Recognition of Frog Chorusing with Acoustic Indices and Machine Learning

Conference Paper · February 2020

DOI: 10.1109/eScience.2019.00019

8 authors, including:

Hongxiao Gan
Queensland University of Technology

Jinglan Zhang
Queensland University of Technology

Michael Towsey
Queensland University of Technology

Anthony Truskinger
Queensland University of Technology

Recognition of Frog Chorusing with Acoustic Indices and Machine Learning

Hongxiao Gan[1], Jinglan Zhang[1], Michael Towsey[1], Anthony Truskinger[1],
Debra Stark[2], Berndt J. van Rensburg[2], Yuefeng Li[1], Paul Roe[1]
[1]*Queensland University of Technology*
[2]*The University of Queensland*
Brisbane, Australia
Email: {h2.gan, jinglan.zhang, m.towsey, a.truskinger, y2.li, p.roe}@qut.edu.au,
{d.stark, b.vanrensburg}@uq.edu.au

***Abstract*—This research explores the recognition of choruses of two co-existing sibling frog species in long-duration field recordings using false-colour spectrograms and acoustic indices. Acid frogs are a group of endemic frogs that are particularly sensitive to habitat change and competition from other species. The Wallum Sedgefrog (*Litoria olongburensis*) is the most threatened acid frog species facing habitat loss and degradation across much of their distribution, in addition to further pressures associated with anecdotally-recognised competition from their sibling species, the Eastern Sedgefrogs (*Litoria fallax*). Monitoring the calling behaviours of these two species is essential for informing *L. olongburensis* management and protection, and for obtaining ecological information about the process and implications of their competition. Considering the cryptic nature of *L. olongburensis* and the sensitivity of their habitat to human disturbance, passive acoustic monitoring is a suitable method for monitoring this species. However, manually processing the large quantities of acoustic data collected using these methods is time-consuming and not feasible in the long-term. Therefore, there is a high demand for automated acoustic recognition tools to efficiently search months of recordings and identify target species. Our research provides more insight on how to choose acoustic features that efficiently recognise species from large-scale field-collected recordings at a larger scale. The experimental results show that these techniques are useful in identifying choruses of the two competitive frog species with an accuracy of 76.7% on identifying four acoustic patterns (whether the two species occurred).**

***Index Terms*—Ecoacoustics, Species Recognition, Acoustic Indices, Machine Learning, Data Mining, Anuran Species, Frog Chorus Recognition**

I. INTRODUCTION

The Wallum Sedgefrog (*Litoria olongburensis*) is a threatened acid frog species endemic to Australia and a recovery plan has been carried out for them [1]. This species exists within the highly specialised environment of wallum wetlands, which are vulnerable to environmental disturbance from surrounding land uses. Environmental disturbances can reduce habitat quality and allow the introduction of competing sibling species previously excluded by physiological limitations; in particular, the Eastern Sedgefrog (*Litoria fallax*). *Litoria fallax* are known to compete with and exclude *L. olongburensis* from potentially viable habitat. While the exact nature, and underlying mechanisms, of this competition are unknown, it is hypothesised to involve competition for acoustic space (e.g. male advertisement calls). As wallum habitats tend to occur within areas prone to residential and urban development, there are many opportunities for species co-existence, and thus, competition. It is imperative that these species are continually monitored within these areas to determine the impact of *L. fallax* presence, if any, on *L. olongburensis* calling behaviours and long-term survival. Traditionally, anuran monitoring has been conducted through manual calling surveys or physical trapping methods, however due to the cryptic nature of *L. olongburensis* and *L. fallax*, and the sensitivity of wallum habitats, passive acoustic monitoring techniques are preferred.

Passive acoustic sensors deployed in the field provide an efficient tool for monitoring species calling activity and spatial distributions. Sound emitted from the landscape is recorded passively and continuously by passive acoustic sensors. Long-duration recordings of the natural environment provide rich information about vocal species activities and the soundscapes of animal habitats. Passive acoustic monitoring has many advantages including being less invasive, more cost effective, evidence preserving, and is effective at scaling spatiotemporally. However, the large volume of data collected by passive acoustic sensors can be an issue, in that it is impossible to manually analyse without a great deal of human effort. Thus, automated recognition tools are required for identifying target species. To recognise species vocalisations, we can search for acoustic patterns, such as oscillations in spectrograms and perform template-matching or build machine learning classifiers. Most solutions fall in the realm of bioacoustics focusing on individual calls. However, for this study we aim to identify species choruses, as this can be a useful tool for improving the efficiency of long-term acoustic studies. The target species is known to have periods of high calling activity, often following heavy rainfall events. Chorus-based recognisers provide a rapid assessment of acoustic data, highlighting these high-activity periods for further in-depth analysis. *Litoria olongburensis* calls can significantly vary due to individual differences, calling locations and weather conditions. Because of the complex noise profile and the significant variations in target signals, this problem requires methods that can be applied at large scales. Ecoacoustics is an ecological discipline that utilises all sources of environmental sound to assess biodiversity and environmental health [2]. One common methodology used is the calculation of acoustic indices which are summary statistics developed to characterise animal acoustic communities by summarising the structure and distribution of acoustic energy in field recordings. Those indices measures different aspects of acoustic data and can be used to create false-colour spectrograms by mapping index values to three primary colours, which helps reveal acoustic patterns in long-duration recordings [3]. Those indices are general features of recordings and do not rely on vocal species present in them.

In this study, we identify frog choruses in complex environments with both visualisation and machine learning methods. We test the performance of false-colour spectrograms on identifying the two target species. Then we use acoustic indices and machine learning techniques to recognise them. The experimental results show that our methods work well achieving average accuracy of 92.29%, 77.99% and 76.69% under different experiment settings. More importantly, our research also provides insight into the application of acoustic indices on recognising animal calls.

In the next section, we discuss related work briefly. In section III, we give details of frog chorusing activity recognition methods, including visualisation and index-based recognition. Section IV shows our recognition results with different index features and machine learning algorithms. Section V concludes the paper and give some thoughts on our future work.

II. Related Work

A. Ecoacoustics

Ecoacoustics is an emerging field that investigates all sources of environmental sound emanating from a landscape [4]. This new ecological discipline provides both theoretical and applied perspectives and contributes to environmental assessment and long-term monitoring [2]. An acoustic event detection system called Ecoacoustic Event Detection and Identification (EEDI) was proposed to evaluate the impact of long term climate change to species and their communities in a certain area [5]. A novel approach for the classification of urban area acoustical patterns was proposed based on study of city audio recordings [6]. Quantitative methods in ecoacoustics were recently summarised with a large emphasis placed on passive acoustic monitoring for environment assessment [2]. It also concludes some issues that remain underdeveloped in ecoacoustics, such as the choice of the most effective index and the most efficient temporal sampling schemes.

B. Acoustic indices and false-colour spectrogram

Studying soundscapes at large scale poses major technical challenges. Researchers are dealing with huge volumes of recordings that contain all sources of sound in the field and desire metrics that can facilitate the process of assessing big acoustic data. Acoustic indices are designed to reflect acoustic properties of audio recordings of the natural environment. There are intensity indices, complexity indices, and soundscape indices, which measure sound level, animal diversity, and the consist of a soundscape, respectively. They are usually calculated from waveforms or spectrograms of those recordings, which indicates characteristics of an ecosystem's acoustic energy across the frequency spectrum and over time [7].

The Acoustic Entropy Index (H) is a commonly calculated index; it is based on the Shannon information theory [8]. A higher H value indicates richer habitats. Acoustic Complexity Index (ACI) is another popular acoustic index. It is designed to produce a quantify the complexity of the soundscape by computing the variability of the acoustic energy registered in audio-recordings, despite the presence of constant human-generated-noise [9]. The Normalized Difference Soundscape Index (NDSI) [10] is used to estimate the level of anthropogenic disturbance on the soundscape by computing the ratio of human-generated (anthrophony) to biological (biophony) acoustic components found in field recordings. It range for -1 to +1, where +1 indicates no anthrophony. There are other acoustic indices that have been developed to assess acoustic activity and further infer other ecological information [11].

There are many applications of acoustic indices on ecological research. For example, ACI is used to evaluate the sonic complexity of the soundscape in the research of the relationship between sonic environment and vegetation structure [12]. NDSI helps visualise the temporal change in soundscape power of a habitat and reveal the acoustic patterns [13]. 14 acoustic indices are calculated to facilitate sampling avian species [14]. The experimental results show that sampling based on combinations of weighted indices outperforms sampling based on single indices. To quickly interrogate months of long-duration recordings, false-colour spectrograms are created by mapping three acoustic indices to three primary colours RGB respectively. Different false-colour spectrograms are generated using different indices and are suitable for exploring different ecological problems. They achieve success in three case studies of detecting Lewins Rail, three anuran species and several bat species. False-colour spectrograms not only can help identify acoustic events from long-duration recordings, but also can help ecologists understand the acoustic pattern changes over time [3].

C. Species acoustic identification

Species recognition is an important task to biodiversity monitoring. When it comes to automated recognition of animal calls, there are several schemes that have been explored. First, we can design a recogniser that scans through the spectrograms of recordings and search for matched target acoustic patterns. This method produces relatively accurate results but is expensive and requires trained experts who possess a great deal of signal processing and computer programming knowledge. Typically such recognisers are species-specific and are hard to generalise to other species.

The second method is template matching, which matches the spectrogram of potential signals against a library of templates using spectrogram cross-correlation [15]. However, it only suits a fraction of species, because it is strongly call-type dependent and requires all variations of vocalisations of the target species to be templates. It does not work for species that have an extensive vocal repertoire.

Yet another technique for animal call recognition is machine learning. Generally, this method includes noise reduction, call detection and isolation, features selection and training classifiers [16]. Various methods for each stage have been proposed. The problem is that there are all too many combinations of methods we can choose due to the large number of methods and parameters in each stage. Huang et al. [17] choose three features (spectral centroid, signal bandwidth and threshold-crossing rate) and kNN (k-nearest neighbours) and SVM (Support Vector Machines) are used to recognise the frog species. Another frog call recognition study [18] uses spectral peak tracks to extract various syllable features to characterise calls, including syllable duration, dominant frequency, oscillation rate, frequency modulation, and energy modulation. Therefore, most machine learning solutions for animal call recognition is species-specific and there is a lack of generalisation.

Acoustic indices are useful as descriptive features and thus classic machine learning algorithms can help discriminate different classes. Zhang et al. [19] use 7 acoustic indices to classify 5 acoustic patterns, including birds, insects, low activity, rain, and wind. Towsey et al. [20] introduces three case studies of the application of long-duration, false colour spectrograms. In one of those case studies, spectral index features are used with a Support Vector Machine (SVM) classifier to build an automated recogniser for the Lewin's Rail. It shows that the recogniser achieves a decent classification performance (precision 80% and recall 67%), but it can be affected by the recording quality (signal-to-noise ratio).

III. Recognition of chorus of Wallum Sedgefrogs and Eastern Sedgefrogs

In this section, we describe our methods of detecting chorus of the two species using acoustic indices. It is the first phase of the whole project which consists of chorusing level recognition and call level recognition. We use long-duration false-colour (LDFC) spectrograms generated using acoustic indices to inspect the general acoustic pattern of continuous recordings from our research site. Then we use acoustic indices as features of segments of recordings to perform machine learning classification on them to detect frog choruses. The results of chorusing level recognition will help narrow down the scope for recognising target calls.

A. Site and Data description

The research site is located at an urban residential development area in coastal south-east Queensland and was selected based on the confirmed co-existence of *L. olongburensis* and *L. fallax*. The data used for this research were collected as part of an ecological study focused on determining species population sizes and spatial distributions to inform restoration and conservation efforts. Song Meter 3 autonomous recorders (Wildlife Acoustics, Concord, MA, USA) connected to a solar panel were used to record continuously over the survey period. For this study, 24-hours of data collected on 6^{th} March 2018 after a period of heavy rainfall were selected due to the high calling activity of target species making the data suitable for exploring acoustic patterns in an intensive-vocalisation condition. Fig. 1 shows the spectrograms of typical calls of *L. olongburensis*, *L. fallax*, and another highly active frog species detected at the survey site, the Wallum Froglet (*Crinia tinnula*).

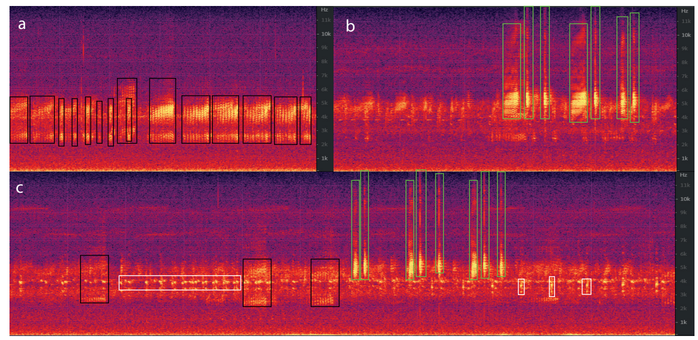

Fig. 1. Spectrograms of 3 frog species on site. (a) *L. olongburensis* calls with two dominant frequency bands in black boxes. (b) *L. fallax* calls in green boxes (with some *L. olongburensis* calling from the distance). (c) Choruses of three species. Wallum Froglet calls are a series of constant oscillating pulses lying in 3000-4000 Hz

To cope with long-duration recordings, we break them into one-minute segments. That gives us 1440 samples for a 24-hour recording. The sampling rate of our recording is 24000 Hz, but they were down-sampled to 22,050 Hz before generating spectrograms and acoustic indices. Spectrograms were prepared using a frame-width of 512, so they have 2,584 frames (frame duration = 23.2 ms) and 256 frequency bins (each with a bandwidth of 43 Hz). We use acoustic indices to describe our data, which include 14 summary indices and 12 spectral indices. We will elaborate on the index features in next section.

B. Calculation of acoustic indices

Our methods of recognising *L. olongburensis* and *L. fallax* includes inspecting false-colour spectrograms and classifying segments of field recordings, both of which rely on the calculation of acoustic indices.

Acoustic indices are proposed to summarise characteristics of recordings. They reflect the diversity, complexity, degree of evenness or ratios of frequencies of the sounds in the recording [4]. They are general features of environmental recordings and do not depend on the vocal species that occur in the recordings. In this study, we use acoustic indices to describe segments of field recordings and try to recognise animal vocalisations by discriminating different segments of recordings instead of directly classifying animal call samples.

There are two type of acoustic indices, spectral indices and summary indices. According to our indices calculation method, a spectral index is a vector of 256 values, one for each frequency bin of a spectrogram. A summary index is a scalar value, in some cases derived directly from the waveform envelope and in other cases from values of the corresponding spectral index. In this study, we choose 14 summary indices and 12 spectral indices, which are commonly used in visualisation, clustering and classification of environmental recordings [19]–[21]. Due to the lack of space, we give a brief introduction of the indices we use. Please refer to the technical report [22] for all the calculation details. All index values are generated by a program of our lab [23]. Tab.II and Tab.I introduce summary indices and spectral indices respectively.

C. Acoustic identification of Wallum Sedgefrogs and Eastern Sedgefrogs with LDFC spectrograms

LDFC spectrograms are developed to visualise the content of environmental recordings. They facilitate navigation of those recordings and help ecologists quickly find acoustic patterns of interest. In this study we choose a 24-hour recording to explore the application of LDFC on detecting acoustic patterns of *L. olongburensis* and *L. fallax*. The purpose is to select periods of interest quickly and narrow down the scope, because we have accumulated several months of recordings. To achieve this goal, we can explore locating the choruses of the two species and identify rain in the recordings, because they respond to rain actively. Fig.2 shows a LDFC spectrogram generated with three acoustic indices (ACI_{sp}, ENT_{sp} and EVN_{sp}).

Fig.2 gives a good overview of acoustic activities in our research site on a rainy day. As we can see from the spectrogram, from 00:00 am to 2:30 am, there is green vertical lines at the top (9500-11025 Hz), synchronising with the pink lines (mixed with some green and yellow lines) below. That indicates the intense chorusing of both two species. The pink hue in 2000-3000 Hz shows when there was intense *L. olongburensis* calls. After 02:30 am, *L. fallax* reduced their calling frequency and *L. olongburensis* dominant the acoustic space. The mixture of frequent *L. olongburensis* chorusing and a few *L. fallax* calls results in the blue hue between 4000 Hz and 6000 Hz. The constant bright dots in 3000-4000 HZ help to identify the third species *Crinia tinnula*, which were chorusing for almost the whole day. As for rains, the heavy period will cause the red and orange hues occupying the whole band and light rains can be identified by the green or orange lines at the bottom. To conclude, we can roughly locate choruses of the two species, except some faint calls mixed with other intense choruses, and rain.

In Fig. 3 (a), Eastern Sedgefrogs called at a loud amplitude, the high frequency power remains a lot, which contributes to green vertical lines above 10000 Hz. The dominant frequency range of Eastern Sedgefrogs lies in 5000-6000 Hz, which displayed in the LFDC spectrogram with bright pink band. The unique low frequency band (2000-3000 Hz) of Wallum Sedgefrog choruses make them stand out in all spectrograms, so the number of pink vertical lines in that range indicate the magnitude of vocal activity of Wallum Sedgefrogs. As for Fig. 3 (b), the constant vocalisations in the range of 3000-4000 Hz are from Wallum Froglets and those are accountable for the two pink horizontal line in that range of LFDC spectrograms. Fig. 3 (b) shows that when the two species are less active, the pink band in (a) will turn to blue band. That is the result from the drop of ACI values, which means less acoustic intensity changes. Different from (a), Wallum Sedgefrogs were calling loud with a wide frequency band, while Eastern Sedgefrogs choruses changed to short band. In this situation, Eastern Sedgefrogs choruses are hard to identify from LFDC spectrograms because they totally overlap with Wallum Sedgefrogs and their call structure is very similar to them, too.

D. Automated recogniser using acoustic indices

Conventionally, acoustic indices refers to summary indices, but in this study we utilise both summary indices and spectral indices. For one index, there is one summary index value for a clip, while a spectral index has 256 values. To build index-based recognisers, we use acoustic indices as features to represent audio clips with fixed length (60 seconds) obtained from long-duration recordings. First, we break 24-hour recordings into 1440 clips and then label them according to which are species presented in the clip. The annotation work is done by listening through the recording by trained ecologists. The annotation result is shown in Fig. 4.

The sub-graph of *L. olongburensis*, *L. fallax* and *Crinia tinnula* provide presence proportion of three species in 1440 clip of recordings. *Crinia tinnula* was not originally a species of interest, however their presence and absence information forms a significant part of background noise of our data. They are very active on that day and their calls appear in 99.03% of all clips. Considering there is a multi-class situation in our data, we choose to create a class label of four classes to depict the presence and absence information of the two target species. Class values A, B, C and D in the last sub-graph represent both *L. olongburensis* and *L. fallax* present, only *L. olongburensis* present, only *L. fallax* present, and neither present respectively. As we can see from the last sub-graph, situation A and B dominant our data and D is rare (only 2.01%). In addition, no matter which situation is in the clips, most of them have *Crinia tinnula* calling in the background.

To prepare training dataset for machine learning classifiers, we use the presence or absence information (A, B, C and D) in each clip as their class label and acoustic indices values of each clip as their feature values. Therefore, we have a summary index dataset with 1440 samples and 14 continuous value features and a spectral index dataset with the same size and 3072 continuous value features.

As for experiment settings, we conduct three sets of experiment to evaluate the ability of indices on discriminating different class of audio clips and identify frog chorusing. The first set of experiment gives us a benchmark of classification performance by using all features and two classic machine learning classification algorithms, including Random Forest (RF) [25] and Support vector machine (SVM) [26].

The second experiment is designed to reveal the impact of different frequency bands and indices on the performance of recognition. First, we try to use index features that derived

TABLE I
SPECTRAL INDICES

Abbreviations	Acoustic indices	Reference
BGN_{sp}	Background Noise	It is the noise profile of the recording and derived from spectrograms. Detailed calculation methods can be found in the same technical report [14].
PMN_{sp}	Power Minus Noise	The value of PMN in frequency bin equals to the maximum decibel value minus decibel BGN value for the corresponding bin [22].
ACI_{sp}	Acoustic Complexity Index	It quantifies the relative change in acoustic intensity in each bin of the spectrogram [9].
CVR_{sp}	Activity	The proportion of cells in each frequency bin of the spectrogram where the spectral power exceeds 3 dB [3].
ENT_{sp}	Temporal Entropy	It is a Shannon index calculated from a waveform envelope, providing information of acoustic concentration in each frequency bin [8].
EVN_{sp}	Events	The number of acoustic events per minute in each frequency bin. An event is counted each time the decibel value in a bin crosses the 3 dB threshold from lower to higher values [22].
$R3D_{sp}$	Ridge 3 Dimensions	Ridge indices measures formants of harmonic structure of animal sounds. R3D is the maximum of RHZ, RPS and RNG [22].
RHZ_{sp}	Ridge Horizontal	Measuring horizontal ridge slope [24]
RNG_{sp}	Ridge Negative	Measuring downward ridge slope [22]
RPS_{sp}	Ridge Positive	Measuring upward ridge slope [22]
RVT_{sp}	Ridge Vertical	Measuring vertical ridge slope [24]
SPT_{sp}	Spectral Peak Tracks	The proportion of local maxima in each frequency bin [22].

TABLE II
SUMMARY INDICES

Abbreviations	Acoustic indices	Description and Reference
BGN	Background Noise	An estimate value of the background noise in an audio segment [14].
SNR	Signal to Noise Ratio	It is calculated from decibel envelop of recordings. It equals to maximum amplitude value minus BGN [14].
ACT	Activity	The ratio of frames where amplitude values that exceed 3 dB [14].
EVN	Events per Second	An acoustic event starts when the signal envelop cross 3 dB from below to above. This index is obtained by dividing the number of events by the total seconds in one segment [22].
LFC	Low-frequency Cover	The proportion of spectrogram cells that grater than 3 dB in low frequency band (below 1000 Hz) [22].
MFC	Mid-frequency Cover	The proportion of spectrogram cells that grater than 3 dB in mid frequency band (1000-8000 Hz) [22].
HFC	High-frequency Cover	The proportion of spectrogram cells that grater than 3 dB in mid frequency band (above 8000 Hz) [22].
ACI	Acoustic Complexity Index	It is the average of the mid-band spectral ACI values we described in spectral indices [22].
ENT	Temporal entropy	It is as the same as the spectral index of ENT, except we calculate it for the whole segment not one frequency bin [22].
EPS	Entropy of the Spectral Peaks	It is the 'entropy' measure for the distribution of local maximum in the mid-band [14].
EAS	Entropy of the Average Spectrum	Similar to EPS, except calculating the average amplitude instead of the number spectral peaks [14].
ECV	Entropy of the Spectrum of Coefficients of Variation	Similar to EPS, except calculating the variance of amplitudes divided by the mean of the energy values instead of the variance of amplitudes [22].
CLS	Cluster Count	The number of distinct spectral clusters in a segment of recordings. It is also calculate only from the mid-band. Clustering details can be found in [22].
SPD	Spectral Peak Density	Similar to the calculation of spectral index SPT, but calculate for the whole segment of recordings instead of one frequency bin [22].

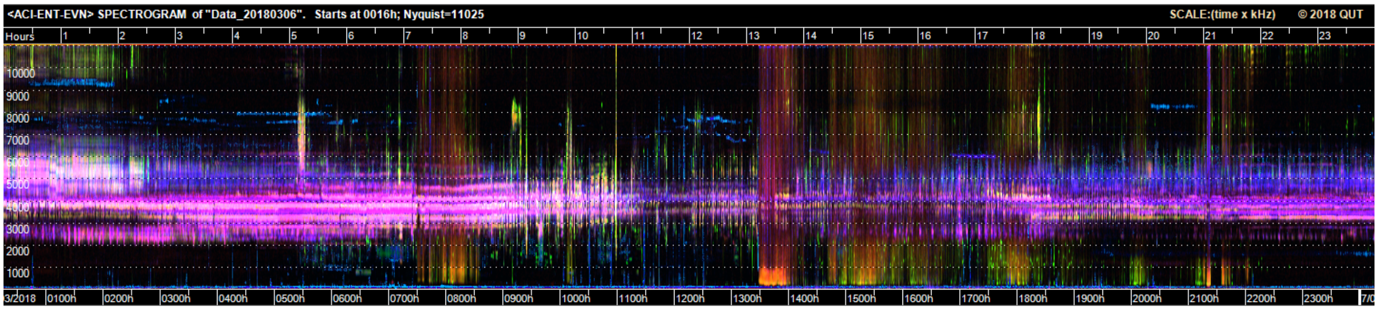

Fig. 2. A 24-hour LDFC spectrogram with a temporal resolution of 60 seconds per pixel

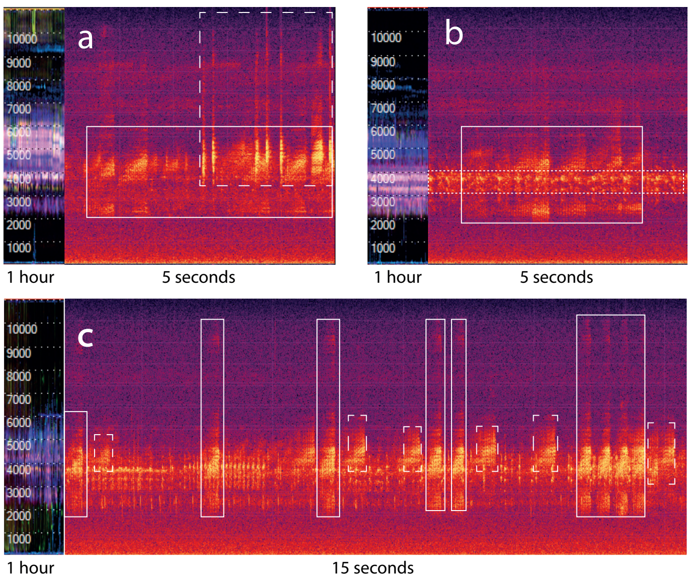

Fig. 3. Three LDFC spectrograms (generated using ACI, ENT and EVN), each of one hour duration, alongside standard spectrogram taken from the same time-period. (a) From 00:00 to 01:00 midnight, the solid box identifies Wallum Sedgefrog chorusing and the dashed box identifies Eastern Sedgefrog chorusing. (b) From 03:00 to 04:00, Wallum Froglets choruses are surrounded by the short-dashed box. (c) From 16:00 to 17:00, same as (a) , except the changes in frequency bands of their choruses

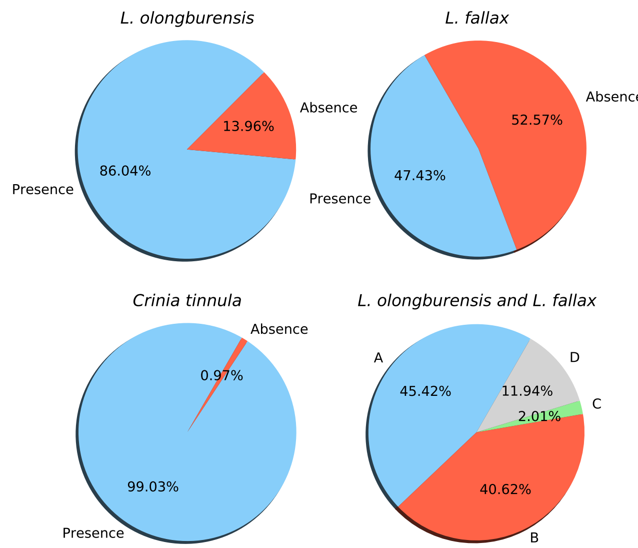

Fig. 4. Annotation results

only from the frequency bands of our target calls. Since we are using field-recorded long-duration recordings, frequency bands of target signals vary a lot due to several reasons, such as different individuals, weather condition and calling locations. As we observe from data, dominant frequency of *L. fallax* calls and the high frequency band of *L. olongburensis* fall in 4000-6000 Hz and the dominant frequency of low frequency band of *L. olongburensis* fall in 2000-3000 Hz. The calls of the third species *Crinia tinnula* calls dominate 3000-4000 Hz. Therefore, we set three frequency bands, 2000-6000 Hz (2k-6k), 2000-3000 Hz and 4000-6000 Hz (2k-3k and 4k-6k) and all bands except 3000-4000 Hz (except 3k-4k), to check if band selection works on our data. Similar to band selection, we remove one index at a time and evaluate the rest data to find which index has a greater impact on our recognition task. The methods in this set of experiment can be classified as manual feature selection.

In the last part of experiment, we use machine learning algorithms to select valuable features instead of manual selection. That will give us new perspectives besides manually selecting bands and indices. We set SVM as our classifier in this section and we use a scheme-specific selection method called classifier subset evaluator to select good features for the SVM classifier [27]. When searching for the best feature set, the classifier subset evaluator will evaluate feature subsets in terms of the performance of a specified classifier, in this case, SVM classifier. The search method used here is *bestfirst*, which searches the space of attribute subsets by greedy hillclimbing augmented with a backtracking facility. The number of consecutive non-improving nodes allowed is set as 5 to controls the level of backtracking done. It starts with the empty set of attributes and search forward. The subsets are evaluated by

10-fold cross validation in the search process.

All experiments are conducted using WEKA Machine learning package [27]. Performance was assessed using ten-fold cross-validation. Given we are dealing with a multi-class problem and all our classifiers are discriminative models, the performance measures we choose are $F1score$, $AverageAccuracy$, $Precision$ and $Recall$ [28]. The average values for four classes are reported. The experiment results and discussion are given in the next section IV

IV. Experiment Results and Discussion

In this section, we give experiment results discussion. The experiment consists of four parts as we describe before, summary indices vs. spectral indices, selection of frequency bands and spectral indices, and feature selection using machine learning.

A. Summary indices Vs. Spectral indices

In this section, we evaluate the classification performance of different classifiers on data represented by different feature sets. There are 14 features (one for each index) in summary index data, while the spectral index data has 3072 features (256 for each index). We also include the comparison of original values and standardised values (z-score) to check if this factor has impact on the classification results. Experimental result is shown in Fig. 5.

As we can observe from Fig. 5, spectral indices have a better performance in terms of all metrics in all settings (8.4% better in terms of accuracy). On one hand, spectral index features contain much more information in frequency domain and it helps better discriminate different class of clips. On the other hand, the huge number of spectral index features makes the time of building the classification model much longer than using summary indices. We notice that $F1score$, $Precision$ and $Recall$ values are relatively low. That is because class C (only *L. fallax*) is very rare (only 29 samples) and none of them are correctly classified. Not enough training data for this class makes those metrics very low. The $AverageAccuracy$ is more informative for our unbalanced data.

Another fact we can observe from the bar charts is that SVM performs better than RF in most cases. Although spectral indices are not ahead of summary ones much, there is a potential of performance improvement in spectral index features selection, because there 12 indices and 256 frequency bins within each of them. Therefore, we conduct a second set of experiment to explore the performance of different combinations of spectral index features on our data.

B. Selection of frequency bands and Spectral indices

In this section, we give experimental results of classification with feature sets consisting of different spectral indices and frequency bands on our data. Due to lack of space, we only report average accuracy here. Fig. 6 gives the experimental results of spectral features from different frequency bands.

As we can observe from Fig. 6, the best performance is achieved by SVM classifier on original data derived from full band except 3000-4000 Hz. It is slightly higher than the previous best (achieved by SVM on all spectral features) about 0.6%. This set of results indicates that SVM performs better than RF on all feature sets whether rescaled or not. For standardised data, the performance of SVM on selected features is not as good as on all features in all three categories. However, performance of RF improves except the last two categories in original data. To conclude, the outcome of band selection varies with classifiers and data rescaling choice.

Besides frequency bands, choice of indices is also important to classification performance. To find out which index has a greater influence on classification performance, we leave one index out each time to construct datasets. Since SVM performs better in all previous experiment, we set it as our classifier later on. Fig. 7 gives the performance of SVM classifier on data with different indices.

The average accuracy values the SVM classifier achieves on all spectral indices are 0.759 (original data) and 0.7542 (Standardised data). In Fig. 7, we can see that original data without PMN$_{sp}$ and BGN$_{sp}$ has a lower performance. As for standardised data, removing indices RHZ$_{sp}$ and BGN$_{sp}$ makes the performance worse. Those indicates that PMN$_{sp}$, BGN$_{sp}$, and RHZ$_{sp}$ are relative useful features for detecting choruses of the two species, because without those information the accuracy drops.

C. Feature selection using machine learning

After two sets of experiment above, we conduct a third experiment to find out which features are identified as informative features by machine learning algorithms. This may give us new perspectives other than manual selection experience. Besides the four-class problem, we include two binary classification problems of the two species (presence and absence). The feature selection results and classification result associated with them are presented in Table. III. The columns are the type of classification problem, the total number of selected features, frequency bin numbers of BGN$_{sp}$, PMN$_{sp}$ and EVN$_{sp}$ and accuracy of classification results with SVM and selected features. All index features selected belong to BGN$_{sp}$, PMN$_{sp}$ and EVN$_{sp}$.

TABLE III
Selected features and performance

	Number of features	BGN$_{sp}$	PMN$_{sp}$	EVN$_{sp}$	Accuracy
L. olongburensis	8	228	3, 50, 79, 113	74, 85, 107	92.29%
L. fallax	15	1, 28, 41, 55, 67, 110	1, 63, 77, 83, 85, 87, 94, 99, 114	N/A	77.99%
Four class	22	0, 9, 36, 84, 126, 129, 207, 224, 228	20, 28, 48, 177, 179, 188, 189, 196, 203, 217, 221	33, 125	76.69%

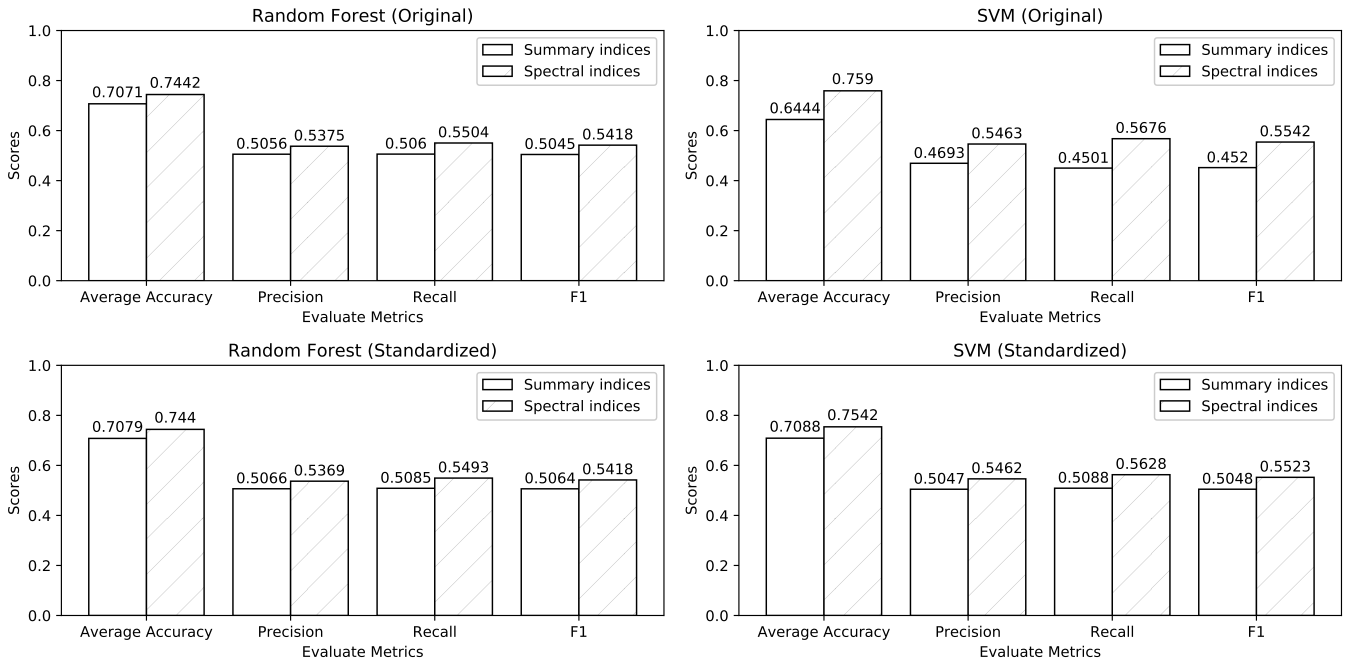

Fig. 5. Classification performance of RF and SVM on summary and spectral indices datasets with original values and standardized values

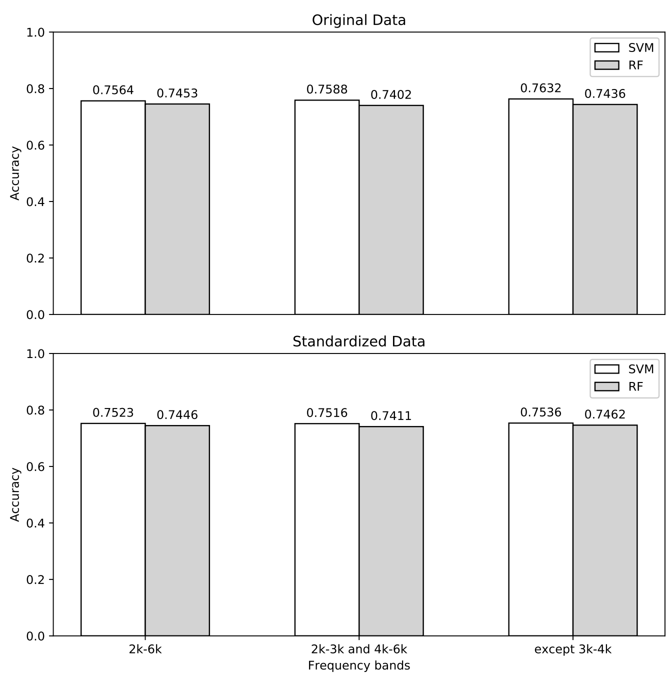

Fig. 6. Classification performance on spectral index data from different frequency bands

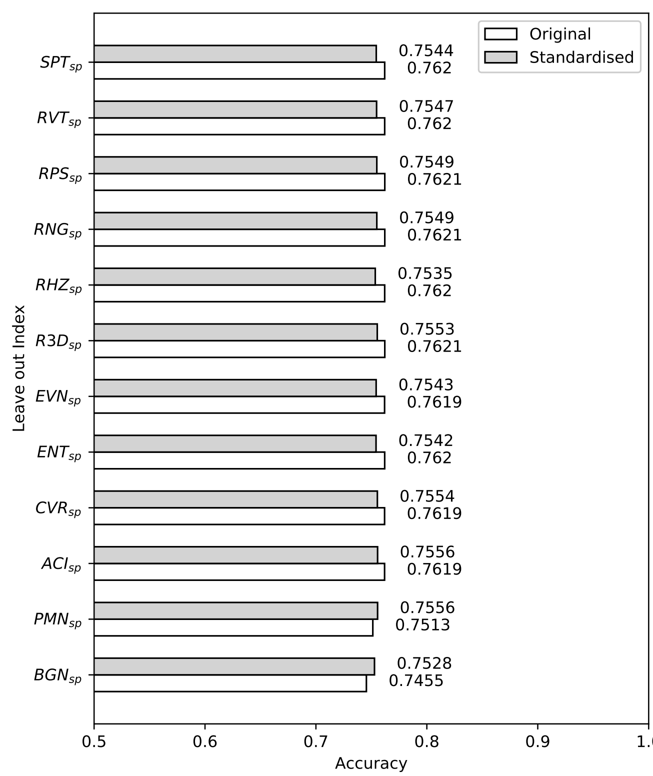

Fig. 7. Classification performance on data without different indices

Table. III shows that BGN$_{sp}$, PWN$_{sp}$ and EVN$_{sp}$ have a greater impact on identifying *L. olongburensis* calls and *L. fallax* calls than other indices. These indices measure background noise profile, signal power minus noise and events (power cross a certain threshold) per frequency bin. As for the number of bins selected in each index, PMN$_{sp}$ features are the most selected features (24) flowed by BGN$_{sp}$ (16) and EVN$_{sp}$ comes the last (5). The four class recognition task requeirs more features than two individual recognition task (22 VS. 8 or 15). Another thing worth exploring is the frequency bins distribution, which will show us which bins are the most effective. Table. IV shows the frequency bins distribution among selected features.

TABLE IV
DISTRIBUTION OF SELECTED FREQUENCY BINS

	0 Hz - 2196 Hz	2196 Hz - 4349 Hz	4349 Hz - 8656 Hz	8656 Hz - 11025 Hz
L. olongburensis	2	3	2	1
L. fallax	4	9	2	0
4-class	7	1	8	6
Total	13	13	12	7

As we can observe from Table. IV, different classification tasks prefer features from different frequency bands. For detecting *L. olongburensis*, index features from different bands do not have much variance. Information from all bands contributes to discriminate their presence or absence. Different from *L. olongburensis*, detecting *L. fallax* mostly relies on acoustic characteristics from 2196-4349 Hz, where their calls are not likely to be seen. At last, recognition of the four acoustic patterns (whether the two species occurred) relies on all frequency bands except 2196-4349 Hz as this is where the three species calls overlap and it cannot provide much distinguishing information.

V. CONCLUSION AND FUTURE WORK

False-colour spectrograms were found to be useful to for detecting choruses of the two species. However, the detectability of *L. fallax* vocalisations declined with lower call amplitudes and in instances where the dominant frequency of the call overlapped with the higher frequency band of simultaneously calling *L. olongburensis* individuals. Compounding this, *L. olongburensis* was found to lose the lower frequency part of their call, making the two species calls indistinguishable on the false-colour spectrogram.

From the three sets of experiment above, we can see that spectral indices outperform summary indices but require more time to train the model. For the spectral indices, manual frequency band selection does not improve classification performance on this 24-hour data. This may be caused by the frequency band overlap and call variability of the two species. From the results of manual index selection and machine index selection, three indices BNG$_{sp}$, PWN$_{sp}$ and EVN$_{sp}$ are the best at discriminating different chorusing activities. Feature selection using machine learning also gives a distribution of frequency bins of selected indices. Detection of *L. olongburensis* requires the least number of index features and those features spread evenly through the whole frequency band. The low-frequency part of *L. olongburensis* calls contributes most to their detection. To properly classify four classes of segments, we need to choose features outside the overlap zone (2000-4000 Hz).

The methods we present in this work have proven to reliably recognise frog choruses. As future work, we plan to test them on larger data sets that over significant period of time, and that vary by season and weather conditions. Further, we plan to extend the application of these methods to other species and develop a standard procedure of recognition of chorusing behaviours of animals. We are also interested in evaluating the performance of deep learning networks on our data.

VI. ACKNOWLEDGMENT

We thank Mark Stephens, Senior Environment and Community Development Manager at Stockland Sunshine Coast, for his support in this research, and Stockland and the Smart Cities and Suburbs Program (Grant ID: SCS59465) for their financial support. Hongxiao Gan acknowledges the financial support from the China Scholarship Council and a CSC Top-Up scholarship from Queensland University of Technology.

VII. REFERENCES

REFERENCES

[1] E. Meyer, *National recovery plan for the wallum sedgefrog and other wallum-dependent frog species*. Queensland. Parks and Wildlife Service, 2006.
[2] A. Farina, "Ecoacoustics: A quantitative approach to investigate the ecological role of environmental sounds," *Mathematics*, vol. 7, no. 1, p. 21, 2019.
[3] M. Towsey, L. Zhang, M. Cottman-Fields, J. Wimmer, J. Zhang, and P. Roe, "Visualization of long-duration acoustic recordings of the environment," *Procedia Computer Science*, vol. 29, pp. 703–712, 2014.
[4] A. Farina and S. H. Gage, *Ecoacoustics: The ecological role of sounds*. John Wiley & Sons, 2017.
[5] A. Farina and P. Salutari, "Applying the ecoacoustic event detection and identification (eedi) model to the analysis of acoustic complexity," *Journal of Mediterranean Ecology*, vol. 14, pp. 13–42, 2016.
[6] M. Rychtrikov and G. Vermeir, "Soundscape categorization on the basis of objective acoustical parameters," *Applied Acoustics*, vol. 74, no. 2, pp. 240–247, 2013.
[7] K. Servick, "Eavesdropping on ecosystems," 2014.
[8] J. Sueur, S. Pavoine, O. Hamerlynck, and S. Duvail, "Rapid acoustic survey for biodiversity appraisal," *PloS one*, vol. 3, no. 12, p. e4065, 2008.
[9] 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 Indicators*, vol. 11, no. 3, pp. 868–873, 2011.
[10] 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.
[11] J. Sueur, A. Farina, A. Gasc, N. Pieretti, and S. Pavoine, "Acoustic indices for biodiversity assessment and landscape investigation," *Acta Acustica united with Acustica*, vol. 100, no. 4, pp. 772–781, 2014.
[12] A. Farina and N. Pieretti, "Sonic environment and vegetation structure: a methodological approach for a soundscape analysis of a mediterranean maqui," *Ecological Informatics*, vol. 21, pp. 120–132, 2014.
[13] S. H. Gage and A. C. Axel, "Visualization of temporal change in soundscape power of a michigan lake habitat over a 4-year period," *Ecological Informatics*, vol. 21, pp. 100–109, 2014.

[14] 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*, vol. 21, pp. 110–119, 2014.

[15] D. Stowell, T. Petruskov, M. lek, and P. Linhart, “Automatic acoustic identification of individual animals: Improving generalisation across species and recording conditions,” *arXiv preprint arXiv:1810.09273*, 2018.

[16] N. Priyadarshani, S. Marsland, and I. Castro, “Automated birdsong recognition in complex acoustic environments: a review,” *Journal of Avian Biology*, vol. 49, no. 5, pp. jav–01 447, 2018.

[17] C.-J. Huang, Y.-J. Yang, D.-X. Yang, and Y.-J. Chen, “Frog classification using machine learning techniques,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3737–3743, 2009.

[18] J. Xie, M. Towsey, A. Truskinger, P. Eichinski, J. Zhang, and P. Roe, “Acoustic classification of australian anurans using syllable features,” in *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*. IEEE, 2015, Conference Proceedings, pp. 1–6.

[19] L. Zhang, M. Towsey, J. Zhang, and P. Roe, “Computer-assisted sampling of acoustic data for more efficient determination of bird species richness,” in *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 2015, Conference Proceedings, pp. 798–805.

[20] M. Towsey, E. Znidersic, J. Broken-Brow, K. Indraswari, D. M. Watson, Y. Phillips, A. Truskinger, and P. Roe, “Long-duration, false-colour spectrograms for detecting species in large audio data-sets,” *Journal of Ecoacoustics*, vol. 2, p. IUSWUI, 2018.

[21] P. Eichinski and P. Roe, “Clustering and visualization of long-duration audio recordings for rapid exploration avian surveys,” in *2017 IEEE 13th International Conference on e-Science (e-Science)*. IEEE, 2017, Conference Proceedings, pp. 168–176.

[22] M. W. Towsey, “The calculation of acoustic indices derived from long-duration recordings of the natural environment,” 2017.

[23] M. Towsey, A. Truskinger, M. Cottman-Fields, and P. Roe, “Ecoacoustics audio analysis software v18. 03.0. 41 (version v18. 03.0. 41). zenodo,” 2018.

[24] X. Dong, M. Towsey, J. Zhang, J. Banks, and P. Roe, “A novel representation of bioacoustic events for content-based search in field audio data,” in *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on*. IEEE, 2013, Conference Proceedings, pp. 1–6.

[25] T. K. Ho, “Random decision forests,” in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, Conference Proceedings, pp. 278–282.

[26] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, “Support vector machine classification and validation of cancer tissue samples using microarray expression data,” *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.

[27] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[28] M. Sokolova, G. J. I. P. Lapalme, and Management, “A systematic analysis of performance measures for classification tasks,” vol. 45, no. 4, pp. 427–437, 2009.