Procedia Computer Science

Volume 29, 2014, Pages 703–712

ICCS 2014. 14th International Conference on Computational Science

Visualization of long-duration acoustic recordings of the environment

Michael Towsey[1], Liang Zhang[1], Mark Cottman-Fields[1], Jason Wimmer[1], Jinglan Zhang[1] and Paul Roe[1]

[1] Queensland University of Technology, Brisbane, Australia.

m.towsey,j.wimmer,jinglan.zhang,p.roe@qut.edu.au;
l68.zhang,m.cottman-fields@student.qut.edu.au

Abstract

Acoustic recordings of the environment are an important aid to ecologists monitoring biodiversity and environmental health. However, rapid advances in recording technology, storage and computing make it possible to accumulate thousands of hours of recordings, of which, ecologists can only listen to a small fraction. The big-data challenge addressed in this paper is to visualize the content of long-duration audio recordings on multiple scales, from hours, days, months to years. The visualization should facilitate navigation and yield ecologically meaningful information.

Our approach is to extract (at one minute resolution) *acoustic indices* which reflect content of ecological interest. An acoustic index is a statistic that summarizes some aspect of the distribution of acoustic energy in a recording. We combine indices to produce false-color images that reveal acoustic content and facilitate navigation through recordings that are months or even years in duration.

1 Introduction

Acoustic recordings of the environment play an increasingly important role in monitoring the biodiversity of terrestrial and aquatic ecosystems. Indeed, recorded audio data can contribute to several kinds of ecological investigation concerning environmental health, threatened species and invasive species. It is fortunate that three major groups of vocal species, birds, insects and frogs are also accepted as important indicators of environmental health (Gregory and Strien 2010). Quite apart from interest in species diversity, 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 (Pijanowski, Farina et al. 2011; Kasten, Gage et al. 2012). From this perspective, the soundscape is a finite resource in which organisms (including humans) compete for bandwidth (Krause 2008).

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

doi:10.1016/j.procs.2014.05.063

Our research group is investigating protocols for the use of environmental recordings (Wimmer, Towsey et al. 2010; Digby, Towsey et al. 2013; Wimmer, Towsey et al. 2013). 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 trained ecologists. Consequently, audio recordings are an attractive methodology for large-scale monitoring (Parsons and Towsey 2014). However, rapid advances in recording and computing technology make it possible to leave unattended acoustic sensors in exposed locations for weeks, even months, of continuous recording. It is clearly impossible for ecologists to listen to even a small fraction of the audio they collect.

The accumulation of environmental recordings presents a classical big-data problem. A single 24-hour recording, even when compressed as MP3, is over 1 GB in size. After six years of collecting recordings from different sites, our lab is now managing 28 TB of data. While data acquisition is easy, data curation, search, analysis and visualization present considerable problems. Standard audio software cannot load long duration recordings (24 hours and more). Their content remains opaque and impenetrable.

The big-data challenge addressed in this paper is how to visualize the content of long-duration audio recordings on multiple scales, from hours, days, months to years. Visualization should enable navigation and should present meaningful information to ecologists. Our approach is to extract *acoustic indices* from audio recordings that reflect content of ecological interest. An acoustic index is a statistic that summarizes some aspect of the distribution of acoustic energy and information in a recording.

There is a growing body of work on the ecological uses of acoustic indices. Some indices are derived from the waveform and others from spectral content. Waveform indices include traditional measures such as signal amplitude and signal-to-noise ratio. More recently, *temporal entropy* (H[t]), a measure of the temporal dispersal of acoustic energy within a recording, has been shown to reflect the number of avian calls in a recording (Sueur, Pavoine et al. 2008).

Spectral indices include *spectral entropy* (H[s]), a measure of acoustic energy dispersal through the spectrum (Sueur, Pavoine et al. 2008), and 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 (Pieretti, Farina et al. 2011). 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.

In this work, we extract the above indices and others at one minute resolution from long duration recordings and use them to produce *false-color* images that reveal acoustic content. We must note the distinction between false-color and *pseudo-color* spectrograms. The latter are produced by mapping one-dimensional (grayscale) spectral power values to color, according to some function. In this work, we display combinations of three indices by mapping them to red, green and blue colors. As long as the three indices capture orthogonal acoustic information, a false-color image of an environmental recording will convey more information than a pseudo-color spectrogram.

Apart from providing ecologically useful information and facilitating navigation, color images can also reveal loss of data integrity. Acoustic sensors in ecological studies are exposed to all kinds of weather and, in practice, management and visualization of long duration data 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 zero-signal, clipping or 'noise' due to wind, rain, planes and traffic.

What constitutes 'noise' in environmental recordings can be a matter of contention. In a non-technical sense, 'noise' is a sound where it is not wanted (adopting the classical definition of a weed). However in the context of soundscape ecology, 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) may all be of ecological significance. 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, given crickets evenly distributed in the landscape around a sensor, there will be a background murmur of crickets but the chirps of those closest to the microphone will register as specific acoustic events. Likewise, wind gusts will stand out as specific low-frequency events within a background of noise generated by constantly moving air.

2 Materials and Methods

2.1 Hardware

All but one recording in this study 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 the environmental background noise at our location (see below) does not fall below -60 dB, even on the coldest winter nights, WAC4 compression does not compromise our recordings despite being a lossy format.

One 24-hour recording described in this paper was obtained using a custom-built acoustic sensor. The sensor was an Olympus DM-420 digital recorder housed in a weatherproof case and powered by four D-cell batteries, providing up to 20 days of continuous recording. The two external microphones were omni-directional electret. Data were stored internally in stereo MP3 format (128 Kbit/s, 22.05 kHz) on high capacity 32GB Secure Digital memory cards. See Wimmer et al. (Wimmer, Towsey et al. 2013) for more detail.

2.2 Data Sets

Recordings were 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*, *E. crebra* (and sometimes *E. 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.

The sensor boxes were attached to a tree at chest height. A road (some 100 meters distant) meant that recordings contained muffled traffic noise in addition to the sounds of airplanes, dog barks, human speech, severe wind gusts and mild to heavy rain.

2.3 Signal Processing

The stereo recordings were divided into one minute segments and mixed down to mono. To reduce subsequent computational burden, recordings were re-sampled at 17,640 samples per second (after filtering to remove content above the Nyquist of 8820 Hz). Frame size was 512 samples and non-overlapping. Thus there were approximately 4,140 frames per one minute of recording. 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 as described in Towsey et al (Towsey, Wimmer et al. 2013). Eight indices were derived from a wave envelope which was itself derived from the maximum absolute value in each frame. 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).

The spectrum was smoothed with a moving average filter (window width = 3). Some indices (ACI, H[t] and H[s]) were derived from amplitude spectrograms, while others (such as background noise) were calculated after converting amplitude values to decibels using

$$dB = 20.\log_{10}(A).$$

2.4 Acoustic Indices

Of the 14 indices described in Towsey et al (Towsey, Wimmer et al. 2013), we investigated the suitability of seven for constructing false-color images. In this report we describe images using only four of the indices, *Background Noise* (BGN), ACI, H[t] and *Acoustic Cover* (CVR). Each index was calculated from a one-minute segment both as a scalar and as a vector. The scalar was a single real value representing the entire one-minute segment. The vector (of 256 values) represented a summary spectrum for the one-minute segment, each element representing the index value for a frequency bin. The summary indices and spectra were calculated as follows.

BGN spectrum: Calculated from the dB spectrogram applying the method of Lamel et al (Lamel, Rabiner et al. 1981) to each frequency bin as described in (Towsey 2013). The values are given in decibels. Recall that our technical definition of *background noise* is that acoustic energy removed using the method of Lamel et al.

BGN index: Estimated from the wave envelope also using the method of Lamel et al. The value is given in decibels.

ACI spectrum: For each frequency bin in a one-minute spectrogram, calculate the average absolute fractional change in spectral amplitude from one spectrum to the next. That is:

$$ACI_f = \sum_i |a_i - a_{i-1}| / \sum_i a_i ,$$

where i indexes over all amplitude values in frequency bin, f, of the spectrogram. $\sum_i a_i$ is a normalizing factor. See (Pieretti, Farina et al. 2011) for more detail.

ACI index: The average of values 15 – 256 in the ACI spectrum - see (Pieretti, Farina et al. 2011).

H[t] spectrum: The entropy of each frequency bin in the amplitude spectrogram. The squared amplitude values were normalized to unit area and treated as a probability mass function (*pmf*). The entropy of the signal was calculated as:

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

where i is an index over all values, 0 – N-1, in the frequency bin and N is the number of spectra in the spectrogram (Sueur, Pavoine et al. 2008).

H[t] index: The entropy of the squared values of the signal envelope and calculated as for the entropy of each frequency bin.

CVR spectrum: The fraction of cells in each frequency bin of the noise-reduced spectrogram where the spectral power exceeds 2 dB.

CVR index: The average of values 15 – 256 in the CVR spectrum. Note that, in the calculation of the ACI and CVR indices, we removed the lowest 14 bins (0 - 482 Hz) in order to avoid traffic and airplane sounds. Non-removal of these low frequency bands meant that the extracted indices were dominated by non-biological acoustic sources.

2.5 Preparation of False Color Images

False-color images were prepared by mapping three indices to the primary colors, red, green and blue (RGB) respectively. Note that H[t] was 'reversed' (1-H[t]) to yield an acoustic *concentration* index rather than an acoustic dispersal index. Although the three indices, ACI, H[t] and CVR, fall in the range [0, 1] by definition, in practice they do not occupy the full interval. ACI seldom takes values less than 0.3 or greater than 0.7. Likewise H[t] seldom takes values less than 0.5 and CVR is seldom greater than 0.3. In order to maximize the color range, we normalized all indices linearly within a

narrower range depending on the index and image. Typical min-max values were: ACI, 0.4 - 0.7; H[t], 0.5 - 0.98; and CVR, 0.0 - 0.3. In all cases, values falling below the minimum or above the maximum were truncated to 0 and 1 respectively. Another approach would be to normalize between the minimum and maximum values for each index in the image but this would make it difficult to compare images derived from greatly different acoustic content. We have also explored the option of non-linear normalization to accentuate extreme values as appropriate.

3 Results

3.1 The Standard Spectrogram

Animal sounds of interest to biologists and ecologists are typically viewed in spectrograms whose time-axis is scaled in seconds. A ten second audio clip, sampled at 17,640 Hz and frame width = 512 samples (non-overlapping), will produce a grey scale spectrogram consisting of 344 spectra and 256 frequency bins. Three such spectrograms, illustrating different bird calls, are shown in Figure 1. The

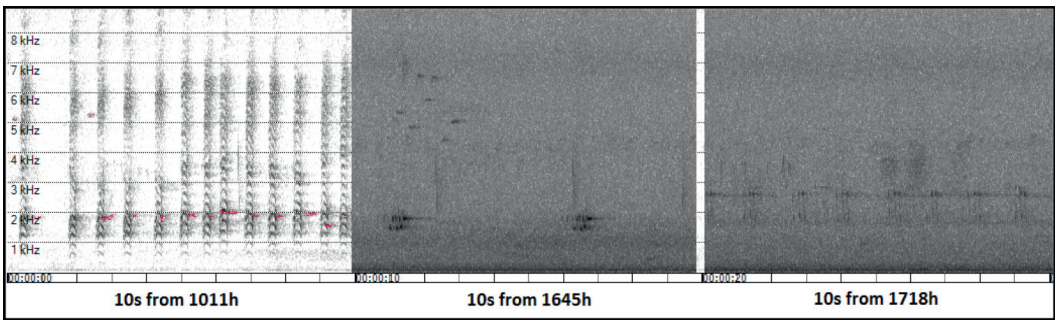

Figure 1: Three 10-second spectrograms selected from different parts of a 24-hour recording, illustrating the calls of three different bird species. All spectrograms have the same time and frequency scales. The left spectrogram has been noise reduced and has grid lines at 1 kHz intervals from 0 - 8.8 kHz. Left: Torresian Crow (*Corvus orru*); Centre: the Olive-backed Oriole, (*Oriolus sagittatus*); Right: the Striated Pardalote (*Pardalotus striatus*).

audio clips were taken from a 24-hour MP3 recording obtained at SERF on the 13th October 2010. Spectrograms can be noise reduced (as in Figure 3, left) to enhance the call of interest but this can also result in loss of detail.

3.2 The 24 hour Spectrogram

A false-color spectrogram of a 24-hour recording (the one from which the spectrograms in Figure 1 were extracted) is illustrated in Figure 2. It was derived by combining three indices, ACI, H[t] and CVR, mapped to RGB respectively. The x-axis extends from midnight to midnight. Since the x-axis scale is now one pixel-column per minute, a greater than 2000x compression is achieved over the standard spectrogram. Note that the frequency scale is unchanged. The bottom image is a grey scale representation of background noise.

A surprising amount of information can be gleaned from these long duration spectrograms and they can be used to navigate a 24-hour recording which would otherwise be opaque and impenetrable. The morning chorus is clearly visible in the top image starting around 0440h while an evening cicada chorus is visible in the lower image starting at 1800h. The sounds made by various species of

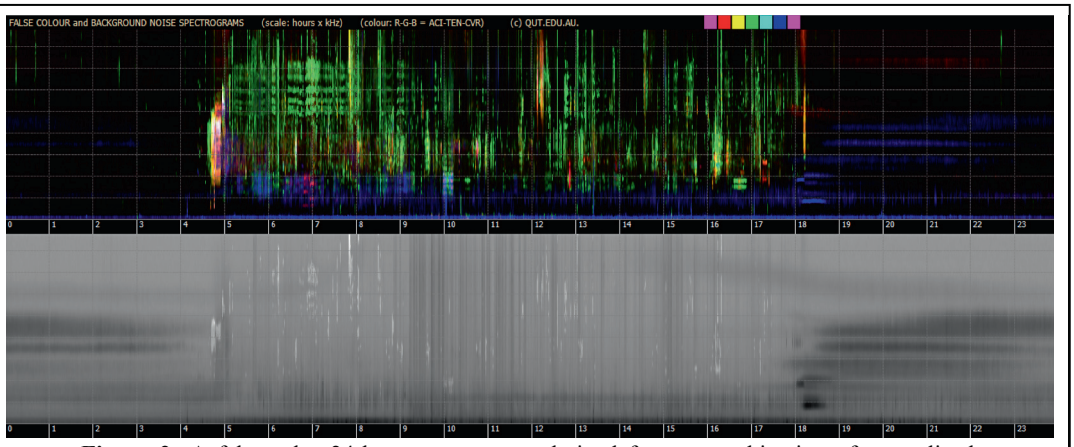

Figure 2: A false-color 24-hour spectrogram derived from a combination of normalized spectrograms for ACI, 1-H[t] and CVR. The vertical gridlines are at one hour intervals, starting and ending at midnight. The horizontal gridlines are at 1 kHz intervals. Note: These images are designed to occupy the width of a computer screen and much detail is lost at publication size. More detail in the above and other images can be obtained if the document is enlarged 200-250%

Orthoptera (grasshoppers, crickets and katydids) appear as tracks during night-time hours. Careful examination reveals that some of the tracks slope downwards in the evening due to the temperature dependence of Orthopteran sounds.

Surprisingly, some calling species can be identified, even at this temporal scale. For example, the crow calls (Figure 1, left), consisting of stacked harmonics, can be identified in the 24-hour spectrogram at 1000-1010h. This is possible because a calling individual leaves a visible trace in consecutive one-minute spectra. One can also observe the arrival and departure of other species (such as the Grey Fantail (*Rhipidura albiscapa*) calling in the 5-7 kHz band during the five hours after dawn; the Yellow-faced Honeyeater (*Lichenostomus chrysops*) calling in the 2-4 kHz band around 0938h; the Olive-backed Oriole (*Oriolus sagittatus*, see Figure 1, center) calling around 1645h; and the Striated Pardalote (*Pardalotus striatus*, Figure 1, right) calling around 1719h. Each of these species leaves visible traces in consecutive spectra.

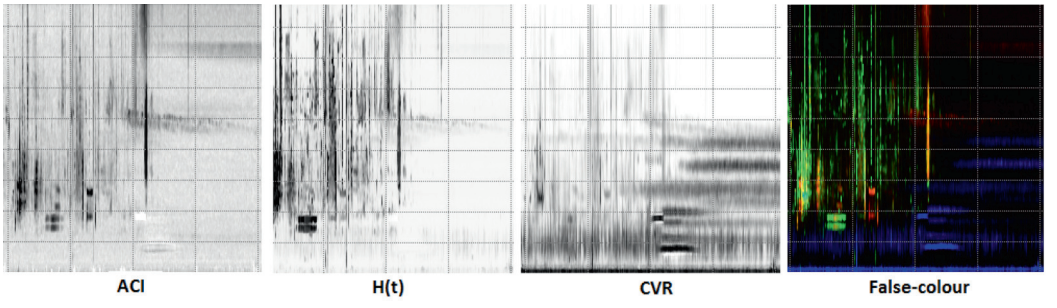

Figure 3: The false-color spectrogram on the right was obtained by combining the ACI, 1-H[t] and CVR spectrograms in red, green and blue colors respectively.

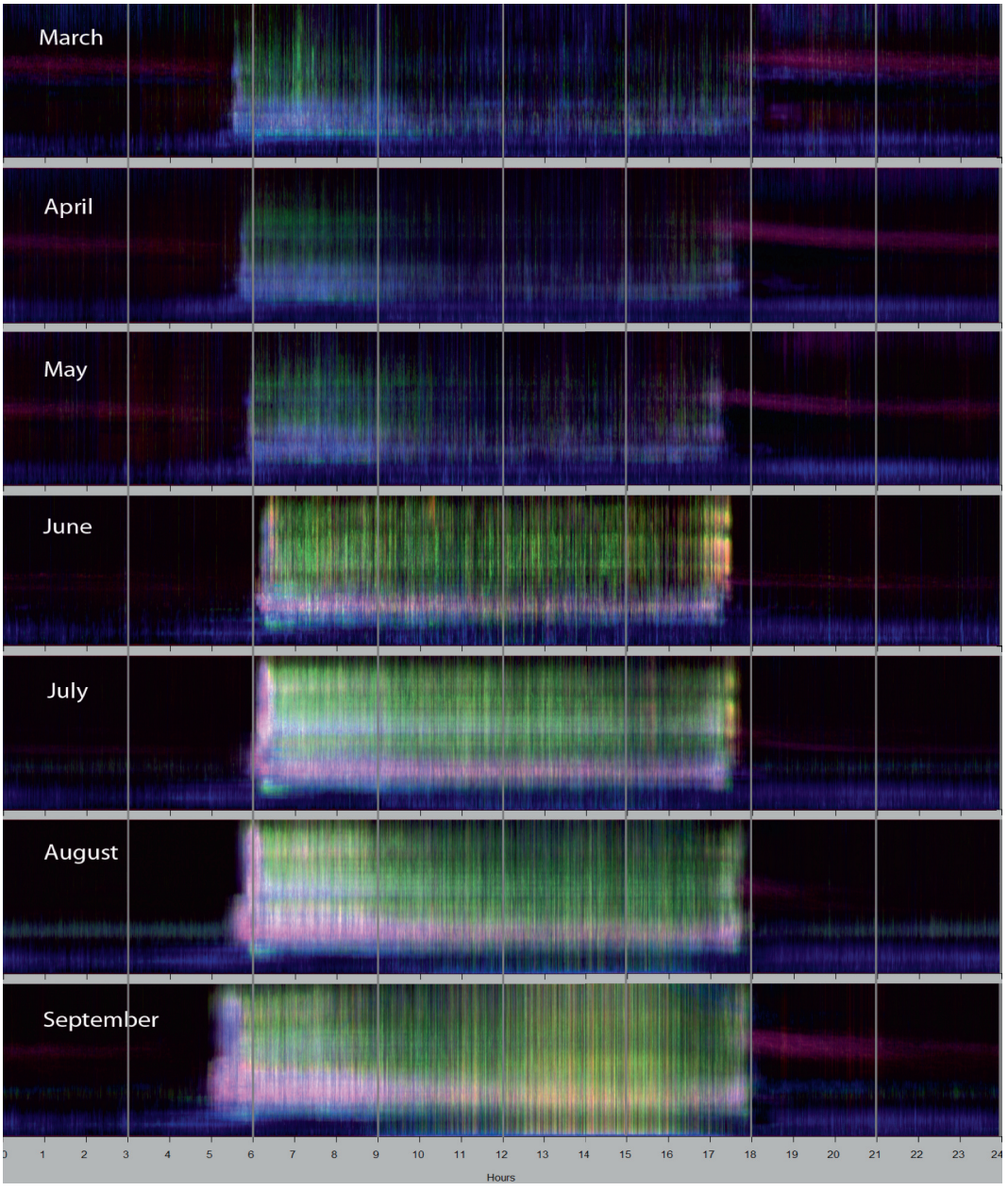

Figure 4: Monthly averaged spectrograms derived from the months March to September, 2013.

It is important that the three indices chosen for a false-color image should convey orthogonal acoustic information (as far as is possible). Figure 3 illustrates the individual spectrograms for ACI, 1-H[t] and CVR for a four hour period (1600 – 2000h in the spectrogram of Figure 2). Note how the indices highlight different features. H[t] responds strongly to calls of the Olive-backed Oriole, because this bird calls only once every 5-6 seconds. Strong, infrequent calls have the effect of concentrating acoustic energy. By contrast, ACI responds more strongly to the calls of the Striated Pardalote and CVR responds to the continuous cicada chorus at 1800hrs. H[t] is particularly useful for picking up infrequent night-time calls.

We have explored various combinations of indices and color-mappings. We have also constructed 'positive' spectrograms in which the zero-activity background is white as opposed to the black background in Figure 2. For examples of other images, see supplementary material (Towsey and Zhang 2014).

Although MP3 recordings at 128kbps have been found suitable for identifying bird calls (Rempel, Hobson et al. 2005), in our experience MP3 compression sometimes generates artefacts in the reconstituted WAV signals. These in turn lead to meaningless index values and spurious shapes in false-color spectrograms (Towsey and Zhang 2014). Consequently, we have discontinued recording in MP3 and use instead Wildlife Acoustics' proprietary WAC4 compression. However the 24-hour recording used to generate the spectrograms in Figures 1, 2 and 3 did not contain MP3 compression artefacts.

3.3 Monthly Average Spectrogram

It is possible to combine acoustic indices into weekly, monthly, seasonal and yearly composites by averaging the values of contributing indices over consecutive days. As an example, Figure 4 illustrates seven false-color spectrograms derived for the calendar months of March to September, 2013. For each acoustic index, we average the available daily (1440×256) matrices for a month, taking into consideration missing days due to equipment failure, etc. The averaged matrices for ACI, 1-H[t] and CVR were assigned to RGB respectively, as described above for 24-hour false-color spectrograms. The monthly spectrograms have a more "washed-out" appearance due to averaging but seasonal changes in the acoustic landscape are clearly visible. The morning chorus is most strong during the late winter, early spring months while night-time Orthopteran sounds are minimal during the three winter months.

3.4 Extended Acoustic Summary Images

To facilitate navigation through months and years of continuous recording, 24-hour spectrograms will not be adequate. For this purpose, we have developed Extended Acoustic Summary (EASY) images. Like the 24-hour spectrograms, EASY images map normalized ACI, H[t] and CVR indices to RGB color values respectively. However these images are not spectrograms because frequency information is now lost. Figure 5 shows an EASY image for the same recordings (14th March 2013 to 10th October 2013, a total of 211 days) used to prepare the monthly average spectrograms in Figure 4. Although they cannot convey spectral information, they have the advantage of being compact and extensible. Additional days, weeks, months can be appended to existing images as long as the values are calculated and normalized in the same way.

The curved, white lines in Figure 5 indicate civil-dawn and civil-twilight through the autumn, winter and spring months. Civil-dawn occurs when the sun is six degrees below the horizon, that is, 24 minutes before sunrise. Likewise civil- twilight occurs 24 minutes after sunset. It is apparent that the morning chorus starts with the onset of civil-dawn and likewise bird and other sounds diminish rapidly with the onset of civil- twilight. The morning chorus is most strong during the months when

the sun rises earlier day-by-day and is absent during months when the sun rises later. Note that the day having the latest sunrise does not coincide with the day having the earliest sunset. Changes of false-color during the daylight hours through the seasons are attributable not only to animal sounds.

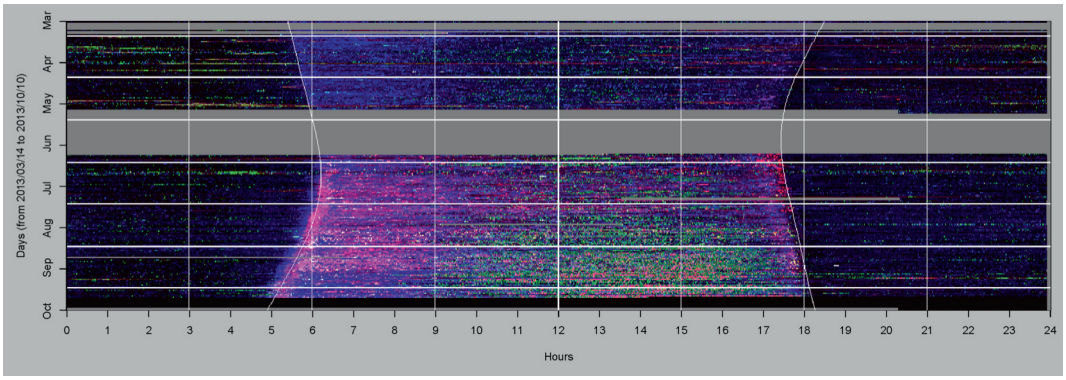

Figure 5: An Extended Acoustic Summary image for the months March to October, 2013. Horizontal grid lines mark the first day of each month. Vertical gridlines are at three hourly intervals. Gray bands indicate missing data.

Seasonal changes in rain and wind also make an impact on the appearance of an EASY image.

4 Conclusion

We have produced false-color spectrograms and EASY images from long duration acoustic recordings of the environment. They have at least two purposes: 1. to expose the acoustic content of long recordings in a meaningful way to ecologists; and 2. to facilitate navigation through long recordings so that an ecologist can 'drill down' and inspect one-minute spectrograms. We have developed software which takes long duration recordings as input, calculates a variety of indices at one-minute resolution and produces a range of different spectrograms. Any one of the spectrograms can be loaded and used for navigation. When the user clicks on the false-color spectrogram, a one-minute grey-scale spectrogram is displayed corresponding to the location of the mouse-click. The one minute segment can be played to verify content. This software is available on request.

We have determined that the practicality of these images for navigation purposes is not strongly sensitive to the color-mapping. However the normalization bounds used to scale the index values do have an effect on the R-G-B balance. We are currently exploring different normalization options. We have adopted a linear frequency scale which tends to give more prominence to the high frequency band than is apparent to the ear. The Mel-scale could be adopted but in our experience it gives too much prominence to the low-frequency band. Birds, which have been the focus of our attention to date, mostly call in the mid-frequency band, so we have persisted with the linear scale.

Finally, it should be noted that many kinds of indices could calculated to highlight audio content of interest. We are presently working on indices that highlight rain, wind and cicada choruses, since for many studies, these are troublesome sources of noise. False-color spectrograms for rain and wind could be used to mask other recording content, depending on the application.

Acknowledgements

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

References

Digby, A., M. Towsey, et al. (2013). "A practical comparison of manual, semi-automatic and automatic methods for acoustic monitoring." Methods in Ecology and Evolution **4**(7): 675–683.

Gregory, R. D. and A. v. Strien (2010). "Wild Bird Indicators: Using Composite Population Trends of Birds as Measures of Environmental Health." Ornithological Science **9**(1): 3-22.

Kasten, E. P., S. H. Gage, et al. (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." Journal of the Audio Engineering Society **56**(1/2).

Lamel, L. F., L. R. Rabiner, et al. (1981). "An improved endpoint detector for isolated word recognition." IEEE Trans. ASSP **ASSP-29**: 777-785.

Parsons, S. and M. Towsey (2014). Report on a workshop to investigate the current status of environmental bio-acoustic monitoring. 8-11 May 2012, (QUT ePrints, http://eprints.qut.edu.au/66572/), Queensland University of Technology, Brisbane.

Pieretti, N., A. Farina, et al. (2011). "A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI)." Ecological Indices **11**: 868–873.

Pijanowski, B. C., A. Farina, et al. (2011). "What is soundscape ecology? An introduction and overview of an emerging new science." Landscape Ecology **26**(9): 1213.

Rempel, R. S., K. A. Hobson, et al. (2005). "Bioacoustic Monitoring of Forest Songbirds: Interpreter Variability and Effects of Configuration and Digital Processing Methods in the Laboratory / Monitoreo bioacústico de aves de bosques: variabilidad en el interpretador y efecto de la configuración del método de procedimiento digital en el laboratorio." Journal of Field Ornithology **76**(1): 1-11.

Sueur, J., S. Pavoine, et al. (2008). "Rapid Acoustic Survey for Biodiversity Appraisal." PLoS ONE **3(12)**(12): e4065.

Towsey, M. (2013). Noise removal from wave-forms and spectrograms derived from natural recordings of the environment. . QUT ePrints, http://eprints.qut.edu.au/61399/. Brisbane, Queensland University of Technology.

Towsey, M., J. Wimmer, et al. (2013). "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.

Towsey, M. and L. Zhang (2014). False-colour spectrograms of long-duration acoustic recordings. Brisbane, Queensland University of Technology. http://eprints.qut.edu.au/66786.

Wimmer, J., M. Towsey, et al. (2010). Scaling Acoustic Data Analysis through Collaboration and Automation. e-Science (e-Science), 2010 IEEE Sixth International Conference on, IEEE.

Wimmer, J., M. Towsey, et al. (2013). "Sampling environmental acoustic recordings to determine bird species richness." Ecological Applications **23**: 1419-1428.