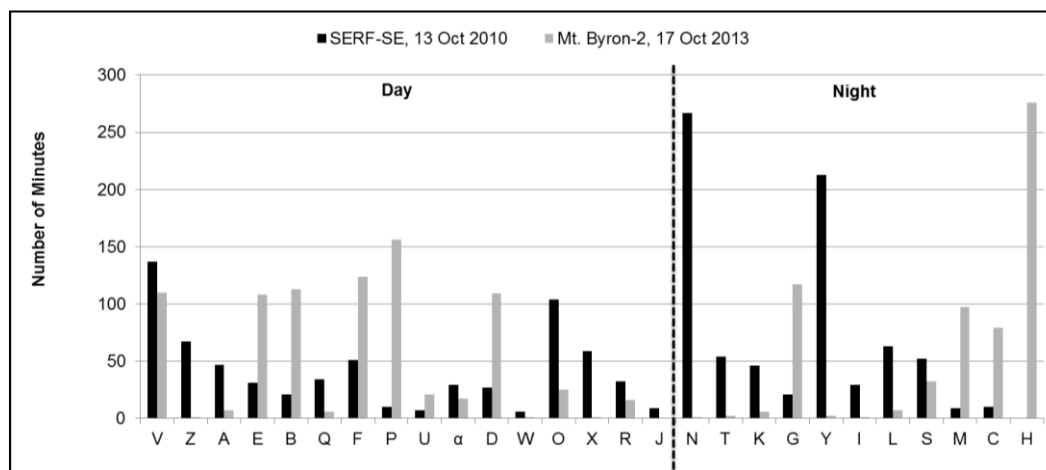


Fig. 2. The 24-hour state occurrence histograms for SERF-SE, 13 Oct 2010 and Mt. Byron-2, 17 Oct 2013. The states/clusters have been arranged in approximate temporal sequence with day on the left and night on the right. Night-time states have the highest occurrence scores. However the dominant night-time states are different for the two recordings (states N and Y in case of SERF-SE; states H and G in case of Mt. Byron-2).

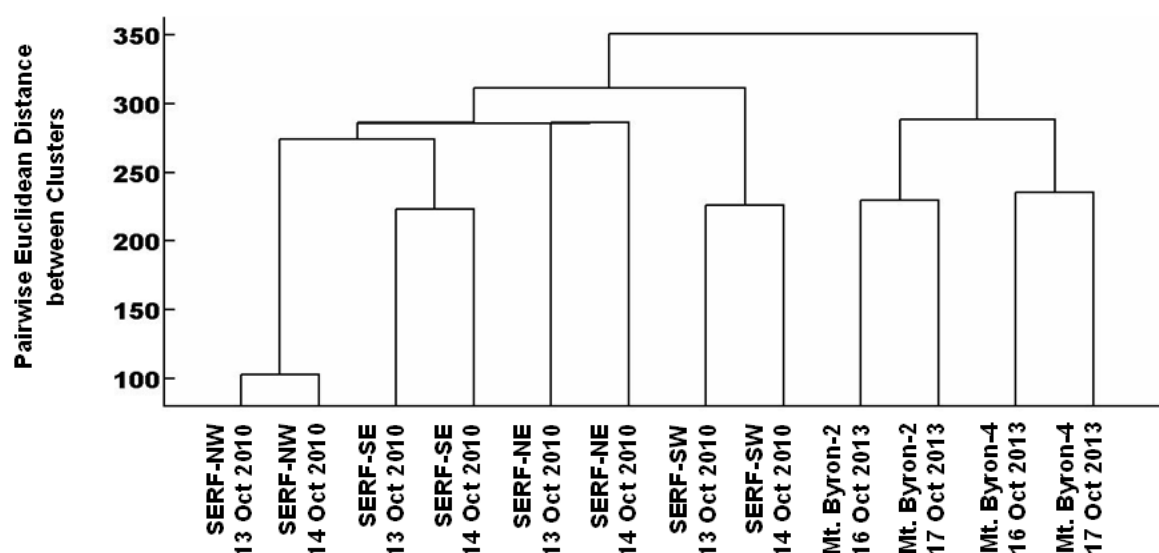


Fig. 3. Agglomerative hierarchical cluster tree of the 12 days of recording based on the 24-hour acoustic 'fingerprints'. The between-day distances at each site are smaller than the between-site distances. In addition, the within-location distances are smaller than the between-location distances.

Note that the ability to visualise a 24-hour period in a single glance is important because many ecological patterns are only apparent on a 24-hour scale.

It is not enough, however, to visualise acoustic data. We also need to interrogate it. Therefore the second challenge addressed in this paper is to make *quantitative comparisons* of different acoustic environments. We have described a 24-hour 'acoustic-fingerprint' which shows preliminary promise to enable comparisons of the acoustic environment at different locations. We deliberately selected locations having similar vegetation structure and species in order to make the 'fingerprinting' task more difficult. The main difference between our two locations is that the Mount Byron sites are located in vegetarian corridors passing through pasture land.

They are therefore more disturbed than three of the four SERF sites, which are 'deep' or 'internal' habitats.

We used spectral indices to address the visualisation challenge because the concept of a spectrogram is well established, despite the greater than 3000-fold compression of the temporal scale. We used summary indices for the 'analytics' part of the big-data problem (*comparisons* of long-duration recordings) because it was believed advisable to reduce the dimensionality of the spectral indices to obtain effective SOM clustering.

An important feature of our approach is that it can scale to big data. Many studies of ecological acoustics manually 'weed out' noisy or corrupted recordings, that is, recordings containing sounds due to wind, rain, airplanes and other



mechanical sources. But manual effort cannot scale. In the work presented here, nothing is excluded from the long-duration spectrograms, the logic being that it is helpful for an ecologist to recognise rain and wind, either to avoid it or perhaps to select rainy periods if frogs are the species of interest. For the calculation of the summary indices, we avoid the low-frequency band (<482Hz) due to the dominance of non-biological acoustic sources (wind and machine noise). For the calculation of the ACI summary index we also avoid the high frequency band (>8820Hz) because mp3 compression artefacts are most likely to occur in this band.

We anticipate three areas of future work:

1. To provide a zooming facility, which enables the users to drill-down from small scale to large scale spectrograms.
2. To monitor sites over an entire year to determine seasonal changes in acoustic 'fingerprints';
3. To search for additional patterns in 24-hour recordings that reflect ecosystem processes;
4. To incorporate other acoustic indices into the learning of the SOM.
5. To explore other aids to the interpretation of environmental audio visualization, such as the addition of temperature, rainfall, sunrise and sunset information as appropriate

#### ACKNOWLEDGMENT

This research was conducted with the support of the QUT Samford Ecological Research Facility (SERF). The authors wish to thank Mark Cottman-Fields for IT support and stimulating discussion.

#### REFERENCES

- [1] A. Gasc, J. Sueur, S. Pavoine, R. Pellens, and P. Grandcolas, "Biodiversity Sampling Using a Global Acoustic Approach: Contrasting Sites with Microendemics in New Caledonia.," *PLoS ONE*, vol. 8(5), p. e65311, 2013.
- [2] E. P. Kasten, S. H. Gage, J. Fox, and W. Joo, "The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology.," *Ecological Informatics*, vol. 12, pp. 50-67, 2012.
- [3] W. Hu, N. Bulusu, C. T. Chou, S. Jha, A. Taylor, and V. N. Tran, "Design and evaluation of a hybrid sensor network for cane toad monitoring," *ACM Transactions on Sensor Networks (TOSN)*, vol. 5, 2009.
- [4] A. Farina, *Soundscape Ecology. Principles, Patterns, Methods and Applications*: Springer, 2014.
- [5] B. C. Pijanowski, A. Farina, S. H. Gage, S. L. Dumyahn, and B. L. Krause, "What is soundscape ecology? An introduction and overview of an emerging new science," *Landscape Ecology*, vol. 26, p. 1213, 2011.
- [6] A. Digby, M. Towsey, B. Bell, and P. Teal, "A practical comparison of manual, semi-automatic and automatic methods for acoustic monitoring," *Methods in Ecology and Evolution*, vol. 4, pp. 675-683, July 2013 2013.
- [7] D. Stowell and M. Plumbley, "Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning.," *PeerJ*, vol. 2:e488, 2014.
- [8] K.-H. Tauchert and K.-H. Frommolt, "Monitoring of booming bitterns (*Botaurus stellaris*) by acoustic triangulation," *Bioacoustics*, vol. 21, p. 83, 2012.

- [9] J. Wimmer, M. Towsey, P. Roe, and I. Williamson, "Sampling environmental acoustic recordings to determine bird species richness," *Ecological Applications*, vol. 23, pp. 1419-1428, 2013.
- [10] J. Sueur, S. Pavoine, O. Hamerlynck, and S. Duvail, "Rapid Acoustic Survey for Biodiversity Appraisal," *PLoS ONE*, vol. 3(12), p. e4065, 2008.
- [11] N. Pieretti, A. Farina, and D. Morri, "A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI).," *Ecological Indices*, vol. 11, pp. 868-873, 2011.
- [12] M. Towsey, J. Wimmer, I. Williamson, and P. Roe, "The Use of Acoustic Indices to Determine Avian Species Richness in Audio-recordings of the Environment," *Ecological Informatics*, <http://dx.doi.org/10.1016/j.ecoinf.2013.11.007>, 2013.
- [13] M. Towsey, L. Zhang, M. Cottman-Fields, J. Wimmer, J. Zhang, and P. Roe, "Visualization of long-duration acoustic recordings of the environment," presented at the The International Conference on Computational Science (ICCS 2014), Cairns, Australia, 2014.
- [14] M. Depraetere, S. Pavoine, F. Jiguet, A. Gasc, S. Duvail, and J. Sueur, "Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland.," *Ecological Indicators*, vol. 13, pp. 46-54, 2012.
- [15] E. Aguado and J. E. Burt, *Understanding Weather and Climate*, Fifth Edition ed. Upper Saddle River, New Jersey: Prentice Hall, 2010.
- [16] L. F. Lamel, L. R. Rabiner, A. E. Rosenberg, and J. G. Wilpon, "An improved endpoint detector for isolated word recognition," *IEEE Trans. ASSP*, vol. ASSP-29, pp. 777-785, 1981.
- [17] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59-69, 1982.
- [18] J. C. Bezdec, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- [19] L. Rokach and O. Maimon, "Clustering methods," in *Data Mining and Knowledge Discovery Handbook*, L. R. Oded Maimon, Ed., ed US: Springer, 2005, pp. 321-352.